Rapid mass spectrometry coupled with machine learning to predict clinically relevant variables in cardiovascular disease

Annabel S. J. Eardley-Brunt

Chemistry in Cells: New Technologies to Probe Complex Biology and Medicine

Medical Sciences Doctoral Training Centre

Keble College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2025

Table of Contents

List of Tables **xi**

List of Figures **xvii**

Abstract **xxxvii**

Acknowledgements **xxxix**

Abbreviations **xli**

1 Introduction **1**

1.1 ST elevated myocardial infarction (STEMI) 2

1.1.1 STEMI pathophysiology 3

1.1.2 STEMI diagnosis 4

1.1.3 STEMI treatment 4

1.1.4 Clinical indicators of STEMI severity and recovery success . . 7

1.1.5 Current STEMI research 9

1.2 Abdominal aortic aneurysm (AAA) 9

1.2.1 AAA pathophysiology 9

1.2.2 AAA diagnosis and monitoring 11

1.2.3 AAA treatment 11

1.2.4 AAA development and growth 12

1.2.5 Current AAA research 13

1.3 Atmospheric solids analysis probe mass spectrometry (ASAP-MS) . . 13

1.3.1 Mass spectrometry 13

1.3.2 Ambient ionisation mass spectrometry 15

1.3.3 Atmospheric solids analysis probe mass spectrometry (ASAP-MS) 16

1.3.4 Application of ASAP-MS to a clinical setting 21

1.4 Machine learning 22

1.5 Aims of this thesis 23

2 Experimental methods **25**
2.1 Instrumentation . . . 25
2.1.1 Standard experimental framework . . . 27
2.2 STEMI samples - The OxAMI study . . . 28
2.3 AAA samples - The OxAAA study . . . 30
2.4 Method development for the analysis of human blood plasma by ASAP-MS . . . 34
2.4.1 Background . . . 34
2.4.2 Materials and methods . . . 37
2.4.3 Results . . . 48
2.4.4 Optimised measurement protocol for ASAP-MS measurement on human plasma . . . 49
2.5 Method development for the analysis of human tissue samples by ASAP-MS . . . 51
2.5.1 Background . . . 51
2.5.2 Materials and methods . . . 52
2.5.3 Parameter optimisation . . . 53
2.5.4 Results . . . 59
2.5.5 Optimised measurement protocol for ASAP-MS measurement on human tissue biopsy samples . . . 60
2.6 Conclusions and outlook . . . 61

3 Data analysis methods **62**
3.1 Processing of ASAP-MS data . . . 62
3.1.1 Extracting the individual measurements . . . 63
3.1.2 Data binning, discretisation and peak selection . . . 64
3.1.3 Background subtraction . . . 66
3.1.4 Data normalisation . . . 66
3.1.5 Data averaging . . . 67
3.1.6 Data processing summary . . . 68
3.2 Machine learning methods . . . 68
3.2.1 Unsupervised machine learning and statistical analysis . . . 68
3.2.2 Supervised machine learning classification methods . . . 70
3.2.3 Symbolic regression analysis . . . 77
3.2.4 Feature selection and reduction methods for machine learning 81
3.2.5 Machine learning assessment . . . 84
3.2.6 Machine learning cross-validation and partitioning . . . 87
3.3 Summary . . . 90

4 Prediction of clinical outcomes for STEMI patients 91
4.1 The prediction of OxAMI clinical variables from coronary aspirate plasma samples . . . 91
4.1.1 Introduction . . . 91
4.1.2 Methods . . . 94
4.1.3 Results and Discussion . . . 101
4.1.4 Identification of peaks using high-resolution ASAP-Q-TOF-MS analysis of blood plasma . . . 112
4.2 Symbolic regression for the prediction of clinical variables . . . 112
4.2.1 Introduction . . . 112
4.2.2 Methods . . . 113
4.2.3 Results . . . 118
4.3 Conclusions and outlook . . . 124

5 Prediction of AAA development and treatment progression 127
5.1 The prediction of AAA patient treatment groups from venous plasma samples . . . 127
5.1.1 Introduction . . . 127
5.1.2 Methods . . . 128
5.1.3 Results and Discussion . . . 131
5.2 The classification of human tissue samples from AAA patients . . . 141
5.2.1 Introduction . . . 141
5.2.2 Methods . . . 141
5.2.3 Results and Discussion . . . 145
5.3 The prediction of AAA growth rate, size, and flow-mediated dilatation 154
5.3.1 Introduction . . . 154
5.3.2 Materials and methods . . . 154
5.3.3 Results . . . 157
5.4 Conclusions and outlook . . . 167

6 Investigation of different ASAP-MS instruments for clinical sample analysis 169
6.1 Introduction . . . 169
6.2 Materials and Methods . . . 170
6.2.1 Biological samples . . . 170
6.2.2 Measurement protocol for Advion CMS instruments . . . 171
6.2.3 Measurement protocol for the Radian instrument . . . 171
6.2.4 Measurement protocol for the ASAP-QTof-MS instrument . . . 172

6.3 Results . . . 173
6.3.1 Subjective comparison of technical features and user observations . . . 173
6.4 Data acquisition methods - quantitative comparisons . . . 176
6.4.1 Inter-instrument comparison 1: Optimum scan time . . . 176
6.4.2 Inter-instrument comparison 2: Optimum acquisition time . . . 177
6.4.3 Inter-instrument comparison 3: Optimum number of repeat measurements . . . 178
6.4.4 Inter-instrument comparison 4: Optimum cone voltage . . . 179
6.4.5 Inter-instrument comparison 5: Optimum capillary temperature 180
6.4.6 Inter-instrument comparison 6: Optimum corona discharge current . . . 181
6.5 Comparison of spectra between instruments . . . 184
6.5.1 Comparison of tissue spectra: Radian and CMS-S . . . 184
6.6 Converting data between instrument formats . . . 188
6.7 Conclusions and outlook . . . 194

7 Investigation of the effect of normalisation, scaling and transformation methods 195
7.1 Introduction . . . 195
7.2 Sources of error and uncertainty in ASAP-MS data . . . 199
7.3 Normalisation, scaling and transformation methods . . . 201
7.3.1 Normalisation methods . . . 202
7.3.2 Scaling methods . . . 205
7.3.3 Transformation methods . . . 206
7.4 Material and methods . . . 206
7.4.1 Method evaluation . . . 207
7.5 Results . . . 209
7.5.1 The effect of normalisation, scaling and transformation methods on clustering efficiency . . . 209
7.5.2 The effect of normalisation, scaling and transformation methods on machine learning classification . . . 212
7.6 Comparison of normalisation assessment methods . . . 214
7.6.1 Batch effect correction . . . 215
7.7 Conclusions . . . 217

8 Conclusions, evaluation and future work 218
8.1 Method evaluation . . . 220

8.1.1 Challenges and opportunities 220
8.2 Comparison with alternative analytical techniques 224
8.3 Future experiments 225
8.3.1 Measurement of healthy human blood 225
8.3.2 Measurement of whole and dried blood by ASAP-MS 226
8.3.3 Measurement of venous blood samples from OxAMI patients . 227
8.3.4 Measurement of samples from a larger AAA patient cohort with increased growth rate monitoring data 227
8.3.5 Measurement automation 228
8.3.6 Development of a single measure of patient risk for use in a clinical setting 228

References **230**

A Appendix: Method development for the analysis of biological samples by ASAP-MS **255**
A.1 Optimisation of tissue acquisition time 256
A.2 Optimisation of tissue repeat measurements 257

B Appendix: Prediction of STEMI patient outcomes using ASAP-MS and ML analysis **258**
B.1 ML classification methods 258
B.2 ML classification results - OxAMI 'All patients' 260
B.3 ML full results for the classification of OxAMI data 264
B.4 ML full results for classification of 'extreme' patients data 267
B.5 Significant m/z peaks determined through FR for STEMI group classification 272
B.6 Identification of ion peaks using ASAP-Q-TOF-MS analysis of blood plasma 273
B.7 PySR hyperparameters 277
B.8 Symbolic regression hyperparameter optimisation 277
B.9 PySR results equations 279

C Appendix: Prediction of AAA patient groups using ASAP-MS and ML analysis **280**
C.1 ML full results for classification of clinical data 281
C.2 Significant peaks determined through RF for AAA plasma classification 283
C.3 Classification of AAA tissue subtypes 285
C.3.1 Statistical analysis of tissue subtype spectral differences 285

C.3.2 Machine learning analysis of tissue subtypes 286
C.3.3 Analysis of peaks found through feature reduction for tissue subtype classification 288
C.4 Tissue classification: Using repeat samples to increase the training data size 290
C.5 Further investigation into the impact of training data size on ML classification results 295
C.5.1 The impact of the number of test patients in the hold-n-out validation method 299
C.6 Classification of AAA patients by clinical variables 302
C.7 Identification of overlapping significant (P<0.05) peaks 304

D Appendix: Investigating different ASAP-MS instruments for clinical sample analysis 305
D.1 Scan speed analysis across Radian and CMS-S for tissue types 306
D.2 Radian plasma measurements with changing cone voltage 307
D.3 Average artery wall measurements on the CMS-S and Radian instrument 308
D.4 Average fat measurements on the CMS-S and Radian instrument . . . 309
D.5 Average thrombus measurements on the CMS-S and Radian instrument 310
D.6 Inter-instrument comparison of tissue sample metrics 311
D.7 Inter and intra instrument reproducibility: correlation and cosine angle 312
D.8 Method for the generation of instrument functions 314
D.9 Instrument function smoothing optimisation 315
D.10 Instrument function centroid distances 318
D.11 Instrument function PCA plots 319
D.12 Instrument function with smoothing results plots 321
D.13 Neural network ASAP-MS resolution mapping 324
D.13.1 Introduction: Using neural networks to map between low-resolution aASAP-MS data and high-resolution ASAP-QTOF-MS data 324
D.13.2 Methods 325
D.13.3 Generation of artificial data 329
D.13.4 Neural network conversion of high-resolution data to low-resolution data 331
D.14 Bayesian optimisation for neural network architecture design - model setup 333
D.15 NN resolution analysis: Results and Discussion 334

E Appendix: Investigation of normalisation methods applied to ASAP-MS data **336**

E.1 Normalisation results across different ML classification methods . . . 337

List of Tables

2.1 Patient demographic, health and lifestyle distributions. 29
2.2 The samples available for AAA patient groups used in the ASAP-MS study. 33
2.3 Ion source settings. The capillary and source offset voltages are applied to the ion transfer capillary and source region respectively (see Figure 2.1). The source voltage is scanned from the initial value at $m/z = 0$ across the range defined by the source voltage span parameter as the mass is scanned. 41
2.4 Instrument scan time for a scan range of 10-1200 m/z. 43
2.5 Allan variance analysis of acquisition time for tissue samples. 57

3.1 Interpretation of κ as levels of agreement, adapted from [267] 87

4.1 Classifications of clinical parameters. 99
4.2 The highest κ scores achieved for each of the nine clinical variables tested with each model setup for all the available patients. The highest score obtained for each variable is highlighted in green. 103
4.3 The highest κ scores achieved for each of the nine clinical variables tested with each model setup for the selections of extreme patients. The highest score obtained for each variable is highlighted in green. . 104
4.4 The best results obtained for each categorical clinical variable, and the data structure required to achieve that result. Variables predicted with over 80% accuracy highlighted green, variables predicted with over 70% accuracy highlighted yellow. 105
4.5 The m/z ion peaks identified as being important for classification by the 3 FR methods for each clinical variable. 108
4.6 The κ scores obtained when the analysis of the undersampled datasets were repeated using only the peaks found to be important for classification in Table 4.5 compared against the best result found using all available peaks. Changes in κ score of less than or greater than 0.05 are highlighted in red and green, respectively. 111

4.7 Analysis of the peaks found within the PySR equations. A '-' entry indicates no peaks identified across that analysis. 125

5.1 Classification of AAA patient treatment groups. 130
5.2 The best results obtained for each patient classification using range of machine learning (ML) models and feature reduction (FR) methods. 134
5.3 The m/z ion peaks identified as being important for classification by the three FS methods for each pair of compared patient groups. . . . 138
5.4 Trends in tentatively assigned metabolite peaks compared to findings by M. Ciborowski *et al.* [324] for small AAA (S. AAA) and large AAA (L. AAA) compared to each other and to healthy volunteers. . . 139
5.5 The type and distribution of tissue samples used in this study. 142
5.6 Distribution of sample numbers within a typical partition. 144
5.7 Variables found to be significantly different ($p < 0.001$) between spectra grouped by tissue type. 148
5.8 Peaks found to be used by FR models with high occurrence ($> 75\%$). 153
5.9 The κ scores obtained when only the peaks found with high occurrence were used as input variables for ML for the tissue type classification. 153
5.10 The number of statistically significant peaks found for each clinical variable at four different significance levels. 158
5.11 The statistically significant peaks ($p < 0.05$) found to overlap in all three clinical variable classifications at different levels of extreme grouping for the growth rate variable. 159
5.12 The statistically significant peaks identified by Pearson's correlation coefficient that were used as inputs in the PySR analysis. 162

6.1 User reported advantages and disadvantages of the Advion CMS ASAP instrument 174
6.2 User reported advantages and disadvantages of the Waters Radian ASAP instrument 175
6.3 The instrument comparison metrics between the Radian and CMS-S instruments for a range of different tissue samples. 187

7.1 Normalisation, scaling and transformation methods 201
7.2 Best normalisation, scaling and transform methods for OXAMI variables assessed by cluster efficiency 210

7.3 The best-performing normalisation, scaling, and transformation methods for ML classification of OxAAA spectra into small aneurysm and HV categories. 213
7.4 The best performing normalisation, scaling and transform methods for OXAMI variables assessed by LDA ML classification accuracy . . 213

B.1 Hyperparameter optimisation for ML classification models. 259
B.2 The full ML results for the classification of OxAMI data by five ML classification algorithms. 264
B.3 The full ML results for the classification of OxAMI data with additional demographic information by five ML classification algorithms. . 264
B.4 The full ML results for the classification of OxAMI data by five ML classification algorithms with χ^2 feature reduction applied. 265
B.5 The full ML results for the classification of OxAMI data with included additional demographic variables by five ML classification algorithms with χ^2 feature reduction applied. 265
B.6 The full ML results for the classification of OxAMI data by five ML classification algorithms with OI feature reduction applied. 265
B.7 The full ML results for the classification of OxAMI data with additional demographic data by five ML classification algorithms with OI feature reduction applied. 266
B.8 The full ML results for the classification of OxAMI data by five ML classification algorithms with statistical feature reduction applied. . . 266
B.9 The full ML results for the classification of OxAMI data with additional demographic data included by five ML classification algorithms with statistical feature reduction applied. 266
B.10 The full ML results for the classification of 'Extreme' OxAMI data by five ML classification algorithms. 267
B.11 The full ML results for the classification of 'Extreme' OxAMI data with additional demographic data included by five ML classification algorithms. 268
B.12 The full ML results for the classification of 'Extreme' OxAMI data by five ML classification algorithms with χ^2 feature reduction applied. 268
B.13 The full ML results for the classification of 'Extreme' OxAMI data with additional demographic data included by five ML classification algorithms with χ^2 feature reduction applied. 269
B.14 The full ML results for the classification of 'Extreme' OxAMI data by five ML classification algorithms with OI feature reduction applied. 269

B.15 The full ML results for the classification of 'Extreme' OxAMI data with additional demographic data included by five ML classification algorithms with OI feature reduction applied. 270
B.16 The full ML results for the classification of 'Extreme' OxAMI data by five ML classification algorithms with statistical feature reduction applied. 270
B.17 The full ML results for the classification of 'Extreme' OxAMI data with additional demographic data included by five ML classification algorithms with statistical feature reduction applied. 271
B.18 The peaks identified using ASAP-QTOF-MS in the same m/z bins as for the CMS-S data for significant peaks identified as significant for OxAMI data classification. The error in ppm between the found and reference masses is shown in column 5. 273
B.19 The PySR generated equations that produced the most accurate prediction of each OxAMI clinical parameter. 279

C.1 ML accuracies for the classification of AAA patients based on clinical groups using all ASAP-MS variables available. 281
C.2 ML accuracies for the classification of AAA patients based on clinical groups using χ^2 FR selected ASAP-MS variables. 281
C.3 ML accuracies for the classification of AAA patients based on clinical groups using OI FR selected ASAP-MS variables. 282
C.4 ML accuracies for the classification of AAA patients based on clinical groups using stats FR selected ASAP-MS variables. 282
C.5 The significant m/z peaks identified for each AAA patient group classification and the relevant statistics. 284
C.6 Percentage of peaks identified as significantly different between groups. 285
C.7 Peaks found to be used by FR models with high occurrence ($> 75\%$). 288
C.8 The κ scores obtained when only the peaks found with high occurance were used as input variables for ML for the tissue type classification. 289
C.9 The ML results for the classification of AAA patients by clinical variables encoded into binary classifications. 303

D.1 Inter-centroid differences for converted data analysis of ASAP-MS instruments 315
D.2 The metrics calculated for the interconversion of plasma data recorded across the four ASAP instruments. 318
D.3 Neural network architecture optimisation hyperparameters that were used within the Bayesian optimisation method. 333

D.4 Neural network architecture optimisation hyperparameters that were used within the Bayesian optimisation method. Range values are input to the Bayesian optimisation as (lower limit ; upper limit) using the range of values defined. 334
D.5 Neural network conversion of data resolution results 335

E.1 Best normalisation, scaling and transform methods for OXAMI variables assessed by KNN ML classification accuracy 338
E.2 Best normalisation, scaling and transform methods for OXAMI variables assessed by SVM ML classification accuracy 339
E.3 Best normalisation, scaling and transform methods for OXAMI variables assessed by NBC ML classification accuracy 340
E.4 Best normalisation, scaling and transform methods for OXAMI variables assessed by RFC ML classification accuracy 341
E.5 Best normalisation, scaling and transform methods for OXAMI variables assessed by the ML classification accuracy, reporting the method that gave the highest accuracy across the five ML models. 342

List of Figures

1.1 The physiology and diagnostic signals of a STEMI. A STEMI is located in a coronary artery, typically the left anterior descending artery, as shown in Figure (a), and causes loss of blood to the myocardium of the left ventricle. The development of a STEMI is caused by the build-up and degredation of atherosclerotic plaques (b), triggering thrombus formation in the coronary arteries. A STEMI often causes a characteristic ST-segment elevation in the ECG trace (c) due to changes in the conductivity of hypoxic cells. Figure (a) was made with www.Biorender.com. Figure (b) was adapted from [13]. 5

1.2 Diagrams of an AAA. Left shows a schematic visual representation of an AAA on the abdominal aorta. Right shows a 2D CT scan showing an abdominal cross section of an AAA patient. The orange arrow shows a section of normal abdominal aorta, with the highlighted region showing the AAA. Image (a) was created with www.Biorender.com. Image (b) was taken from the open access radiology database www.Radiopaedia.org (case 36) [80, 81]. 10

1.3 An example of an ASAP ionisation source, based on the Advion CMS-S instrument setup [148]. In other instruments, the key components are the same, but exact values for pressure and orientations may differ. 16

1.4 The layout and voltages applied to a quadrupole mass filter. The four parallel rods are configured in parallel (a) such that opposite rods have the same RF and DC voltages applied (b). When ions of different m/z travel through the quadrupole, their trajectories are affected by the electric fields such that only ions of a certain m/z have a stable trajectory (green) and reach the detector (c). Adapted from [172]. 20

2.1 A schematic of the Advion CMS-S ASAP-MS instrument, see text for description. Adapted from the Advion CMS User Manual [161], and [180]. 26

2.2 The study design and workflow for the OxAMI project when initiated in 2011. The coronary aspirate samples were collected at the time of PCI. Image replicated with permission from The OxAMI study webpage [199]. 30
2.3 The study design and workflow for the OxAAA project, showing the healthy volunteer (HV, green), surveillance stream (S1, yellow), and surgical streams (S2, red). Note that S1 patients will move to S2 once the AAA size has reached the surgical threshold. 32
2.4 A schematic showing the types of biopsy samples that were taken during AAA repair surgery from the area in and around the AAA. The diagram represents a cross-section of the abdominal aorta within the abdominal cavity, with the proximal end (towards the heart) on the left, and the distal end (towards the legs) on the right. Sampling locations are indicated with (*). 34
2.5 An example of an ideal chromatogram for a plasma sample. The red crosses represent each individual time point in the data acquisition where a capillary measurement was taking place. The spectrum at each time point is summed to generate the TIC signal for that time point. 38
2.6 Examples of non-optimal chromatograms for OxAAA plasma samples arising as a result of various measurement errors. The resulting problems are highlighted by red circles: (a) Highly variable background and noise arising when the ion source needs to be cleaned; (b) Signal saturation due to introduction of too much sample (no sharp peak observed, and peak intensity remains high as long as the probe remains in the spectrometer); (c) Variable peak profiles and intensities due to inconsistent sampling and introduction of too much sample to the spectrometer; (d) Inconsistent acquisition times and insufficient time between measurements for the total ion count to return to background levels. Data recorded in 2020 by E. Ngetich. 39
2.7 Mass spectra recorded for a human plasma sample using the eight different combinations of ion source settings listed in Table 2.3. The spectra were recorded with a scan time of 500 ms and an acquisition time 20 s. The spectra shown are an average over five repeats, each averaged over the 40 time-resolved spectra recorded during each acquisition time. ‘LT’, ‘MT’, and ‘HT’ refer to low, medium, and high temperature, respectively, while ‘LF’, ‘MF’, and ‘HF’ refer to low, medium, and high fragmentation. 42

2.8 The average signal-to-noise ratio (blue) and number of peaks identified (red) for scan times in the range from 200 to 1000 ms. Data were averaged over five repeat measurements, with the exception of the 800 ms data, which were averaged over four repeats due to an anomalous measurement. One-standard deviation error bars in the signal-to-noise are shown. 44
2.9 Allan variance as a function of acquisition time t for two plasma samples, (a) and (b). Three repeat measurements for each sample are shown by the blue, yellow, and red lines. 46
2.10 The change in average normalised intensity (ΔI) across all m/z peaks as a function of number of repeats averaged (N). One-standard-deviation error bars are shown. 47
2.11 (a) Chromatograms for a set of ten repeat measurements (five each with two different capillaries) on an OxAAA sample recorded using the optimised method; (b) Individual mass spectra recorded with the second capillary; (c) Mean (black dashed line) and standard deviation (grey solid region) of the coefficient of variance in m/z peak intensities, determined as described in the text. The average coefficient of variance across the entire spectrum is 39.9%, shown as a red dashed line on the plot. 50
2.12 Images of selected OxAAA tissue samples for the wall, thrombus, and fat samples used. Images were taken with a standard USB camera. Samples were left in the standard sample tubes. The base of each tube is approx 10 mm in diameter. 54
2.13 The absolute change in spectra intensity with number of repeats for homogenised and unhomogenised tissue for a) abluminal thrombus, b) luminal thrombus, c) distal wall, d) maximal wall, e) preperitoneal fat, and f) subcutaneous fat. 55
2.14 The scan time optimisation plots for a) thrombus, b) artery wall, c) fat, showing signal-to-noise ratio against peak count. Data is averaged for each tissue type over patients 71, 73 and 77. Error bars show the signal-to-noise sample error. 56
2.15 Allan variance analysis of (a) abluminal thrombus, and (b) maximal wall. Each colour represents one of 5 individual repeat measurements. Note the distinctive minimum between 20 and 30 seconds in (b), which is not present in (a). 58

2.16 Analysis of the change in ΔI as the number of repeats averaged over increases for (a) thrombus, (b) wall, (c) and fat tissues. Each coloured line represents a tissue samples analysed. 59
2.17 The CV% calculated for each tissue subtype calculated over 20-30 patients. The plasma CV% calculated previously is shown in orange. 60

3.1 A flowchart showing the data processing steps developed for the analysis of ASAP-MS data in preparation for use in ML or statistical analysis in the format of a single spectrum per patient. The green trapezia show the size of the data after each step. Two commonly used end points were to use all the repeats for a sample measurement (red endpoint), or to average over the repeats to generate a single spectrum per sample (yellow endpoint). 63
3.2 A plot of the total ion count (integrated over m/z) against time (s) recorded during the analysis of a blood plasma sample. The five repeat acquisitions can be observed as five regions of raised intensity, each lasting approximately 25 seconds. Each red cross is a time point has been determined to be part of a sample measurement through the method described in section 3.1.1. 64
3.3 A visualisation of the transformation of a data set into 2d PC space. The variance in the data set along the x axis in the original data set is significantly less than the variance in PC1, allowing for PC1 to be used to describe a difference between the yellow circles and blue squares. 69
3.4 A visualisation of the 2D classification of an uncategorised data paint (green star) into blue or orange categories using the KNN algorithm with a $K = 3$ limit, leading to a blue label, and a $K = 6$ limit, leading to an orange label. 72
3.5 A simplified visualisation of an SVM hyperplane separating the blue and orange classes. 73
3.6 A simplified visualisation of an LDA projection of the data on the left to linear discriminant space on the right. 75
3.7 An example of a random forest algorithm consisting of three very small decision trees, for the classification of patients for cardiovascular disease risk. Two trees vote for high risk, whilst one tree votes for low risk, therefore, the patient is classified as high risk. 77

3.8 An example of a symbolic regression nodal syntax tree representation of an equation. The outcome y is related to a data set with variables a, b, c via an unknown function. A function that uses variables a and c has been determined through the algorithm. 78
3.9 The overlap integral feature selection method shown for an m/z peak at 46. The violin distribution (a) shows a clear difference between the distribution of class 0 and class 1, which is reflected in the low overlap integral value and low overlap area in (b) with respect to the non-overlapping areas of the orange and blue distributions. 83
3.10 A confusion matrix, showing a high accuracy prediction outcome for a model. Note the dark blue of the TP and TN boxes showing the high proportion of values in these quadrants, compared to the light orange hue, which shows that there are few false classifications. The accuracy of this model is 0.75%, and κ score is 0.49. 85
3.11 The effect of increasing partition number on the final κ obtained for patient mortality (upper) and peak troponin level (lower) by KNN analysis. The κ obtained after each new partition is generated is shown in a. and c., and the first derivative $\frac{\delta\kappa}{\delta n}$ from the previous score is shown in b. and d. 89

4.1 Histogram plots of the patient population distributions across each of the clinical parameters of interest: (a) MVO measured six months after the pPCI; (b) IMR measured during the pPCI; (c) TScore; (d) ITime prior to the pPCI; (e) peak troponin level measured prior to pPCI; (f) creatinine level. 97
4.2 Box and whisker plots showing the distribution of normalised peak intensities within the 'All patients' data set (except for MVO, which used the MVO-20 data set) for m/z peaks identified by the FR analysis as being of interest for each patient group and clinical parameter investigated: (a) patient mortality; (b) heart failure diagnosis (see Figure 4.3 for the same data on a magnified scale); (c) MVO (MVO-20 data set); (d) IMR; (e) TScore; (f) ischemic time; (g) troponin; (h) creatinine. In each plot, the central line represents the data median, with the inner box representing the inter-quartile range between the first and third quartiles. The whiskers extend from the box to the farthest data point lying within 1.5× the inter-quartile range from the box. Significance was measured by an independent t-test. 109

4.3 Box and whisker plots showing the distribution of normalised peak intensities within the 'All patients' data set for m/z peaks identified by the FR analysis as being of interest for the heart failure diagnosis clinical parameter at a magnified scale. 111
4.4 The relationship between PySR equation complexity and accuracy by r^2 analysis applied to both the training (blue) and test (red) data. The results are shown for the prediction of MVO without feature reduction (top), IMR with statistical feature (middle), and Creatinine with statistical feature reduction (lower), showing best generalisation at low, intermediate, and high complexity equations, respectively. . . 118
4.5 The accuracy obtained for each optimum PySR output, run over the nine clinical variables tested. The accuracy is shown as the r^2 fit between the predicted and true values for each variable, with the lower of either the test or training fit shown. The colours represent the data input method, with no feature reduction shown in green, statistical feature reduction in blue, and PCA feature reduction in red. The colour hues show the repeat results over multiple partitions. 121
4.6 The PySR predicted values for AIS (a) and FIS(b) plotted against the true measured clinical values. The training data is shown in blue, and the test data in red. Both variables are reported as values with units LV%. The training and test values are generated by applying the PySR output equation to the input data partitioned the same as for input to PySR. 122
4.7 The correlation between measured AIS and measured FIS as reported in the clinical dataset. Both AIS and FIS are reported as LV%. Note the slight trend towards higher AIS values, showing that some patients recover more cardiac function that expected. 123
4.8 The PySR predicted values for the highest accuracy prediction of MVO (a) plotted against the true measured clinical values. Removal of the outlier in plot (a) results in plot (b), showing a clear bias in the test data to low MVO values. Plot (c) shows a repeat analysis conducted with a more evenly distributed partition. The training data is shown in blue, and the test data in red. MVO is reported with units LV%. 124

5.1 132

5.2 The averaged spectra for patient groups: (a) healthy volunteer, (b) small AAA, (c) large AAA, and (d) post-surgery. Spectra are displayed as normalised intensity plotted against m/z. Averaged spectra are calculated as the average of all patient spectra in each class. . . . 132

5.3 Difference spectra for patient classes showing only the peaks that show a significant ($p < 0.001$) difference between patient classes by an independent t-test. Spectra shown are S1-TPS1 - HV for small AAA - healthy volunteer patients (a), S2-TP3 - HV for large AAA - healthy volunteer patients (b), S2-TP3 - S1-TPS1 for large AAA - small AAA patients (c), S2-TP5 - HV for post-surgery - healthy volunteer patients (d), S1-TPS1 - S2-TP5 for small AAA patients - post-surgery (e), and S2-TP3 - S2-TP5 for large AAA - post-surgery patients (f). 133

5.4 The κ obtained from the ML classification analysis of six different OxAAA group comparisons. Verrtical bars show the κ obtained for each ML model for that clinical variable, horizontal coloured bars show the level of agreement with *no agreement* in grey, *low agreement* in purple, *fair agreement* in light red, *moderate agreement* in yellow, and *strong agreement* in green. Results are shown for when all peaks were used (a), the FR FS method (b), the OI FS method (c), and stats FS method (d). 135

5.5 Boxplots showing the distribution of normalised peak intensities for peaks found to be important in over 75% of models for the classification of each of the classifications of AAA patient groups: small AAA and HV (a), large AAA and HV (b), post surgery and small AAA (d), post surgery and large AAA (e), and post surgery and HV (f). The peaks in (c) for large AAA vs small AAA were identified with a lower threshold, in over 50% of models. The central line represents the data median, with the inner box representing the inter-quartile range between the first and third quartiles. The whiskers extend from the box to the farthest data point lying within 1.5x the inter-quartile range from the box. All peaks were statistically different at $p < 0.001$. 140

5.6 Comparison of the spectra for the averaged tissue types, and a heatmap representation of each repeat for each tissue type. The mass spectra on the left show the normalised intensity averaged over each repeat and each sample for each tissue type: fat (a), wall (c), and thrombus (e). The heatmaps shown the normalised intensity for each repeat measurement of each sample, plotted as m/z value (x), sample and repeat number (y), and measured intensity (shown by colour bar). The samples are plotted by type, showing fat (b), wall (d), and thrombus (f). 147

5.7 PCA analysis of tissue sample spectra. PCA cluster plots are shown on the left, with the first three principal components plotted, and the data coloured by class. The figures on the right show the distribution of data variance with principal components. By row, the analysis shown is for fat vs thrombus (a and b), fat vs wall data (c and d), and wall vs thrombus data (e and f). PCA clusters and variance for tissue type analysis. 149

5.8 The κ scores achieved for each analysis of the tissue type classification by combinations of five ML models and seven FR methods applied to the mass spectra, averaged for each sample. The errors shown are the standard deviation in κ score achieved across each partition. The colourmap indicates the accuracy and level of agreement of the results by κ score. 151

5.9 The κ scores achieved in ML classification of tissue by tissue type. Bars show the κ score from the average ML algorithm results, error bars represent standard deviations. Results are plotted as the mean κ score achieved across all ML algorithm results (yellow), or the maximum κ score achieved with a single ML algorithm and FR method (red). The analysis was repeated using peaks found to be significant by the FR analysis, and the maximum κ score determined for the data shown (blue). 152

5.10 Histogram plots of the patient population distributions across each of the clinical parameters of interest: (a) AAA size in mm; (b) AAA growth rate as determined by ultrasound measurement in mm year^{-1}; and (c) FMD score baseline measurement %. The median value for each distribution is shown by the red dotted line. 155

5.11 The significant peaks at the lowest confidence interval used ($p<0.05$) found in across three different patient group comparisons: pre vs post surgery (green), mental fat vs abluminal thrombus (blue), and slow vs fast growth (red), where the slow vs fast growth analysis has been conducted on extreme grouped patients at $n = 10$ per class (a), $n = 20$ per class (b), and $n = 30$ per class (c). 160
5.12 The change in normalised intensity with AAA growth rate for the most significant correlation found ($p<0.05$) at $m/z = 319$ (a), and the most significant correlation found from the overlap analysis peaks shown in Table 5.11 plotted against FMD (b). The correlation plot in (c) is taken from [201] and shows the correlation of the molecule attractin with AAA growth rate. Note the poor correlation in all three cases. 161
5.13 The accuracy obtained for each PySR output for the three clinical variables: FMD, Growth rate, and AAA size. The accuracy is shown as the r^2 fit between the predicted and true values for each variable, with the lower of either the test or training fit shown. Colours represent the data input method, with no feature reduction (green), statistical feature reduction (blue), and PCA feature reduction (red). Colour hues show the repeat results over multiple partitions. 164
5.14 The fit of Equation 5.2 for FMD using the stats approach and the first partition. 164
5.15 Plots of the fit of the best two equations generated using PySR for FMD (upper), growth rate (centre), and AAA size (lower). Training data is shown in blue and test data in red. The partition version is represented by $-n$ at the top of each plot. 166

6.1 Analysis of scan speed averaged across the tissue samples for patient 77. Average signal to noise is plotted in blue on the left y axis, and number of peaks identified above the threshold intensity is plotted in red on the right y axis. 177
6.2 Allan variance plotted against τ in seconds. The plots shown are for abluminal thrombus, maximal wall, and preperitoneal fat from left to right, and for the Advion CMS-S and Radian instruments in the upper and lower plots respectively. 178

6.3 Analysis of the number of repeat measurements averaged to generate a patient spectrum. A set of 8 tissue samples were taken from patient 77. These were run for 50 repeat acquisitions on both the (a) Radian and (b) CMS-S instruments. The figures show the error in the mean change, error bars show the standard deviation in ΔI. 179
6.4 Averaged spectra of human blood plasma recorded at increasing source gas temperature from 300-600 $°C$ using the Waters Radian ASAP-MS. The spectra are normalised to unit area under the curve. 182
6.5 Averaged spectra for human blood plasma recorded at increasing corona discharge currents from 2 – 6 μA using the Waters Radian ASAP-MS. The spectra are normalised to unit area under the curve. . 183
6.6 Mass spectra for Patient 77: thrombus recorded on the Waters Radian (left, a,c) and the Advion CMS-S (right, b,d) representing the different thrombus tissue subtypes of abluminal (top), and luminal (lower) thrombus. 186
6.7 The Pearson's Correlation matrix (left) and Cosine angle matrix (right) showing inter- and intra-instrument sample variation for Sample 1, recorded on the four different ASAP-MS instruments. Darker colours represent lower correlation values or larger cosine angles plotted as 1-cosine angle. The x and y axes are split into the four instrument datasets. Each instrument dataset contains 10 repeat measurements, each represented by a pixel square in the diagram. 189
6.8 The CV values for each instrument (CMS-S in (a), CMS-L in (b), Radian in (c), QTOF in (d)) averaged over the recorded plasma data. CV is shown for each m/z value as a spectrum in black, with the standard deviation in CV added as grey shading to $\pm 1\sigma$. The average CV over the spectrum is shown as the red dashed line. 190
6.9 Plots of instrument spectra transformed using PCA, showing the 2D and 3D projections of PCs to determine clustering behaviour for the conversion of plasma data from one instrument (B) to another instrument (A). Plots show 2D projections of CMS-L to CMS-S converted data in (a), and Radian to CMS-S converted data in (b), and 3D projections of the same conversions in (c) and (d), respectively. Data shown is the training data recorded on instrument A (blue), training data recorded on instrument B (yellow), test data recorded on instrument A (green), test data recorded on instrument A (magenta), and the test data recorded on instrument B converted to instrument A (cyan). 192

6.10 Plots show the instrument functions plotted as mass spectra. The CMS-L to CMS-S instrument function in is shown in (a), Radian to CMS-S instrument function in (b). Note the significant mass dependence of the function, which will be a result of differences in ionisation, fragmentation, and transmission of ions through the quadrupole. . . . 193

7.1 Identified sources of error that may occur in the ASAP-MS measurement of human blood samples, grouped into biological sample-related error (orange), sample loading-related error (blue), and ASAP instrument-related error (grey). 196

7.2 An illustration of the impact of AUC normalisation on the apparent and relative intensities of three mass peaks. The raw intensities are shown for three peaks (A, B and C) in two spectra at the top, and after the application of AUC normalisation underneath. The raw intensity of peak C changes drastically. Note the resulting change in normalised intensity of peaks A and B, even though no change in raw intensity was recorded. 198

7.3 A diagram showing the histogram normalisation function, showing the distribution of histogram bins. Bin segments (B) are shown separated by the red lines at locations $P\%$, and the number of bins in each segment, n, shown by the coloured bars. 205

7.4 A 2d graphical representation of the cluster ratio analysis. Samples from two classes: class 0 in blue, and class 1 in orange. The two cluster centroids are shown by x_0 and x_1 respectively. Data is shown with respect to two axis, a and b, which represent two m/z values, but in reality the dimensional space will be much larger. 208

7.5 Comparison of normalisation, scaling and transform methods for OXAMI clinical variables evaluated by cluster efficiency. Cluster efficiency when no normalisation, scaling and transformation is used is shown in blue, where AUC normalisation with no other scaling or transformations is shown in orange, and the best normalisation, scaling and transformation combined method is shown in grey 210

7.6 The cluster size to centroid ratio values for the OxAAA data clustered into HV and S1 patient groups. The colourmap represents the C_R value from low (blue) to red (high). 211

7.7 Comparison of normalisation, scaling and transform methods for AAA S1 vs HV classification evaluated by κ score. The ML clustering κ score using the five different ML models when no normalisation, scaling and transformation is used is shown in blue, where AUC normalisation with no other scaling or transformations is shown in orange, and the best normalisation, scaling and transformation combined method is shown in grey. Error bars show the standard deviation in κ over 50 partitions. 214
7.8 Comparison of normalisation, scaling and transform methods for OXAMI clinical variables evaluated by κ score. The ML clustering κ score using an LDA model when no normalisation, scaling and transformation is used is shown in blue, where AUC normalisation with no other scaling or transformations is shown in orange, and the best normalisation, scaling and transformation combined method is shown in grey. Error bars show the standard deviation in κ over 50 partitions. 215
7.9 Comparison of normalisation evaluation methods for AAA S1 and HV patient classification (a) and patient mortality (b). The cluster ratio is plotted against ML accuracy as κ score. No correlation was observed in either case. 216

A.1 The Allan variance analysis of three tissue types: abluminal thrombus (upper), maximal wall (centre), and preperitoneal fat (lower) across four patients. Each colour represents one of five individual repeat measurements. 256
A.2 The change in ΔI_N as a function of number of repeats averaged for three differrent tissue subtypes: abluminal thrombus (top), distal wall (centre), and subcutaneous fat (lower) across three different patients: patient 00071 (left), patient 00073 (centre) and patient 00077 (right). 257

B.1 The κ scores obtained from analysing 'All patients' data set with: (a) no feature reduction; (b) feature reduction with overlap integral method; (c) feature reduction with statistical method; and (d) feature reduction by χ^2 method. Bars show the κ score obtained for each ML model, shading shows the level of agreement, with *no agreement* in grey, *low agreement* in purple, *fair agreement* in light red, *moderate agreement* in yellow, and *strong agreement* in green. 260

B.2 The κ scores obtained from analysing 'All patients' data set with the inclusion of demographic data, and: (a) no feature reduction; (b) feature reduction with overlap integral method; (c) feature reduction with statistical method; and (d) feature reduction by χ^2 method. Bars show the κ score obtained for each ML model, shading shows the level of agreement, with *no agreement* in grey, *low agreement* in purple, *fair agreement* in light red, *moderate agreement* in yellow, and *strong agreement* in green. 261

B.3 The κ scores obtained from analysing 'Extreme patients' data set with: (a) no feature reduction; (b) feature reduction with overlap integral method; (c) feature reduction with statistical method; and (d) feature reduction by χ^2 method. Bars show the κ score obtained for each ML model, shading shows the level of agreement, with *no agreement* in grey, *low agreement* in purple, *fair agreement* in light red, *moderate agreement* in yellow, and *strong agreement* in green. . . 262

B.4 The κ scores obtained from analysing 'Extreme patients' data sets with inclusion of clinical data and: (a) no feature reduction; (b) feature reduction with overlap integral method; (c) feature reduction with statistical method; and (d) feature reduction by χ^2 method. Bars show the κ score obtained for each ML model , shading shows the level of agreement, with *no agreement* in grey, *low agreement* in purple, *fair agreement* in light red, *moderate agreement* in yellow, and *strong agreement* in green. 263

B.5 Assessment of the impact of increasing `n_cycles` on the accuracy of PySR models for predicting IMR, Troponin and MVO. Troponin was not assessed above 1000 cycles as the troponin dataset was much larger resulting in prohibitively long run times. 277

B.6 Assessment of the impact of increasing `n_iterations` on the accuracy of PySR models for predicting IMR, Troponin and MVO. 278

B.7 Assessment of the impact of increasing `population_size` on the accuracy of PySR models for predicting IMR, Troponin and MVO. . . . 278

C.1 The κ scores achieved for each analysis of the tissue subtype classification by combinations of five ML models and seven FR methods applied to the mass spectra, with the spectra averaged across each sample. The errors shown are the standard deviation in κ score achieved across each partition. The colourmap indicates the accuracy and level of agreement of the results by κ score. Tissue subtypes are abbreviated to Subcut. (subcutaneous), Om. (Omental), Prep. (Preperitoneal), Lum. (Luminal), Ablum. (Abluminal), Dist. (Distal), Max. (Maximal), and Prox. (Proximal). 287

C.2 The κ scores achieved for each analysis of the tissue type classification by combinations of five ML models and seven FR methods applied to the mass spectra, with all repeats of each sample used. The errors shown are the standard deviation in κ score achieved across each partition. The colourmap indicates the accuracy and level of agreement of the results by κ score. 291

C.3 The κ scores achieved for each analysis of the tissue subtype classification by combinations of five ML models and seven FR methods applied to the mass spectra, with all repeats of each sample used. The errors shown are the standard deviation in κ score achieved across each partition. The colourmap indicates the accuracy and level of agreement of the results by κ score. Tissue subtypes are abbreviated to Subcut. (subcutaneous), Om. (Omental), Prep. (Preperitoneal), Lum. (Luminal), Ablum. (Abluminal), Dist. (Distal), Max. (Maximal), and Prox. (Proximal). 292

C.4 The κ scores achieved in ML classification of tissue by tissue type. Bars show the κ score from the average ML algorithm results, error bars represent standard deviations. Results are shown for analysis where repeats were averaged to one spectrum per patient (yellow), or repeats were used as separate samples (red) are plotted as the mean and standard deviations of all ML algorithm results respectively. The best κ score achieved with a single ML algorithm and feature reduction method was plotted for both the repeat averaged (purple) and repeats used (green). Where appropriate, the analysis was repeated using peaks found to be significant by the FR analysis, and the maximum κ score achieved for the data shown (blue) 293

C.5 The κ scores achieved in ML classification of tissue by tissue subtype. Bars show the κ score from the average ML algorithm results, error bars represent standard deviations. Results are shown for analysis where repeats were averaged to one spectrum per patient (yellow), or repeats were used as separate samples (red) are plotted as the mean and standard deviations of all ML algorithm results respectively. The best κ score achieved with a single ML algorithm and feature reduction method was plotted for both the repeat averaged (purple) and repeats used (green). Where appropriate, the analysis was repeated using peaks found to be significant by the FR analysis, and the maximum κ score achieved for the data shown (blue). 294

C.6 The κ scores achieved for changes in n repeats with all patients used for wall and thrombus. Error bars show the standard deviation of κ score over the partitions used. The first group of bars, labelled 'Avg', represents the data obtained when all 10 repeats are averaged over. . 297

C.7 The % accuracy obtained by the best ML model for the tissue type analysis conducted whilst varying the number of repeats used and the number of patients used. The number of repeat measurements used for each patients is represented by the changing colour of the plotted points. Plots on the right (a, c, e) show the change in accuracy as the number of patients used changes. Plots on the left (b, d, f) show how the accuracy changes as a function of absolute training data size. The % accuracy shown is the best accuracy obtained across the five ML models used for each n repeats and m patients combination. . . 298

C.8 The % accuracy obtained by the best ML model for the tissue subtype analysis conducted whilst varying the number of repeats used and the number of patients used. The number of repeat measurements used for each patients is represented by the changing colour of the plotted points. Plots on the right (a, c, e) show the change in accuracy as the number of patients used changes. Plots on the left (b, d, f) show how the accuracy changes as a function of absolute training data size. The % accuracy shown is the best accuracy obtained across the five ML models used for each n repeats and m patients combination. . . 300

C.9 The effect of changing the hold-n-out cross validation number on the κ score obtained for four different subtype classifications: subcutaneous vs omental fat (a), subcutaneous vs preperitoneal fat (b), preperitoneal vs omental fat (c), and abluminal vs luminal thrombus (d), using the Chi^2 FR method. 301

C.10 The highest ML accuracy κ scores across the five ML algorithms used for the classification of patients by measured AAA clinical variables FMD, AAA growth and AAA size. Error bars show the standard deviation in κ achieved across the 100 partitions. 302
C.11 The highest ML accuracy κ scores across the five ML algorithms used for the classification of patients by measured AAA size using four different FR algorithms: stats, OI, and χ^2 FR. Error bars show the standard deviation in κ achieved across the 100 partitions. 302
C.12 The significant peaks at the lowest confidence interval used ($p<0.05$) found in across four different patient group comparisons: slow vs fast growth (blue), omental fat vs luminal thrombus (red), pre vs post surgery (green), and large vs small aneurysm (cyan), where the slow vs fast growth analysis has been conducted on extreme grouped patients at $n = 10$ per class (a), $n = 20$ per class (b), and $n = 30$ per class (c). 304

D.1 Analysis of scan speed for multiple tissue samples for patient 77 across the Radian and Advion CMS-S instruments. The data is highlighted in green for scan speed 0.9 s, which was the chosen optimal scan speed. 306
D.2 The averaged plasma mass spectrum recorded from 3 – 21 V cone voltage on the Waters Radian instrument. The spectra are normalised to unit area under the curve. 307
D.3 Mass spectra for Patient 77: artery wall recorded on the Waters Radian (left, a,c,e) and the Advion CMS-S (right, b,d,f) representing the different artery wall tissue subtypes of proximal (top), maximal (centre) and distal (lower) wall. Peaks at m/z=42 and 369 have been removed. 308
D.4 Mass spectra for Patient 77: fat tissue recorded on the Waters Radian (left, a,c,e) and the Advion CMS-S (right, b,d,f) representing the different fat tissue subtypes of omental (top), subcutaneous (centre) and preperitoneal (lower) fat. Peaks at m/z=42 and 369 have been removed. 309
D.5 Mass spectra for Patient 77: thrombus recorded on the Waters Radian (left, a,c,e) and the Advion CMS-S (right, b,d,f) representing the different thrombus tissue subtypes of abluminal (top), and luminal (lower) thrombus. Peaks at m/z=42 and 369 have been removed. . . 310

D.6 Bar charts showing the similarity of spectra for tissue samples recorded on the Radian and CMS-S instruments. Values shown are the average over all repeat comparisons, with the standard deviation shown as error bars. The bar colours represent the six different patient IDs for each sample type. Metrics calculated are percentage difference (a), Pearson's r (b), and Cosine angle (c). 311
D.7 Pearson's correlation (a) and Cosine angle (b) inter-instrument analysis for plasma sample 2. 312
D.8 Pearson's correlation (a) and Cosine angle (b) inter-instrument analysis for plasma sample 3. 312
D.9 Pearson's correlation (left) and Cosine angle (right) inter-instrument analysis for plasma sample 4. 313
D.10 Pearson's correlation (left) and Cosine angle (right) inter-instrument analysis for plasma sample 5. 313
D.11 The effect of increased Gaussian smoothing on the conversion of CMS-L data to CMS-S data. The plots show instrument spectra transformed using PCA, showing the 2D projections of PCs to determine clustering behaviour for the interconversion of plasma data. Training data recorded on instrument A is shown in blue, training data recorded on instrument B is shown in yellow, test data recorded on instrument A is shown in green, test data recorded on instrument A is shown in magenta, and the test data recorded on instrument B converted to instrument A is shown in cyan. The converted data is highlighted with a cyan ellipse. 317
D.12 Plots of instrument spectra transformed using PCA, plotted as PC1 and PC2 showing the interconversion of plasma data from one instrument (B) to another instrument (A). Plots show CMS-L to CMS-S converted data in (a), Radian to CMS-S converted data in (b), Radian to CMS-L converted data in (c), QTOF to CMS-S converted data in (d), QTOF to CMS-L converted data in (e), and QTOF to Radian converted data in (f). Training data recorded on instrument A is shown in blue, training data recorded on instrument B is shown in yellow, test data recorded on instrument A is shown in green, test data recorded on instrument A is shown in magenta, and the test data recorded on instrument B converted to instrument A is shown in cyan. 319

D.13 Plots of instrument spectra transformed using PCA, plotted as PC1, PC2 and PC3 showing the interconversion of plasma data from one instrument (B) to another instrument (A). Plots show CMS-L to CMS-S converted data in (a), Radian to CMS-S converted data in (b), Radian to CMS-L converted data in (c), QTOF to CMS-S converted data in (d), QTOF to CMS-L converted data in (e), and QTOF to Radian converted data in (f). Training data recorded on instrument A is shown in blue, training data recorded on instrument B is shown in yellow, test data recorded on instrument A is shown in green, test data recorded on instrument A is shown in magenta, and the test data recorded on instrument B converted to instrument A is shown in cyan. 320

D.14 Plots show the raw instrument functions plotted as mass spectra. The CMS-L to CMS-S instrument function in is shown in (a), Radian to CMS-S instrument function in (b), Radian to CMS-L instrument function in (c), QTOF to CMS-S instrument function in (d), QTOF to CMS-L instrument function in (e), and QTOF to Radian instrument function in (f). Note the significant peak dependence of the functions. 321

D.15 Gaussian blurr applied to the instrument functions using the smoothing parameters defined in Table D.1. The CMS-L to CMS-S instrument function in is shown in (a), Radian to CMS-S instrument function in (b), Radian to CMS-L instrument function in (c), QTOF to CMS-S instrument function in (d), QTOF to CMS-L instrument function in (e), and QTOF to Radian instrument function in (f). . . . 322

D.16 Savitzky-Golay filter applied to the instrument functions using the smoothing parameters defined in Table D.1. The CMS-L to CMS-S instrument function in is shown in (a), Radian to CMS-S instrument function in (b), Radian to CMS-L instrument function in (c), QTOF to CMS-S instrument function in (d), QTOF to CMS-L instrument function in (e), and QTOF to Radian instrument function in (f). . . . 323

D.17 The data processing protocol for the ASAP-QTOF-MS data from the raw data output of the Waters Xevo G2 XS instrument into a single spectrum per sample that can be used in further analysis. 326

D.18 The TOF reference spectrum generator, showing a Gaussian peak example. The peak in (a) shows the full-width half maximum (FWHM) of a typical peak with a peak maximum shown at 257.2475 m/z. The spectrum in (b) shows a zoomed-in view of the red box in (a), showing the spectral resolution signal at each m/z instance, showing how the peak width is spread over the mass range. 328
D.19 The percentage error in normalised peak intensity calculated for plasma spectra as measured using the ASAP-QTOF-MS (a), and Advion CMS-S instrument (b). The spectra have been matched to the reference spectrum, cleaned to only the reference peaks, and normalised. The high m/z in the CMS-S spectrum is due to the lack of high-intensity peaks at these values. 331

E.1 Comparison of normalisation, scaling and transform methods for clinical variables evaluated by KNN classification accuracy κ score. Classification accuracy is shown with no normalisation, scaling and transformation (blue), with AUC normalisation with no other methods (orange), and the best normalisation, scaling and transformation combined method (grey). Errors show the standard deviation in κ over partitions. 337
E.2 Comparison of normalisation, scaling and transform methods for clinical variables evaluated by SVM classification accuracy κ score. Classification accuracy is shown with no normalisation, scaling and transformation (blue), with AUC normalisation with no other methods (orange), and the best normalisation, scaling and transformation combined method (grey). Errors show the standard deviation in κ over partitions. 338
E.3 Comparison of normalisation, scaling and transform methods for clinical variables evaluated by NBC classification accuracy κ score. Classification accuracy is shown with no normalisation, scaling and transformation (blue), with AUC normalisation with no other methods (orange), and the best normalisation, scaling and transformation combined method (grey).Errors show the standard deviation in κ over partitions. 339

E.4 Comparison of normalisation, scaling and transform methods for clinical variables evaluated by RFC classification accuracy κ score. Classification accuracy is shown with no normalisation, scaling and transformation (blue), with AUC normalisation with no other methods (orange), and the best normalisation, scaling and transformation combined method (grey). Errors show the standard deviation in κ over partitions. 340

E.5 Comparison of normalisation, scaling and transform methods for clinical variables evaluated by the best ML model across the five models used, shown as the classification accuracy κ score. Classification accuracy is shown with no normalisation, scaling and transformation (blue), with AUC normalisation with no other methods (orange), and the best normalisation, scaling and transformation combined method (grey). Errors show the standard deviation in κ over partitions. . . . 341

Abstract

Cardiovascular diseases create a large burden on global healthcare systems. Treating patients with cardiovascular diseases is challenging, as not all patients respond well to the same treatments. Identifying patients who need more intensive treatments, surgery, or monitoring is important in order to enable more personalised healthcare plans. Personalised healthcare, or tailoring medical treatment plans to the specific needs of a patient, has the potential to improve the lives of many patients, with the additional benefit of reducing the burden on healthcare systems through reducing readmission rates. Tools to allow this sort of care need to be developed.

The atmospheric solids analysis probe mass spectrometer (ASAP-MS) has been investigated for its potential for use as a clinical tool for predicting patient outcomes based on biological sample measurements. We have investigated this in the context of two cardiovascular diseases: ST elevated myocardial infarctions (STEMI), and abdominal aortic aneurysms (AAA).

We have developed an optimised protocol for the measurement of human blood plasma and tissue samples with ASAP-MS, discussed in Chapter 2. Using this optimised method, we can measure a metabolic fingerprint of biological samples. These fingerprints are complex to interpret, but can be mapped to clinical measurements of interest using machine learning methods, which are good at pattern recognition within complex data.

Using ASAP-MS measurements of blood plasma samples taken soon after STEMI patients first presented at the hospital, we were able to predict prognostically relevant information using machine learning methods. Highlights include the prediction of patient mortality with over 85% accuracy, and microvascular obstruction with over 83% accuracy. In the AAA study, patient groups with AAAs of different stages

of development were classified with extremely high accuracy, at over 95%. AAA size was predicted with high accuracy at 93%, and indicators related to AAA growth were also found. Using both feature reduction methods and symbolic regression analysis, we have been able to start to identify the underlying relationships between specific metabolite peaks and clinical outcomes.

In the final chapters of this thesis, other aspects of the ASAP-MS method were investigated, including a comparison of the commercial instrumentation available and assessment of the impact of normalisation methods on ASAP-MS data.

Through this investigation, we have provided a proof-of-concept study showing that combining ASAP-MS and machine learning techniques can provide clinically useful information for the rapid prediction of disease outcomes. This provides further evidence to support the growing use of techniques based on ambient ionisation mass spectrometry in clinical environments.

Acknowledgements

Firstly, I would like to say a huge thank you to my supervisor Professor Claire Vallance. I have learned so much from you Claire, your positivity and relentless energy is infectious and has made doing this DPhil a truly enjoyable experience. I will really miss working with you. Thank you so much for putting up with all my spelling mistakes, for always knowing what to say, and for having a million brilliant ideas of things to try if we ever get stuck.

I have been very privileged to have the opportunity to work with two clinical studies, the OxAMI study and the OxAAA study. The clinicians and staff on these projects do amazing work driving research in cardiovascular medicine, and I am very grateful to all of them, particularly my co-supervisor Gian, for taking the time to teach me about their work, and for giving me access to some of their samples. Without this, none of this project would have been possible.

I have had the privilege of working alongside some really fantastic people in the Vallance group. Thanks to Liwen, who is a wonderful friend and has been so helpful in discussing ideas and working on the ASAP projects together. Thanks to Alex, I could not have asked for a better desk neighbour; his dry wit has always made me smile. Thanks to David and Patrick for being so kind, clever and helpful whenever I needed anything, be that help with science, or just for a chat. Thanks to Zhihao, Usant and James, for being excellent company and brilliant co-workers, and thank you to Dean for being the best lab partner I have ever had and teaching me so much about being a scientist. I want to express particular thanks to all the Part II students (Master's year project students) that I was lucky to work with, particularly the four that I worked with on this project: Tom, Anna, Lloyd, and Mani. Together, we were able to collect a very large amount of experimental data

using the ASAP-MS instrument. Their hard work, positive attitude, and dedication to their projects made my time as a DPhil student very enjoyable.

The Chemistry in Cells DPhil programme has been really supportive during my time here, and I am very grateful to them for allowing me to complete this project, and for funding me alongside the Wellcome Trust. Particular thanks to Tracey Marr the programme administrator who was so welcoming in my first year.

Thank you to my wonderful husband, Rupert, who is the kindest, most supportive person I know. You keep me sane, make me laugh, and I don't know where I would be without you. Thanks to my amazing family who have been so supportive throughout my life and university, even though you didn't really understand what I was doing half the time. I am very lucky to have you. I appreciate everything you have done for me, and I really enjoyed being able to go back to the farm and spend time with you in the fresh air whenever I got tired of being in the city. I hope that you might read this one day, even if only this page!

Finally, I need to thank our cats: Lyra and Timmy. Lyra sadly passed away recently, but was such a huge part of our lives and always made me smile. She was a big fan of sitting on whatever piece of paper you were trying to write on or read from (which was never annoying at all), and was a very good listener when practising presentations. Now we have Timmy (Timothy) the kitten, who is wonderful and has made thesis writing very fun by trying to play with the keyboard or the mouse on the computer screen.

Abbreviations

AAA	abdominal aortic aneurysm
ACS	acute coronary syndrome
AI	artificial intelligence
AIMS	ambient ionisation mass spectrometry
AIS	acute infarct size
APCI	atmospheric pressure chemical ionisation
ARC	Advanced Research Computing
ASAP	atmospheric solids analysis probe
ASAP-MS	atmospheric solids analysis probe mass spectrometry
AUC	area under the curve
BMI	body mass index
BO	bayesian optimisation
BSA	body surface area
CBS	coated blade spray
CV	coefficient of variation
CVD	cardiovascular disease
DART	direct analysis in real time
DC	direct current

DESI	desorption electrospray ionisation
DESI-MS	desorption electrospray ionisation mass spectrometry
ECG	electrocardiogram
FIS	final infarct size
FMD	flow-mediated dilatation
FR	feature-reduction
GAN	generative adversarial network
HFD	heart failure diagnosis
HMDB	Human Metabolome Database
HPC	high performance computing
IMR	index of microcirculatory resistance
ITime	ischaemic time
KNN	K-nearest neighbours
LAESI	laser ablation electrospray ionisation
LCMS	liquid chromatography mass spectrometry
LDA	linear discriminant analysis
LGE	late gadolinium enhancement
LNO-CV	leave-n-out cross-validation
LV	left ventrical
MDRD eGFR	modification of diet in renal disease equation for estimated glomerular filtration rate

MI	myocardial infarction
ML	machine learning
MS	mass spectrometry
MRI	magnetic resonance imaging
MRMR	minimum redundancy maximum relevance
MSI	myocardial salvage index
MVO	microvascular obstruction
NBC	Näive Bayes classifier
NHS	National Health Service
NN	neural network
OI	overlap integral
OxAAA	Oxford abdominal aortic aneurysm
OxAMI	Oxford acute myocardial infarction
PC	principal component
PCA	principal component analysis
pPCI	primary percutaneous coronary intervention
PSI	paper spray ionisation
REIMS	rapid evaporative ionization mass spectrometry
RelF	ReliefF ranking algorithm
RF	radio frequency
RFC	random forest classifier

RMSE	root mean squared error
STEMI	ST elevated myocardial infarction
SVM	support vector machines
TIC	total ion current
TOF	time-of-flight
TScore	thrombus score

1

Introduction

Personalised healthcare is increasingly recognised as a vital approach in modern medicine, aiming to tailor prevention, diagnosis, and treatment strategies to the individual characteristics of each patient. This shift away from a one-size-fits-all model is particularly important in complex conditions such as cardiovascular disease (CVD). CVDs are the leading cause of death globally, with the World Health Organisation (WHO) reporting that they account for approximately 17.9 million deaths per year [1]. This presents a significant burden on both patients and healthcare systems. The diverse range of risk factors, symptoms, phenotypes, and treatment responses seen across individuals highlights the urgent need for more precise and patient-specific clinical tools. Clinicians need access to new tools to help them differentiate between different types of patients, identifying individuals who may benefit from more intensive treatments earlier in the treatment pathway.

Recent advances in analytical instrumentation have led to the development of smaller, faster, and more cost-effective methods for analysing biological samples. These improvements enable clinicians and pathologists to profile patients based on fundamental chemical signatures. Among these technologies, mass spectrometry has emerged as a leading method for biomedical sample analysis. In particular, the development of ambient ionisation techniques has significant potential to improve the utility of clinical mass spectrometry. These methods allow for the direct analysis of samples with minimal or no preparation, and in some cases, can be applied directly *in vivo* [2, 3].

Such analytical approaches generate large volumes of data rapidly, necessitat-

ing data processing and analysis that goes beyond the current state of the art in order to provide efficient and clinically meaningful interpretation. Machine learning algorithms are well-suited to handling complex, high-dimensional data, and can facilitate the translation of raw measurements into clinically relevant metrics. Integrating mass spectrometry data with machine learning models offers the potential to support clinical decision-making by providing interpretable outputs, for example, presenting clinicians with a simplified 'traffic light' system indicating low to high patient mortality risk.

This thesis presents the development of a novel clinical tool that combines ambient ionisation mass spectrometry with machine learning, designed for use in clinical and/or surgical settings to predict patient outcomes, disease progression, or response to treatment. We will discuss our approach in the context of two cardiovascular diseases - ST elevated myocardial infarction (STEMI) and abdominal aortic aneurysm (AAA). Such tools offer the potential to improve patient outcomes, with additional benefits of reducing healthcare costs and readmission rates.

1.1 ST elevated myocardial infarction (STEMI)

Globally, over 7 million individuals are diagnosed with acute coronary syndrome (ACS) annually [4]. A STEMI, or 'major heart attack', is the most severe form of ACS, which covers conditions that are characterised by a sudden decrease in coronary perfusion (the ability of the heart to supply oxygenated blood to its own muscles) as a result of complete blockage of one or more of the coronary arteries [4, 5]. Figure 1.1(a) shows the impact of such an occlusion of the heart. The onset of ischaemia (inadequate blood supply to the heart muscles) triggers the symptoms of the myocardial infarction (MI), commonly acute chest pain, shortness of breath, and pain in the left side, with complete cardiac arrest in the most severe cases [6, 7].

STEMI is a common clinical condition, responsible for 30,000 to 40,000 hospital

admissions in the UK each year, and continues to be associated with a high risk of patient mortality [8, 9]. From 2021 to 2022, over 85,000 confirmed acute myocardial infarction (AMI) cases were reported by the National Cardiac Audit Programme in the UK, 36% of which were the higher-risk STEMI, and the remaining 64% being the lower-risk NSTEMI (a MI caused by the incomplete blockage of a coronary artery, similar to that shown in the central image in Figure 1.1(b)) [10, 11]. The risk factors for STEMI development are shared by many other diseases of age and cardiovascular diseases: obesity, smoking, male sex, hypertension, high blood cholesterol, diabetes, and family history of the condition [4, 12].

1.1.1 STEMI pathophysiology

Coronary occlusion is caused by the development of an arterial thrombus, which usually forms at the site of arterial wall damage (shown in Figure 1.1(a-b)) caused by the rupture of a vulnerable fatty plaque (called a thin-cap fibroatheroma, described below) [13, 14]. Development of a STEMI typically starts with atherlosclerosis, the depositing of fibroatheromas or lesions within the wall of the coronary arteries [13, 15]. This process is illustrated in Figure 1.1(b). These lesions consist of an accumulation of lipids, inflammatory cells, smooth muscle cells, and necrotic cell debris [16]. They form within the intimal space between the smooth muscle of the blood vessel and the lining endothelial cells, underneath a monolayer of endothelial cells that line the interior vessel wall [16]. Over time, lesions develop a fibrous cap, consisting of a layer of fibrous connective tissue (collagen, elastin, smooth muscle cell bundles, and immune cells) that stabilises the plaque. Movement of blood through the vessel gradually causes thinning of the cap, creating a vulnerable plaque. If this plaque ruptures, the thrombolytic contents of the plaque are exposed to the blood supply [17]. This triggers blood clotting via the coagulation cascade, due to exposure of tissue factor in the endothelium to plasma factor VIIa in the lumen of the blood vessel, and platelet activation [18, 19]. The thrombus occludes the already narrowed coron-

ary artery, preventing the perfusion of blood to the myocardium distal to the clot. The hypoxic tissue becomes necrotic over time, characterised by loss of myocardial structure and cell death. This tissue will begin to release the biochemical signals that are indicative of cellular distress reactions: immune activation, inflammation, cardiac fibrosis, tissue repair and remodelling, angiogenesis (the development of new blood vessels to circumvent the blockage), scar tissue formation, and the release of intracellular contents of necrotic cells [20, 21].

1.1.2 STEMI diagnosis

A STEMI is typically identified by a new (or increased) and persistent ST-segment elevation on an electrocardiogram (ECG) in at least two contiguous leads (ECG leads that view adjacent areas of the heart) [6, 22]. An ST-elevation in an ECG, shown in Figure 1.1(c), is indicative of changes to the re-polarisation of the ventricles. This is caused by ischaemic tissue having abnormal electrical conductivity relative to healthy cells [15, 23]. In addition to the detection of elevated levels of cardiac troponin, a regulatory protein found in cardiomyocytes that is released upon damage to myocardial tissue, this abnormal ECG trace is the typical method of diagnosing STEMI in patients presenting with the symptoms of an MI [15, 22].

1.1.3 STEMI treatment

The treatment of STEMIs is robust and well-developed, and has lead to a drastic decrease in patient mortality over the last 50 years, with the UK in-hospital mortality rate following ACS falling from 20% to 5% [24–26]. This improvement has been attributed to improved timely access to treatments, expanded treatment options tailored to patients, and an increasing number of pharmacological treatment options [26]. However, there remains a high number of cases in which mortality occurs in the two years post-presentation due to complications including subsequent heart failure. In addition, patient readmission as a result of cardiovascular episodes and

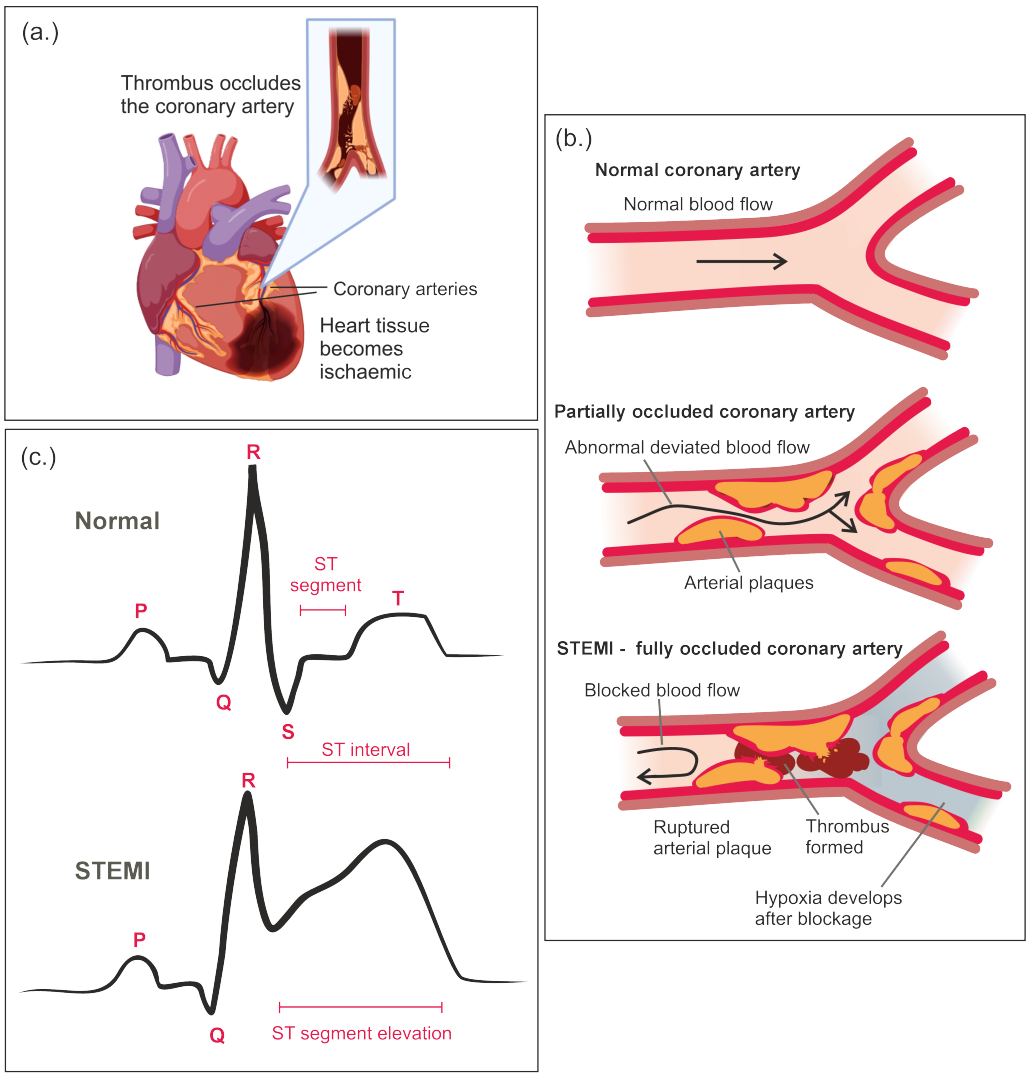

Figure 1.1 The physiology and diagnostic signals of a STEMI. A STEMI is located in a coronary artery, typically the left anterior descending artery, as shown in Figure (a), and causes loss of blood to the myocardium of the left ventricle. The development of a STEMI is caused by the build-up and degredation of atherosclerotic plaques (b), triggering thrombus formation in the coronary arteries. A STEMI often causes a characteristic ST-segment elevation in the ECG trace (c) due to changes in the conductivity of hypoxic cells. Figure (a) was made with www.Biorender.com. Figure (b) was adapted from [13].

related diseases remains a significant burden on healthcare systems [27–29]. Successful treatment requires re-vascularisation of the hypoxic tissue through relieving or removing the thrombus in the coronary artery, allowing blood to re-perfuse the hypoxic tissue [30]. This is done though a primary percutaneous coronary intervention (pPCI) procedure. The procedure should be completed as soon as STEMI is confirmed, to prevent long-term and irreversible damage to the heart and other organs [31].

During the pPCI, the blood flow within the heart is mapped via a coronary angiogram, which uses contrast dye and x-ray imaging to visualise the cardiac vasculature. The occluding thrombus or multiple thrombi are detected, as well as other areas of severe vessel narrowing [32]. A catheter and guide wire are then inserted into a blood vessel in the arm or groin of the sedated patient, and guided through the heart to the target vessel under x-ray guidance. A stent is guided into position at the site of the thrombus, and expanded using a balloon in order to re-open the occluded vessel. The balloon and guide wires are then removed, and the area monitored for re-perfusion. A number of stents may be required to fully re-vascularise the myocardial tissue [32–34].

One complication that can occur from this procedure is a distal embolism, which occurs due to fragments of the clot blocking smaller vessels distal to the original thrombus [35, 36]. To avoid this, many stent implants are now preceded by thrombus aspiration in a thrombectomy procedure, in which the thrombus is extracted from the occlusion site through an aspiration catheter [36].

Research shows that pPCI should be performed within 120 min of first medical contact where possible, with 50% of mortality occurring in the first 2 hours post-STEMI [31, 37]. In the UK, the 'door-to-balloon-time' target is < 90 mins. This is achieved in the majority of cases, and recent years have seen a decrease in mortality rate [25]. After 120 min, the outcomes of pPCI are worse, and fibrinolysis may be recommended in addition to the pPCI [25, 38]. Fibrinolysis involves the

administration of strong fibrolytic medication that activates the fibrinolytic system, a system of enzymes responsible for the breakdown of fibrin, the protein that creates the insoluble mesh framework of a clot [39].

The available pharmaceutical interventions for MIs are extensive, and often very effective when tailored to the patient in addition to the surgical interventions. Pharmaceutical treatment of STEMIs focuses on antiplatelet and antithrombotic medication to prevent the development of new clots and break down existing ones, pain management, and medication to help control blood pressure and maintain a normal heart rhythm [40–42].

In some patients, pPCI is not suitable or is not able to successfully re-vascularise, and so coronary artery bypass surgery is recommended. This may be due to a very severe MI causing prolonged hypoxia and cardiogenic shock, severe coronary artery disease damaging the coronary artery beyond the stage for which stents will be beneficial, or abnormal coronary artery anatomy [43, 44]. The bypass surgery requires a blood vessel from another part of the body (typically the arm or leg) to be extracted and attached to the coronary artery as a graft, allowing blood to circumvent the damaged region [44]. A coronary artery bypass is an open heart surgery, and has a significantly higher mortality rate than the pPCI procedure, and so is typically only used in around 5% of cases [43].

1.1.4 Clinical indicators of STEMI severity and recovery success

A number of clinical measurements and scores have been defined to help assess the severity of STEMIs and the recovery of the heart post re-vascularisation. In this work, four of these key scores will be the focus of patient risk analysis: the index of microcirculatory resistance (IMR), thrombus score (TScore), ischaemic time (ITime), and microvascular obstruction (MVO). These will be described further in Chapter 4, and are briefly defined as follows:

1. **IMR** - A measure of resistance to blood flow in the coronary micro-circulation assessed using a pressure sensor/thermistor-tipped guide-wire, or thermodilution catheter [45]. IMR is calculated as the difference in blood pressure between the coronary artery and the coronary vein, divided by the coronary artery flow rate. A high IMR measurement post STEMI indicates poor circulation in the smallest blood vessels in the heart muscle, and is associated with worse outcomes post MI [46, 47].

2. **TScore** - A visual measure of thrombus burden (size of the thrombus in relation to the size of the blood vessel) graded from 0-5, with 0 indicating no thrombus and 5 indicating a total occlusion. The majority of STEMI patients have a TScore of 4 or 5 [48].

3. **ITime** - A measure of the duration from onset of symptoms (such as chest pain) to the point at which reperfusion is achieved. Shorter ITimes typically result in better patient outcomes [49]. The ITime measurement is subjective, as the onset of symptoms can be difficult to ascertain in some patients. For example, female STEMI patients often have a different presentation to male patients, without the characteristic 'crushing' chest pain [7, 49]

4. **MVO** - The persistence of blockage or poor vascularisation of the microvasculature in the heart muscle despite successful restoration of flow. This is often associated with the no-reflow phenomenon, which occurs when, despite successful reperfusion of the larger vessels, there is no or very slow blood flow in the smaller vessels. MVO is thought to be caused by plaque debris blocking the microvascular vessels [50, 51], and is measured by cardiac MRI after the STEMI has been treated. Any MVO greater than zero indicates that some heart function has been irreversibly lost due to the STEMI [52].

1.1.5 Current STEMI research

Recent research into STEMI treatment and pathology has focussed on the prediction of long-term patient outcomes, with a particular focus on expanding the recently published and shared large-scale global patient data collections, and utilising machine learning (ML) techniques in the analysis of such large data sets [53–57]. Other areas include the development of novel surgical devices and catheters for improved surgical safety [58–61], and the implications and interactions of novel drug therapies in STEMI patients [62–65].

1.2 Abdominal aortic aneurysm (AAA)

An AAA occurs when there is an irreversible swelling or ballooning of the abdominal aorta, the major artery carrying blood from the heart to the organs of the abdomen and lower body. The swelling is defined as an AAA when the aorta reaches an internal diameter of over 3 cm, compared to the average normal diameter of 2.46 cm for men and 2.26 cm for women (see Figure 1.2(a)) [66, 67]. Studies on the epidemiology of AAAs state that this disease impacts approximately 5% of the population, with the majority of those affected being males over the age of 65 [66, 68, 69], potentially contributing to 1-3% of all deaths among men aged 65–85 years in developed countries [66, 70].

1.2.1 AAA pathophysiology

Unlike for STEMIs, the cause and development of AAAs is relatively poorly understood. AAAs were once thought to be the result of atherosclerosis, but a growing body of literature now supports the hypothesis that an AAA is in fact a local manifestation of a systemic vascular abnormality, the cause and development of which is still unclear [69]. The phenotype of this disease is characterised by three key features: proteolysis (degradation) of the structural proteins responsible for blood vessel wall

integrity and elasticity (namely elastin and collagen); inflammation of the aortic wall and surrounding tissues; and apoptosis of smooth muscle cells [69]. Studies of blood vessels distant from the abdominal aorta have found biochemical changes similar to those present in the wall of the AAA, supporting the systemic disease theory [69, 71–74]. Mild dilatation and reduced distensibility (changes in the ability of blood vessels to rebound after stretching under pressure) are observed in AAA patients in distant blood vessels compared with controls [69, 75, 76]. This has been investigated further by R. Lee *et al.* in the Oxford abdominal aortic aneurysm (OxAAA) study (described in this thesis) using flow-mediated dilatation (FMD), a measurement that assesses of the elasticity of the brachial artery in the forearm, and has been shown to be weakly correlated with AAA growth rate [77–79].

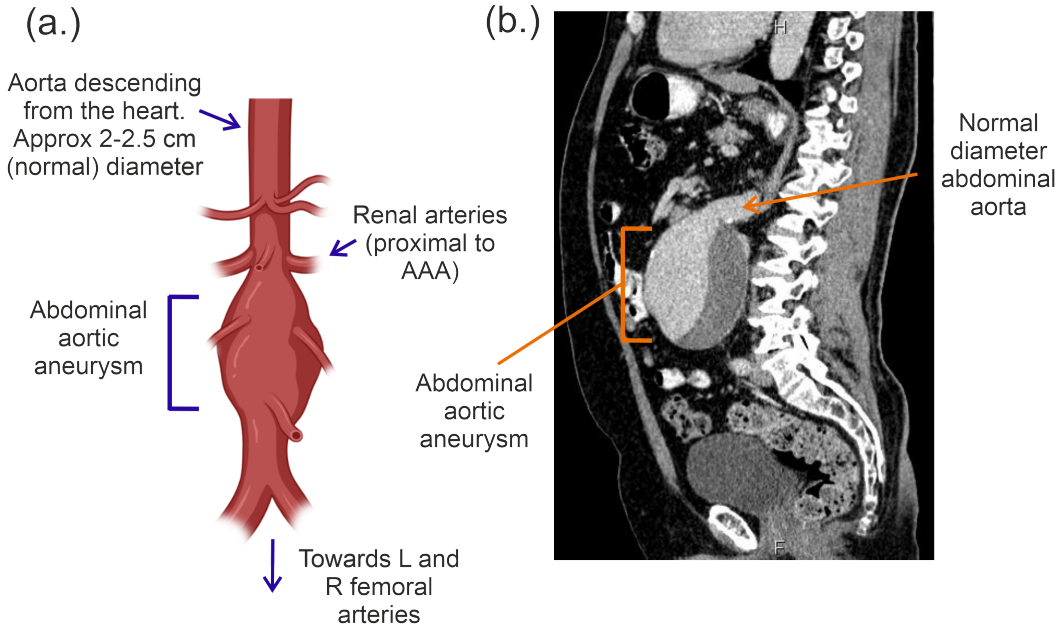

Figure 1.2 Diagrams of an AAA. Left shows a schematic visual representation of an AAA on the abdominal aorta. Right shows a 2D CT scan showing an abdominal cross section of an AAA patient. The orange arrow shows a section of normal abdominal aorta, with the highlighted region showing the AAA. Image (a) was created with www.Biorender.com. Image (b) was taken from the open access radiology database www.Radiopaedia.org (case 36) [80, 81].

In the majority of patients, the artery wall within the AAA is covered by an intra-luminal thrombus, composed of a dense fibrin network, inflammatory cells,

red blood cells and platelets, to form a thick incompressible layer, which serves to thicken and stiffen the weakened wall [82, 83]. Release of metabolically active molecules from the AAA, thrombus, and abnormal vasculature into the surrounding tissue and circulation may have further systemic effects [84].

1.2.2 AAA diagnosis and monitoring

Currently, the majority of AAAs are identified either by risk group ultrasound screening of known at-risk demographic groups, with the UK being one of the first to adopt this type of screening programme in the 1990's, or as an incidental finding during the investigation of another disease [70, 85]. Further visualisation and monitoring of the AAA is frequently conducted using CT and MRI scans, which give clear visualisations of the AAA size and volume, as shown in Figure 1.2(b) [86]. Risk factors for the development of AAAs are male sex, smoking, hypertension, atherosclerosis, and a family history of AAAs [87, 88]. Death can occur due to the aneurysm rupturing, which occurs as pressure exerted by abnormal blood flow through the aneurysm eventually overcomes the strength of the artery wall, which has been weakened by the dilation of the AAA [89, 90]. This leads to catastrophic internal bleeding, disrupting blood flow to the heart and other major organs. AAA rupture has a very high mortality rate of up to 80% [91, 92].

1.2.3 AAA treatment

To prevent rupture, AAAs are surgically repaired once the size has surpassed a surgical threshold of 5.2 cm and 5.5 cm for female and male patients, respectively [67, 93]. Surgical repair involves strengthening the aorta using an internal graft, using either open surgery or a keyhole endovascular procedure. In both procedures, a graft is inserted into the lumen of the aorta within the aneurysm region, and once completed, blood flow resumes through the graft within the AAA. Therefore, the AAA is not truly removed, but strengthened structurally so that the pressure of

blood flowing through the vessel is diverted to the graft and is no longer exerted on the walls of the aorta [90]. No pharmaceutical treatments are available specifically targetting AAA currently, but some patients will be treated for co-morbidities that may increase rupture risk, such as high blood pressure or cholesterol, or conditions that may prevent surgery from being considered, such as lung and kidney disease [90]. The AAA repair surgery is a high-risk surgical procedure, with a 'major complication post-operation' risk of approximately 10% for the open-repair surgery [90], and high patient mortality rates of $> 5\%$ [94, 95]. Risks include infection, blood clots, stroke, heart attack, kidney injury, and loss of circulation to the intestines or the legs. Many AAA patients have co-morbidities, a factor both of the age demographic and high incidence of systemic circulatory diseases such as heart disease [96–100].

1.2.4 AAA development and growth

AAAs are often asymptomatic, however when the AAA becomes very large some patients experience lower back pain and abdominal pain, caused by the AAA putting pressure on the spine and organs (the close proximity of the AAA to the spine can be clearly seen in Figure 1.2) [101, 102]. Over time, an AAA increases in diameter, but the rate at which this growth occurs can vary significantly between patients [103]. Patterns in AAA growth rate are non-uniform, with studies in the literature showing a range of dependencies including linear [104, 105], staccato or discontinuous [106–108], and exponential [109–112] growth patterns [113]. The growth rate of AAAs is usually slow at <3 mm per year, but growth rate appears to increase with AAA size [103, 113].

More recent studies have focused on the development of large-scale mathematical models and using AI techniques to predict AAA growth [108, 114–116]. For many patients, the growth of the AAA is so slow that it never requires surgical intervention, and the patient continues to live with the AAA asymptomatically [98, 117]. One of the major challenges of treating this disease is identifying patients for whom surgical

intervention is required, as opposed to those patients for whom the risk of rupture is very low or symptoms never occur, and so can be treated with continued monitoring and lifestyle changes [117, 118].

1.2.5 Current AAA research

Current research trends appear to focus on improvements to AAA repair surgery for complex patients [119, 120], potential pharmaceutical treatments [121–123], druggable targets identified through metabolite screening and proteomics studies [124, 125], and applications of ML in AAA patient screening and care [126–128]. Also under investigation is the incidence and treatment of AAAs in women, for whom the presentation of AAAs can differ, and for whom unfortunately very little research has previously been conducted [129–131].

1.3 Atmospheric solids analysis probe mass spectrometry (ASAP-MS)

1.3.1 Mass spectrometry

In chemical and biochemical analysis, few techniques rival the versatility and precision of mass spectrometry (MS), which has become indispensable in a wide range of fields. MS is an analytical technique that forms ions from atoms or molecules within a sample and then measures their relative signal intensities and mass-to-charge ratios (m/z) [132]. A mass spectrometer typically consists of three core components: an ionisation source; a mass analyser; and a detector [133]. These components function to measure the mass-to-charge ratios (m/z) of analyte ions present in a sample [134].

The ionisation source converts neutral analyte molecules into gas-phase ions, which can then be magnetically or electrically manipulated within the high vacuum

regions of the mass spectrometer to allow measurement of a mass spectrum [132]. Ionisation can be achieved through a range of techniques, including electron ionisation (EI), electrospray ionisation (ESI), matrix-assisted laser desorption/ionisation (MALDI), and chemical ionisation (CI) [135]. The choice of ionisation method influences the type, proportion, and stability of ions produced, and is selected based on the chemical nature and physical properties of the analytes under investigation, and sometimes on other aspects of the analytical pipeline [132].

Following ionisation, the ions are directed into the mass analyser by an electric field, sometimes supplemented by a pressure differential [136]. The mass analyser separates ions according to their mass-to-charge ratio (m/z) [133]. Several types of mass analysers are employed in modern MS, each offering distinct advantages in terms of resolution, sensitivity, and scan speed. These include quadrupole analysers, time-of-flight (TOF) analysers, ion trap analysers, and orbitrap analysers [132, 136].

Once ions have been separated based on their m/z, they are detected and quantified by a detector, which converts the physical impact of an incident ion into a current, which is recorded electronically via appropriate acquisition software [134]. Common types of detectors include electron multipliers, photomultipliers, and Faraday cups. The resulting ion signal is then processed and visualised through a software interface, enabling interpretation and analysis of the mass spectrum [134, 136].

A variety of different MS instruments have been developed, coupling together a range of features and components that allow devices to be tailored to different needs. For the present application, we require an instrument that is easy to use, has a small footprint, can rapidly detect metabolic fingerprints of clinical samples in their native state, without the need for the addition of solvents, and requires minimal sample preparation. The atmospheric solids analysis probe (ASAP) mass spectrometer is an ambient ionisation mass spectrometry (AIMS) instrument which meets these specifications.

1.3.2 Ambient ionisation mass spectrometry

AIMS is a relatively new development in the field of mass spectrometry, first described in 2004 with the introduction of desorption electrospray ionisation (DESI) by G. Cooks *et al.* [137]. Since then, a variety of ionisation sources have been developed that can be coupled with a range of mass spectrometers, all with the common feature of having the capability to measure samples in their native state without the need for much sample preparation. AIMS techniques can be categorised into three main classes: plasma based techniques (e.g. ASAP, DESI, and direct analysis in real time (DART)); solid-liquid extraction based techniques (e.g. coated blade spray (CBS) and paper spray ionisation (PSI)); and laser-based techniques (e.g. rapid evaporative ionization mass spectrometry (REIMS) and laser ablation electrospray ionisation (LAESI)) [2, 138, 139]. Plasma-based techniques utilise electric discharges to produce a reactive plasma of electrons, radicals and excited-state (metastable) ions. They are arguably the most applicable to a wide range of sample types in their native state, including biological samples, and have lower costs and barriers to use than other alternatives [2].

The ability of AIMS techniques to provide rapid detection and identification of samples has led to it being utilised across many fields, including detection devices for explosives, illicit substances and counterfeit medicines [140–143]; in surgical settings for tumour delineation and diagnostics [144, 145]; and in chemical reaction monitoring [2, 146]. However, many of these applications, particularly in the medical and pharmaceutical fields, have yet to move beyond the research phase. A number of barriers have been cited that have prevented the translation of this technology, either alongside or to replace existing 'gold standard' techniques (commonly LC-MS or GC-MS). Professor Patrick Sears (head of the Ambient Ionisation special interest group of the British Mass Spectrometry Society (BMSS), 2025) noted recently that *"Arguably, this area of science (AIMS) could be described as being in its formative years and recommendations have been made for closer investigation of sampling*

heterogeneity, ion suppression effects, and a critical evaluation of areas for improvement" [147]. These barriers, and their implications in relation to the applications discussed in this work, will be discussed in detail in the final chapter of this thesis (Chapter 8).

1.3.3 Atmospheric solids analysis probe mass spectrometry (ASAP-MS)

ASAP-MS is a form of plasma-based AIMS that utilises an atmospheric pressure chemical ionisation (APCI) process to ionise a sample after introduction to the ion source on the tip of a hand-held insertable probe. A schematic of the ion source housing, including the ASAP probe, is shown in Figure 1.3. This relatively new ionisation method has become increasingly popular in recent years due to its ease of use and ability to analyse a large range of samples with minimal sample preparation.

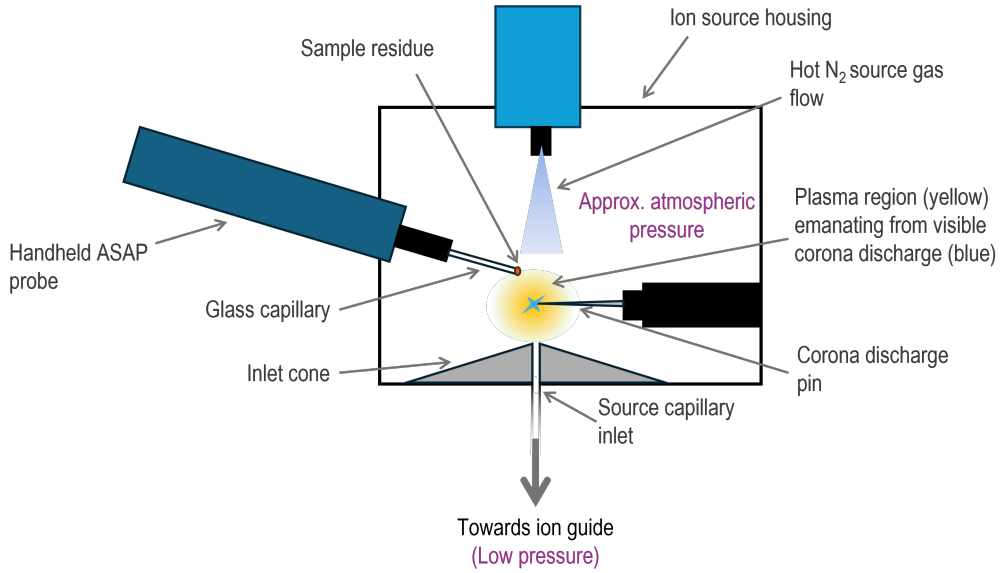

Figure 1.3 An example of an ASAP ionisation source, based on the Advion CMS-S instrument setup [148]. In other instruments, the key components are the same, but exact values for pressure and orientations may differ.

Applications have been broad, from the identification of counterfeit food and drink [149–151], to the detection of explosives and illegal substances [152–155], and chemical reaction monitoring [156, 157]. The simplest and most common implement-

ations couple the ASAP ion source with a single quadrupole mass spectrometer, but as the ion source is essentially an add-on feature it could be coupled with most types of mass analyser [148, 158, 159]. The instrument that was predominantly used in this work is the Advion Expression CMS-S ASAP mass spectrometer [160]. The key components of this instrument are the ASAP ion source, a hexapole ion guide, a quadrupole mass analyser, and an electron multiplier detector.

1.3.3.1 ASAP ionisation

ASAP utilises an APCI process to ionise a sample after introduction to the ion source on the tip of a glass capillary [161]. Thermal desorption of the analyte from the tip of the glass capillary is facilitated by the flow of nebulising nitrogen gas [162, 163]. The rate of desorption is governed by a range of factors, and can be modelled simply in a similar way to temperature-programmed desorption experiments, using the Polanyi-Wigner equation:

$$-\frac{d\theta}{dt} = \theta^n \nu_o e^{(-\frac{\Delta E}{k_B T})} \tag{1.1}$$

Equation 1.1 relates the rate of change in surface coverage ($-\frac{d\theta}{dt}$), to the total surface coverage θ, the desorption order n, the pre-exponential factor ν_o, the activation energy of desorption ΔE, the Boltzmann constant k_B, and the temperature T [164]. The model assumes vacuum conditions. However, as the ion source is under atmospheric pressure, the rate of desorption will also be dependent on the atmospheric pressure in the source housing and the heated gas flow.

During APCI, a high voltage is applied to the corona discharge pin, creating a strong electric field at the tip of the pin. This induces ionisation of the surrounding gas molecules, mainly the N_2 carrier gas and atmospheric O_2 and H_2O, in a corona discharge represented by the yellow region in Figure 1.3 [165, 166]. In positive mode, this ionisation process generates positive ions and free electrons. The electrons collide with further neutral gas molecules, displacing an electron out of the gaseous

molecule to generate a positive ion and two scattered secondary electrons. The atmospheric pressure in the source housing promotes frequent collisions between the ions and other gaseous molecules, to produce a highly energised plasma region of ionic species, including N_2^+, H_2O^+, H_3O^+, N_2H^+, and N_4^+ (Equations 1.3-1.6) [167, 168]. The gaseous analyte molecules (A) are ionised in collisions with these ions, generating AH^+ (Equation 1.7), rather than $A^{\bullet+}$, due to the abundance of water in the ion source. In summary, ionisation occurs in the following methods:

$$\text{Initial electron-molecule collisions: } N_2 + e^- \rightarrow N_2^+ + 2e^- \tag{1.2}$$

$$\text{Example ion collision reactions: } N_2^+ + 2N_2 \rightarrow N_4^+ + N_2 \tag{1.3}$$

$$N_2^+ + H_2O \rightarrow H_2O^+ + N_2 \tag{1.4}$$

$$N_4^+ + H_2O \rightarrow H_2O^+ + 2N_2 \tag{1.5}$$

$$H_2O^+ + H_2O \rightarrow H_3O^+ + HO^{\bullet} \tag{1.6}$$

$$\text{Analyte ionisation: } H_3O^+ + A \rightarrow H_2O + AH^+ \tag{1.7}$$

In positive-ion mode, electrons generated via ionisation are quickly lost to the positive electrode, and most analyte ionisation proceeds through ion-molecule and ion-ion reactions. In negative ion mode, electrons generated by the initial ionisation of surrounding gas molecules are accelerated away from the corona discharge pin. This promotes electron cascade reactions, facilitating the formation of AH^- ions, due to the presence of water [161, 169]. Negative ion mode was not used in this study due to poor signal levels, and so only positive ionisation will be discussed.

The positive ions are accelerated towards the capillary inlet cone by a pressure differential and electric field; for the CMS-S instrument, the pressure changes from 3 bar in the source housing to 2 mbar in the source block, and the electric field is determined by the user settings (discussed in Chapter 2). Collisions between ions in this region facilitate in-source fragmentation, the extent of which is controlled by a potential applied to the inlet cone [161, 170]. Higher potentials increase the

kinetic energy of ions, increasing the extent of fragmentation due to ion-ion and ion-molecule collisions.

1.3.3.2 Hexapole ion guide and quadrupole mass analyser

The ions generated in the ion source housing exit the capillary inlet and enter the source block, where they are directed using voltages that act as an ion lens towards an extraction electrode and into the hexapole. The hexapole is an ion guide that focuses the ion beam and transfers ions from the source to the quadruple [161]. The pressure in the hexapole, at 5×10^{-3} mbar in the CMS-S instrument, is lower than in the source block, which further reduces fragmentation. Ions accelerating from the source block to the hexapole along the pressure gradient undergo collisional relaxation of vibrational and rotational degrees of freedom. This narrows their speed distributions, resulting in fewer collisions and preventing further fragmentation after this point. Within the hexapole, radio frequency (RF) voltages are applied to six parallel metal rods arranged with a hexagonal cross-section, generating an electric field with near-cylindrical symmetry. The hexapole field condenses the ion beam onto a central axis, generating a collimated ion beam travelling along the centre of the hexapole, the z axis, with low velocity components along in the x, y axes. [171].

The resulting ion beam is transmitted into the quadrupole mass analyser region, with a further drop in pressure (to 5×10^{-6} mbar in the CMS-S instrument). The quadrupole consists of four parallel metal rods, connected pairwise to a combination of direct current (DC) and RF voltages [133]. The voltages applied to the rods are configured in pairs, A and B, where both A and B rods are opposite each other, as shown in Figure 1.4. These rods have the following applied potentials:

$$\text{Potential A} = U + V\cos(\omega t) \tag{1.8}$$

$$\text{Potential B} = U - V\cos(\omega t) \tag{1.9}$$

where U is a DC voltage, and $V\cos(\omega t)$ is an RF voltage defined by the voltage V,

the angular frequency ω, and time t [172].

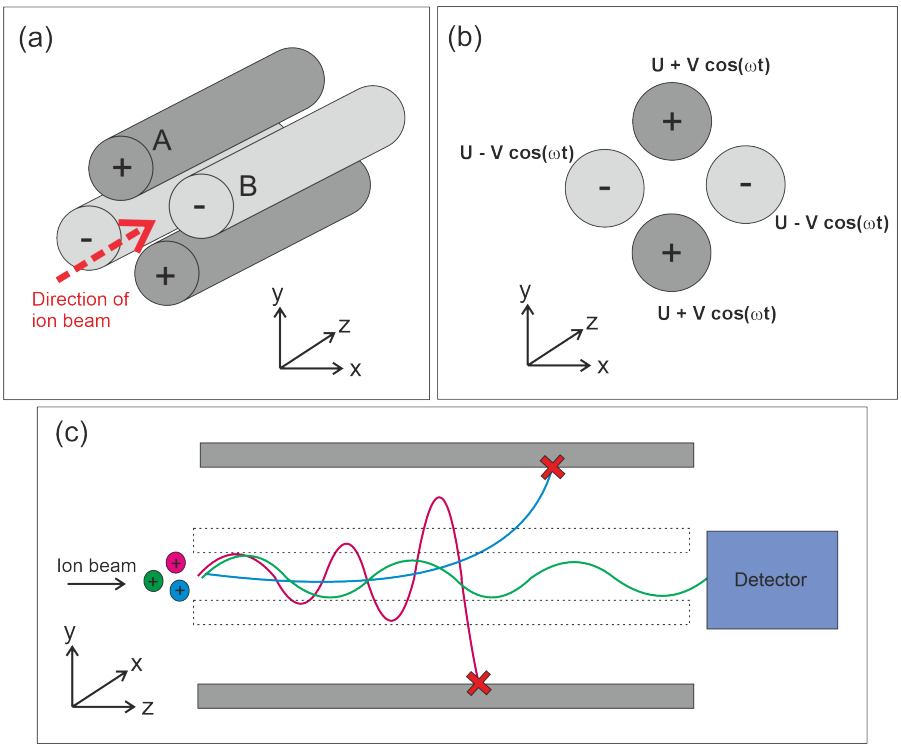

Figure 1.4 The layout and voltages applied to a quadrupole mass filter. The four parallel rods are configured in parallel (a) such that opposite rods have the same RF and DC voltages applied (b). When ions of different *m/z* travel through the quadrupole, their trajectories are affected by the electric fields such that only ions of a certain *m/z* have a stable trajectory (green) and reach the detector (c). Adapted from [172].

The applied voltages generate a hyperbolic electric field that affects the trajectory of ions travelling down the z axis between the four rods. For a given combination of DC and RF voltages, only ions of a certain mass-to-charge ratio have a stable trajectory to pass through the quadrupole filter. All other ions deviate from the central z axis path and either collide with one of the rods or escape the quadrupole entirely [173]. This is shown in Figure 1.4(c), with the green ion having a stable trajectory and reaching the detector, but the red and blue ions being lost. A mass spectrum is obtained by measuring the current generated by ions that exit the quadrupole filter as the voltages on the rods are varied, corresponding to scanning through the m/z values that have stable trajectories.

Two methods of scanning the quadrupole voltages exist: varying ω at a constant

U and V (varying the RF component), or varying U and V with the U/V ratio fixed for a constant ω (varying both DC and RF) [174]. The stability of an ion is described by the Mathieu equations, which describe the x and y motion of an ion with a particular m/z [172, 173, 175].

1.3.3.3 Ion detectors

The detector used in the Advion CMS-S instrument features a conversion dynode and an electron multiplier [161]. Incoming ions are accelerated toward the conversion dynode, which is a plate made of secondary-emissive material. On impact of a positive ion, secondary electrons are emitted. In negative mode, an incoming negative ion is converted to a positive particle on impact, before impacting the next dynode at which secondary electrons are emitted. These are directed towards the electron multiplier, which amplifies the signal before detection. Electron multipliers contain dynodes that emit secondary electrons when bombarded with electrons of sufficient kinetic energy. The secondary electrons recurrently impact the further dynode surfaces along the trajectory pathway, amplifying the signal through a cascade effect. Electron multipliers can be discrete or continuous, determined by whether they contain a set number of individual dynodes in series (for the former), or a single continuous dynode in a horn shape (for the latter) [133, 176]. The CMS-S instrument features a continuous dynode electron multiplier. At the end of the electron multiplier is the collection plate, at which the current is converted to a signal and displayed on the user interface. This is displayed as signal intensity as a function of scan voltage (which is converted to the associated m/z value), and time.

1.3.4 Application of ASAP-MS to a clinical setting

Previous work has shown that ASAP-MS can be used successfully for the analysis of biological samples, particularly blood and urine [177–180]. ASAP-MS has the capacity for real-time analysis of analytes, making it an ideal candidate to use in clinical

settings, where the speed, reliability of results, and ease of use are of the utmost importance. It is important to note that the mass spectra generated by the single quadrupole mass analyser ASAP instruments are untargeted, meaning that the identity and origin of the ion peaks detected are unknown. The technique therefore offers an untargeted 'molecular fingerprint' of the sample, enabling the use of pattern recognition methods such as machine learning for biomarker discovery. Untargeted approaches are becoming a valuable method that complements traditional targeted approaches of biomarker discovery in several ways. Firstly, it does not require detailed knowledge of the biochemical pathways or specific molecules involved in the disease, thereby reducing both the time and cost typically associated with multi-omics analyses. Secondly, it allows for the simultaneous measurement of numerous biomolecules, which can help uncover subtle interactions between metabolites that might otherwise go unnoticed, potentially expanding biomarker discovery beyond the currently recognised molecular landscape [181, 182].

1.4 Machine learning

ML is a type of artificial intelligence that automates the process of building analytical models for data analysis [183]. This is achieved through identifying patterns in data sets and using those patterns to make predictions [184]. ML methods can be classified as supervised or unsupervised. Supervised models are given the true label of each sample alongside the training data, with the aim of identifying features of the data that enable the prediction of those labels successfully. Unsupervised models are not provided with the data labels; instead, the algorithm identifies internal patterns within the dataset, independent of the labels [185].

The use of ML techniques in medical research has increased rapidly in the last few years [186]. When applied correctly, machine learning tools can be used to uncover complex patterns in data that were otherwise inaccessible. In cardiovascular disease, this has provided new methods of predicting variables such as patient progression

and outcomes, improving patient screening, uncovering biomarkers, and predicting response to treatment [56, 126, 127, 187–189]. These approaches have the potential to advance personalised healthcare, identifying and enabling treatment of patients with the most appropriate methods more rapidly, reducing the strain on healthcare systems from hospital readmission, and saving or improving patients' lives.

ML tools have also been utilised for the interpretation of untargeted MS data [190–193]. Treating these data as a molecular fingerprint, as opposed to a set of metabolites with relative intensities that must be individually interpreted, has proved to be a powerful method of interpreting biological data, which is fundamentally very noisy [192]. Using interpretable ML techniques, these results can then be probed to identify biomarkers that can be investigated further.

1.5 Aims of this thesis

The aim of this work is to provide initial proof-of-concept results showing how ASAP-MS and ML techniques can be combined in order to predict patient outcomes. This work will utilise samples from across two Oxford-based clinical studies, the Oxford acute myocardial infarction (OxAMI) project, and the Oxford abdominal aortic aneurysm (OxAAA) project (both described in Chapter 2), to demonstrate the use of untargeted metabolomics for cardiovascular diseases. The use of ASAP-MS in the analysis of biological samples required significant method development, which is described in Chapter 2. Chapter 3 describes the data analysis methods that were employed throughout this work, including the machine learning models that were employed. Chapter 4 discusses the results obtained for the prediction of STEMI patient outcomes from blood plasma using ASAP-MS and ML. Chapter 5 discusses the results obtained for the prediction of AAA patient progression from ASAP-MS analysis of blood plasma and tissue samples with ML. Chapter 6 details a study into the comparison of four different ASAP-MS instruments for the analysis of blood plasma and tissue. Chapter 7 describes a study comparing the different methods of

data normalisation that can be applied to ASAP-MS data, and the impact of this on further ML analysis. Chapter 8 discusses the key results of this project, providing an evaluation of the method, and describes the work that is still required and the challenges that lie ahead in creating an in-clinic ASAP-MS tool.

We would like to highlight that, unfortunately, there is no current framework for reporting the error, accuracy, and uncertainty in untargeted MS data, or data with multiple peaks (the specifics and origins of these errors are discussed in Chapter 7). This makes analysis of this type challenging, particularly when comparing results to literature data. We have reported a number of different measurements of error and variation in this study, none of which are necessarily the 'best' or 'correct' way to report this data, but in all instances, we have tried to report everything in the most appropriate, understandable, and transparent way possible. If the MS community comes together in the future and determines a framework for this analysis type, this would be very powerful.

2

Experimental methods

In the following, we describe the development and optimisation of atmospheric solids analysis probe mass spectrometry (ASAP-MS) methods for the analysis of human blood plasma and tissue samples. Aspects of this chapter have been adapted from our recent publication [180].

2.1 Instrumentation

The analytical measurements discussed throughout this thesis were primarily conducted on an Advion expression® compact mass spectrometer (CMS-S) ASAP-MS instrument (Advion Ltd, Harlow, United Kingdom)[148]. A detailed description of how this instrument functions to record mass spectra for a sample can be found in Section 2.1. A small number of measurements were conducted on other ASAP-MS instruments, which are all discussed within Section 6.

A schematic diagram of the spectrometer is shown in Figure 2.1. To make a measurement, a glass capillary tip mounted into the ASAP-MS is exposed to the solid or liquid sample and inserted into the ASAP-MS ion source housing. As described in Chapter 1, the capillary is heated and the sample is vaporised by a flow of hot N_2 nebulising gas, provided by the house gas supply at 3 bar. The resulting vapour is ionised via APCI initiated by a corona discharge. Ions are drawn into the ion inlet skimmer by an electric field and pressure differential at the atmospheric pressure interface. Collision-induced fragmentation can occur at this interface, the extent of which may be controlled by altering the inlet potential ('Capillary offset

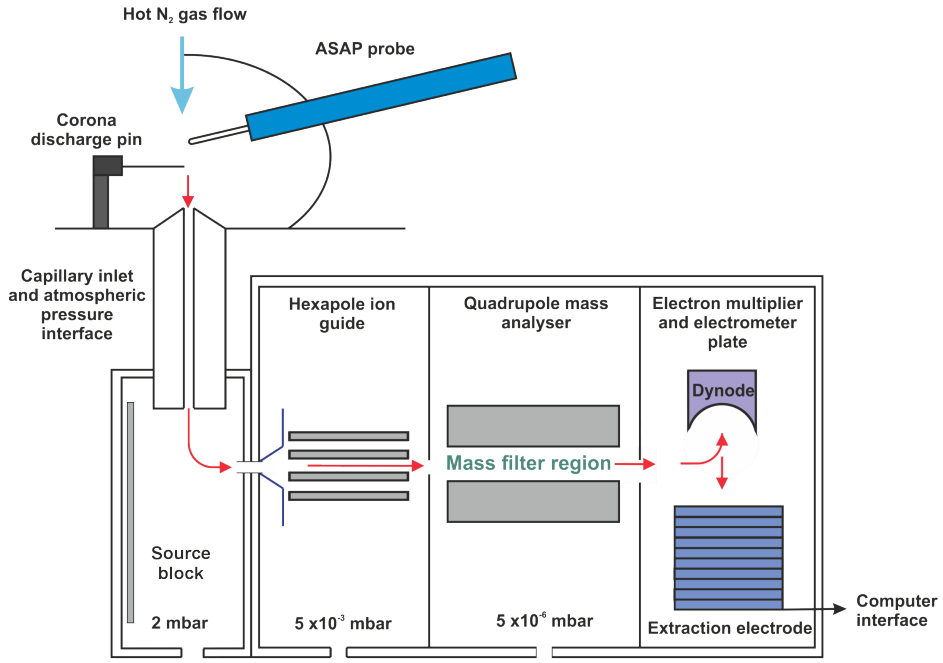

Figure 2.1 A schematic of the Advion CMS-S ASAP-MS instrument, see text for description. Adapted from the Advion CMS User Manual [161], and [180].

voltage') and therefore the collision energy with background gas. The ions travel into the source block, where an ion guide directs the ions towards the inlet skimmer of the hexapole region. The 'source voltage' applied here facilitates further collision-induced fragmentation. A voltage offset can also be applied, which increases the voltage applied at the ion guide when the quadrupole is set to allow the passage of high mass ions, increasing the fragmentation of those higher mass ions. The ions entering the inlet skimmer are collimated within a hexapole ion guide and directed to the quadrupole mass filter. The quadrupole voltages are scanned repeatedly to record a series of time-resolved mass spectra over a user selected range of mass-to-charge ratios, m/z [148, 161, 194, 195]. Automated instrument calibration was conducted daily in positive ion mode using Advion APCI calibration/tune standard mix. [148].

Data acquisition was controlled using Advion Mass Express software (version 6.9.38.1) and Advion Data Express data manipulation software (version 6.9.38.1). The instrument can record spectra in positive or negative ion mode, or can record

both within a single measurement. Early tests showed that very little signal was observed for plasma samples in negative ion mode, so all data reported here employed positive ion mode only. User settings that were optimised as part of this work include ion source temperature and fragmentation settings, mass range to scan over, and time per individual scan.

2.1.1 Standard experimental framework

Optimisation of sampling technique and determination of the optimum number of individual measurements to average over were key elements of present study. A standard experimental framework was followed for running measurements of all sample types on the ASAP-MS instrument; this was adapted and optimised as described in Sections 2.4 and 2.5 for specific biological matrices:

The glass capillaries used for sample loading (Advion ASAP S01 short) were sterilised by heating in an oven for 30 minutes at 250 °C and were then stored in a desiccator. After mounting a capillary into the probe tip and before recording spectra for a sample, a background spectrum of the clean glass capillary was recorded by inserting the probe into the probe slot within the ion source housing of the mass spectrometer for 30 s. The probe was then removed from the source housing, cleaned and cooled with methanol, and dried using a Thorlabs MC-5 lens tissue. For plasma measurements, frozen 1 mL biological samples were thawed and vortexed immediately prior to measurement. For each measurement, a glass capillary was mounted into the tip of the ASAP probe, a small amount of sample was transferred to the capillary tip, and the probe was inserted into the mass spectrometer for the chosen acquisition time, during which a series of repeat mass spectra was recorded. The probe was then removed and cleaned with methanol and lens tissue before the next repeat measurement of the sample. Based on an evaluation of repeat background measurements made during a test series of sample measurements, each capillary was used for five repeat measurements and then replaced with a clean capillary.

The mass spectrometer is generally run continuously while recording background and sample spectra for an individual sample, with the spectrometer outputting the time-dependent mass spectra as a 2D array of intensities for each m/z ratio and time point in the acquisition series.

Data sets were analysed using MATLAB R2022a and Python 3.7 software written in-house. The data analysis process is described in further detail in Section 3, but an overview is provided here to allow an understanding of the method development in this section. The first step of data processing is to extract from the large data file the individual measurements (background measurements and individual sample measurement repeats), corresponding to the measurement times when the probe was mounted in the ion source housing. This is an automated process based on analysis of the total ion signal as a function of time. Following background subtraction from each sample repeat, various additional data processing steps can be carried out as desired. These include averaging the data over the acquisition time, data binning (e.g. to unit m/z to reduce file sizes), and averaging over repeat measurements.

2.2 STEMI samples - The OxAMI study

The OxAMI study is a prospective observational cohort study, with voluntary patient enrolment taking place on admission to hospital when a patient was in need of a pPCI. The samples used in this report were from $n = 283$ STEMI patients who underwent pPCI at the John Radcliffe Hospital, University of Oxford, Oxford, UK, between 2010 and 2021. The OxAMI study and patient selection criteria has been described in a number of publications. Within the study criteria, STEMI was defined as chest pain continuing for > 30 min and ST elevation > 2 mm in > 2 contiguous leads. Exclusion criteria were symptom duration > 12 h, the presence of severe haemodynamic instability, and contraindication to the use of adenosine. The study was conducted in accordance with the Declaration of Helsinki. All participants gave written informed consent (REC number 10/H0408/24) [196–198]. Coronary aspirate

blood samples were obtained from OxAMI patients during pPCI to restore coronary artery blood flow. Coronary aspirate is the blood removed from the coronary artery when the thrombus causing the blockage is aspirated. The plasma component of each sample was obtained by mechanical separation using a centrifuge. The plasma was snap-frozen in dry ice and stored in a freezer at $-80°C$ until measurements were made.

A large amount of measured clinical data was provided for each patient in the study. This anonymised database provided over 100 variables per patient, and contained standard medical demographic information (e.g. age, sex), health and lifestyle indicators (e.g. weight, smoking, diabetes, blood pressure), and measured clinical parameters relevant to the presentation and treatment of the STEMI. Table 2.1 shows the distribution of patients across a number of these demographic and health indicators.

Table 2.1 Patient demographic, health and lifestyle distributions.

Clinical parameter	Group 1	Group 2	% missing
Patient sex	Female, 48	Male, 217	6.36
Previous cardiological history	No, 214	Yes, 106	6.36
Current smoker	No, 160	Yes, 106	6.01
Ex-smoker	No, 117	Yes, 106	29.68
Smoker or ex-smoker	No, 78	Yes, 184	7.42
Hypertension	No, 154	Yes, 112	6.01
Diabetes	No, 226	Yes, 40	6.01
Hypercholesterolemia	No, 165	Yes, 100	6.36
Family history of MI	No, 157	Yes, 109	6.01
Previous MI	No, 249	Yes, 16	6.36
Previous pPCI	No, 250	Yes, 15	6.01
Previous vascular disease	No, 134	Yes, 5	50.88
COPD or Asthma	No, 125	Yes, 16	50.88
Age over 60	No, 108	Yes, 155	7.07
Body mass index (BMI) over 25	No, 64	Yes, 174	15.90

It can be observed that there are highly uneven distributions with respect to some of these parameters, including gender and body mass index (BMI). This often stems

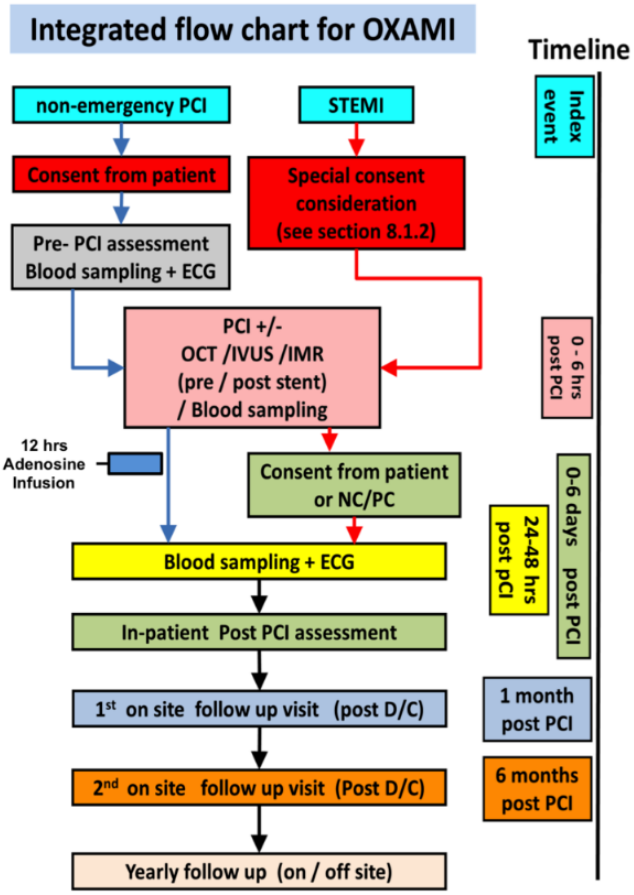

Figure 2.2 The study design and workflow for the OxAMI project when initiated in 2011. The coronary aspirate samples were collected at the time of PCI. Image replicated with permission from The OxAMI study webpage [199].

from the existence of common risk factors for cardiovascular diseases, such as male sex, high BMI, and past or present smoker. The OxAMI study design is shown in Figure 2.2, and shows the study timelines for patient enrolment and measurements. Not all data was available for all patients, therefore, different selections of mass spectra were used to investigate different parameters.

2.3 AAA samples - The OxAAA study

The OxAAA study (John Radcliffe Hospital, University of Oxford, Oxford, UK) is a prospectively enrolled cohort study of AAA patients who are undergoing routine National Health Service (NHS) management for AAA (Ethics Ref: 13/SC/0250).

The aims of the OxAAA study are to gain a better understand how an AAA grows and progresses over time and the underlying pathophysiology of the disease, and to identify potential biomarkers that may enable better patient monitoring and treatment [200]. As discussed previously, patients are currently elected for surgical repair of AAA based on AAA size, which may not be the best indicator of rupture risk in some patients. Identifying biomarkers related to AAA growth may enable more personalised treatment plans to be developed. Further information can be found on the clinical study website [79], and in a number of papers published by the clinical leads Regent Lee and Ashok Handa [78, 114, 201].

Each participant in the study gave written informed consent. Patients identified as having an AAA were screened according to the recruitment criteria for the study, and then allocated to either the surveillance stream (S1) or surgical stream (S2), as shown in the flow chart in Figure 2.3. The surveillance stream (S1, yellow track in 2.3) consisted of patients recruited with infrarenal AAAs with an aortic diameter greater than 30 mm and less than the surgical threshold of 50 mm (S1 = 30-50 mm, described in this study as a 'Small AAA'). Patients in the surgical stream (S2, orange track in 2.3) consisted of those with AAAs with an aortic diameter greater than the surgical threshold (S2 = $>$ 50 mm, described in this study as a 'Large AAA'). Post-surgery, patients routinely returned for a post-operative assessment in the outpatient clinic between 8 and 12 weeks, when the research assessment was also repeated.

Clinical measurements of AAA size and growth are recorded when patients attend the clinic for monitoring, and additional venous blood samples and measurements of FMD from the brachial artery are recorded [77]. A third stream of 'healthy volunteers' was also included in the study (HV, green track in 2.3). These were patients who went through the AAA screening programme, but no AAA was found. These volunteers therefore matched the common demographic for AAA, but the absence of other disease was not a selection criteria for healthy volunteer patients,

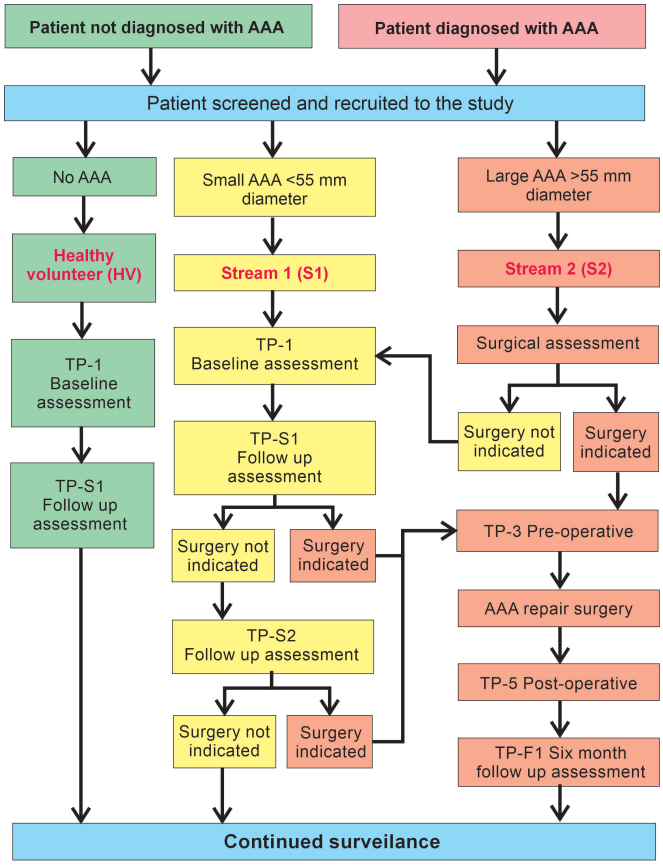

Figure 2.3 The study design and workflow for the OxAAA project, showing the healthy volunteer (HV, green), surveillance stream (S1, yellow), and surgical streams (S2, red). Note that S1 patients will move to S2 once the AAA size has reached the surgical threshold.

and so 'healthy' simply refers to the absence of AAA. AAA size was determined by ultrasound measurement, and was defined by the maximum antero-posterior (AP) diameter (outer to outer). The annual growth rate of AAA during surveillance was calculated using the following equation:

$$\text{Annual growth rate } = \frac{\Delta\text{APD}/\text{ APD baseline}}{n \text{ days lapsed } /365} \tag{2.1}$$

where ADP is the anteroposterior diameter (outer-wall to outer-wall measurement) of the AAA [77]. Patient groups included in this study are defined in Table 2.2.

Fasted venous blood samples were collected at each time point in EDTA tubes. Platelet-poor plasma was prepared immediately using a two-stage centrifugation, and then stored at -80 °C. If a patient underwent surgical intervention to repair

Table 2.2 The samples available for AAA patient groups used in the ASAP-MS study.

Patient treatment stream	Time point	Explanation	Patient samples available
S1	TP-1	Initial baseline assessment	24
S1	TP-S1	Surveillance point 1	89
S1	TP-S2	Surveillance point 2	7
S1	TP-S3	Surveillance point 3	5
S1	TP-S4	Surveillance point 4	7
S1	TP-S5	Surveillance point 5	4
S1	TP-3	Surgery pre-operative	4
S1	TP-5	6 weeks post-surgery	3
S1	TP-F1	Long-term follow up 1	1
S1	TP-F3	Long-term follow up 2	1
S2	TP3	Surgery pre-operative	99
S2	TP5	6 weeks post-surgery	44
S2	TP-F2	Long-term follow up 1	1
HV	-	Healthy volunteer	20

the AAA (patients in the S2 stream), intraoperative biopsy samples were taken from up to eight different tissue types around the AAA, the locations of which are described below and shown schematically in Figure 2.4. These were taken to enable mechanistic investigations into AAA growth and development. The distribution of plasma samples available across the different time-points is shown in Table 2.2.

As described in [77], tissue samples were processed immediately in the operating theatre. Tissue samples were rinsed thoroughly with normal saline to remove contaminating blood traces, and then stored at -80° until analysis. Biopsy samples were taken from the wall of the abdominal aorta, fat from the regions surrounding the AAA, and thrombus if present. Aortic wall samples were collected at proximal, maximal and distal locations with respect to the AAA. Fat samples were taken from subcutaneous, preperitoneal and omental fat deposits shown in Figure 2.4. Subcutaneous fat is the fat found directly under the skin, preperitoneal fat is found between the abdominal muscle and the peritoneum (the lining that encloses the abdominal organs), and omental fat is found in the abdominal cavity surrounding the internal organs. In a large number of AAA patients, a thrombus is found within the aneurysm, as described in Chapter 1. In cases where this is found, samples of

luminal (inner) and abluminal (outer) thrombus are taken. The thrombus is commonly removed during AAA repair. Unfortunately, there were very few patients for which plasma samples across different time points were collected, or for which all eight tissue samples were collected, meaning that most analysis done on the OxAAA samples were un-matched for patients.

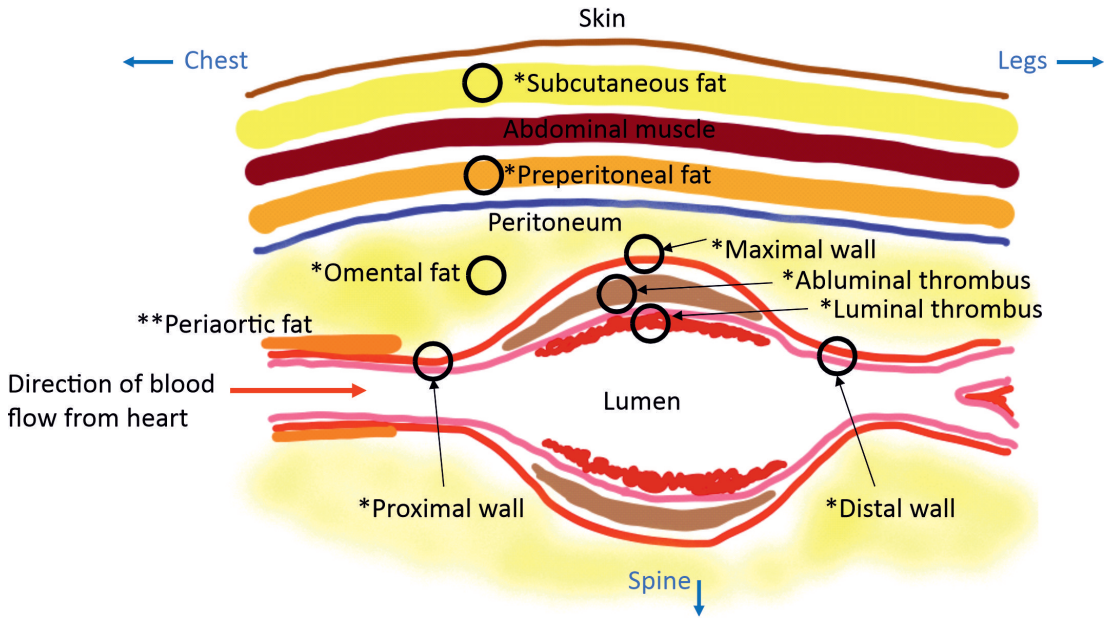

Figure 2.4 A schematic showing the types of biopsy samples that were taken during AAA repair surgery from the area in and around the AAA. The diagram represents a cross-section of the abdominal aorta within the abdominal cavity, with the proximal end (towards the heart) on the left, and the distal end (towards the legs) on the right. Sampling locations are indicated with (*).

2.4 Method development for the analysis of human blood plasma by ASAP-MS

2.4.1 Background

Determining the optimal approach to recording ASAP mass spectra for plasma samples requires optimisation of a variety of instrument settings, as well as consideration of a number of other factors. As noted earlier, the primary aims are

to collect high-quality and reproducible mass spectra, with the secondary objective of developing a protocol that is sufficiently short and simple to be suitable for adaptation to a clinical settings.

Key parameters to optimise are:

- **Amount of sample:** the amount of sample introduced to the mass spectrometer on each measurement;
- **Ion source settings:** The temperature and voltage settings employed within the ion source region;
- **Scan time:** The time period over which the quadrupole voltages scan through the m/z range in an individual scan;
- **Scan range:** The m/z range scanned by the quadrupole;
- **Acquisition time:** The time period for which the ASAP probe is inserted into the mass spectrometer for each sample measurement;
- **Number of repeat measurements:** The number of repeat measurements to take for each sample. These are averaged to generate the final mass spectrum for the sample for use in later data analysis.

Before working through the optimisation of each of these parameters in turn, there are few other important experimental considerations worth noting. Even with an optimised protocol, over time residual sample will build up on surfaces within the ion source. This can lead to considerable variability in subsequent measurements (see example in 2.6(a)). This can be prevented by cleaning the source according to the manufacturer's instructions. The frequency with which the ion source should be cleaned is dependent on both the sample type and the frequency of use. A general rule was applied that cleaning is required when the background signal is observed to rise by 10%, or when artefacts begin to appear in the chromatograms.

During method development, any consumables and solvents used should be assessed for the potential to introduce unwanted additional peaks into the mass spectra. Sample vials, solvents, ASAP capillaries, and cleaning wipes should all be considered and tested. The selection of consumables and cleaning regimen should be developed so that the minimum possible number of additional peaks are introduced. Baking the glass capillaries used in the ASAP probe for 30 mins at 250°C was found to be a suitable method of cleaning prior to their use in sample delivery. It was also found that the same capillary can be used for sequential repeat measurements of the same sample if it is cleaned sufficiently well in between measurements. A suitable method of cleaning the capillary is to wipe the tip with methanol and a lens tissue. The capillary can then be further cleaned if necessary by inserting it into a heat source such as a butane flame or the ASAP-MS source housing for a few seconds. Once the capillary has cooled, it can then be used to conduct the next repeat measurement on the sample.

Finally, some sources of variation between measurements must be dealt with during the data processing and analysis phase. Key amongst these is correction for the amount of sample introduced into the spectrometer on each measurement via an appropriate normalisation process. Spiking the sample with an internal standard is one approach, but adds complexity to the sample preparation protocol, and is generally not desirable. Various different data scaling and normalisation methods are available, and in general, it is worth trying several of these to find the most appropriate approach for a given study. An investigation into the impact of changing normalisation methods is described later in Chapter 7.

The number of freeze-thaw cycles that a sample has been through can also impact the spectra observed. An assessment of this using plasma samples concluded that the total time for which a sample was thawed resulted in a greater change in the spectra observed than the number of freeze-thaw cycles a sample had been through. Therefore, samples were held on ice during measurements, and re-frozen as quickly

as possible if they were to be re-used.

2.4.2 Materials and methods

Human blood plasma samples used in this study were from the OxAAA and Ox-AMI studies as described in Sections 2.2 and 2.3. Mass spectrometry analysis of plasma samples was conducted using an Advion CMS-S ASAP mass spectrometer. Parameter optimisation was conducted using experiments based on the standard experimental framework described in Section 2.1.1, with any altered experimental variables highlighted in the following descriptions of the various optimisation procedures. The practical work in this section was conducted with the help of Thomas Mills and Anna Jones, who were Part II students in the group.

2.4.2.1 Method development experiment 1: Amount of sample

While ASAP-MS has the advantages of allowing very quick and easy measurements, with minimal sample preparation required, the manual transfer of sample to the capillary tip of the probe has the consequence that the amount of sample introduced into the mass spectrometer varies on each measurement. Optimising the amount of sample transferred to the capillary tip is extremely important in order to obtain high-quality mass spectra. The most common error is to transfer too much sample to the probe tip, so one should generally err on the side of not introducing enough sample to the spectrometer. In the present work on plasma samples following the standard experimental framework laid out in Section 2.1.1, the thawed samples were first vortex mixed to ensure homogeneity, and then rather than dipping the probe into the plasma, sample was transferred to the probe tip by wiping it against the inside surface of the sample container approximately 5 mm above the sample surface. The thin layer of residual sample coated onto the surface after vortexing allows a small enough sample to be introduced into the spectrometer, whereas sampling from the bulk introduces far too much sample, as discussed further in the following. During

a measurement, one generally monitors the total ion current recorded by the mass spectrometer as a function of time, known as a chromatogram. An example of a set of well recorded chromatograms over five repeat measurements of the same sample, recorded using the sampling technique just described, which is shown in Figure 2.5.

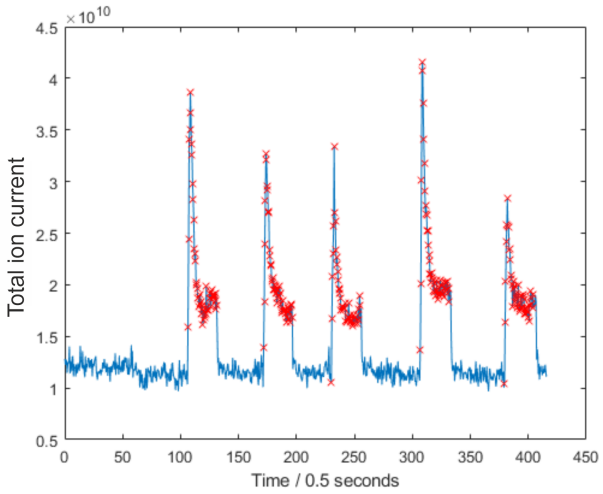

Figure 2.5 An example of an ideal chromatogram for a plasma sample. The red crosses represent each individual time point in the data acquisition where a capillary measurement was taking place. The spectrum at each time point is summed to generate the TIC signal for that time point.

The ion signal is strongly time dependent following insertion of the probe on each repeat. The total ion signal initially rises very rapidly as the most volatile molecules are vapourised, then falls off more slowly over a few tens of seconds until all of the sample has been desorbed. It is worth noting that the intensities of individual m/z peaks will show different time dependences, depending on factors such as the molecular enthalpy of desorption, vapourisation, or sublimation, the concentration of the molecule at the surface, and any interactions with other molecules in the sample. In contrast to the expected behaviour just described, Figure 2.6(b) shows a chromatogram recorded under conditions in which too much sample was introduced into the spectrometer. On insertion of the probe, the ion current rises to a high value and does not degrade with time, a clear sign of instrument saturation. Not only does introducing too much sample lead to highly unreliable m/z peak intensit-

ies, it will also result in rapid contamination of the ion source with residual sample and a corresponding need for frequent cleaning. Too much sample can also lead to highly variable peak profiles and intensities, including 'double peaks' in the chromatogram; an example of these effects is shown in Figure 2.6(c). While the peaks in the chromatogram vary slightly in intensity between measurements even with the best sampling technique, with some practice it should be possible to obtain repeat chromatogram peaks that are similar in shape and intensity.

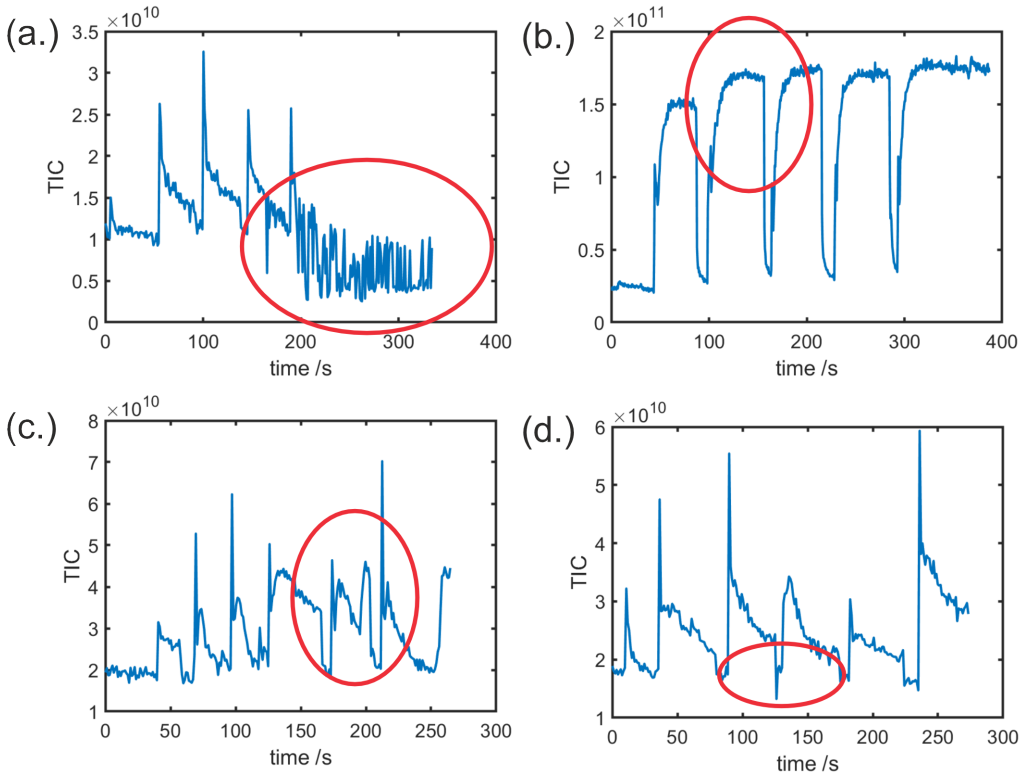

Figure 2.6 Examples of non-optimal chromatograms for OxAAA plasma samples arising as a result of various measurement errors. The resulting problems are highlighted by red circles: (a) Highly variable background and noise arising when the ion source needs to be cleaned; (b) Signal saturation due to introduction of too much sample (no sharp peak observed, and peak intensity remains high as long as the probe remains in the spectrometer); (c) Variable peak profiles and intensities due to inconsistent sampling and introduction of too much sample to the spectrometer; (d) Inconsistent acquisition times and insufficient time between measurements for the total ion count to return to background levels. Data recorded in 2020 by E. Ngetich.

2.4.2.2 Method development experiment 2: Ion source settings

The composition of the ion cloud entering the mass analyser is strongly dependent on both the temperature of the N_2 gas flow that desorbs sample from the probe tip, and on the potential through which the ions are accelerated into the inlet of the hexapole ion guide. At low temperatures only the most volatile molecules are desorbed, while at higher temperatures a wider range of molecules enter the gas phase. The Advion ASAP-MS instrument has pre-set temperature settings of 'low' (LT, 135°C), 'medium' (MT, 200°C), and 'high' (HT, 250°C). Increasing the acceleration potential of the ions as they enter the ion transfer capillary increases the energy of the collisions they undergo with the background gas. Increasing the collision energy increases the amount of collisional dissociation, in which larger ions break apart into smaller fragments. This leads to a reduction in the number of parent ions and an increase in the number of daughter or fragment ions observed in the mass spectra. The Advion ASAP-MS instrument has three pre-set acceleration potentials, accessed via the 'high fragmentation', 'medium fragmentation', and 'low fragmentation' ion source settings. When performing metabolic profiling, it is generally advantageous to minimise the extent of fragmentation in order to maximise the chances of matching mass spectrometric biomarkers to known molecular species within the sample. In addition, many metabolites dissociate to give one or more common fragments, which can make interpretation more challenging and obscure correlations within the data.

Given the above considerations, we predicted that the 'high temperature, low fragmentation (HT/LF)' setting would be optimum for plasma metabolic profiling: the high temperature would yield the largest variety of gas-phase molecules, while the low fragmentation setting would minimise dissociation so that most species are detected as intact parent ions. To test this, data were recorded using experimental protocol 2.1.1 for all eight available combinations of ion source settings, as listed in Table 2.3. The results are shown in Figure 2.7.

As the fragmentation setting is increased from 'low' to 'high', a greater number

Table 2.3 Ion source settings. The capillary and source offset voltages are applied to the ion transfer capillary and source region respectively (see Figure 2.1). The source voltage is scanned from the initial value at $m/z = 0$ across the range defined by the source voltage span parameter as the mass is scanned.

Ion source setting	Capillary Temperature (°C)	Capillary voltage (V)	Voltage Offset (V)	Voltage Span (V)	Gas Temperature (°C)	Corona Discharge (μA)
LT/LF	135	120	20	0	250	5
LT/MF	135	160	30	5	250	5
LT/HF	135	180	40	20	250	5
MT/LF	200	120	20	0	350	5
MT/HF	200	180	40	20	350	5
HT/LF	250	120	20	0	400	5
HT/MF	250	160	30	5	400	5
HT/HF	250	180	40	20	400	5

of lower mass ions are observed, as expected. As the temperature is increased, the total ion signal is seen to increase, and a higher number of higher-mass ions, formed from less volatile molecules within the sample, are observed. As predicted, the ion source settings of HT/LF were determined to be optimal, as this maximised both the intensity and the variety of intact molecules detected from the samples. These settings were used in all subsequent measurements.

2.4.2.3 Method development experiment 3: Scan time

The scan time chosen to scan across the m/z range of interest determines the number of mass spectra recorded while the sample is being desorbed. A slow scan results in detection of many ions per m/z step, and therefore a high signal-to-noise ratio, but risks missing some of the most volatile species if the scan is longer than the time period over which they desorb from the sample surface. In contrast, a fast scan yields high time resolution, but lower signal-to-noise. Scan time is a product of the rate at which the quadrupole mass analyser scans the DC and RF voltages, generating electric fields that control the passage of ions based on their m/z ratio. As noted above, each chemical species within the sample has its own characteristic desorption curve, with different species yielding a peak signal intensity at different times after

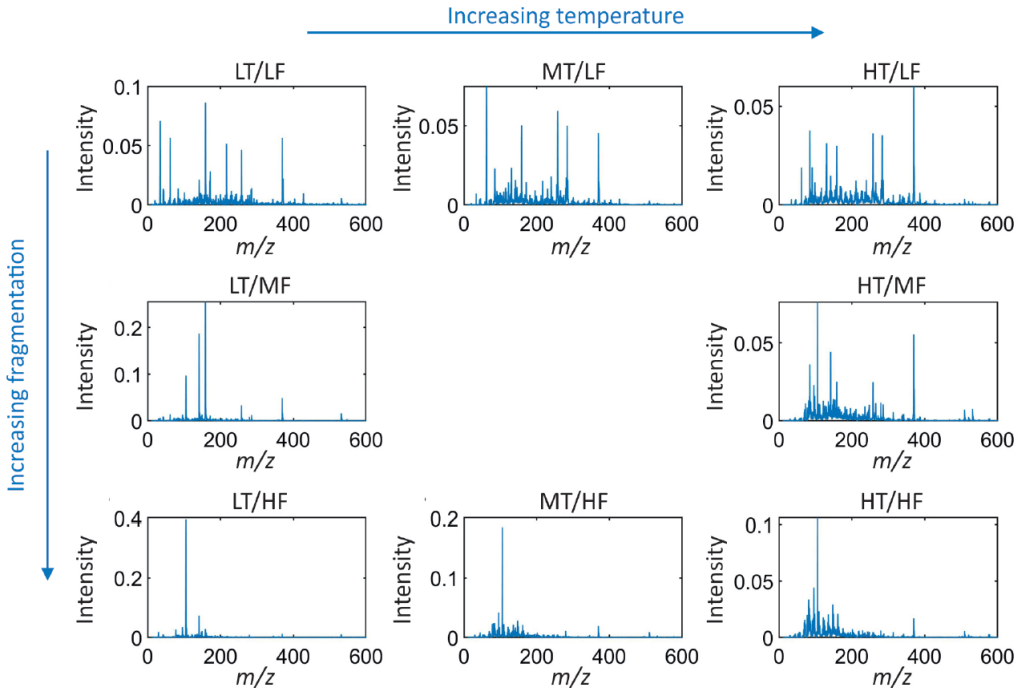

Figure 2.7 Mass spectra recorded for a human plasma sample using the eight different combinations of ion source settings listed in Table 2.3. The spectra were recorded with a scan time of 500 ms and an acquisition time 20 s. The spectra shown are an average over five repeats, each averaged over the 40 time-resolved spectra recorded during each acquisition time. 'LT', 'MT', and 'HT' refer to low, medium, and high temperature, respectively, while 'LF', 'MF', and 'HF' refer to low, medium, and high fragmentation.

insertion of the probe into the ion source. In general, the signals associated with lighter, more volatile molecules tend to peak at earlier times, while those from heavier and less volatile molecules peak at later times. The optimum scan time to choose is therefore sample dependent, and is a key parameter to be investigated in any ASAP-MS study. To determine the optimum scan time, measurements on human plasma were performed using scan times ranging from 200 ms to 1000 ms at 100 ms intervals. For each scan speed, five repeat measurements were recorded using experimental protocol 2.1.1, with the probe inserted into the mass spectrometer for 20 s on each measurement. Scans were performed over the full 10-1200 m/z mass range of the instrument. The corresponding scan speeds (in seconds per unit m/z) are shown in Table 2.4. The effect of employing different scan speeds was assessed by considering both the signal-to-noise ratio and the number of m/z peaks observed,

with the aim being to maximise both of these quantities. For this purpose, a 'peak' was identified as any signal more than three times the standard deviation of the baseline, while the signal-to-noise ratio was defined as the average ratio between the intensity of each identified peak and the standard deviation of the baseline. Figure

Table 2.4 Instrument scan time for a scan range of 10-1200 *m/z*.

Scan time /ms	Scan speed / $(m/z)s^{-1}$
200	5950
300	3967
400	2975
500	2380
600	1983
700	1700
800	1488
900	1322
1000	1190

2.8 shows that the scan time that provided the highest values for both quantities of interest was 900 ms, though the performance is similar across a range of scan times from 600-1000 ms. A scan time of 900 ms was used in subsequent measurements.

2.4.2.4 Method development experiment 4: Scan range

As noted previously, the Advion ASAP-MS instrument has a full scan range spanning 10 to 1200 *m/z*. However, in our measurements on blood plasma, no significant signal was observed above 1000 *m/z*, so the scan range in subsequent measurements was reduced to 10-1000 *m/z*. Reducing the scan range has the result that for a given scan time, the dwell time on a given *m/z* is increased, increasing the number of ions detected and improving the signal-to-noise ratio. We note that the ionisation efficiency and transport of ions through the instrument, and therefore the signals detected as a function of mass, can vary significantly depending on the details of the instrument design and settings, so the appropriate mass range over which to scan should be determined separately for each new instrument and sample type.

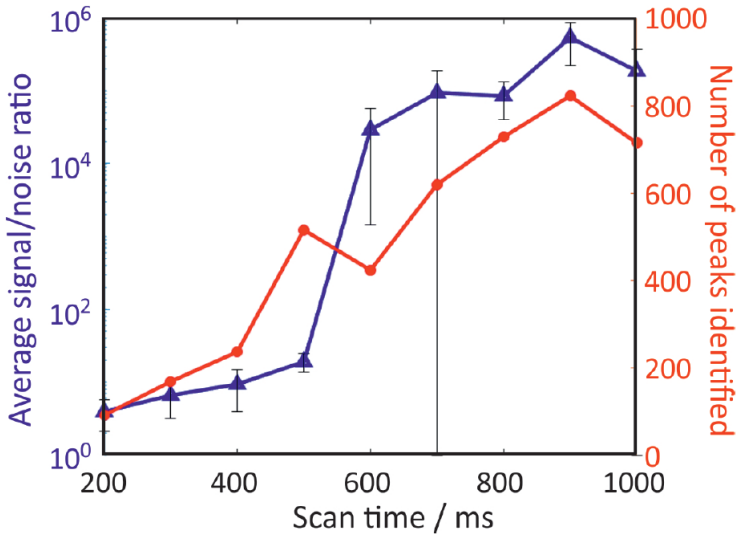

Figure 2.8 The average signal-to-noise ratio (blue) and number of peaks identified (red) for scan times in the range from 200 to 1000 ms. Data were averaged over five repeat measurements, with the exception of the 800 ms data, which were averaged over four repeats due to an anomalous measurement. One-standard deviation error bars in the signal-to-noise are shown.

2.4.2.5 Method development experiment 5: Acquisition time

Having chosen a scan speed, the acquisition time, i.e. the amount of time for which the probe is inserted into the mass spectrometer, determines the number of spectra that will be recorded during each measurement. One of the goals of the present work is to develop a rapid protocol suitable for making measurements on large numbers of samples in a clinical setting, and choosing a suitably short acquisition time is therefore a key consideration. However, the acquisition time should be long enough to capture signals from both volatile and non-volatile species within the sample, and to generate enough spectra to ensure that averaging the spectra over the acquisition time yields an acceptable signal-to-noise ratio. It is also worth noting that if the acquisition time is too long then the sample can start to burn on the surface of the capillary, after which artefacts may appear in the spectra, and any further information gained will be minimal. The optimum acquisition time was determined via an Allan variance [202, 203] analysis, as follows. Allan variance is

an analysis that takes into account the frequency stability of time-resolved analysis, and is often used in characterising clocks, oscillators and amplifiers. To generate a suitable data set for the analysis, measurements were made on two plasma samples. For each sample, three repeat measurements were made using experimental protocol 2.1.1 with an acquisition time of 60 s for each repeat. For these measurements the scan range was set to 10-1200 m/z and the scan time was 500 ms. These are slightly different from the previously determined optimum settings for these two parameters. The scan range was set to the full mass range of the instrument in order to pick up any new high-mass peaks appearing as a result of sample degradation, and the scan time was reduced in order to increase the number of scans per acquisition cycle for the purposes of the Allan variance analysis. The Allan variance, $\sigma_x^2(t)$, is used to determine the intrinsic noise in a measurement of variable x (in our case x represents the m/z peak intensities) as a function of the averaging time, t:

$$\sigma_x^2(t) = \frac{1}{2(N-1)} \sum_{i=1}^{N-1} (x_{i+1} - x_i)^2 \tag{2.2}$$

where N is the total number of measurements made during the acquisition time and i is the index of a given measurement. Plotting the Allan variance as a function of the acquisition time (or number of measurements) quickly reveals the optimum acquisition time. At short times the variance is dominated by measurement-to-measurement noise, and therefore reduces rapidly as the measurement time is increased and a large number of measurements are averaged. At long times the variance becomes dominated by experimental drift, and begins to increase again. The optimum measurement time is the time at which the curve passes through a minimum. Allan variance plots for each repeat measurement on the two samples are shown in Figure 2.9. Based on the plots, the optimum acquisition time was determined to be between 20 and 30 s, and an acquisition time of 25 s was therefore chosen for all subsequent measurements. We note that in addition to standardising the acquisition time, it is important to leave enough time in between repeat meas-

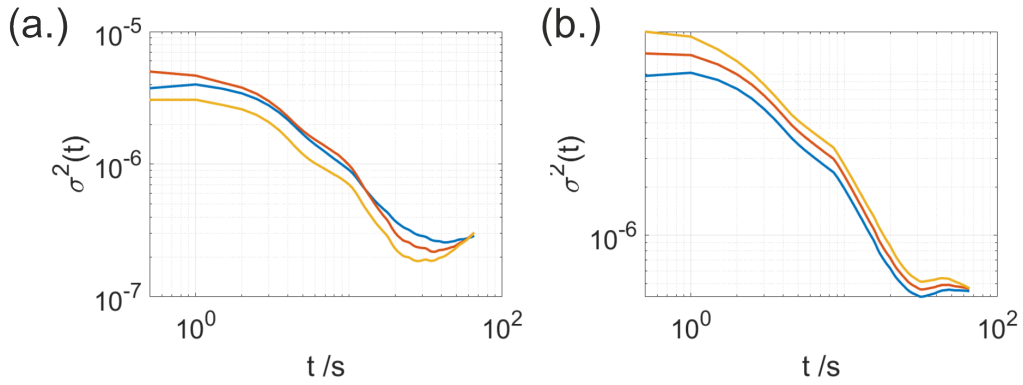

Figure 2.9 Allan variance as a function of acquisition time t for two plasma samples, (a) and (b). Three repeat measurements for each sample are shown by the blue, yellow, and red lines.

urements for the signal to fall back to baseline levels. Failing to do this can lead to problems in automating the data processing, as well as possible contamination of spectra with residue from previous samples. Figure 2.6(d) shows chromatograms for a data set in which such problems arise.

2.4.2.6 Method development experiment 6: Number of repeats

Averaging the mass spectra over a number N of repeat measurements for each sample improves both the signal-to-noise and the reproducibility of the measurement. For a given sample, while we expect some variation between individual repeat measurements, we expect the spectra to converge to a constant form as N is increased. To determine the number of repeats after which this convergence is achieved, a data set was generated using experimental protocol 2.1.1 comprising 50 repeat measurements on a single sample. The spectra were recorded over a mass range of 10-1000 m/z, with a scan time of 500 ms and an acquisition time of 25 s. For each value of N from 2 to 50, a set of N spectra were chosen randomly from the 50 measurements, and averaged. This process was repeated 100 times to obtain a set of 100 averaged spectra for each value of N. For each N, the mean peak intensity $\bar{I}_{m/z}(N)$ and standard deviation across the 100 averaged spectra were determined for each m/z peak. The mean intensities were used to determine the summed fractional change ΔI_N in peak

intensity across all mass peaks when N is incremented by one, according to the following expression:

$$\Delta I_N = \sum_{m/z} \frac{\bar{I}_{m/z}(N) - \bar{I}_{m/z}(N-1)}{\bar{I}_{m/z}(N-1)} \tag{2.3}$$

The plot of Equation 2.3 against N shown in Figure 2.10 reveals how the averaged spectra change with the addition of further measurements.

We see that the mean intensity variation converges towards a constant value, and the standard deviation in those measurements decreases. The decrease in mean intensity change is rapid in the range from $N = 1$ to $N = 10$, and then remains reasonably constant for larger N.

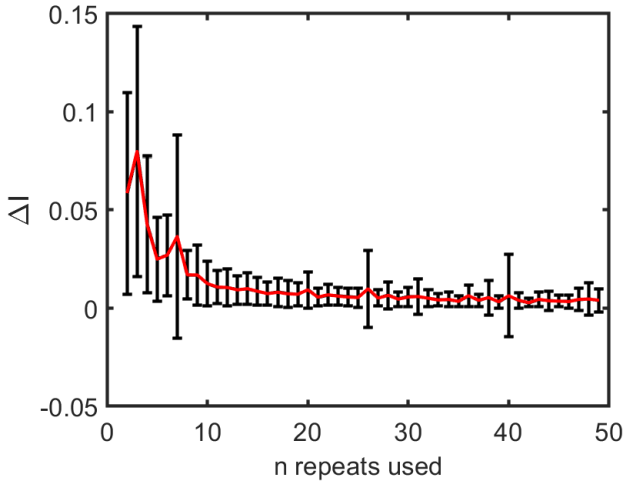

Figure 2.10 The change in average normalised intensity (ΔI) across all *m/z* peaks as a function of number of repeats averaged (N). One-standard-deviation error bars are shown.

Based on this analysis, we conclude that averaging the data over 10 repeat measurements for each sample provides a good balance between optimising the reproducibility and quality of the measurement and maintaining a rapid measurement protocol. As noted earlier, after five repeat uses of the same capillary, residue on the capillary tip begins to affect the background measurement. Consequently, the ten repeat measurements were obtained as two cycles of five repeat measurements, with a change of capillary in between the two cycles. The complete optimised meas-

urement protocol can be found in Section 2.4.4.

2.4.3 Results

A sample data set recorded for an OxAAA plasma sample using the optimised measurement protocol (described in Section 2.4.4) is shown in Figure 2.11. Chromatograms for all measurements are included, with mass spectra for the second set of five measurements also shown to provide a qualitative indication of the reproducibility of individual measurements. Considering first the chromatograms, we see that during the background measurement at the start of each trace, the intensity is low and stable. The ten repeat measurements on the sample yield well defined spikes in signal, with an initial sharp rise and a slower decay. Not all biological samples give rise to the same peak shape in the ion chromatogram. In particular, samples with a high fat content tend to yield more variable peak shapes. The initial sharp rise is largely associated with desorption of cholesterol, which tends to be one of the most intense and variable peaks in the mass spectra, appearing at m/z 369 for the dehydrated ion (cholesterol parent peak at m/z 387 $-H_2O$). In between each measurement, and after the final measurement, the total ion count returns rapidly to the baseline intensity level. This indicates that sample ions have cleared the spectrometer and that the instrument is clean. The clear, well-separated signals in the chromatogram make it very straightforward to automate the extraction of individual measurements from the data file.

Looking at the individual mass spectra shown in Figure 2.11(b), we see that though there are two intense peaks whose intensity varies considerably (plotted in a lighter colour), overall the mass spectra are reasonably reproducible from measurement to measurement. To gain a more quantitative insight into the degree of reproducibility, ten repeat measurements were made on each of 20 plasma samples donated by healthy volunteers as part of the OxAAA study. For each measurement, the time-resolved mass spectra were averaged over the 25 s acquisition time to give a single

mass spectrum, yielding 10 mass spectra for each of the 20 samples. The 10 mass spectra were used to calculate the coefficient of variation (CV) for each sample as a function of m/z, defined as the ratio of the standard deviation to the mean across the 10 repeat measurements for each m/z value above a chosen threshold intensity in the mass spectrum. An intensity threshold of 5×10^5 counts was chosen based on visual inspection of peaks in the spectra. The mean and standard deviation of the resulting CV (m/z) functions over the 20 different samples then provide a reasonable measure of the typical variability in the measurements. Figure 2.11(c) shows the result of this analysis. Inspecting the plot of the averaged coefficient of variation, we see that CV is fairly constant across all peaks, perhaps increasing somewhat with m/z. Values for individual m/z peaks range from 20-50%, with an average value of 39.9%. While we were not able to find any other ASAP-MS studies on human plasma with which to compare our data, the CV determined in the present work is similar to that reported by Qian et.al for LC-MS measurements of blood plasma in a label-free quantification experiment [204, 205], suggesting that our approach provides data of at least comparable quality to more sophisticated measurement techniques.

2.4.4 Optimised measurement protocol for ASAP-MS measurement on human plasma

To summarise, for optimum performance as determined by the processes just described, the ASAP-MS instrument is operated in positive ion mode with the ion source settings set to "high temperature, low fragmentation" mode. The mass range is selected to be 10-1000 m/z, and the scan time is set to 900 ms. The acquisition run is started. A clean capillary is mounted into the tip of the ASAP instrument probe. The probe is inserted into the probe slot in the ion source housing for 30 s, and a background or blank spectrum is recorded for the clean capillary during this time. The background spectrum is subtracted from each sample measurement during the

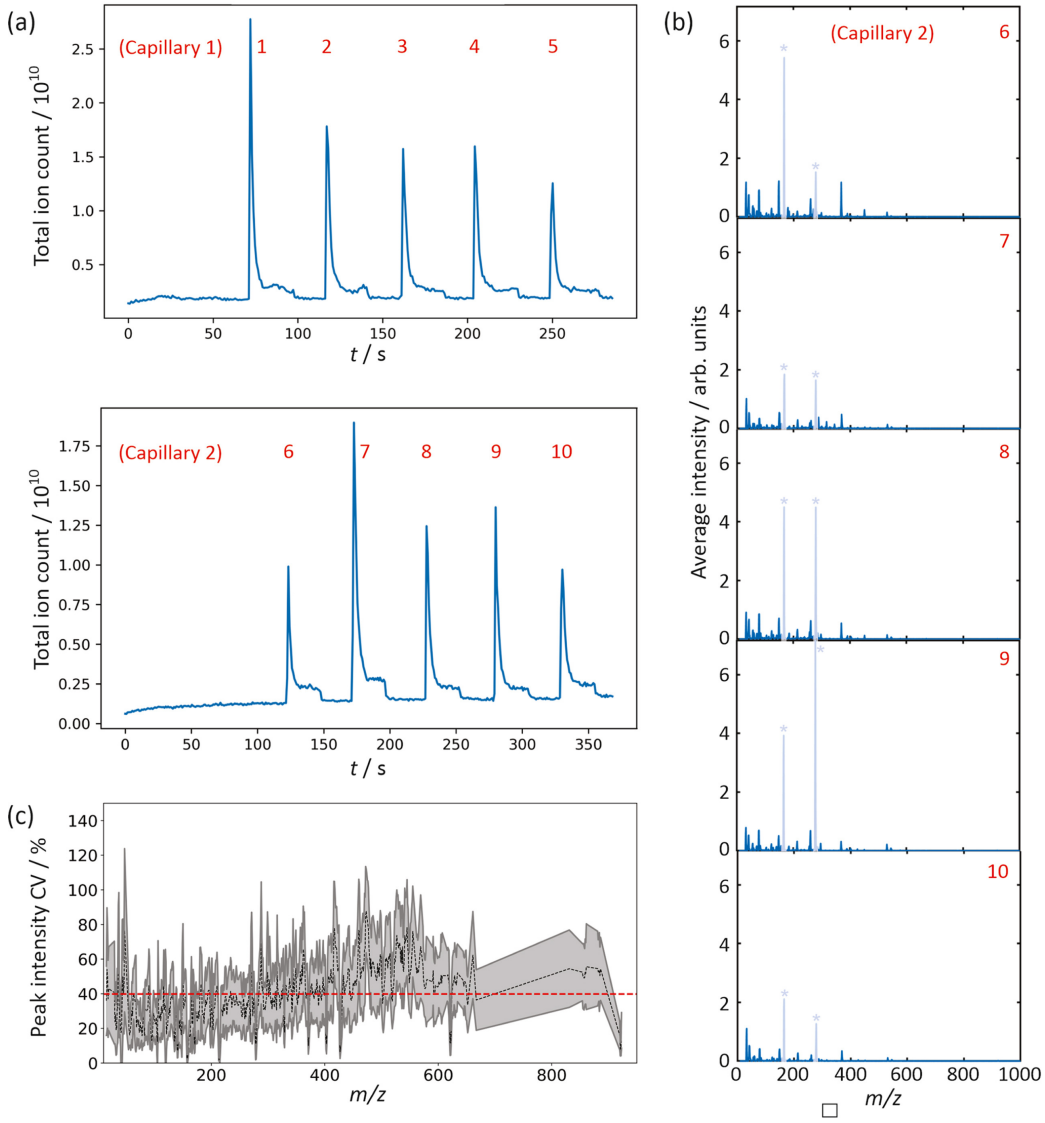

Figure 2.11 (a) Chromatograms for a set of ten repeat measurements (five each with two different capillaries) on an OxAAA sample recorded using the optimised method; (b) Individual mass spectra recorded with the second capillary; (c) Mean (black dashed line) and standard deviation (grey solid region) of the coefficient of variance in *m/z* peak intensities, determined as described in the text. The average coefficient of variance across the entire spectrum is 39.9%, shown as a red dashed line on the plot.

data processing. The probe is then removed from the spectrometer, cleaned and cooled with methanol, and dried on a lens tissue. The plasma sample to be measured is thawed and vortex mixed. The probe is then used to wipe the inside surface of the sample container approximately 5 mm above the sample surface. The probe is then reinserted into the mass spectrometer for a period of 25 s, during which a series of mass spectra are recorded for the sample. The probe is then removed from the mass spectrometer and cleaned with methanol and lens tissue ready for the next repeat measurement. Five repeat measurements are made with the same capillary. After this the complete set of measurements, including background measurement, are repeated with a new clean capillary, for a total of ten measurements per sample.

Having developed an optimised measurement protocol for plasma samples, we move on to consider similar measurements on tissue biopsy samples.

2.5 Method development for the analysis of human tissue samples by ASAP-MS

2.5.1 Background

Analysis of tissue samples is a valuable way to obtain information about localised pathology of a disease, but introduces further levels of complexity with regards to sample handling for MS analysis compared to plasma. In order to develop an experimental protocol suitable for tissue samples, the optimisation experiments described in Section 2.4 for plasma samples were repeated on tissue samples from the OxAAA study, along with other tissue-specific optimisation.

In contrast to human plasma samples, tissue samples cannot be assumed to be structurally homogeneous. There is a large degree of heterogeneity within a single tissue sample, in addition to the variation between different tissue types and subtypes. Homogenisation of tissue samples may be required in order to achieve reproducible spectra between repeat measurements. However, this does introduce addi-

tional sample preparation steps into the method, which is disadvantageous within a clinical setting. Part of the required optimisation is therefore to investigate whether direct sampling of whole tissue samples can produce comparable results to measurements on homogenised tissue samples.

As discussed previously, the conditions in the ion source impact the degree of desorption, ionisation, and fragmentation of sample molecules. The quadrupole scan speed may also impact the observed spectra through changing the signal-to-noise ratio and signal intensity. The primary objectives are similar to those for plasma measurements, i.e. to identify a measurement protocol that enables the collection of high-quality and reproducible data across the different tissue samples, with the secondary objectives that the protocol be short and require very little sample preparation so that it might be suitable for adaptation to a clinical setting. Ideally, the optimum conditions for the measurement of each tissue type would be the same as those found for the measurement of plasma samples, making comparison between the matrix types easy to interpret. The optimised method developed for plasma was therefore used as a starting point for the investigation into measurements on tissue biopsy samples. The parameters to optimise for human tissue samples were scan time, acquisition time, and number of repeat measurements. The effect of freeze-thaw cycles and homogenisation must also be considered.

2.5.2 Materials and methods

Biopsy samples avaliable from the OxAAA study were described in Section 2.3. Samples of the eight different tissue types (abluminal thrombus, luminal thrombus, subcutaneous fat, omental fat, preperitoneal fat, distal wall, proximal wall and maximal wall) were collected from four different patients. Mass spectrometry analysis of the samples was conducted using the same Advion CMS-S ASAP mass spectrometer employed for the previously described plasma measurements, and was based on the standard experimental protocol described in Section 2.1.1. The experiments in this

section were completed with the assistance of Lloyd Smith, a Part II student in the group in 2023-2024. Homogenisation was conducted using a ™BeadBug Mini Homogeniser and 3 mm zirconium homogeniser beads in pre-filled sterile 2 mL sample tubes. The method for loading sample onto the ASAP probe tip was adapted for tissue samples. The tissue sample to be measured was thawed, and the sample tube kept cool in an ice bucket whilst analysis was conducted in order to prevent sample degradation. For whole tissue samples, the probe was swabbed across the surface of the sample, and then gently wiped on a piece of clean lens tissue to remove excess tissue residue, preventing saturation of the instrument. For homogenised samples, the same residue sampling protocol employed for plasma was followed. The probe was then inserted into the mass spectrometer for a period of 25 s, during which a series of mass spectra are recorded for the sample. This was then repeated for ten repeat measurements, with five repeats per capillary.

2.5.3 Parameter optimisation

2.5.3.1 Method development experiment 7: Tissue homogenisation

Assessment of whether homogenised and whole tissue samples give comparable ASAP-MS spectra was conducted over the different tissue subtypes. Tissue samples showed a high degree of variation in composition, size, and colour within each sample type. Images of a small number of these samples are shown in Figure 2.12. The most consistent samples were within the fat tissue class, which were often pale and contained little liquid. The aneurysm wall and thrombus samples were less consistent in both composition and size, with some sample tubes also containing whole blood residues. Samples were sectioned and measured both whole and homogenised. Each tissue was defrosted to room temperature, and then 0.05 g of tissue was cut from the sample, placed in the homogeniser tube with 0.1 mL of deionised water, and homogenised with 3 × 30 s bursts at a frequency of 350 rpm, with a 30 second break in between each 30 s burst. The homogenised samples were analysed by ASAP-MS

using the experimental protocol employed to assess the optimum number of repeat measurements, as described in Section 2.4.2.6.

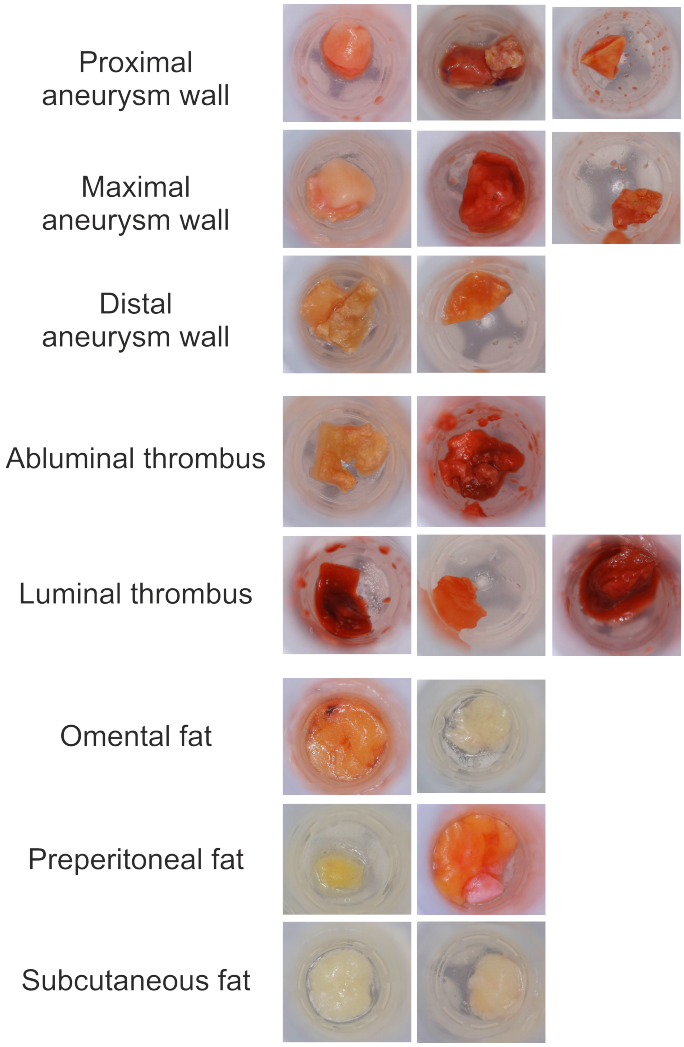

Figure 2.12 Images of selected OxAAA tissue samples for the wall, thrombus, and fat samples used. Images were taken with a standard USB camera. Samples were left in the standard sample tubes. The base of each tube is approx 10 mm in diameter.

The remaining tissue was measured using the same experimental protocol but using the tissue surface sampling method. An assessment of signal convergence with number of repeats, as described previously in Section 2.4 was performed, with the results shown in Figure 2.13. It was expected that ΔI_N would converge across fewer repeats for the homogenised samples than for the non-homogenised samples. However, no significant difference was found between the whole tissue and homogenised samples. In a small number of cases, notably the 'maximal aneurysm wall' tissue (see

Figure 2.13(d)), the 'whole tissue' average spectrum stabilised with fewer repeats than the homogenised sample. Further investigation revealed that a notable dilution

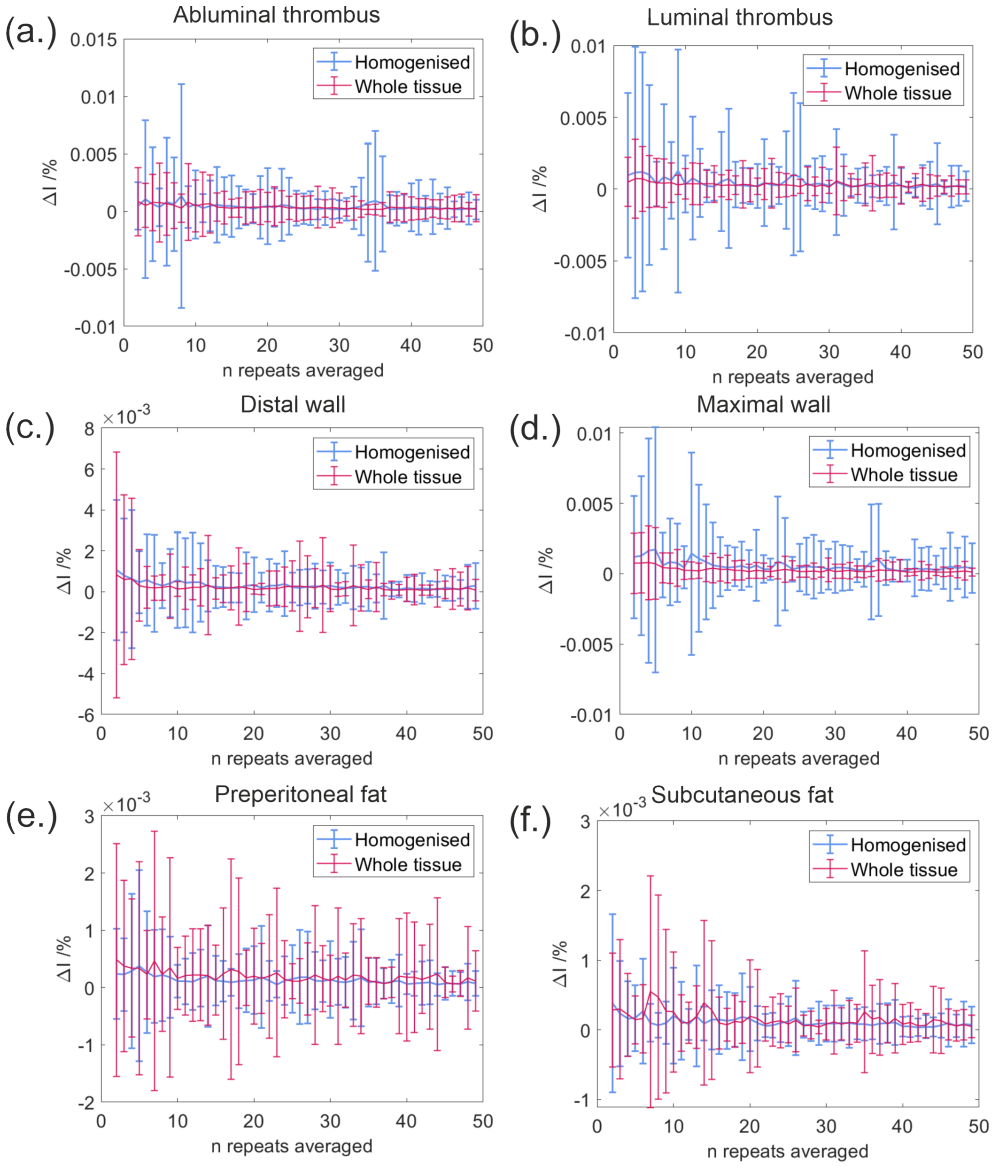

Figure 2.13 The absolute change in spectra intensity with number of repeats for homogenised and unhomogenised tissue for a) abluminal thrombus, b) luminal thrombus, c) distal wall, d) maximal wall, e) preperitoneal fat, and f) subcutaneous fat.

effect was present in the homogenised samples due to the addition of water, and a number of mass peaks were suppressed to such an extent that they were no longer visible in the sample spectrum. As no clear benefits of homogenisation were found, and noting the inconvenience that homogenising would bring to an intra-operative or other clinical setting, all subsequent tissue measurements were conducted on whole

tissue samples.

2.5.3.2 Method development experiment 8: Scan time

The optimum scan time was assessed using the method described in Section 2.4.2.3 and by maximising the number of observed peaks in each mass spectrum, as well as the corresponding signal-to-noise ratio. This analysis was conducted for each of the three tissue types (all wall, thrombus and fat tissue subtypes averaged into the three tissue classes) from three patients, and averaged to generate the results shown in Figure 2.14. It was observed that mass spectra corresponding to fat samples were

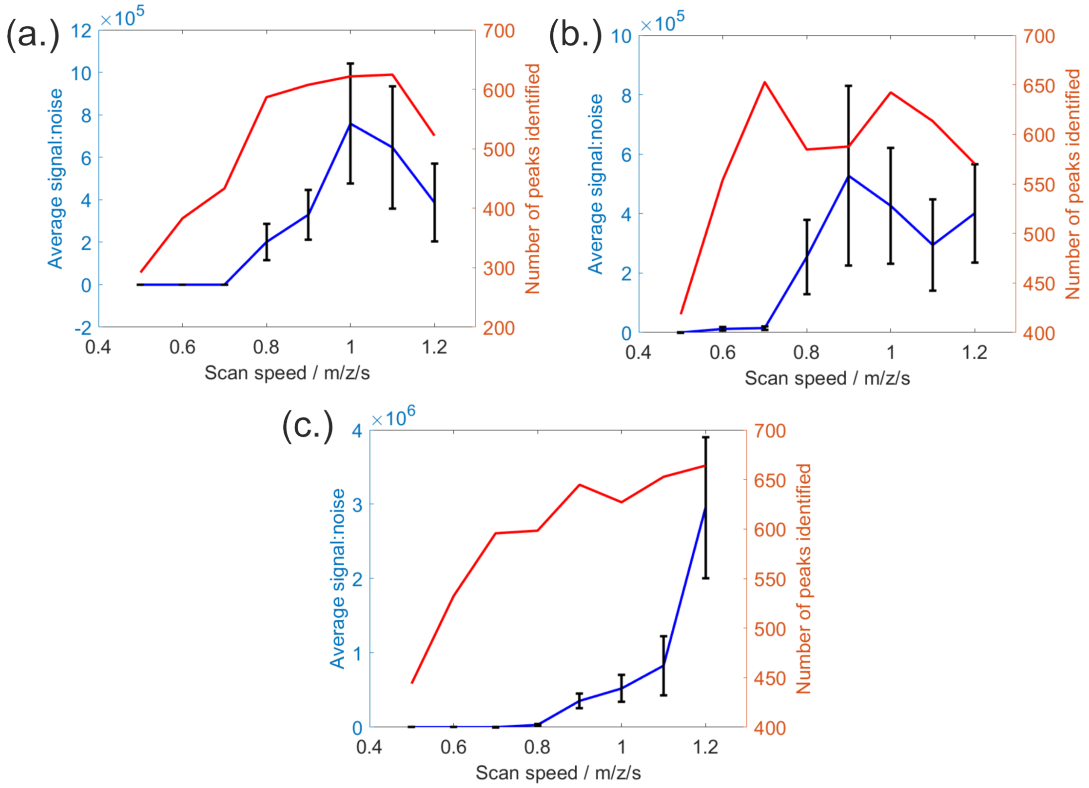

Figure 2.14 The scan time optimisation plots for a) thrombus, b) artery wall, c) fat, showing signal-to-noise ratio against peak count. Data is averaged for each tissue type over patients 71, 73 and 77. Error bars show the signal-to-noise sample error.

optimal at higher scan times, yielding the highest signal-to-noise and peak count. The metabolite profile for fat is largely dominated by high-mass lipids and glycerides, which have lower desorption rates on average than metabolite smaller ions. The ideal scan time for both thrombus and wall measurements demonstrated favourable peak

counts and signal-to-noise ratios for scan times of 0.9 and 1.0 seconds. For fat, the optimal scan time was observed at 1.2 s, but favourable values for both peak count and signal-to-noise were observed from 0.7 - 1.2 s. To allow for straightforward comparisons to be made between tissue types, and to reduce the possible loss of signal from small transient molecules, a scan time of 0.9 s was used for subsequent measurements.

2.5.3.3 Method development experiment 9: Acquisition time

The optimum acquisition time for tissue measurements was assessed using the method described in Section 2.4.2.5, with a scan time of 0.9 s. An Allan variance analysis as a function of scan time was calculated for each different tissue type from three patients. The locations of the minima in the resulting Allan variance curves, indicating where noise and drift were simultaneously minimised, are shown in Table 2.5 for each tissue type.

Table 2.5 Allan variance analysis of acquisition time for tissue samples.

Sample	Patient	Allan Deviation Minimum /s
Pre-peritoneal fat	71	20 - 30
Pre-peritoneal fat	73	20 - 30
Pre-peritoneal fat	77	N/A
Abluminal thrombus	71	N/A
Abluminal thrombus	73	N/A
Abluminal thrombus	77	N/A
Maximal wall	71	20 - 30
Maximal wall	73	20 - 30
Maximal wall	77	N/A

The maximal wall and preperitoneal fat samples showed distinctive minima in most analyses at 20-30 s acquisition time, which is inline with that found in similar measurements on human plasma in Section 2.4.2.5. The Allan variance for abluminal thrombus did not reach a minimum in any analysis. Example plots revealing a distinctive minimum for maximal wall, and no minimum for abluminal thrombus, are shown in Figure 2.15. Further example results can be found in Appendix A.1. A possible explanation for this difference in behaviour is that the presence of whole

blood in thrombus samples, which was susceptible to burning on the capillary tip, results in changes in the ions detected as the acquisition progresses. An acquisition time of 25 s was implemented following this analysis, which was optimal for fat and aneurysm wall samples, and was also sufficiently short to minimise the effects of sample burning during thrombus measurement.

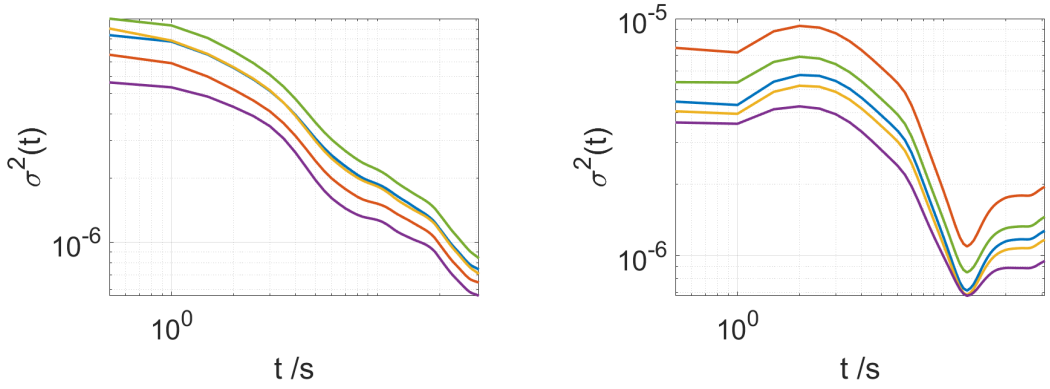

Figure 2.15 Allan variance analysis of (a) abluminal thrombus, and (b) maximal wall. Each colour represents one of 5 individual repeat measurements. Note the distinctive minimum between 20 and 30 seconds in (b), which is not present in (a).

2.5.3.4 Method development experiment 10: Number of repeat measurements

The optimum number of repeat measurements to acquire and average over was assessed using the method described previously for plasma measurement in Section 2.4.2.6. Data were recorded for all tissue types over three different patients. The quantity ΔI_N is plotted as a function of number of repeats in Figure 2.16 for each tissue type. The data are shown with associated errors for each individual patient in Appendix A.2. For each sample, the average ΔI_N is observed to converge at approximately 10 repeats. This is consistent with the previous measurement on plasma. Again, an optimum of 10 repeats was chosen as a compromise between maximising data quality and minimising the analysis time in accordance with the overarching research aims. The number of times that a capillary can be re-used for measurement of biological samples before metabolite build-up occurs was determined

previously to be five, so to analyse a sample over 10 repeats, two acquisitions of five repeats each were conducted.

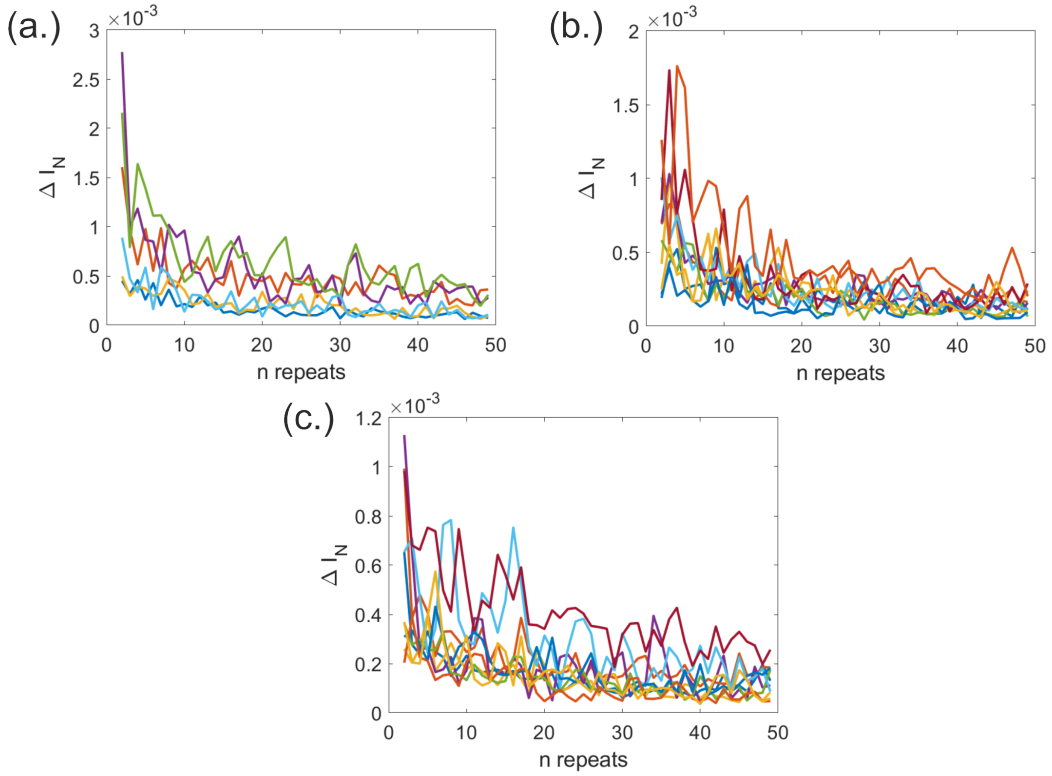

Figure 2.16 Analysis of the change in ΔI as the number of repeats averaged over increases for (a) thrombus, (b) wall, (c) and fat tissues. Each coloured line represents a tissue samples analysed.

2.5.4 Results

The optimised method (described below) was used to record data for 20-30 patients for each different tissue subtype. The CV for each tissue type was calculated, the average of which is shown in Figure 2.17. The CV for plasma, calculated in the previous section to be 39.9%, is shown for comparison in orange. The CV values for thrombus and wall samples were found to be similar to those for plasma, between 20 and 40%, which is in the range similar to that described in the literature for desorption electrospray ionisation mass spectrometry (DESI-MS) analysis of human tissue, if slightly higher [205–207]. A slight increase is expected for ASAP-MS when compared to DESI-MS given that the method of sample introduction is generally

automated and therefore more reproducible in DESI-MS. The CV for fat samples was found to be higher, at $\geq 60\%$. A possible explanation for this is the decreased sensitivity and reproducibility of the ASAP mass spectrometer for measurements made in the higher mass range of > 700 m/z. This results in higher variation in signals arising from high-mass lipids, which are found in high concentrations in fat samples.

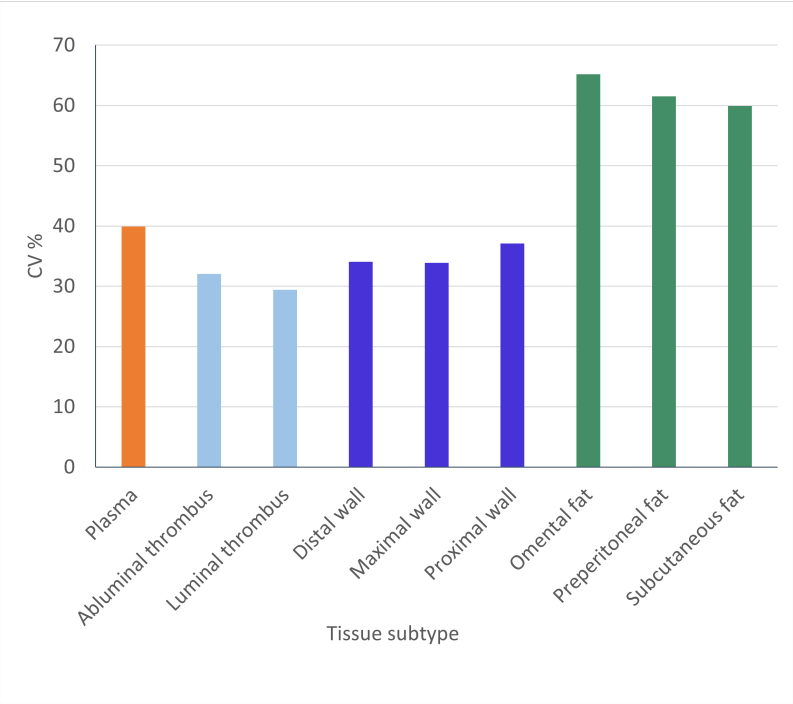

Figure 2.17 The CV% calculated for each tissue subtype calculated over 20-30 patients. The plasma CV% calculated previously is shown in orange.

2.5.5 Optimised measurement protocol for ASAP-MS measurement on human tissue biopsy samples

The final optimised method used for the recording of ASAP-MS is as follows: The ASAP-MS instrument is set up in positive-ion mode with the ion source settings set to "high temperature, low fragmentation". The mass range is selected to be 10-1000 m/z, and the scan time is set to 900 ms. The acquisition run is started. A clean glass capillary is mounted into the tip of the ASAP instrument probe. The probe is inserted into the probe slot in the ion source housing for 30 s, and a

background or blank spectrum is recorded for the clean capillary during this time. The background spectrum is subtracted from each sample measurement during the data processing. The probe is then removed from the spectrometer, cleaned and cooled with methanol, and dried on a lens tissue. The tissue sample to be measured is thawed and held in an ice bath (within its sample tube) during sampling. The probe tip is used to swab the top surface of the sample, and is then gently wiped on a piece of clean lens tissue to remove excess sample residue. The probe is then reinserted into the mass spectrometer for a period of 25 s, during which a series of mass spectra are recorded for the sample. The probe is then removed from the mass spectrometer and cleaned with methanol and lens tissue ready for the next repeat measurement. Five repeat measurements are made with the same capillary. After this the complete set of measurements, including background measurement, are repeated with a new clean capillary, for a total of ten measurements per sample.

2.6 Conclusions and outlook

Optimised methods have been developed for the measurement of human blood plasma, and tissue, using ASAP-MS. This method will enable untargeted metabolomics profiles to be recorded for patients with a range of diseases. The optimised protocol allows rapid measurements to be made by a variety of operators, and does not require extensive prior experience or training in mass spectrometry. The coefficient of variation between measurements is within the range commonly observed when using other techniques to analyse human plasma. Based on our experience to date, we believe that ASAP-MS is a promising technique for translation into clinical settings, and has considerable potential to be used in clinical diagnostic and prognostic assessments.

3

Data analysis methods

This chapter begins by describing the data processing methods employed to extract the relevant spectra from each measurement file, remove any unwanted background signal, perform appropriate signal averaging, and carry out normalisation and additional pre-processing steps. All of these steps are designed to preserve as much chemical information as possible within the data, with the aim of this to prepare the data for input to machine learning algorithms, which will be applied to predict clinical data. The remainder of the chapter will describe the analytical methods used to explore correlations between the molecular information in the mass spectra and the clinical data available for the patient groups, with the eventual goal of classifying patient groups according to the measured mass spectra for their clinical samples.

3.1 Processing of ASAP-MS data

Data files from the ASAP-MS instrument are exported as large (approx 90 MB) comma-separated value (CSV) files. These contain a series of mass spectra with m/z resolution of 0.05 Da recorded at regular time intervals over the duration of the acquisition run. A purpose-written MATLAB script removes the MS instrument settings from the end of the files, making them more easily readable by our automated pre-processing software, and the resulting file is saved with a *.txt* suffix. As will be described in detail in the following, the files for each acquisition run are then processed to extract the individual measurements, background-subtracted, peak-aligned or binned, and normalised to unit area under the spectrum. The overall data pro-

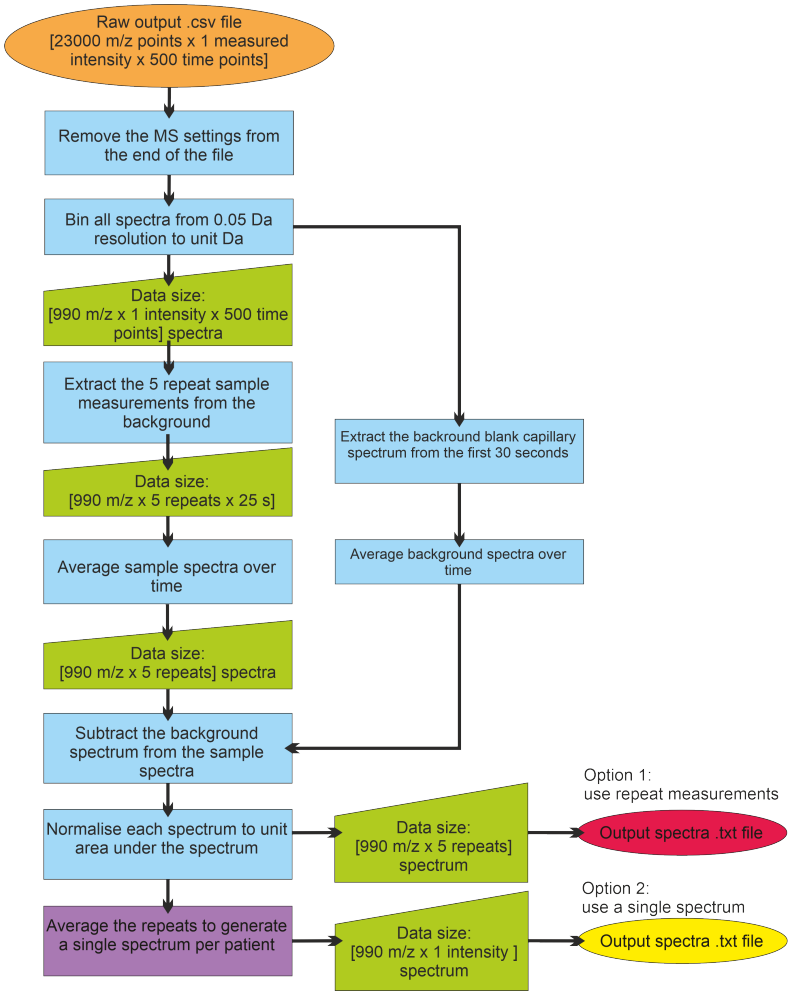

Figure 3.1 A flowchart showing the data processing steps developed for the analysis of ASAP-MS data in preparation for use in ML or statistical analysis in the format of a single spectrum per patient. The green trapezia show the size of the data after each step. Two commonly used end points were to use all the repeats for a sample measurement (red endpoint), or to average over the repeats to generate a single spectrum per sample (yellow endpoint).

cessing workflow that has been determined for biological sample measurements is shown in the flow chart in Figure 3.1.

3.1.1 Extracting the individual measurements

The individual measurements (recorded each time the ASAP probe is inserted into the mass spectrometer) are extracted from the complete acquisition run for each patient sample by finding the time points at which consecutive total ion current (TIC) values are greater than the average background intensity by a factor of 3σ or more,

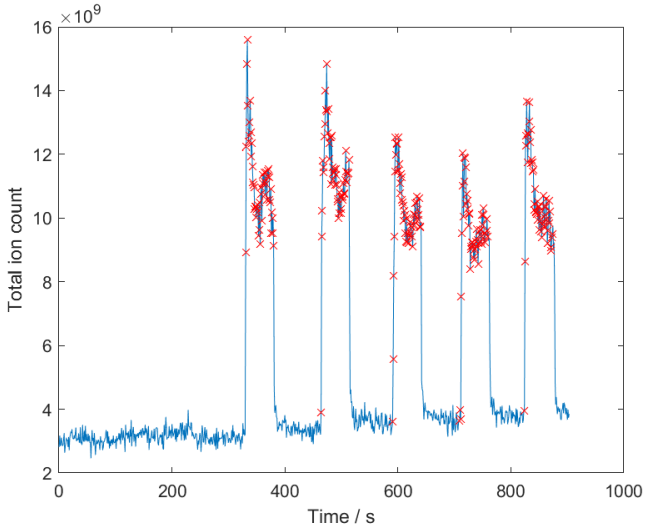

Figure 3.2 A plot of the total ion count (integrated over *m/z*) against time (*s*) recorded during the analysis of a blood plasma sample. The five repeat acquisitions can be observed as five regions of raised intensity, each lasting approximately 25 seconds. Each red cross is a time point has been determined to be part of a sample measurement through the method described in section 3.1.1.

where σ is the standard deviation of the background TIC. The background TIC was defined as the average TIC observed in the first 30 seconds of the data acquisition, whilst the clean capillary was inserted into the spectrometer. An example of a data acquisition run in which the time points corresponding to the individual measurements have been automatically identified is shown in Figure 3.2.

3.1.2 Data binning, discretisation and peak selection

The mass spectra in the raw data files are exported at a resolution of 0.05 m/z intervals, which is higher than the mass resolution of the instrument of 0.6 m/z. Each ion peak therefore spans around ten or so m/z values within the data file. In addition, peak positions observed in the mass spectrum can drift by up to 0.3 m/z or more, both as a function of time within the same acquisition run and across multiple acquisition runs. The peaks must therefore be aligned across all of the spectra so that a peak corresponding to a given compound always appears at the same m/z value. For a mass range of 10-1000 m/z, a resolution of 0.05 m/z leads

to a dimension of $> 23,000$. This is unfavourable in terms of computer memory usage when conducting calculations, but also is unsuitable for use in ML models as individual m/z intensities do not represent the intensity of an entire peak. This number of m/z values (variables) also gives the data set very high dimensionality (a 'wide' data set), and so requires a higher number of measurements on which to train the ML algorithms over.

Data binning, or discretisation, is commonly used to achieve both peak alignment and data reduction. Two different methods were initially trialled to achieve this: peak alignment against a reference spectrum, and binning the masses to unit m/z. The data binning method was used for all final analysis, as it better represented the data recorded on the instrument. Peak alignment against a reference spectrum can lead to misrepresentation of recorded data through treating multiple superimposed peaks as one peak value with a particular mass accuracy. Binning to unit Da is the most common approach used to describe single-quadrupole MS data in the literature, and does not misrepresent the mass resolution in the same way as the raw data might. For example, reporting data at $m/z = 51$ suggests that there is low mass resolution, rather than reporting at $m/z = 51.02$ such that the reader may mistakenly assume that there is 0.01 Da resolution. A further limitation of the 'peak alignment to reference' method is that the generation of a suitable reference spectrum requires a large amount of data to be collected prior to analysis of the data, with comparison of data between different data sets becoming challenging if new peaks are present, for example, comparing coronary aspirate plasma with venous plasma.

Each integer mass bin m/z contains the summed signal for the mass range m/z +/- 0.5 from the raw data file. Using the mass range m/z 10-1000 this leads to a reduction in file size from $>$ 23000 data points per mass spectrum in the unbinned data, to 991 in the binned data file. This method is favourable for ML methods as it standardises the number of m/z bins across each spectrum, and sparsely populated

peaks can be observed.

3.1.3 Background subtraction

Background signals in the mass spectra arise from ionisation of gaseous molecules present in the atmospheric pressure ion source that do not arise from the sample. This includes any contaminant molecules such as residual calibration mixture and plasticisers in the mass spectrometer components. The background spectrum recorded during the first 30 seconds of the acquisition (as explained previously) must also be discretised or binned using the method described in Section 3.1.2. The average intensity for each peak in the background spectrum is then subtracted from each peak in the sample spectra. Known background peaks should be monitored throughout data acquisition. Two peaks, appearing at m/z 42.3 and 214.0, respectively, were found consistently to have large and variable intensities in both the background and the sample measurements. They appear to be unrelated to the sample and to arise instead from contaminants within the mass spectrometer ion source, acetonitrile and plasticiser, respectively. Since the variable intensity of these peaks makes their removal by background subtraction difficult, we have chosen simply to remove these peaks from the mass spectra in order to ensure that the subsequent normalisation step is not adversely affected by their presence.

3.1.4 Data normalisation

The total intensity in each mass spectrum recorded will vary, and this is particularly true in ASAP-MS due to the manual sampling method. It is important to normalise the spectra so that measurements can be compared independently of the exact amount of sample that was introduced into the ion source. A large number of normalisation, scaling and transformation techniques exist, and there is no general consensus on the best approach to use for normalising mass spectrometry data; literature reviews reveal that a large range of methods are used. The method used has

the potential to impact the final analysis conducted on the data, as normalisation methods affect the shape and distribution of a data set [208–211]. A commonly used method for the analysis of MS data is to display each intensity relative to the total area under the spectrum, the unit area under the curve (AUC) method. An assessment of a large number of normalisation, scaling and transformation methods was conducted for this ASAP-MS clinical data, assessed against the accuracy of clustering and machine learning classification of plasma data. This analysis is discussed further in Section 7. The unit AUC normalisation method was determined to be appropriate for this type of analysis, and so this method was applied to the resulting peak-picked and background-subtracted spectra, and was used in all subsequent data processing.

3.1.5 Data averaging

The processed mass spectra may be averaged in a number of ways:

1. The spectra may be averaged over the range of times corresponding to an individual measurement. When the capillary bearing a sample is inserted into the ASAP ion source there is significant time-dependence to the mass spectra, with the most volatile molecules appearing at early times, and less volatile molecules appearing at later times once the capillary has reached higher temperatures. Averaging over the acquisition time yields a single mass spectrum for each of the N individual measurements containing the signals from all detected molecules. Following this averaging the complete data set for the sample comprises an $(n \times N)$ array, where n is the number of mass peaks and N the number of measurements.

2. To study the time dependence of the signals for particular compounds, we can instead average the time-dependent signals over the multiple individual measurements. This improves the signal-to-noise ratio of the time-dependent signals, and results in an $(n \times m)$ array, where n is the number of mass peaks

and m the number of time points (red outcome in Figure 3.1). Data was not commonly used in this format as it was found that the time dependence of peaks was not useful for further analysis of patient data.

3. We may also choose to average over both time and individual acquisitions in order to obtain a single mass spectrum (dimension n) for each patient sample (yellow outcome in Figure 3.1).

3.1.6 Data processing summary

The data processing steps described above have been implemented in both MATLAB and Python scripts and optimised specifically for the analysis of blood plasma and tissue samples, but could readily be adapted for use with other types of data.

3.2 Machine learning methods

3.2.1 Unsupervised machine learning and statistical analysis

Unsupervised machine learning (defined in the introduction) can be an efficient way of exploring patterns and relationships in complex and high-dimensional data prior to further analysis [212]. It is used frequently in mass spectrometry analysis, particularly for identifying clustering patterns [213–215]. A number of unsupervised ML algorithms exist, with some of the more commonly used being principal component analysis (PCA), K-means clustering, uniform manifold approximation and projection for dimension reduction (UMAP), t-distributed stochastic neighbour embedding (t-SNE), and linear discriminant analysis (LDA) [213, 216]. The method that was used most frequently in this work was PCA.

PCA is an unsupervised ML technique commonly used for dimensionality reduction, compressing data while retaining its variance [217]. It linearly transforms the data onto a new set of orthogonal axes [212], generating 'basis' mass spectra

(principal components) and expressing each original spectrum as a linear combination of these, with coefficients known as 'principal component scores'. A covariance matrix captures feature relationships, with eigenvalues and eigenvectors indicating the variance explained and the direction of each principal component (PC), respectively. PCs are ranked by eigenvalue, with the first explaining the most variance [218]. Each data point receives a score indicating its contribution from each PC, and the data can be reconstructed from the scores and PCs. PCA is useful for visualising clusters and patterns, and for reducing input variables for ML algorithms [212]. A 2D illustration is shown in Figure 3.3.

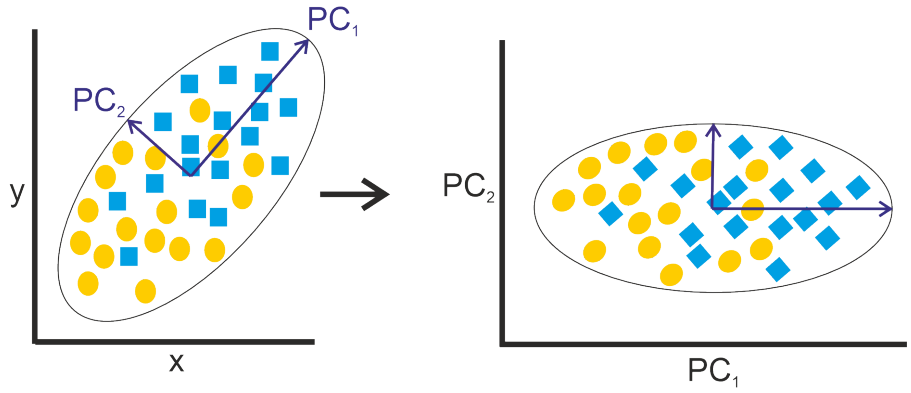

Figure 3.3 A visualisation of the transformation of a data set into 2d PC space. The variance in the data set along the x axis in the original data set is significantly less than the variance in PC1, allowing for PC1 to be used to describe a difference between the yellow circles and blue squares.

However, PCA may not always suit ML tasks, as it emphasises variation, not necessarily relevant to classification. Emphasising irrelevant variation can obscure patterns and cause underfitting [219]. PCA also assumes linear variance and similar amplitude across variables, assumptions often violated in biological MS data without preprocessing. Thus, while PCA is a valuable tool for dimensionality reduction, it may be less effective than using all input features or applying feature selection algorithms discussed later in this chapter.

Within this study, unsupervised ML techniques were used to search for data clustering and as a dimensional reduction technique. Clustering was rarely identified

via PCA for the analysis conducted, and the use of PCA for dimensional reduction was generally found to be detrimental to any further analysis. However, many other statistical and supervised ML analyses did prove successful. This implied that the disease-related patterns in the ASAP-MS data were generally due to intensity variations in low-intensity peaks with low variance, and so unsupervised ML algorithms were not frequently used.

3.2.2 Supervised machine learning classification methods

A number of supervised ML classification methods were employed for the classification of patient groups within the mass spectral data set. A large number of ML classification methods exist, well over 30 distinct methods, and many more if hybrid models are included [184]. Selecting an optimal model can be time-consuming; therefore, for this proof-of-concept study, a set of well-established and widely used machine learning (ML) classifiers—commonly referred to as classical ML methods—was employed [220, 221]. Furthermore, these models are suitable for use with sample numbers found in this work.

Five classical algorithms were chosen, each representing a distinct approach to data representation: vector- or spatial-distribution-based methods, including K-nearest neighbours (KNN) and support vector machines (SVM) [222, 223]; probabilistic methods based on Bayes' theorem, such as LDA and Näive Bayes classifier (NBC) [224–226]; and a decision tree-based method, random forest classifier (RFC) [227]. Each of these algorithms is discussed in turn below. Employing a range of machine learning classification methods, each with distinct approaches to interpreting the data, provides a more comprehensive assessment of the model's generalizability than using any single method alone. If only a single type of algorithm yields successful results, it may indicate potential overfitting or that the underlying patterns in the data are not robust.

3.2.2.1 K-nearest neighbour (KNN) classifier

The KNN algorithm clusters a data set by representing labelled data points as vectors in n-dimensional space, and calculates the distance metric between data points to determine a suitable class label for each data point [228]. The unclassified test data is then assigned a class through considering the closest K neighbouring data points, where K is a number defined within the hyperparameter settings. The distance metric used is also a defined hyperparameter [229]. A number of distance metrics exist, the most commonly employed of which is the Euclidean distance; this gives the shortest straight line distance ($d(a, b)$) between two points, a and b, in n-dimensional space [230].

$$d(a,b) = \sqrt{\sum_{j=1}^{n} (a_j - b_j)^2} \tag{3.1}$$

Figure 3.4 shows a visualisation of this method in 2D. If $K = 3$, the star is classified as a blue point, whereas the same point is classified as orange when $K = 6$.

Advantages and Disadvantages: A key advantage of the KNN classification method is its simplicity, making it fast to apply with low computational requirements for small data sets, and giving it very few hyperparameters to optimise. As this method makes no assumptions about the relationship between the input variables, it is also non-parametric, so can be applied across most data sets. Disadvantages of the KNN classification method are that it scales very poorly with large or high-dimensional data sets due to the scaling of vector space. It is also prone to overfitting due to the use of distance metrics, with the result that data with high uncertainty and noise will also prove more challenging to classify with this method [230, 231].

3.2.2.2 Support vector machine (SVM) classifier

The SVM classifier is designed to separate classes in n-dimensional space using $n - 1$ dimensional boundaries known as hyperplanes. The algorithm works to find

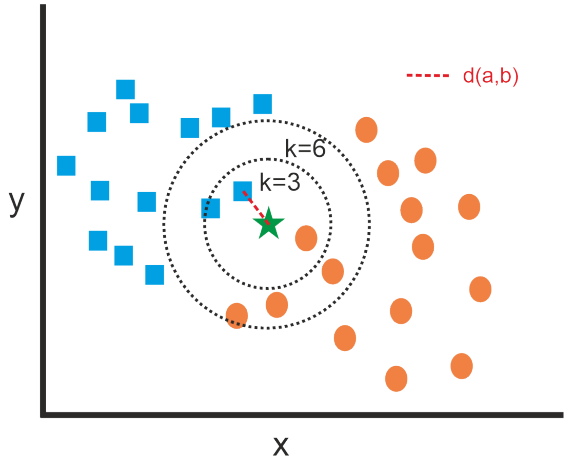

Figure 3.4 A visualisation of the 2D classification of an uncategorised data paint (green star) into blue or orange categories using the KNN algorithm with a $K = 3$ limit, leading to a blue label, and a $K = 6$ limit, leading to an orange label.

the hyperplane that maximises the margin, or distance between the plane and the closest data points from each class, known as the 'support vectors' [226, 232]. A visualisation of this is shown in Figure 3.5. Hyperplanes are typically linear, and so to fit a hyperplane to data that requires a non-linear separation, kernels are applied. Kernels are functions that map the data into higher higher-dimensional space, allowing a linear hyperplane to be fitted [185]. When the data can be perfectly separated, a 'hard margin' is applied to separate the classes completely. If the data is noisy or classes are overlapping, a 'soft margin' is used, where mixing is allowed but controlled by a regularisation parameter that balances maximising the margin and minimising classification errors [233]. Once the hyperplane is learned, new points are classified based on which side of the hyperplane they fall on. A hyperplane is defined by Equation 3.2, where x is the input feature vector (data point), w is the weight vector, or the direction of the hyperplane, and b is the bias term that defines where the hyperplane will cross the axis with respect to the origin.

$$w \cdot x + b = 0 \tag{3.2}$$

When $f(x) = w \cdot x + b$ is evaluated for a data point x, if $f(x) \geq 0$ the class is

$+1$, and if $f(x) \leq 0$ the class is -1 [234].

Advantages and Disadvantages: SVMs are very effective in high-dimensional spaces through the use of kernels, allowing complex spatial relationships to be handled more easily. This, and the use of soft boundaries, also allows them to deal with noisy boundaries and outliers very well when compared to other classifiers. They are also still effective on small data sets. They are, however, slow to develop for very large data sets, and sensitive to very noisy data. SVMs have a large number of tunable hyperparameters, which take time to tune, and incorrect tuning can lead to overfitting [233, 235].

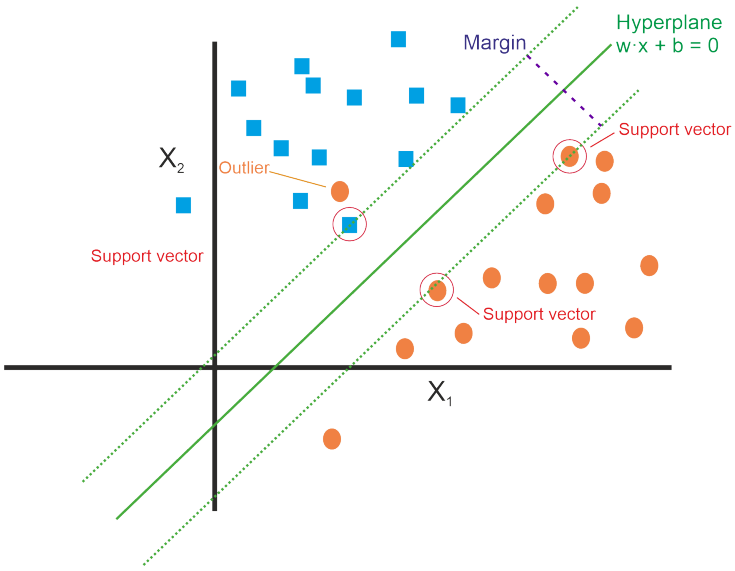

Figure 3.5 A simplified visualisation of an SVM hyperplane separating the blue and orange classes.

3.2.2.3 Näive Bayes classifier (NBC)

The NBC is a probabilistic classifier based on Bayes' Theorem. It assigns a class label to a sample based on the posterior probability that the sample belongs to a particular class, given its features [185]. Bayes' theorem is defined in Equation 3.3:

$$P(y|X) = \frac{P(X|y) \cdot P(y)}{P(X)} \tag{3.3}$$

where $P(X|y)$ is the likelihood of observing the features given the class y (ob-

serving the current evidence), $P(y)$ is the prior probability of class y, (how common the class is based on previous evidence), $P(y|X)$ is the posterior probability of class y given the features X (the updated probability of an event by combining the past and observed evidence), and $P(X)$ is the evidence, or the total probability of the features [224]. The algorithm is näive as it makes the assumption of conditional independence; it assumes that all features are independent given the class label. In this instance, this would assume that the concentration of each metabolite in a sample is unrelated to the concentration of all others, which is unrealistic given the integrated nature of the human metabolome. An unseen data point is assigned a class by computing the posterior probability for each class, and assigning the class with the highest value [236]. Hyperparameters to be optimised for an NBC model include the type of fitting algorithm to use, such as multivariate, Gaussian, and multinomial distribution, depending on the shape and structure of the input data, and the smoothing parameter which allows for zero probability classes to be handled appropriately.

Advantages and Disadvantages: The NBC model performs well for large and high-dimensional data, and has an interpretable probabilistic outcome. It does, however, rely on the strong independence assumption, which is often unrealistic, especially for metabolite data. It is also very sensitive to irrelevant features that have a high variability, which can distort the results and lead to poor outcomes [237].

3.2.2.4 Linear discriminant analysis (LDA) classifier

LDA aims to project the data onto a lower-dimensional space while maximising class separability, and hence it can be used for both supervised classification, or unsupervised dimensional reduction similar to PCA [225]. The chosen projection will be the dimension that minimises the within-class scatter matrix (S_W, the spread of the data within a class), and maximises the between-class scatter matrix (S_B, the

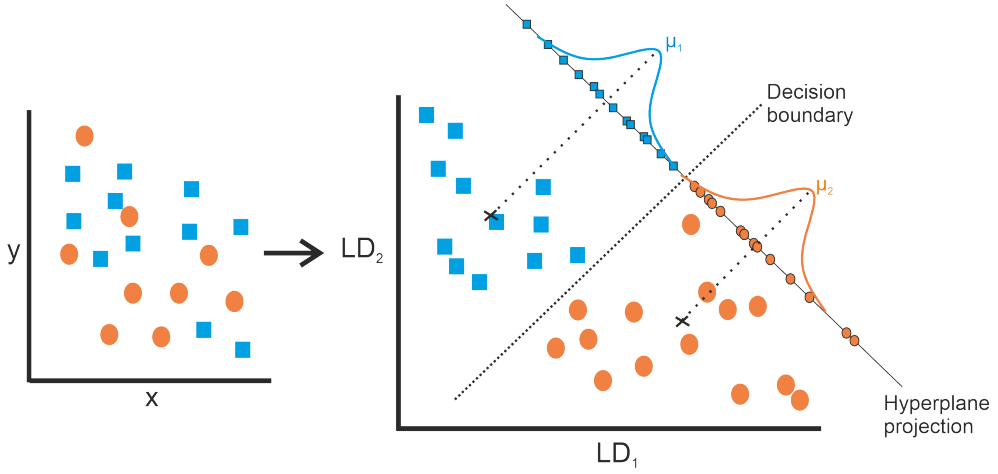

Figure 3.6 A simplified visualisation of an LDA projection of the data on the left to linear discriminant space on the right.

distance between the class means), through solving the following eigenvalue problem:

$$S_W^{-1} S_B \nu = \lambda \nu \tag{3.4}$$

This generates the eigenvectors (ν, the discriminant directions) and eigenvalues (λ, amount of separability in that direction) required to project the data into discriminant space [185, 238]. The largest eigenvalues correspond to the directions that best separate the classes.

A linear decision boundary is then determined using Bayes' Theorem, assuming that each class can be described using a Gaussian distribution centred around the class mean [185]. This is visualised in Figure 3.6. In addition, the class variances are assumed to be identical, known as the homoscedacticity assumption, making this a parametric method (i.e. one that should be used only for normally distributed data) [239]. Unseen data points are labelled by determining which side of the decision boundary they fall under. Hyperparameters to optimise include the solver, which determines the algorithm used to compute the LDA decision boundary, number of components, which controls the number of linear discriminates, and the shrinkage parameter, which helps if the data set is small or variables are highly correlated [240].

Advantages and Disadvantages: The results of LDA are highly interpretable, and features that contribute the most to class separation can be easily identified. LDA is very efficient, and so can handle moderately sized data sets easily. The most significant disadvantage of LDA is the assumption of normally distributed homoscedastic data. As this is rarely the case, the performance of LDA can be low for unparametric data. LDA is also sensitive to outliers, which distort class means [241].

3.2.2.5 Random forest classifier (RFC)

The random forest algorithm is an ensemble model (a model consisting of a combination of models that are combined to deduce an outcome), that uses combinations of decision trees to achieve class prediction [242]. A decision tree is a tree-like model that makes decisions through splitting data at node positions into branches based on feature values [243]. An example of this is shown in Figure 3.7. The root node represents the entire data set, which then splits along branches into internal nodes representing feature-based splits, and terminates at leaf nodes representing the class labels. At each node, the algorithm chooses the best feature to split the data [227]. It uses a splitting criterion to decide which feature gives the most useful information, the exact method used being a tunable hyperparameter. The network, or 'forest' of decision trees vote for the final class decision. This technique uses randomly selected subsets of the training data to train each decision tree, known as bootstrap aggregation, or 'bagging'. When splitting a node, each tree randomly samples a subset of features instead of considering all features. This makes the trees more diverse and reduces the risk of trees becoming correlated. Each tree is trained independently and will grow until a stopping criterion is met, such as maximum depth or maximum samples per leaf, which are tunable hyperparameters [244].

Advantages and Disadvantages: Due to the use of bootstrapping, the RFC classifier is highly robust to overfitting. However, RFC can be computationally

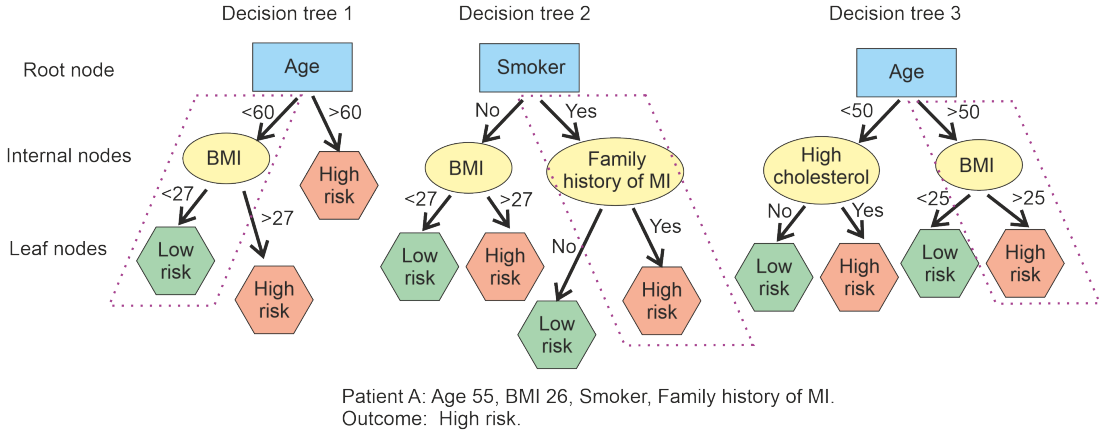

Figure 3.7 An example of a random forest algorithm consisting of three very small decision trees, for the classification of patients for cardiovascular disease risk. Two trees vote for high risk, whilst one tree votes for low risk, therefore, the patient is classified as high risk.

expensive when dimensionality is high, and complex to interpret due to being an ensemble method, compared to single decision trees that are easily interpreted [244].

3.2.3 Symbolic regression analysis

Symbolic regression is a type of machine learning evolutionary algorithm that can find a function relating a set of variables to an outcome value [245]. Rather than just optimising the parameters of a fixed model structure (such as linear or polynomial regression), symbolic regression does not assume a predefined equation. Instead, it searches for the equation by combining mathematical operators to build a model that best fits the data. One of the major advantages of this technique is that the solutions are highly interpretable, with a model being interpretable if "*the relationship between the input and output of the model can be logically or mathematically traced in a succinct manner*" (Makke and Chawla, 2024 [245]). The algorithm develops an encoded structure that represents the solution, commonly represented in the form of a nodal syntax tree (represented in Figure 3.8). This tree structure is similar to a decision tree and represents the search space mapping between the variable space and the output, where nodes (represented by white circles) are mathematical operators,

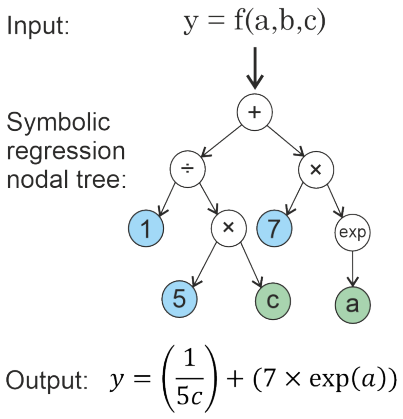

Figure 3.8 An example of a symbolic regression nodal syntax tree representation of an equation. The outcome y is related to a data set with variables a, b, c via an unknown function. A function that uses variables a and c has been determined through the algorithm.

connected by branches that terminate at leaves that represent either constants (blue circles) or features of the input data (green circles). The mathematical operators can be unary (accepts a single argument, e.g. cosine) or binary (accepts two arguments, e.g. subtraction) [246].

Symbolic regression is capable of handling small data sets, and as there are no assumptions made about the function type or which features of the data set to include, this is potentially an ideal method for use with mass spectral data [247]. To solve a problem, the algorithm initiates a set of functions randomly by combining operators and features, and then these functions are optimised using a search algorithm. Commonly, search algorithms are based on genetic algorithms (discussed below), but can also use other evolutionary programming methods such as grammatical evolution, as well as Bayesian methods and neural networks [248, 249].

A genetic algorithm is an optimisation technique that is used to solve complex problems with multiple unknown variables. The optimisation progresses through iteratively evolving (making small changes to) a population of individuals (a group of potential solutions) in a manner similar to how organisms evolve adaptations over generations in response to a stimulus. Initially, a random population of solutions is initiated, with each individual represented as a binary sequence, termed a chro-

mosome. A fitness function is then used to evaluate the strength of the solution, in a process termed a competition. The fitness function is determined by the user, a common example being root mean squared error (RMSE) between a predicted value and the true value. The strongest individuals are selected to be used for crossover, a process in which a new generation of individuals is created through combining features (genes) of the fittest individuals from the previous population. This process continues over a large number of iterations, over which the fitness of the overall populations improves. At each crossover, mutations are introduced, which are small changes applied to solutions in the population. This prevents the algorithm from getting stuck within a local minimum. The algorithm continues over a number of generations until a stopping criterion is reached, and the best individual is put forward as the solution to the problem. Stopping criteria are set by the user, with examples including the number of iterations, or requirements of the solution such as the absolute change in fitness function [250].

For the optimisation of a symbolic regression problem, each individual in a population is a nodal syntax tree, representing a potential equation. At each competition stage, the equations are evaluated for a portion of the input data, and the prediction accuracy assessed against the true outcome. Subsequent generations of equations will consist of features taken from the most accurate equations in the previous generation, as well as mutations and expansions to allow for equation flexibility. The equations generated will vary in complexity, which is a composite metric that increases with equation length, number of input variables included, and number of mathematical functions [248, 251].

Due to the scale of the search space analysed during a symbolic regression analysis, the 'correct' function is unlikely to be found (if such a function exists at all) unless the problem is very small. However, functional relationships that are present in the data can be found, and this gives useful information about the data set being assessed. One way to tackle this is to perform the same symbolic regression analysis

several times, and to look for underlying trends and relationships in the resulting functions together, rather than in isolation [252].

A mathematical function can be developed to fit almost any data set to any outcome perfectly if given enough freedom, even if the data set is comprised only of noise. This means that symbolic regression will tend to overfit if steps are not taken to restrict the algorithm and appropriately to assess the true accuracy [253]. Restricting the model hyperparameter space to decrease the risk of overfitting is required in most cases. The best method to use to do this will depend on the symbolic regression model being used and the data being modelled. Constricting equation complexity is often also required. A general rule is that a simpler equation is preferred to a complex equation, as a simple relationship is more likely to be a real feature of the generalised data set, and a complex function is more likely to be overfitting of noise. As symbolic regression is a type of machine learning, all models are tested on unseen test data, rather than assessing the fit of the training data, and deviation between a strong fit to the training data and a weak fit to the test data will give a clear indication of overfitting.

In the analysis of medical and population data, r values of > 0.15, which may be considered low in other fields, may be considered meaningful. This is due to large degrees of variation present in the populations, and measurement uncertainty is often large [254]. However, a point of note is that when searching for relationships within large data sets, there is always a chance of random correlations being present that are not 'real', but occur by statistical chance. Searching over a large data set with a powerful regression method, such as symbolic regression, will likely find such spurious correlations. This is commonly termed 'data dredging' or 'p-hacking', and should be avoided using methods typical to good scientific practice: define the hypotheses before analysing any data, carefully design experiments to study these hypotheses, and use cross-validation to test all findings [255, 256].

3.2.4 Feature selection and reduction methods for machine learning

Feature selection is an additional step that can be conducted prior to training a machine learning algorithm, and is intended to reduce the number of input variables to those most likely to be important for successful machine learning. Each feature selection algorithm has different search criteria, and so may produce different information about the 'importance' of each variable, in this case the m/z peaks that are potentially the most useful for classification. Therefore, a range of methods were used in each case, and the resulting important peaks critically assessed. The feature selection algorithms investigated include statistical and distribution-based methods, machine learning-based methods, and PCA.

To prevent data leakage, all of these feature selection methods must be applied to the training data only, and then the relevant transformation applied to the test data after model training has taken place. If the feature selection was conducted on the entire data set, the model would have an unfair advantage in that important features of the test data may skew the training process. Feature selection was conducted separately over every partition (discussed in Section 3.2.6, a partition is the process of dividing a data set into small subsets that are non-overlapping so that an algorithm can be trained and assessed for accuracy and generalisation), and the most important features used to conduct the further machine learning analysis. The features used each time were reported, allowing for analysis of the features most used by each algorithm.

3.2.4.1 Statistical test feature reduction

A statistical test is the simplest form of feature reduction. The correlation coefficient between each variable, an m/z peak in this instance, and the outcome variable, the clinical variable of interest, is calculated, and the variables that show a significant correlation with the outcome are used as input variables in further analysis. Stat-

istical tests can be parametric, used for data that follows a specific (often normal) distribution, or unparametric, where no assumptions about normality or distribution are made. Parametric statistical tests often have higher statistical power than unparametric tests, and so should be used if possible. A common test for normality is the Sharpiro-Wilkes test. If a large proportion of the data variables are shown to be normally distributed, above a set percentage (typically 95%), a parametric Pearson correlation coefficient test can be used for feature reduction [257]. The unparametric version of this method is a Spearman correlation. Initial tests showed that within the ASAP-MS data for the OxAMI samples, 98.9% of mass peaks were determined to be normally distributed by the Shapiro-Wilkes test with $p < 0.05$, making the parametric test appropriate for this analysis. The Pearson correlation coefficient (r) measures the strength and direction of the relationship between two variables x and y:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3.5}$$

where x_i is the intensity of peak for sample i, y_i is the clinical variable of interest for sample i, and n is the total number of samples [258, 259].

3.2.4.2 Analysis of overlap integrals for peak intensity distributions

It was observed that if the distribution of peak intensities for each m/z was plotted, split into two distributions by the two classes in the clinical variable classification, the distributions were quite different in some instances. This behaviour could be tested using one of several statistical tests, such as a t-test, but could also be studied by determining the overlapping area (labelled overlap integral (OI)) of the distribution histograms.

For each clinical parameter, the normalised intensity distributions for each m/z peak were determined for the mass spectra assigned to the two defined patient groups ('low' and 'high' or 'yes' and 'no' for a clinical parameter). An example of this is shown in Figure 3.9. This yields two intensity distributions per m/z peak,

corresponding to 'class 0' and 'class 1' in the figure. Calculating the OI between these two intensity distributions results in a value close to one in a distribution for which there is little or no intensity variation between the two groups, and this peak is unlikely to contain information on the clinical parameter of interest, while those for which there is a significant intensity difference will return a lower value, and are more likely to correlate with the clinical parameter.

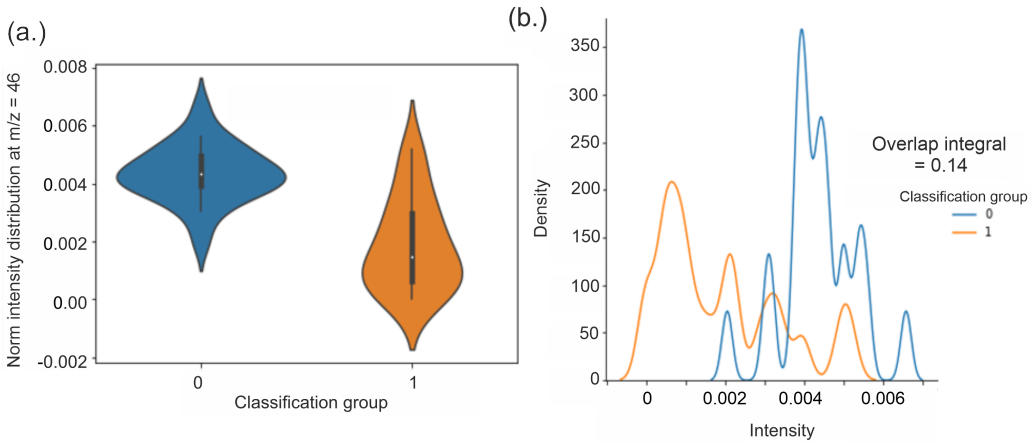

Figure 3.9 The overlap integral feature selection method shown for an *m/z* peak at 46. The violin distribution (a) shows a clear difference between the distribution of class 0 and class 1, which is reflected in the low overlap integral value and low overlap area in (b) with respect to the non-overlapping areas of the orange and blue distributions.

3.2.4.3 Chi^2 classification ML feature ranking

The Chi^2 (χ^2) classification feature ranking method uses the χ^2 statistical test to identify the dependence of one variable on another [260]. This is achieved by quantifying the deviation of a distribution compared to the expected distribution if each variable were independent of the outcome variable (class) [261]. This is calculated in the following:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{3.6}$$

where O_i is the observed value, and E_i is the expected value, for each peak (m/z value) in the data set i. The expected value in this case would be that the mean intensity of a peak across the samples would be the same for both classes, assum-

ing that peak intensity is independent of class (the null-hypothesis). The peaks that show the highest χ^2 statistic show the most deviation in distributions between classes, and so are used in further ML analysis as input variables.

3.2.4.4 Minimum redundancy maximum relevance ML feature ranking

Minimum redundancy maximum relevance (MRMR) is an algorithm used to find the smallest number of different features that represent variation in the data set that correlates with the output class. When iterating over the input variables, the algorithm evaluates both the correlation to the output, and correlation to the remaining variables. The algorithm selects for those features that have a high correlation to the former (high relevance), and low to the latter (low redundancy) [262, 263].

3.2.4.5 ReliefF ML feature ranking

The ReliefF ranking algorithm (RelF) algorithm estimates the relevance of features (m/z values) by assigning each feature a weighting based on its ability to distinguish between similar sample spectra that are in different classes [264]. Spectra are selected at random, and the K-nearest neighbours of the same class (near-hits), and different class (near-misses), are found. The weighting for each variable in that spectrum is increased if the feature value is different from near-miss instances, and so helps to separate classes. This is continued over many iterations until the weightings converge. The features with the highest scores (highest ranks) are used in further analysis as input features [265].

3.2.5 Machine learning assessment

Interpreting ML model results in order to determine model accuracy and predictive power can be done through examination of a number of different performance metrics. In this study, we used five commonly used metrics: sensitivity, specificity, percentage accuracy, Cohen's κ score, and F_1 score. These metrics are determined

through comparing the four possible outcomes of a classification:

- True positive (TP): The model correctly identifies data belonging to class zero (the positive class);
- False positive (FP): The model incorrectly identifies data into class zero that should be in class one (the negative class);
- True negative (TN): The model correctly identifies data belonging to class one;
- False negative (FN): The model incorrectly identifies data into class one that should be in class zero.

These ML model outcomes are commonly visualised in a confusion matrix, shown in Figure 3.10.

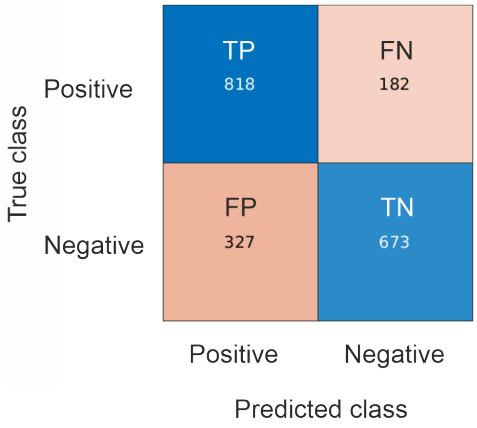

Figure 3.10 A confusion matrix, showing a high accuracy prediction outcome for a model. Note the dark blue of the TP and TN boxes showing the high proportion of values in these quadrants, compared to the light orange hue, which shows that there are few false classifications. The accuracy of this model is 0.75%, and κ score is 0.49.

3.2.5.1 Sensitivity and specificity

The sensitivity of a model is the proportion of true positives correctly identified as such (also called the recall), and specificity is the proportion of true negatives

correctly identified as such [266].

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{3.7}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3.8}$$

These metrics are very important for predicting health data. For example, it is usually far safer to mistakenly predict that a patient is at high-risk of mortality when they are not than to incorrectly assess them as low-risk and prematurely discharge or end treatment. In the latter case, a patient with an undetected high mortality risk could miss out on crucial treatment and monitoring. In this instance, if high mortality is defined as the 'positive' class, high sensitivity would be the target of a predictive model, over high specificity, but in a truly robust model both metrics would be high. Equally, a high sensitivity is not useful if there are a large number of false positives, as this can lead to a large number of unnecessary and potentially highly damaging medical interventions that serve no useful purpose.

3.2.5.2 Accuracy, F_1 score, and Cohen's κ score

The 'accuracy' of a model is often reported as the percentage accuracy;

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{3.9}$$

This determines the proportion of classifications made by the model that were 'correct'. There are no adjustments within this metric to account for the correct prediction of classes due to chance. To account for this, the F_1 score and κ score metrics are used.

The F_1 Score is the harmonic mean of recall (sensitivity) and precision (the proportion of TP out of all positive classifications). Therefore, a model can only

achieve high F_1 scores when both precision and recall are high.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.10}$$

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{TP}{TP + \frac{1}{2}FN(FP + FN)} \tag{3.11}$$

The Cohen's κ score (κ) is arguably the most robust metric to determine the performance of ML classification models. This metric takes into account the expected probability (P_e) of the model classifying correctly only due to chance, and adjusts the accuracy of the model (P_0) by this value.

$$\kappa = \frac{P_0 - P_e}{1 - P_e} = \frac{2 \times (TP \times TN - FN \times FP)}{(TP + FP) \times (FP + TN) + (TP + FN) \times (FN + TN)} \tag{3.12}$$

This metric varies between -1 and 1, and is interpreted in this study using the 'Levels of Agreement' as defined by McHugh and reproduced in Table 3.1 [267].

Table 3.1 Interpretation of κ as levels of agreement, adapted from [267]

κ value	Level of Agreement	Estimated equivalent % accuracy
< 0.2	No agreement	$< 60\%$
$0.2 - 0.4$	Weak agreement	$60 - 70\%$
$0.4 - 0.6$	Fair agreement	$70 - 80\%$
$0.6 - 0.8$	Moderate agreement	$80 - 90\%$
$0.8 - 0.9$	Strong agreement	$90 - 95\%$
> 0.9	Very strong agreement	$> 95\%$

3.2.6 Machine learning cross-validation and partitioning

Cross-validation and partitioning are techniques used in machine learning to assess how well a model generalises to unseen data. It is essential for identifying whether a model is overfitting (a model learns irrelevant features of the training set so closely that the model fails to make correct predictions on new data), and for assessing

the model accuracy and error. Cross-validation involves splitting the data set into training and testing partitions, so that the model performance can be assessed on an unseen set of test data not used for model training. This is particularly important with small, noisy, and highly dimensional data sets, for which the risk of a model 'learning' noise is high. A number of different cross-validation and partitioning methods exist which are suitable for different types of machine learning problems and models, including a fractional train-test split, and leave-n-out cross-validation (LNO-CV).

During a train-test split, a data set is divided into two groups, a training set and a test set, by a particular partition fraction, the most commonly used fractions being 80:20 or 90:10 for training:test ($Tr{:}Te$) data, respectively. These allocations are made randomly, and can be repeated for a data set of N samples k times using the Binomial Theorem rule, where:

$$k < \frac{N!}{Te!\,(N - Te)!} \tag{3.13}$$

The final performance of the model is the average score across all k partitions.

LNO-CV is a very similar approach, but instead of a fractional partition of the data set chosen, a set of n samples is reserved as the training data, regardless of the data set size. This process can be repeated k times using Equation 3.13, where N is the total size of the data set, and Te is the number of test samples chosen. The LNO-CV method with $n = 1$ is often used for very small data sets ($N \lessapprox 100$).

The number of cross-validation partitions to use for a particular data set can be optimised. Using a large number of partitions improves the accuracy of the final model assessment, as well as allowing for the calculation of error or precision. However, running an algorithm for many partitions is computationally expensive, and for a given data set and size, the accuracy and precision should approach constant values after a particular number of partitions.

The number of partitions to use in the present work was optimised by determining

the point at which the κ obtained over each classification analysis converged on an approximately constant value. A study was conducted in which each of the five ML models were used to assess each categorical clinical variable over 500 partitions. Each partition and model combination generates a confusion matrix.

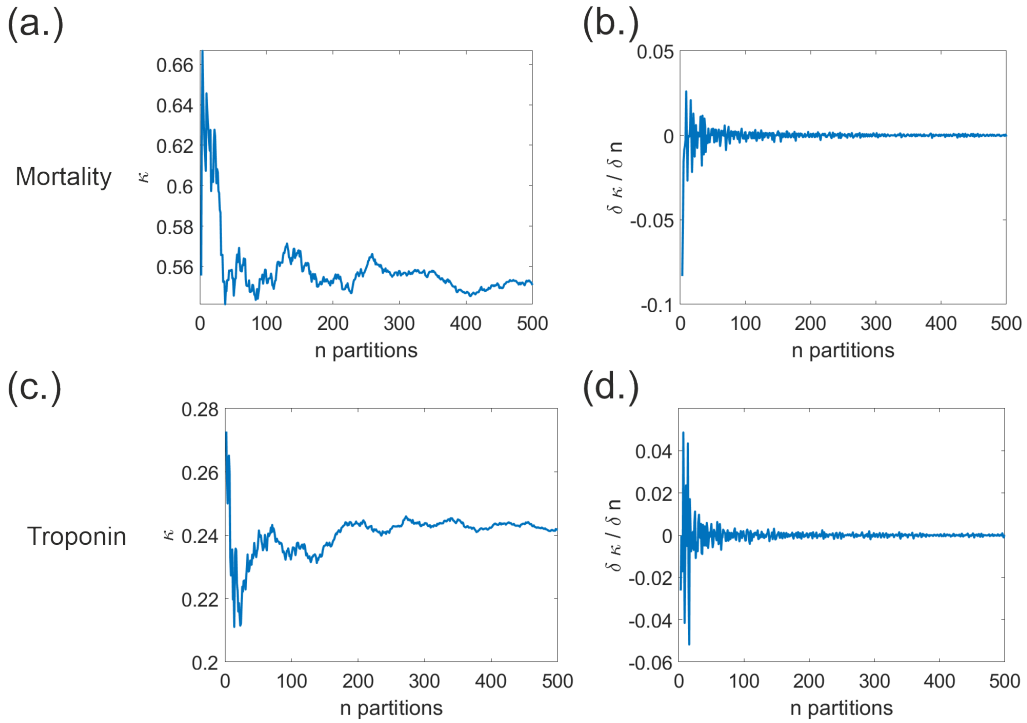

Figure 3.11 The effect of increasing partition number on the final κ obtained for patient mortality (upper) and peak troponin level (lower) by KNN analysis. The κ obtained after each new partition is generated is shown in a. and c., and the first derivative $\frac{\delta\kappa}{\delta n}$ from the previous score is shown in b. and d.

A κ score was determined for the confusion matrix resulting from the addition of each new partition, and this score was averaged across the five different ML models. For each different ML model across each clinical parameter, convergence in κ was found to occur between 200 and 300 partitions. A pair of examples are shown in Figure 3.11 for a small sample set (where $N < 50$, patient mortality 3.11 a-b) and a large sample set (where $N > 200$, peak troponin 3.11 c-d), both of which will be defined in the next chapter. Both the κ and first derivative $\frac{d\kappa}{dn}$ values for mortality and troponin data sets showed similar trends in both visualisations, with the κ value stabilising after approximately 200 partitions. As most of the data sets we analysed

had similar sizes, this trend was seen across a number of other data sets tested. Therefore, 200 partitions were used in most ML algorithms.

3.3 Summary

To summarise, a large number of analytical methods, including standard data processing protocols, statistical analysis, and machine learning algorithms, have been explored for application to the ASAP-MS data collected for the clinical samples studied. As there is no standardised protocol for handling data of this type, much of this analysis was selected as it appeared appropriate for the type of data analysis we wanted to achieve. Similar approaches to mass spectral data analysis have been implemented within commercial software packages such as Metaboanalyst [268]; however, ASAP-MS data format is yet to be integrated into these programmes.

4

Prediction of clinical outcomes for STEMI patients

Analysis of small-molecule metabolites found in blood plasma of patients undergoing treatment for STEMI has the potential to be used as a clinical diagnostic and prognostic tool, capable of predicting disease progression, risk of negative outcomes, and response to treatment. In this Chapter, we discuss the application of ASAP-MS and machine learning techniques to predict outcomes for STEMI patients. Section 4.1 is adapted from our paper in Analyst: 'Prediction of clinical outcomes of ST-elevated myocardial infarction patients using atmospheric solids analysis probe mass spectrometry and machine learning' [269].

4.1 The prediction of OxAMI clinical variables from coronary aspirate plasma samples

4.1.1 Introduction

As discussed in Chapter 1, identification of 'high risk' patients earlier on in their treatment journey would allow a more targeted response to their initial STEMI, employing more aggressive treatments that are withheld from the general patient

population due to a high risk of complications, cost, or unpleasant side effects. The aim of the present study is to determine whether biological markers present in coronary aspirate blood plasma of STEMI patients can prove predictive for their progression and recovery trajectory. Classification of patients into risk categories for disease progression will allow for more personalised, targeted treatments that may prolong or enhance the lives of patients with cardiovascular disease.

Due to recent improvements in the availability, collection and publication of large clinical datasets, an influx of studies have been published over the last three years that utilise ML and multivariate statistics to predict outcomes for STEMI patients [53–57, 270]. Some of the models developed within studies include data for thousands of patients, in the form of both demographic information (age, sex, family history) and clinical parameters (TScore, IMR, troponin), and can predict negative patient outcomes such as mortality with high accuracy. The OxAMI data set has been used previously to develop models of this type, with varied results [271–278]. These various studies have led to the development of a range of functions and composite scores that aim to quantify the severity of the cardiac event; examples include the age-thrombus burden index of microcirculatory resistance (ATI) score and the resistive reserve ratio (RRR), both of which show prognostic value for STEMI patients [276–278]. Based on these results, it is clear that indicative predictions of probable patient outcomes and response to treatment are possible based on early patient information. Within the present study, we take this relationship closer to the pathology of the disease by linking the prediction of patient outcomes directly to a measurement of metabolites in the blood.

Achieving an accurate representation of the biochemical composition of complex biological matrices such as blood plasma is challenging. Most analytical instruments report on a selected fraction of the matrix with known physicochemical attributes, e.g. 'lipophilic molecules', or require the use of various mechanical and chemical separation steps involving the addition of solvents that may alter the sample. As

described in Chapter 1, ASAP-MS provides an untargeted fingerprint of metabolites and lipids in a sample. Measurements are completed in a matter of seconds, making these compact instruments a potential candidate for use in a clinical setting. The method provides an untargeted 'molecular fingerprint' of the sample, allowing for a pattern recognition approach to biomarker discovery to be conducted. This approach complements conventional specific biomarker discovery methods in a number of ways. Firstly, there is no requirement for comprehensive knowledge of the biochemical pathways and specific molecules underpinning the disease, saving time as well as costs associated with multi-omics analysis studies. Secondly, many biomolecules may be measured simultaneously, providing an opportunity to identify subtle relationships between metabolites that may otherwise be missed, and potentially opening studies up to include biomarkers beyond the current known molecular space. In the case of STEMI metabolomics, an untargeted approach may shed some light on the relationship between metabolic pathways for lipophilic and non-lipophilic molecules, with (amongst others) both glucose and lipid metabolism known to play a crucial role in myocardial ischemia-reperfusion injury [279].

The pattern recognition process required to link mass spectrometric fingerprints with clinical outcomes may be performed in a number of ways, with some of the most promising being machine learning (ML) approaches [280]. A variety of ML algorithms have been developed to conduct pattern recognition or classification analysis on multivariate data sets, including collections of mass spectra [280–283]. These can be employed in the present context to classify patients into risk groups based on correlations between the mass spectra and a number of clinical parameters known to be linked with further complications or patient mortality. Machine learning methods can also be combined with feature-reduction techniques in order to determine which features of the mass spectra, i.e. which molecular signals, are most important in the classification process. Such approaches identify which metabolite mass peaks show significant differences between the patient groups, and might therefore be worthy of

further investigation as potential biomarkers.

In the present study, we have investigated the use of ML to analyse mass spectra of coronary aspirate blood plasma from STEMI patients enrolled on the OxAMI clinical study [197, 284], with the goal of stratifying patients according to their future cardiac risk. The coronary aspirate samples were collected during the pPCI undergone by patients on admission to hospital, and were chosen for this initial proof-of-concept study as we hypothesised they were likely to contain the highest concentrations of metabolites related to the coronary event. The coronary aspirate is a waste product of the cardiac catheterisation procedure, and is usually disposed of. We will focus on the ability of ML methods to classify patients into appropriate risk groups based on the mass spectrometric data rather than on biomarker discovery.

4.1.2 Methods

4.1.2.1 Study patients and criteria

The 283 coronary aspirate plasma samples used in this study were acquired from STEMI patients enrolled in the Oxford Acute Myocardial Infarction (OxAMI) clinical study, described previously in Section 2.2. The practical work in this section was conducted with the help of Thomas Mills, a Part II student in the group in 2021-2022. Within the OxAMI study, samples and data were collected for STEMI patients who were identified over a 10 year period in the catchment area who gave consent to join the study. Overall, the OxAMI study aims to investigate treatment options and trends in STEMI patients. We were granted access to a set of samples available for patients within the study to identify if the ASAP-MS instrument would provide a new way of predicting STEMI outcomes in an ancillary analysis on top of the existing clinical investigations within the study. The aim of our study is to determine if this method is worth investigating further. This is hypothesis-generating work, conducting pilot metabolomics fingerprinting analysis as a secondary use of these samples, and is not the primary clinical study objective. We accept the impact

of this study is limited by the low patient numbers.

4.1.2.2 Sample measurement and data processing

Frozen blood plasma samples, described previously in Section 2.2, were obtained from the Oxford Heart Centre, John Radcliffe Hospital. Mass spectra of blood plasma were recorded on an Advion expression CMS-S ASAP-MS, using the measurement protocol defined in Section 2.1.1, and published in [180]. Measurement considerations that impact data reproducibility have also been studied for ASAP-MS analysis of biological samples in our previous work [285], and steps to limit the impact of these effects were implemented in these measurements. The resulting mass spectra were processed as described in Chapter 3, generating a single mass spectrum per sample. The background-corrected spectra were then normalised to unit area under the spectrum, and averaged to generate a single mass spectrum for each sample. A detailed study into the impact of a large number of different normalisation methods applied to this data, including more complex methods such as quantile, vector, and Pareto scaling, has been conducted for this dataset, and will be described in Chapter 7 [286]. The results of this showed that the impact of changing normalisation methods was negligible, and the AUC method is appropriate for this analysis type [286].

4.1.2.3 Data analysis

As noted above, the data analysis was designed to investigate correlations between the 'molecular fingerprints' contained within the plasma mass spectra and a number of clinical parameters of interest [148, 196–198, 272, 274], as detailed in Table 4.1.

The five standard supervised ML classification models described in Chapter 3 were employed in order to investigate correlations between the plasma mass spectra and relevant clinical parameters, based on an evaluation of each algorithm's robustness, suitability across a range of data types, suitability for data sets of this size,

and ease of use. Further details of the ML models can be found in the Appendix (Appendix B.1), including the model setup and description of hyperparameters. All data analysis was conducted using custom MATLAB R2022a and Python 3.7 software. ML analysis was run on CPU nodes within the University of Oxford Advanced Research Computing (ARC) high performance computing (HPC) cluster.

The number of patients involved in the OxAMI study, and therefore the number of samples in the mass spectral data set, were too low to perform the regression analysis required in order to predict numerical values for each continuous clinical variable. Instead, patients were assigned as having 'high' or 'low' values for each clinical parameter of interest, based either on thresholds defined in the literature, or using the median measured value.

Most of the clinical parameters investigated are continuous. The individual values for each variable have been grouped into classes (mostly by defining two groups corresponding to 'high' and 'low' values for each parameter) so that classification algorithms can be employed. The distribution of patients across each continuous parameter are shown in Figure 4.1. The threshold values separating the groups for each variable have been chosen either by using definitions from the literature (ITime and IMR), or using the median measured value where no threshold is defined (peak troponin, and creatinine on admission). TScore was compressed from a five-class grouping (TScore 1, 2, 3, 4, and 5) to a binary grouping (TScore 1-3 and TScore 4-5) due to low populations in the low TScore classes. ITime was grouped by ITime $<$6 hours and ITime $>$6 hours as defined by J. A. Rodrigues *et al.* [287]. The IMR is grouped by IMR$<$ 40 and IMR$>$ 40 as determined by W. Fearon *et al.* and G. L. De Maria *et al.* [47, 198]. The remaining clinical variables, heart failure diagnosis (HFD) and patient mortality, are already binary variables.

The thresholds employed for each clinical variable are shown in Table 4.1. Having defined these thresholds, we were then able to investigate the extent to which ML classification algorithms could assign patients correctly to each group based on the

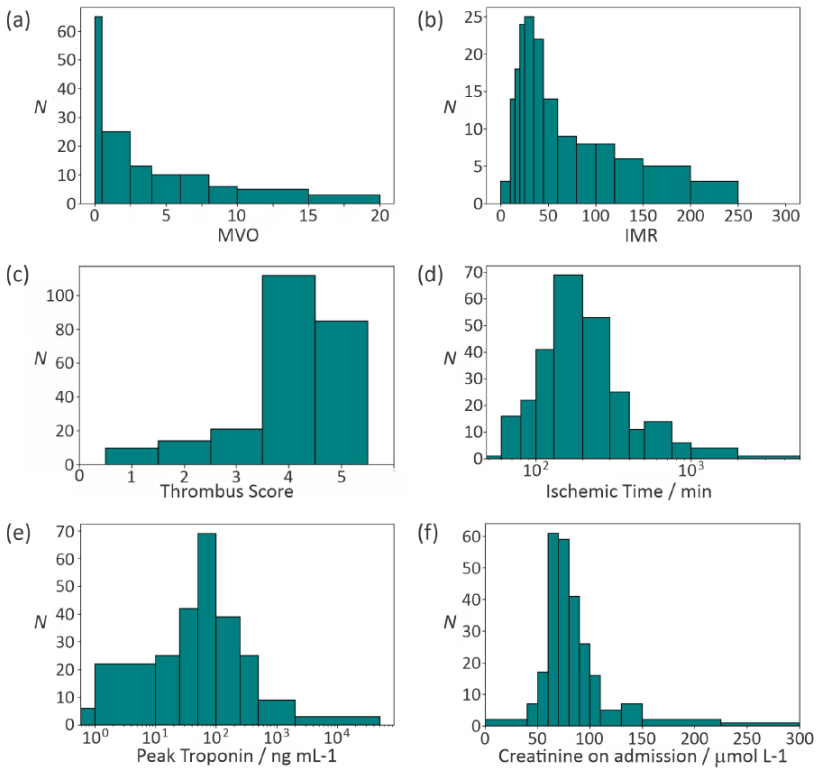

Figure 4.1 Histogram plots of the patient population distributions across each of the clinical parameters of interest: (a) MVO measured six months after the pPCI; (b) IMR measured during the pPCI; (c) TScore; (d) ITime prior to the pPCI; (e) peak troponin level measured prior to pPCI; (f) creatinine level.

mass spectra for their respective plasma samples.

To ensure that the ML algorithms use features of the mass spectra, i.e. m/z peaks, to perform the classification, rather than 'learning' based on the balance of probabilities, it is important to ensure that the data set used to train the algorithm contains a similar number of mass spectra for each group in the classification. For example, our data set contained samples from 25 patients that were diagnosed with heart failure (HFD-positive), and 200 without HFD (HFD-negative), as well as 58 for whom this was not recorded. Constructing a classifier to distinguish between patients that were HFD-positive and those that were HFD-negative on the measured mass spectra would mean that we should use a data set comprising mass spectra from 25 HFD-positive patients and 25 HFD-negative patients, rather than using all 200 mass spectra available for the HFD-negative patients. This way, the algorithms

are forced to use features of the mass spectra to predict HFD, rather than 'learning' that simply guessing 'HFD-negative' will give a correct result over 85% of the time. The number of patients per class is shown in Table 4.1 for each of the clinical parameters investigated. The population distributions for each clinical variable, Figure 4.1, show frequently skewed distributions.

The available mass spectra were partitioned into training and test data in an 80:20 ratio, with the training data used to train the ML models, and the test data used to evaluate the performance of the models. In order to preclude over or underperformance of the models resulting from the specific choice of training and test data, a cross-validation procedure was employed: the training and evaluation process was repeated multiple times with different selections of mass spectra allocated to the training and test sets on each run.

Since setting a threshold between 'high' and 'low' for a clinical parameter that can take continuous values is unlikely to result in clearly separated groups, in addition to the complete mass spectral data set (referred to in the remainder of the paper as 'All patient' data) we also prepared 'Extreme patient' data sets comprising subsets of mass spectra for patients with extreme high and low values for each variable (summarised in the final column of Table 4.1). In the context of this early pilot study, this approach offers the best chance of identifying mass spectral differences correlating with the clinical variable of interest. If successful, further investigation with larger data sets, broader distributions, and potentially more sophisticated machine learning approaches may be warranted. However, if classification is unsuccessful for 'extreme' patient groups, then we should accept that we have reached a 'dead end' and that further investigation is probably not warranted.

In addition to the 'All patient' and 'Extreme patient' data sets described above, two modifications were employed to investigate whether the performance of the ML models could be improved any further. In the first, we included clinically relevant demographic details that would be known at the point of treatment (the continuous

Table 4.1 Classifications of clinical parameters.

Clinical parameter	[0] conditions	[1] conditions	n patients per class
MVO	no MVO	MVO present	64
TScore	TScore $1-3$	TScore $4-5$	45
ITime	<6 hours	>6 hours	46
IMR	IMR<40	IMR>40	63
HFD	no HFD	HFD	25
Patient mortality	no	yes	23
IMR	IMR<40	IMR>40	63
Peak troponin	<50	>50	110
Creatinine on admission	<75	>75	121

parameters age, weight, height, BMI, and body surface area (BSA), and categorical parameters of present/past smoking status, and presence of hypertension, diabetes, or previous cardiological history) in the training data set. Only patients for which a complete set of these parameters was available were included in this analysis, which reduced the data set sizes somewhat further. We chose not to include patient sex as a parameter when constructing the 'Extreme patient' data sets due to the low number of female patients enrolled in the OxAMI study. The range for each continuous clinical parameter was rescaled to the median intensity of the mass spectra before integration into the data set in order to avoid demographic parameters from skewing the ML process. ML algorithms prove most successful when each individual variable is standardised to vary on the same scale, sometimes known as z-score normalisation [288].

Finally, we employed feature-reduction (FR) methods [289, 290] to construct data sets comprising a smaller number of m/z peaks. The FR process identifies mass spectral peaks that correlate significantly with the clinical variables of interest, potentially enabling us to focus the ML process on the most relevant parts of the measured molecular fingerprints, and also potentially revealing mass spectrometric biomarkers for clinical parameters of interest. FR was conducted on the training data within each 80:20 data partition prior to ML analysis. Three different approaches to FR, which were described in Section 3.2.4, were explored: statistical tests, analysis

of overlap integrals (OI), and Chi^2 ML FR.

i **Statistical tests:** The correlation coefficent between each peak and the clinical variable of interest was calculated. A parametric Pearson correlation coefficient test was used, as 98.9% of mass peaks were determined to be normally distributed by the Shapiro-Wilkes test with $p < 0.05$. Significant peaks (with $p < 0.05$) within the training data were then used as the input features, as opposed to the 990 m/z peaks used initially.

ii **Analysis of overlap integrals for m/z peak intensity distributions:** For each clinical parameter, the OI FR method described in Section 3.2.4 was used, and the 40 peaks with the smallest OIs were used to generate training data sets for the ML models.

iii χ^2 **machine learning-based feature ranking:** Feature ranking analysis was conducted using the χ^2 classification feature ranking method [260]. The 40 features that were determined to have the highest 'importance' were used to generate training sets for the ML models.

The ability of each model to predict patient categories accurately was assessed though comparison of the % Accuracy and Cohen's kappa score (κ score) described previously in Section 3.2.5. The mean and standard deviation of the values for the accuracy and κ score obtained across 200 partitionings of the data into test and training sets are reported. We note that for the Extreme patient selections of $n = 20$ and $n = 30$, only κ scores of $\kappa \geq 0.6$ and $\kappa \geq 0.5$, respectively, are considered statistically significant ($p < 0.05$) [291, 292]. This also applies to analysis with low sample numbers such as HFD and mortality.

4.1.3 Results and Discussion

4.1.3.1 Prediction of clinical parameters

Tables 4.2 and 4.3 show the highest κ scores obtained for prediction of each of the clinical parameters (patient mortality, heart failure diagnosis, ischemic time, TScore, MVO, IMR, and troponin and creatinine levels) by the various trained ML models described previously. Table 4.2 shows the results for models trained on mass spectra from all available patients, while Table 4.3 shows results for models trained using mass spectra only for the Extreme groups of patients with the 20 or 30 highest and lowest values for each clinical variable. Full results tables and results distributions for each model can be found in the Appendix (Sections B.2-B.4). Before considering the results for each clinical parameter, we can make some general observations.

i For a given clinical parameter, the κ score, and therefore predictive power of the model, is reasonably consistent across all ML methods and FR methods explored. Interestingly, reducing the data set so that only patients with the 20 or 30 highest and lowest values for the clinical parameter of interest were included led to a significant improvement in performance for prediction of MVO, creatinine, and troponin, but worse performance for prediction of IMR, ischemic time, and TScore.

ii When comparing the performance of different FR methods, we note that in the models trained on all patient data, the highest scores were isolated to data sets for which no FR had been performed (for patient mortality, heart failure diagnosis, TScore, and creatinine), or for which FR had been achieved via statistical analysis (for ischemic time, MVO, IMR, and troponin).

iii Inclusion of clinical and demographic parameters within the training data did not improve the results in most cases.

iv The uncertainties returned from the models were found to be largest in the prediction of patient mortality and heart failure diagnosis. This is consistent with the very small available data sets for these parameters. The kappa scores achieved for patient mortality and HFD were higher than the limit of statistical significance (k>0.6) in many cases, but we recognise that more samples would be needed to confirm this.

The best results achieved for each clinical variable across all models explored are summarised in Table 4.4. Despite the small data set, the most successful prediction was achieved for patient mortality (85% accuracy, κ 0.71), with classification of 'high' and 'low' MVO in extreme patient groups also predicted with a high accuracy of 83%, κ 0.66. Accuracies ranging from 70-80% were achieved for the remaining clinical parameters under investigation. The most successful ML models out of the five considered were LDA and SVM, which between them accounted for all of the best-performing models. LDA tends to perform well when the two classes are well represented by normal distributions, while SVM can cope more easily with outliers.

Both models deal better with noisy and high-dimensionality data sets than the KNN and NBC models. The high accuracies achieved for prediction of patient mortality and heart failure diagnosis after treatment for STEMI come with the caveat that since the number of patients available for the analysis was small (n = 46 and 50, with 23 and 25 patients per class, respectively), the accuracy may be somewhat exaggerated due to over-fitting. However, given the clinical importance of these predictions - the ability to identify patients at risk of fatality or serious complications post-STEMI at the point of presentation may enable such patients to be targeted with more aggressive treatments earlier on in the treatment pathway, potentially saving or prolonging life - further exploration is certainly warranted should samples become available.

Table 4.2 The highest κ scores achieved for each of the nine clinical variables tested with each model setup for all the available patients. The highest score obtained for each variable is highlighted in green.

Data inputs	**Analysis setup**							
Patients used	All	All	All	All	All	All	All	All
Clinical variables included	No	Yes	No	Yes	No	Yes	No	Yes
FR used	None	None	FS	FS	OI	OI	Stats	Stats
Creatinine	0.158±0.134	0.186±0.136	0.164±0.135	0.182±0.137	0.154±0.14	0.108±0.125	0.154±0.132	0.158±0.143
Mortality	0.706±0.212	0.565±0.269	0.552±0.265	0.59±0.262	0.536±0.296	0.576±0.289	0.676±0.236	0.685±0.239
ITime	0.448±0.214	0.481±0.223	0.506±0.179	0.503±0.171	0.487±0.196	0.409±0.199	0.536±0.189	0.531±0.172
HF	0.544±0.257	0.417±0.266	0.505±0.243	0.487±0.25	0.454±0.267	0.506±0.265	0.51±0.257	0.493±0.237
MVO	0.008±0.169	0.031±0.165	0.007±0.177	0.031±0.197	0.035±0.174	0.027±0.162	0.073±0.178	0.079±0.167
Troponin	0.265±0.14	0.279±0.134	0.304±0.129	0.29±0.138	0.316±0.147	0.303±0.122	0.319±0.128	0.305±0.127
IMR	0.324±0.17	0.305±0.17	0.31±0.168	0.334±0.171	0.27±0.165	0.243±0.173	0.333±0.188	0.346±0.155
TScore	0.54±0.184	0.498±0.175	0.49±0.192	0.465±0.177	0.446±0.202	0.464±0.184	0.467±0.2	0.497±0.18

Table 4.3 The highest κ scores achieved for each of the nine clinical variables tested with each model setup for the selections of extreme patients. The highest score obtained for each variable is highlighted in green.

Data inputs	**Analysis setup**							
Patients used	Extreme	Extreme	Extreme	Extreme	Extreme	Extreme	Extreme	Extreme
Clinical variables included	No	Yes	No	Yes	No	Yes	No	Yes
FR used	None	None	FS	FS	OI	OI	Stats	Stats
Creatinine-20	0.373±0.294	0.37±0.629	0.293±0.317	0.06±0.685	0.412±0.326	0.295±0.547	0.383±0.323	0.44±0.623
Creatinine-30	0.35±0.239	0.302±0.342	0.327±0.277	0.162±0.363	0.33±0.255	0.101±0.341	0.333±0.241	0.219±0.332
IMR-20	0.388±0.314	0.029±0.324	0.16±0.324	0.139±0.426	0.18±0.322	0.133±0.442	0.191±0.329	0.101±0.44
IMR-30	0.333±0.296	0.085±0.344	0.198±0.323	0.22±0.56	0.404±0.337	0.26±0.494	0.324±0.331	0.165±0.509
MVO-20	0.646±0.259	0.602±0.294	0.574±0.278	0.524±0.29	0.654±0.274	0.562±0.301	0.663±0.27	0.592±0.258
MVO-30	0.239±0.283	0.194±0.306	0.125±0.273	0.131±0.267	0.194±0.274	0.175±0.313	0.152±0.27	0.132±0.269
ITime-20	0.081±0.335	-0.063±0.38	-0.045±0.305	-0.078±0.418	-0.048±0.307	-0.18±0.378	-0.035±0.291	-0.136±0.329
ITime-30	0.034±0.27	0.022±0.369	0.104±0.283	-0.088±0.35	0.09±0.278	-0.071±0.268	0.035±0.266	-0.034±0.324
Troponin-20	0.51±0.28	0.522±0.355	0.57±0.271	0.498±0.366	0.579±0.279	0.57±0.397	0.561±0.284	0.54±0.351
Troponin-30	0.46±0.23	0.491±0.307	0.394±0.224	0.395±0.32	0.443±0.234	0.45±0.339	0.457±0.242	0.399±0.348
TScore-20	0.03±0.315	0.229±0.489	-0.015±0.338	0.071±0.562	0.011±0.353	0.246±0.453	0.03±0.353	0.219±0.467
TScore-30	0.14±0.242	0.355±0.358	0.078±0.286	0.192±0.354	0.05±0.253	0.34±0.368	-0.055±0.229	0.288±0.353

Table 4.4 The best results obtained for each categorical clinical variable, and the data structure required to achieve that result. Variables predicted with over 80% accuracy highlighted green, variables predicted with over 70% accuracy highlighted yellow.

Clinical variable	Accuracy	κ score	κ error	Dataset used	FR method used	Clinical input used	ML model
Creatinine	72%	0.44	0.62	Extreme 20	Stats	Yes	SVM
Mortality	85%	0.71	0.21	All patients	None	No	LDA
HFD	77%	0.54	0.26	All patients	None	No	LDA
IMR	70%	0.40	0.34	Extreme 30	OI	No	LDA
ITime	77%	0.54	0.19	All patients	Stats	No	LDA
MVO	83%	0.66	0.27	Extreme 20	Stats	No	SVM
Troponin	79%	0.58	0.28	Extreme 20	OI	No	SVM
TScore	77%	0.54	0.26	All patients	None	No	SVM

MVO is a known complication of pPCI in which the myocardial microcirculation remains obstructed despite re-establishment of blood flow through the affected coronary artery [293]. In the OxAMI study, the extent of MVO is measured via a cardiac magnetic resonance imaging (MRI) scan at two days and six months after the STEMI. The majority of patients are measured to have either zero or very low MVO, reflecting successful treatment, and only a relatively small number have high values, indicating a poor cardiac outcome.

Given the threshold of zero for distinguishing between the 'high MVO' and 'low MVO' patient groups, and the fact that the patient distribution is highly skewed towards very low, near-zero, MVO values (see Figure 4.1), it is perhaps not surprising that ML models trained on the 'All patient' ($n = 128$) data were unable to assign patients correctly to these groups, even when FR methods were employed. However, a drastic improvement was found when the models were trained on the 'Extreme patients' data. A κ score of 0.66 and an accuracy of 83% was achieved with the smallest patient selection of $n = 20$ per class, reducing to $\kappa = 0.24$ and an accuracy of 62% in the larger $n = 30$ 'Extreme patients' data set. These results imply that the ML models are capable of identifying patients at most risk of poor outcomes, opening up potential avenues for future investigation.

The IMR, measured via an invasive procedure while the patient is still cathet-

erised, provides an assessment of cardiac microvascular function immediately post-pPCI. IMR is often used as a proxy for MVO, with values higher than 40 known to be associated with an adverse clinical outcome. However, in around a third of cases the two parameters can be surprisingly discordant [294], perhaps due to resolution of reversible symptoms associated with edema relatively early on in the course of the patient's recovery [295]. Given the success of this ML approach at identifying patients with high MVO, it is perhaps surprising that there was very little success in predicting IMR. Using the 'All patients' data set ($n = 126$), the trained models all performed equally poorly, achieving no-to-low agreement, with only a slight improvement to a best accuracy of 67% ($\kappa = 0.346$, low agreement) when feature selection was implemented. When using the 'extreme patients' data the performance of the models was more varied, with the best result (70% accuracy, $\kappa = 0.40$) obtained for the SVM model trained on the IMR-30 data set with OI FR.

Ischemic time was predicted with low accuracy using the 'All patients' ($n = 92$) data set, improving to fair agreement (77% accuracy, $\kappa = 0.54$) when feature selection was implemented. Limiting the data set to 'Extreme value' patients led to a drastic reduction in predictive power to no better than a random guess, implying that a larger data set is beneficial when characterising this parameter. Since ischemic time is often relatively subjective, and usually known at the time of patient admission, the ability to 'predict' this quantity is probably not useful in developing diagnostic or prognostic indicators for heart health following STEMI. However, our results do show that there are detectable metabolic changes in plasma that correlate with ischemic time, which with further investigation may be beneficial in understanding physiological changes associated with coronary artery occlusion.

TScore (or thrombus burden) is measured during pPCI by passage of a flow guide wire, and measures the extent of occlusion to the culprit coronary artery [271, 296, 297]. The performance of the ML models in relation to TScore was very similar to that for ischemic time, showing fair (77%) agreement ($\kappa = 0.54$) when

trained on the 'All patients' data, and no predictive power when trained only on the 'Extreme patients' data. The ability to predict TScore implies that thrombus size has a measurable effect on the metabolic profile of the patient that could potentially prove useful in making clinical decisions.

The final two clinical parameters, peak troponin and creatinine levels, are somewhat different from the other parameters in that they relate to concentrations of specific identifiable chemical species. Both are biomarkers for STEMI, with troponin being the more sensitive and specific, and therefore more widely used in clinical practice. Despite the fact that creatinine is a small molecule lying within the m/z range of our mass spectrometer, none of the ML models performed particularly well in classifying patients to 'high' or 'low' groups. No model performed significantly better than a random guess when presented with the data set for all 242 patient samples, and only low agreement was achieved when the data set was limited to the 20 patients with the highest and lowest values for creatinine, with FR and inclusion of demographic variables improving the situation slightly to yield a κ score of 0.44 and accuracy of 72% obtained using the SVM model. This metabolite will be discussed further in the next section when considering the m/z peaks of interest revealed through the FR process.

Troponin is a protein with a molecular mass outside of the range of our mass spectrometer; hence plasma troponin cannot be measured directly in the present study, but we may expect to be sensitive to changes in the plasma metabolome resulting from alterations in biological pathways involving troponin. Using the 'All patients' ($n = 220$) data set the best result was an accuracy of 66% ($\kappa = 0.319$), with similar results across all ML models trained on the feature selected data. Using the data set comprising the patients with the 20 highest and lowest troponin values improved the accuracy to 78% and κ to 0.58, with the best results for the SVM model. Performance was fairly consistent across all models apart from RFC, which performed poorly in all cases.

4.1.3.2 Identification of potential biomarkers by FR

Interrogating the results of the various FR methods allows us to identify which *m/z* peaks are most important in distinguishing between the patient groups of interest. It was generally found that there was some overlap between the peaks identified by the three different FR methods employed, but also substantial differences. In a first attempt to identify potential mass spectrometric biomarkers associated with each clinical variable, we applied the criterion that qualifying *m/z* peaks should be identified in more than 75% of the data training partitions for all three FR methods employed. Table 4.5 shows the results of this analysis, which returned between one and six *m/z* peaks of interest for each clinical variable. Figures 4.2 and 4.3 show the intensity distributions for these peaks within the different patient groups associated with each clinical parameter, in the form of box and whisker plots. The *t statistic* and *p score* determined for each *m/z* peak can be found in the Appendix (Section B.5).

Table 4.5 The *m/z* ion peaks identified as being important for classification by the 3 FR methods for each clinical variable.

Clinical variable	**ID of *m/z* ion peaks found with > 75% occurrence across 3 FR methods**	**ID of *m/z* ion peaks found with > 75% occurrence across 3 FR methods in extreme grouped datasets**
Creatinine	114	114, 62
Mortality	129, 169, 156, 197	NA
ITime	387, 45, 46	None
HFD	208, 219, 532, 189	NA
MVO	None	355, 387, 531, 532, 533
Troponin	370, 371, 45, 46	49, 576, 45, 46
IMR	45, 46	538
TScore	45, 46	115

The present study is focused on a pattern recognition approach, and the resolution of our ASAP-MS instrument is not high enough for us to provide a definitive identification of the molecules giving rise to the peaks identified by the various FR methods. While metabolite databases such as the Human Metabolome Database (HMDB) [298, 299] can identify some potential candidates, more sophisticated

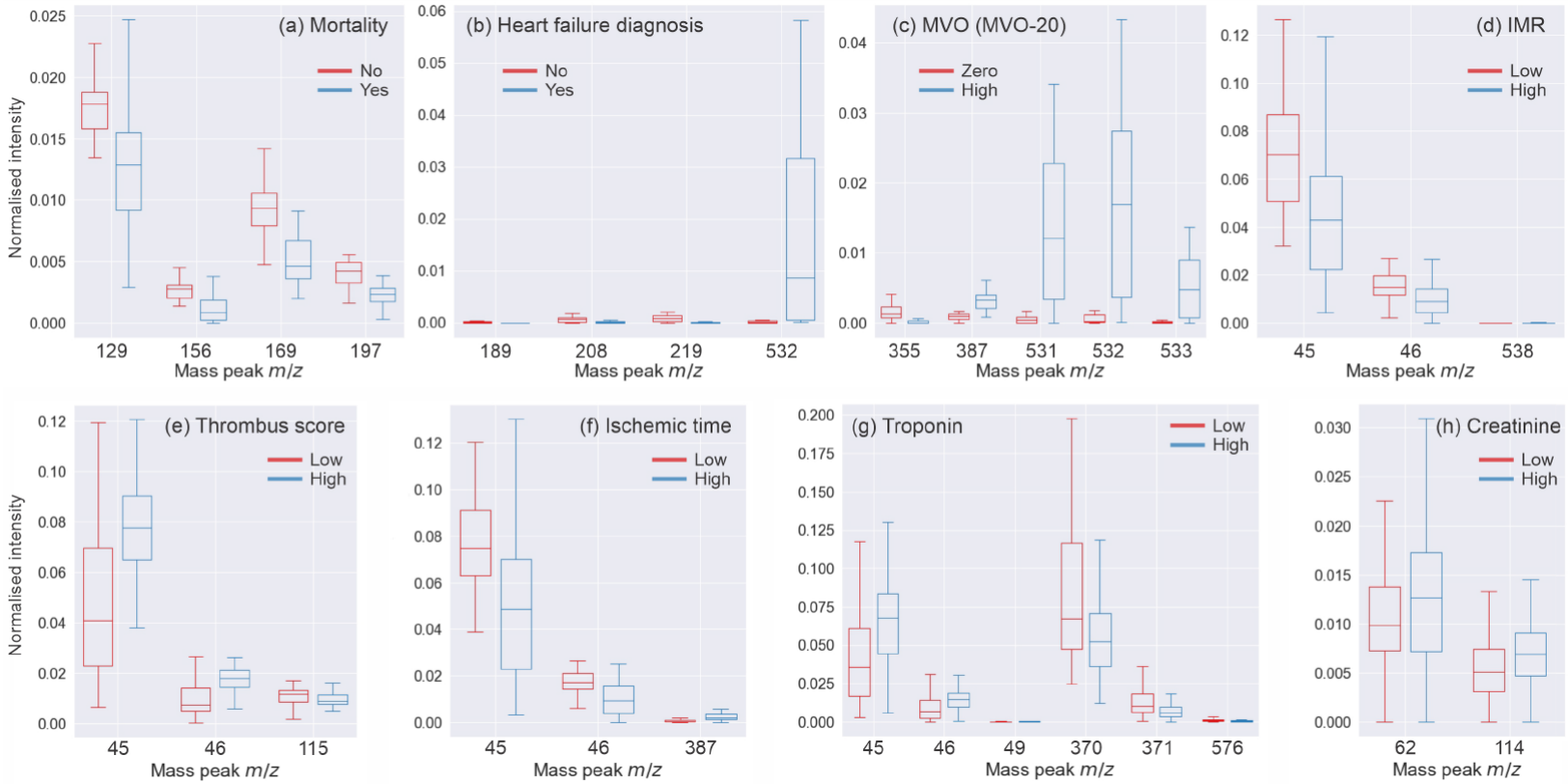

Figure 4.2 Box and whisker plots showing the distribution of normalised peak intensities within the 'All patients' data set (except for MVO, which used the MVO-20 data set) for *m/z* peaks identified by the FR analysis as being of interest for each patient group and clinical parameter investigated: (a) patient mortality; (b) heart failure diagnosis (see Figure 4.3 for the same data on a magnified scale); (c) MVO (MVO-20 data set); (d) IMR; (e) TScore; (f) ischemic time; (g) troponin; (h) creatinine. In each plot, the central line represents the data median, with the inner box representing the inter-quartile range between the first and third quartiles. The whiskers extend from the box to the farthest data point lying within 1.5× the inter-quartile range from the box. Significance was measured by an independent t-test.

methods such as liquid chromatography mass spectrometry (LCMS) will be needed in order to progress work on identifying molecules of interest. However, with this caveat in place, there are a few masses worth commenting on.

The pair of m/z values 45 and 46 were found to be important for four of the clinical parameters - ischemic time, IMR, TScore, and troponin - always as a pair. The ASAP-MS instrument tends to generate protonated positive ions, so assuming a protonated parent ion, potential molecular candidates are CO_2 (carbon dioxide), C_2H_4O (acetaldehyde), and C_2H_7N (dimethylamine or ethylamine). All of these molecules are common metabolites relevant to numerous anabolic and catabolic pathways that are relevant in many disease states. Alternatively, the m/z 45 and 46 peaks may be fragment ions generated from larger metabolites.

Two peaks were found to be important for successful classification of creatinine, one of which was at m/z 114. As creatinine has a molecular mass of 113.12 g/mol, it is highly likely that this peak corresponds to the protonated parent ion. However, the accuracy with which creatinine is predicted from the mass spectra is low, and the intensity of this peak shows no correlation with the clinically reported value for creatinine. There are several possible reasons for this discrepancy. The peak creatinine level measured in the clinic is recorded at a different time from the plasma sample collected from the coronary artery during pPCI. The rate of creatinine excretion via the kidneys differs significantly between patients [300, 301], leading to considerable inter-patient variability in the correlation between the two creatinine levels. It is also possible that blood plasma that has been stagnating in the region occluded by the coronary thrombus is not representative of the systemic blood creatinine. Finally, mass spectrometric analysis is only semi-quantitative, in comparison with the highly quantitative clinical measurement of creatinine [208, 211].

Table 4.6 shows the results of training the ML models using only the very small number of mass peaks identified in Table 4.5, compared against the best results obtained using all m/z peaks. We see that, for the most part, the accuracies are similar

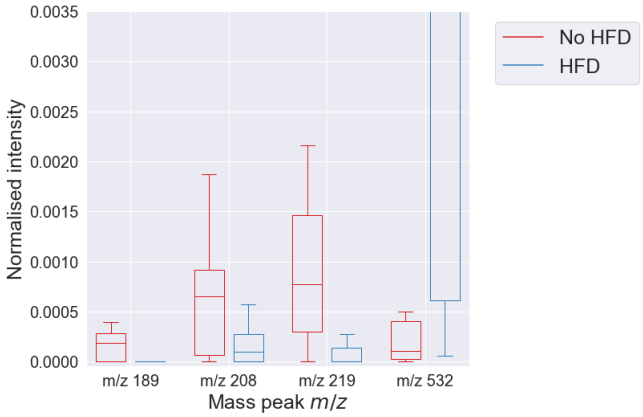

Figure 4.3 Box and whisker plots showing the distribution of normalised peak intensities within the 'All patients' data set for *m/z* peaks identified by the FR analysis as being of interest for the heart failure diagnosis clinical parameter at a magnified scale.

Table 4.6 The κ scores obtained when the analysis of the undersampled datasets were repeated using only the peaks found to be important for classification in Table 4.5 compared against the best result found using all available peaks. Changes in κ score of less than or greater than 0.05 are highlighted in red and green, respectively.

Clinical variable	**Best κ score**	
	All peaks	**Identified peaks only**
Creatinine	0.186±0.136	0.238±0.127
Mortality	0.706±0.212	0.559±0.249
ITime	0.536±0.189	0.521±0.191
HFD	0.544±0.257	0.659±0.212
MVO	0.079±0.167	0.059±0.175
Troponin	0.319±0.128	0.383±0.111
IMR	0.346±0.155	0.386±0.148
Tscore	0.54±0.184	0.488±0.191

in the two analyses, implying that the peaks found by the FR algorithms account for much of the variation between the patient groups of interest. For creatinine, HFD, troponin, and IMR, the κ scores obtained using the feature-selected m/z peaks were slightly higher than those obtained using all peaks, while for mortality, TScore, ITime, and MVO the κ score was lower, though still in the 'fair agreement' range. The fact that a very small number of m/z peaks seem to be of high importance when classifying patients is a positive outcome for any future clinical applications, since focusing on the quantification of a small number of molecules is likely to improve the efficiency of any analysis considerably.

4.1.4 Identification of peaks using high-resolution ASAP-Q-TOF-MS analysis of blood plasma

One of the next steps in this analysis is the identification of some of the ion peaks identified in this analysis. This would take us a step further towards mapping biomolecular changes in the human metabolome to the differences observed in the recovery and prognosis of STEMI patients. A pilot study was conducted in which plasma samples were measured using an ASAP-Q-TOF-MS instrument, the methods and instrument setup for which are discussed in Chapter 6. Unfortunately, without the use of internal standards and separation techniques such as chromatography, ion peak identification was not possible. A number of potential peaks were identified through comparison with the HMDB, details of which are shown in Appendix B.6. Experiments to improve ion identification capabilities are described in the Future work section.

4.2 Symbolic regression for the prediction of clinical variables

4.2.1 Introduction

We have shown that machine learning classification can be used successfully to predict patient outcomes using ASAP-MS data. However, if the clinical variable is continuous, the mass spectra must first be encoded into appropriate binary groupings (e.g. 'low' and 'high'). Ideally, continuous data would be predicted more accurately using regression analysis. However, there are many challenges that prevent conventional regression approaches from working using our currently available data. Firstly, the number of samples available is very low, and the patient samples that are available are unevenly distributed across the clinical range. Secondly, the functional relationship between the intensities of peaks within the mass spectra and

the clinical outcome of interest is unknown (e.g. linear, non-linear, exponential to suggest some simplest cases), and so choosing an appropriate model to fit is not straightforward. The features (mass peaks) to include in a model are not known, and given the high dimensionality of this data type, the inclusion of all peaks is not feasible. There is also error associated with both the mass spectral data (which can be approximately characterised), and the clinical measurements (which have an unknown error). Symbolic regression, described previously in Section 3.2.3 is an approach to regression that may be able to cope with these challenges.

4.2.2 Methods

4.2.2.1 Symbolic regression methods

A large number of open-source symbolic regression codes exist in a range of programming languages for use in research environments. In this study, seven such tools were evaluated for their suitability in performing regression on the OxAMI dataset: PySR (Python and Julia) [251, 302, 303], QLattice (Python) [252, 304, 305], Evolutionary Forest (Python) [306–308], HeuristicLab (C#) [309, 310], GP-GOMEA (C++ and Python) [311, 312], and dCGP (C++ and Python) [313–315]. Among these, PySR and QLattice emerged as the most promising candidates due to ease of use, availability of accessible documentation, and interpretability. Ultimately, PySR was selected due to its superior explainability in both methodology and results, the large range of hyperparameters to optimise for exploring the evolution space, as well as its higher accuracy and generalisability in preliminary tests. The PySR model applies a genetic algorithm approach to the optimisation of nodal tree structures set up to describe the data, as explained previously in Section 3.2.3.

4.2.2.2 OxAMI clinical variables for prediction

A number of clinical measurements and outcomes were chosen for prediction by symbolic regression. The continuous variables that were previously converted to binary

groupings in Section 4.1 and classified were chosen: MVO (left ventrical (LV)%), IMR (no units), peak troponin (μg/L), ischaemic time (minutes), and creatinine on admission (μg/L). Another four variables were also added to the prediction analysis, which were chosen due to their relevance in assessing STEMI patient prognosis. These were acute infarct size (AIS) (units LV%, n=138), final infarct size (FIS) (LV%, n=102), myocardial salvage index (MSI) (%, n=85), and modification of diet in renal disease equation for estimated glomerular filtration rate (MDRD eGFR) (μg/L, n=240). These variables were not classified previously due to low patient numbers, missing data, and skewed data distributions. AIS, also known as infarct volume, is defined as the volume of cardiac tissue that is impaired by the STEMI, measured as the LV% by late gadolinium enhancement (LGE) cardiac MRI at the time of pPCI [316, 317]. FIS is the same measurement taken at 6 months after the STEMI pPCI, which indicates how much of the heart function has been permanently impaired [318]. MSI was defined as the difference between potential and actual infarct size, or the difference between the AIS area and the FIS measurement [275]. The MDRD eGFR measurement is a widely used tool for assessing kidney function [319]. The equation incorporates creatinine measurements, age, sex, race and BSA to estimate a standardised measure of renal dysfunction. Renal dysfunction has been shown to have a strong relationship with poor outcomes in STEMI patients, as well as having important implications for drug dose adjustments and contrast media due to reduced clearance ability [319, 320].

4.2.2.3 Application of the PySR model

The PySR model was applied to the OxAMI data using a similar approach to that used for the classification models. The patients relevant to the clinical variable of interest were selected, and split into a training and test set using an 80:20 split, and an averaged and normalised ASAP-MS spectrum was recorded for each patient. We then explored preparation of the resulting data set in three different ways for

input to the PySR model, similar to the approaches explored in the classification analysis. In approach (i), the PySR model was trained using the complete mass spectra with all 990 m/z variables used. In approach (ii), a 10-component PCA feature reduction was applied to the training data, before using it to train the PySR model. In approach (iii), a statistical test was applied to the continuous data, and the statistically significant peaks were used for PySR training. In the latter approach, a parametric Pearson correlation coefficient test was used as described in the previous section for classification analysis. Significant peaks were determined at the $p < 0.001$, $p < 0.005$, $p < 0.01$, and $p < 0.05$ levels for the training data. If significant peaks were found at the highest confidence level ($p < 0.001$), then those peaks were used for the PySR analysis. If no peaks were found at that level, peaks were found for the next confidence level down, and repeated until significant peaks were identified.

The PySR hyperparameters were set as shown in Appendix B.7 (see reference [321] for hyperparameter definitions). A number of these hyperparameters were optimised as discussed in Section 4.2.2.5. The PySR analysis was run three times for each clinical variable outcome and input method, respectively, each time with a new training-test data partition. The average training run time was approximately 12 hours, with variables for which a large number of patient samples were available (Creatinine, MDRD, and Troponin), having average run times of up to 24 hours. PySR training was conducted using a custom Python 3.7 software implementation of the PySR function, and was run on CPU nodes within the University of Oxford ARC HPC cluster.

Once the model had been trained, the output equation was applied to the test data. If feature reduction was used, the feature reduction algorithm used on the training data was applied to the test data. The accuracy of the model predictions was then assessed by Pearson's r correlation (defined previously in Section 3.2.4.1 Equation 3.5), and the prediction error described using RMSE.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Predicted_i - Actual_i)^2}{n}} \tag{4.1}$$

In cases for which the algorithm generated multiple equations of differing complexity, the equation with the highest r^2 value when applied to the test data is reported as the output equation.

4.2.2.4 Reducing overfitting

As mentioned previously, symbolic regression is prone to overfitting as it can model noise perfectly if given the freedom to explore a sufficiently large space of allowed functions. A number of actions can be taken to help reduce this behaviour: optimising the hyperparameters to restrict the allowed function space; restricting the model complexity; data partitioning; assessment by test data fit; and selection of appropriate data to predict in order to avoid data dredging. These actions have been implemented in the present analysis, but overfitting of the training data may still occur.

4.2.2.5 Hyperparameter optimisation

Within the PySR algorithm, there are a large number of hyperparameters that can be altered to adjust the symbolic regression evolution space and complexity. Three of the most important hyperparameters were explored and optimised in order to improve the accuracy and generalisation of the models. These parameters were `n_iterations` (the number of iterations over which to run the genetic algorithm), `ncycles_per_iteration` (the number of mutations to include per 10 samples per population per iteration of the algorithm), and `population_size` (the number of individual 'test' functions in each population) [321]. To assess each hyperparameter, the analysis was repeated with `ncycles_per_iteration` and `n_iterations` set to 10, 50, 100, 500, 1000 and 5000, and `populations_size` set to 25, 50, 75, 100, 250, 500, 750 and 1000. The aim of this investigation was to understand if changes to

these values would reduce the degree of overfitting. No clear relationship was identified between model accuracy or overfitting for any of these three hyperparameters (further details of these results are shown in the Appendix, Section B.8). Therefore, the values used for these hyperparameters were set to those shown in the Table in Appendix B.7, the values of which were a compromise between allowing increased model evolution, and efficient use of computational resources.

The complexity of the equation generated, which is a combination of equation length, number of mathematical operators, and number of input variables to use, can vary between 1 and 50 within the PySR algorithm. This was also assessed to determine the relationship between equation fit and complexity for each variable. For each analysis, the `Maxsize` parameter, which defines the maximum complexity allowed, was set to the maximum of 50. After each evolution of the algorithm, the best performing equations across the range of complexities are reported. The final equations reported show the best accuracies generated across the complexity range, so the user can observe the impact of increasing complexity on the fit of the data.

For some parameters, lower complexity equations are optimal for reducing overfitting (observed for MVO in Figure 4.4(a)), whereas for others, increasing equation complexity is beneficial for both the training and test data fit. This behaviour was observed in the exploration of the OxAMI variables, examples of which are shown for three variables in Figure 4.4. The point at which the training and test data give similar accuracies is the point at which the model is generalised and not overfitting the training data. It is observed that the training data accuracy will always increase with increased complexity (see rising blue line in each example in Figure 4.4). However, if the test data accuracy consistently falls lower than the training accuracy, the model is overfitting the training data. This is one of the disadvantages of symbolic regression, as overfitting very easily occurs. In general, a simpler equation is preferred when using symbolic regression, as the equation is more likely to be generalised, hence reducing the chance of overfitting, and is more likely to be a 'real'

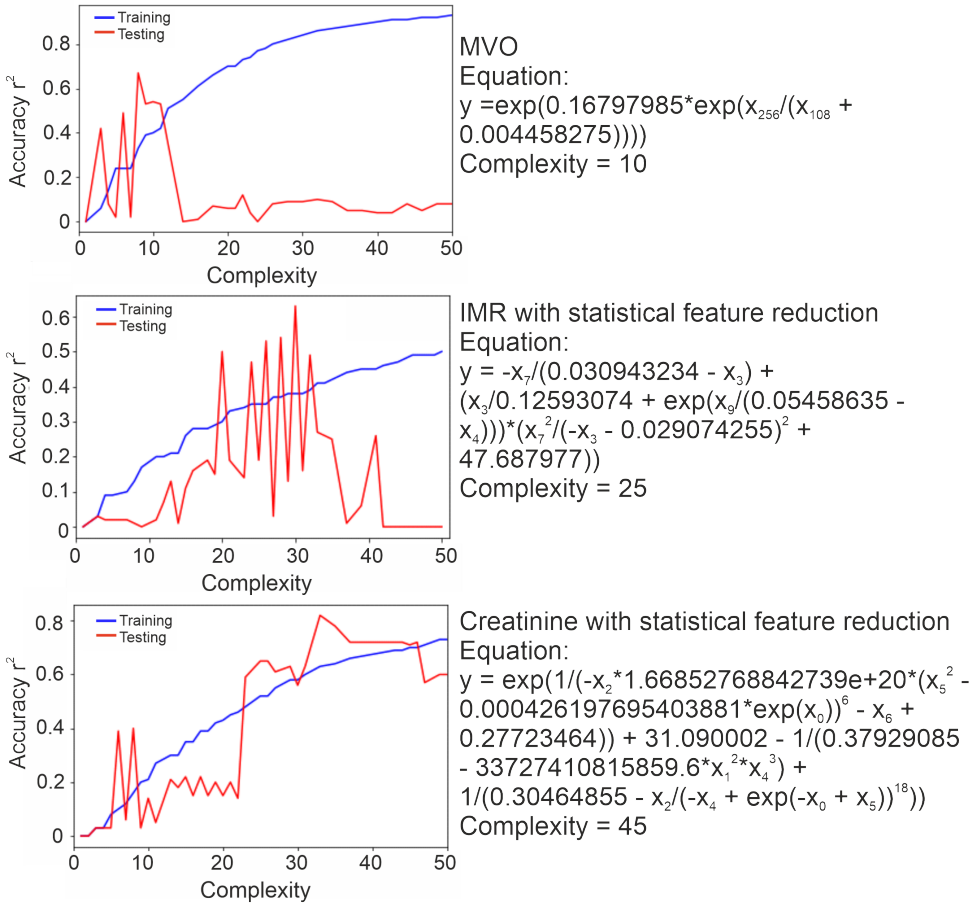

Figure 4.4 The relationship between PySR equation complexity and accuracy by r^2 analysis applied to both the training (blue) and test (red) data. The results are shown for the prediction of MVO without feature reduction (top), IMR with statistical feature (middle), and Creatinine with statistical feature reduction (lower), showing best generalisation at low, intermediate, and high complexity equations, respectively.

correlation between the two datasets rather than an artefact of the measurement. Due to this relationship, the complexity of optimised equations for different clinical variables is expected to vary.

4.2.3 Results

Using the PySR symbolic regression model to predict OxAMI clinical variables was moderately successful, and identified some interesting features of both the data and the analysis method. The results for this analysis are plotted in Figure 4.5, which

shows the accuracy of the prediction across each training run for each variable. The accuracy is shown as the minimum of either the training or test data r^2 value, whichever was lower. This was done to report an unbiased representation of the minimum accuracy. Repeat partitions are shown by changing hues of the same colours. Some key observations from these results are as follows:

i The accuracy observed across repeat runs of the algorithm with different partitions was inconsistent. This was first thought to be due to partitioning dependence, but was then shown to be run dependent, as running the same algorithm on the same data produced different results. This is in line with other analyses conducted using symbolic regression, where many runs are completed using the same model to try to find a number of 'good' equations [252].

ii There was no obvious optimum data preparation approach out of the three methods tested, although the statistical feature reduction method (blue bars in Figure 4.5) provided high accuracy results on numerous occasions.

iii The overall accuracy of the trained models was low, which is consistent with the errors present in the ASAP-MS and clinical data. However, a number of equations were generated with $r^2 > 0.4$, which is considered potentially clinically significant.

iv Some variables, particularly AIS and FIS, were predicted with high accuracy, and regression plots appear to show the equations generated have true predictive capability.

v Some results appeared to show high accuracy predictions but were found to be misleading when plots of the predicted and true clinical values were studied.

As the present study employs data relating to human health and behaviour, measurements with r^2 values of > 0.2 can be reported as meaningful in some cases [254, 322]. There were many equations found that resulted in $r^2 > 0.2$ across the

range of clinical variables studied, showing that there were many potential relationships found between the clinical data and peaks in the MS data. However, due to the stochastic (statistical and random probability distribution of initial search parameters across a large search space), and non-deterministic (not able to give the same result every time) properties of symbolic regression algorithms, it was rarely possible to find the same relationships repeated in the output equations.[a] Increasing the accuracy in this analysis may be possible with more sophisticated algorithms or more data, but the most effective way to increase the accuracy would be to reduce the error in both the mass spectrometry data and the clinical measurements. Therefore, only tentative conclusions will be made from this analysis.

Results for the prediction of AIS and FIS showed some relatively accurate predictions. The plots in Figure 4.6 show the prediction of both the training and test data, with $r^2 > 0.3$ for AIS and $r^2 > 0.4$ for FIS, importantly with very good generalisation between the two datasets. The equations associated with these results, Equations 4.2 and 4.3, were generated using statistical feature reduction, and so the mass peaks used in the analysis can be probed. Equations 4.2 and 4.3 are written with respect to the mass peaks used:

$$AIS_p\ LV\% = \frac{0.00092929037}{(-I_{56}/(I_{44}/0.207251 - 0.3582202) + I_{63} - I_{70} + I_{73})} + \frac{0.002034326}{(I_{73} - 0.0028416575)} + \frac{1}{(I_{44} + 0.031837903))} \quad (4.2)$$

$$FIS_p\ LV\% = \frac{0.5738643}{(I_{44} + ((I_{164})^2/(I_{239})^2) + I_{239} + 0.017887488)} + \frac{2 \times I_{60}}{I_{370} - 0.089101024} \quad (4.3)$$

where I_n represents the intensity of the peak at $m/z = n$. Both equations were in the low to medium complexity range at 26 and 19, respectively, which is ideal for

[a]For example, an equation may be produced that gives high accuracy with a reliance on peaks x, y, z, and another equation with a similar accuracy may not feature any of those variables.

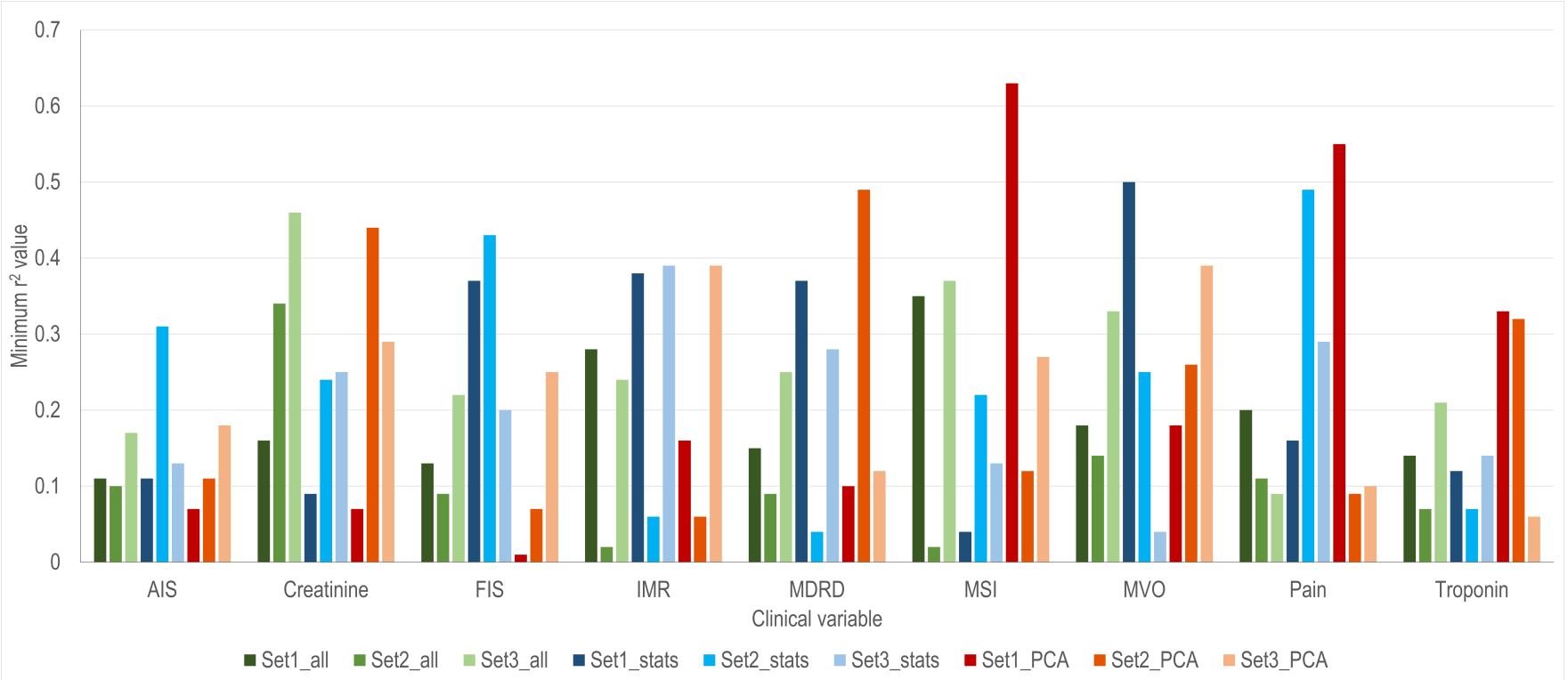

Figure 4.5 The accuracy obtained for each optimum PySR output, run over the nine clinical variables tested. The accuracy is shown as the r^2 fit between the predicted and true values for each variable, with the lower of either the test or training fit shown. The colours represent the data input method, with no feature reduction shown in green, statistical feature reduction in blue, and PCA feature reduction in red. The colour hues show the repeat results over multiple partitions.

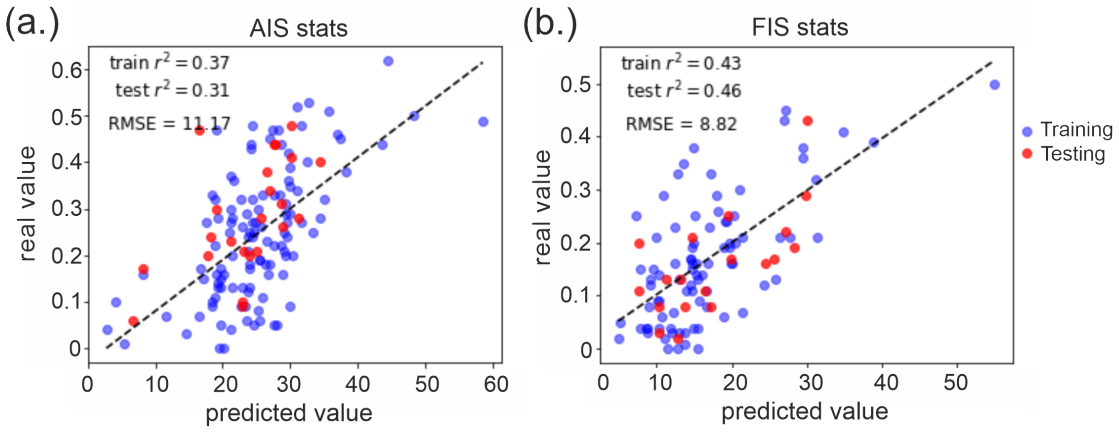

Figure 4.6 The PySR predicted values for AIS (a) and FIS(b) plotted against the true measured clinical values. The training data is shown in blue, and the test data in red. Both variables are reported as values with units LV%. The training and test values are generated by applying the PySR output equation to the input data partitioned the same as for input to PySR.

this analysis as it gives confidence that these relationships are, to some degree, 'real'. A further indicator that these equations represent real relationships in the data is the strong correlation between the AIS and FIS clinical measurements, shown in Figure 4.7.

Patients with a high AIS and FIS have a worse prognosis due to the extent of cardiac damage. Only one mass peak was shared between these two equations, at $m/z = 44$, which has been tentatively assigned previously as formaldehyde ion. Furthermore, the peaks appearing in the AIS equation were all in the low mass region, whereas the FIS peaks spanned further into the higher masses. It is interesting that there were not more peaks in common given the strong relationship between these variables, as shown in Figure 4.7. The ability of symbolic regression analysis to find relationships between the ASAP-MS data and the AIS and FIS measurements gives yet further support to the use of ASAP-MS to probe metabolomic data at a high level, and shows a strong link between blood plasma metabolic profiles and STEMI prognosis.

The equations resulting from the symbolic regression analysis for the other variables are shown in the Appendix (Section B.9). Many of the other 'high accuracy'

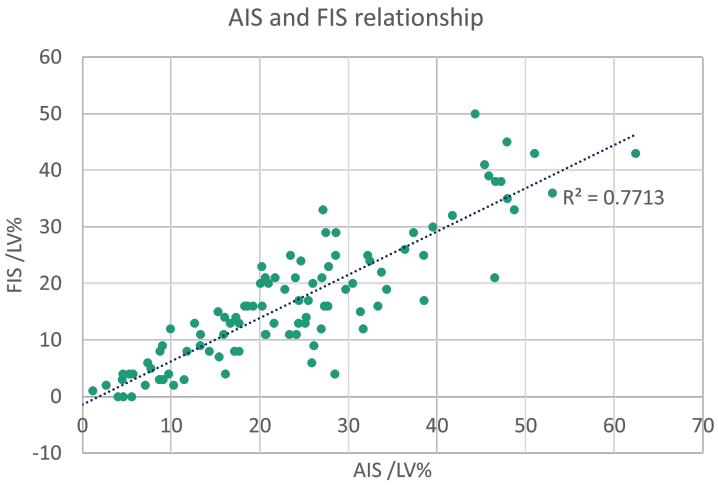

Figure 4.7 The correlation between measured AIS and measured FIS as reported in the clinical dataset. Both AIS and FIS are reported as LV%. Note the slight trend towards higher AIS values, showing that some patients recover more cardiac function that expected.

results were found to be either misleading or the result of poorly distributed training and test data. An example of this is shown for MVO in Figure 4.8. The initial plot of true vs predicted MVO shows what appears to be a highly correlated dataset with one anomalous result. On removal of that data point, results show that the training data has high accuracy, but the test data partition was restricted only to the low MVO region and so is not a representative selection. When the analysis was repeated on a partition that had a better test data distribution, the accuracy of the resulting output was very low. This demonstrates the importance of taking a 'human in the loop' type approach to this machine learning method, in which an 'expert' helps make decisions to guide suitable analysis [323].

Due to the explainability of this method, important peaks can be extracted from the statistical FR and no-FR analysis, and compared with the peaks found in the previous classification analysis. The equations generated across partitions were examined, and the MS peaks used in those equations were reported. Table 4.7 shows these results. Any peaks identified in both the no-FS and statistical FS equations were highlighted (column 4). Although not all of those equations produced high-accuracy results, it was useful to understand whether particular peaks were

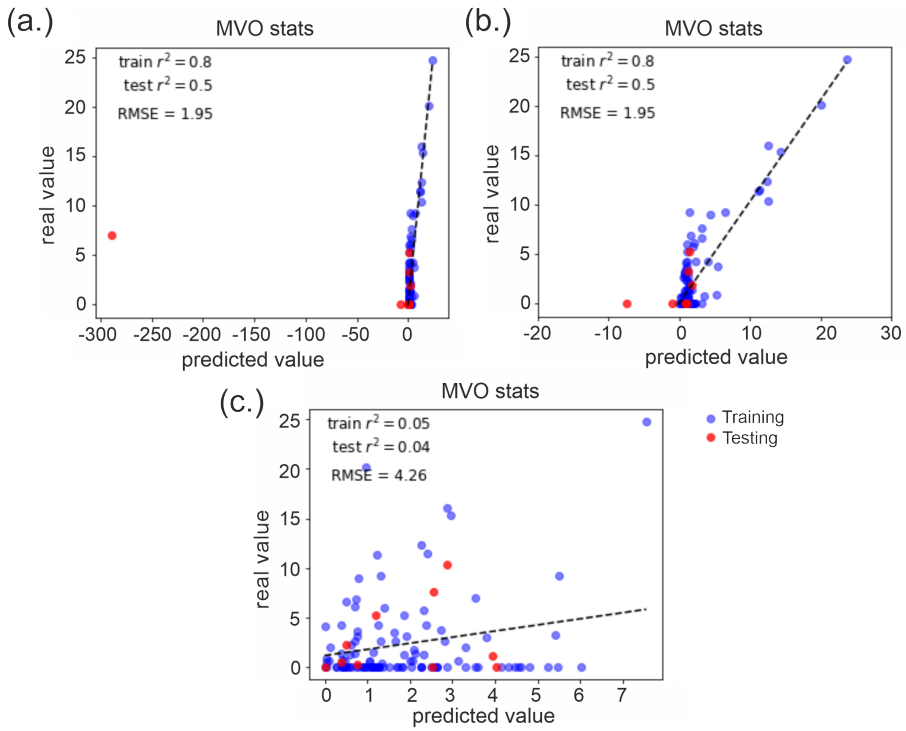

Figure 4.8 The PySR predicted values for the highest accuracy prediction of MVO (a) plotted against the true measured clinical values. Removal of the outlier in plot (a) results in plot (b), showing a clear bias in the test data to low MVO values. Plot (c) shows a repeat analysis conducted with a more evenly distributed partition. The training data is shown in blue, and the test data in red. MVO is reported with units LV%.

being selected by the algorithms repeatedly. Several peaks were identified as being used across all partitions, including for AIS and FIS. The peak at $m/z = 44$, the formaldehyde fragment, was found in both AIS and FIS analysis. A number of peaks identified in this analysis also overlapped with peaks identified in the classification analysis discussed in the previous Section. This includes the peak at $m/z = 114$ in the creatinine analysis, and the peak at $m/z = 387$ for MVO.

4.3 Conclusions and outlook

We have presented results from a proof-of-concept investigation into the use of a combined ASAP-MS and ML analysis of blood plasma for the prediction of clinically relevant pathologies and outcomes for patients participating in the OxAMI study

Table 4.7 Analysis of the peaks found within the PySR equations. A '-' entry indicates no peaks identified across that analysis.

Clinical variable	Peaks used in statistical FS equations in all partitions	Peaks used in no-FS equations in all partitions	Overlap between peaks found in statistical FS equations and no-FS equations (any partition)	Overlap with peaks found in previous classification analysis
AIS	44	44	44	NA
Creatinine	114	61	61, 114	114
FIS	60, 164	44	44, 164	NA
IMR	45, 239	-	522	45
ITime	-	-	18	-
MDRD eGFR	62, 603	62	62, 69, 603	62
MSI	44	-	-	NA
MVO	387	-	-	387
Troponin	-	-	45, 370, 46	45, 46, 370

into ST-segment elevation myocardial infarction. Prediction of the nine clinical variables was completed through analysis of ASAP-MS measurements of coronary aspirate plasma, with patient mortality and MVO predicted with over 80% accuracy, HFD, ITime, peak troponin and TScore predicted with over 76% accuracy, and creatinine and IMR predicted with over 70% accuracy. Feature-reduction methods were applied to try to improve the speed and accuracy of the models used, but no single method was found to give particularly significant improvements to model accuracy. It was found that the ML method requirements were specific to the variable being analysed. Using the feature-reduction methods, peaks found to be important for classification were identified. The biomolecules responsible for these peaks will be investigated in further studies, with the aim of learning more about the pathology of STEMI patients and potentially developing clinical tools for identifying high-risk STEMI patients. The patient numbers used in this analysis were low, which may contribute both to over-fitting by ML models, or to reduced accuracy due to sample representation. Ideally, this study will now be repeated with larger patient numbers, potentially from a number of studies similar in set-up to the OxAMI study.

Analysis of coronary aspirate plasma was completed in this early investigation,

as this resource was not deemed to be as valuable to the biological database as other samples. As discussed previously, coronary aspirate blood has been in some cases sat stagnant in the occluded coronary artery for unknown periods of time, and will be contaminated to differing degrees with saline solution and medication administered to aid the pPCI procedure. The reproducibility in terms of composition, consistency, and representation of the patient for coronary aspirate blood is likely to be low compared to plasma samples taken from a standard blood draw. This increased variation will have a large impact on the results obtained using ASAP-MS.

The use of symbolic regression to find functional relationships between ASAP-MS data and clinical data appears to offer a potentially powerful tool. A number of relationships were found between mass spectra and clinical variables that could prove to be prognostically useful. The impact of the relatively large uncertainties in the ASAP-MS data was clear in this analysis, with the predictive accuracies of the regression models being generally low. Overfitting is a common occurrence in symbolic regression, but we have shown that comparing test and training predictions allows this to be detected fairly quickly. Further work could focus on running many more repeats of each analysis and combining the results to assess the type of equations that are being found, the predictive accuracy that can be obtained, and the associated uncertainties. The high computational cost of symbolic regression means that care should be taken to ensure meaningful tests are performed in order to reduce wasted resources, and adequate memory should be allocated to each run.

The results of the present study have determined that our approach offers predictive power that may be useful in developing new clinical tools for the assessment of patient risk and prognosis. The analysis will now be repeated on a range of biological samples that are likely to be more consistent than coronary aspirate samples, such as venous blood samples and coronary sinus plasma.

5

Prediction of AAA development and treatment progression

5.1 The prediction of AAA patient treatment groups from venous plasma samples

5.1.1 Introduction

Patients with abdominal aortic aneurysms (AAA) exhibit distinct biochemical profiles at various stages of disease progression and treatment. Differences in metabolic plasma signatures between individuals with small and large aneurysms have been reported [324, 325], along with variations in tissue gene expression [326]. These molecular patterns offer insights into the underlying mechanisms of AAA development and expansion.

High-resolution mass spectrometry has been a key tool in identifying metabolite differences across patient cohorts. However, this process is time and resource intensive, and often requires prior identification of a single or small group of biomolecules to study. Ambient ionisation mass spectrometry presents a rapid and promising approach for qualitatively profiling the small-molecule composition of blood plasma across a much larger, untargeted selection of molecules. This technique could un-

cover novel methods of rapidly classifying patients, as well as relationships between metabolites that are not apparent through targeted analyses.

By adopting an untargeted fingerprinting strategy combined with machine learning classification, the present study aims to reveal systemic biochemical alterations associated with AAA size and treatment status. The goal of the study is to evaluate whether ambient surface analysis probe mass spectrometry (ASAP-MS) can detect circulating metabolic changes in AAA patients. Using machine learning to classify patients by aneurysm size and treatment stage, we aim to assess the potential of this technique in distinguishing biochemical fingerprints related to disease progression.

This research does not seek to replace current ultrasound-based AAA monitoring protocols, but instead aims to enhance our understanding of how AAA influences systemic biochemistry, and whether disease progression can be better understood or predicted through metabolic signatures.

5.1.2 Methods

5.1.2.1 Study patients and criteria

The study employed blood plasma samples from the OxAAA study, as described in Section 2.3, taken from 280 prospectively enrolled AAA patients. Plasma samples were distributed across the four patient streams described previously: demographically matched healthy volunteers (HV), 'small AAA' patients (S1 TP1), 'large AAA' patients pre surgery (S2 TP3), and large AAA post-surgery (S2 TP-5).

5.1.2.2 Materials and Instrumentation

ASAP-MS analysis of blood plasma was conducted according to the optimised method described in Section 2.4.4 using the Advion CMS-S instrument. A set of 10 repeat capillary measurements was recorded for each sample, and averaged to generate a single spectrum per patient. The practical work in this section was conducted with the help of Anna Jones, a Part II student in the group in 2022-2023. Data ana-

lysis was performed using the methods described in Chapter 3, implemented using custom MATLAB R2022a and Python 3.7 software. Machine learning analysis was conducted on CPU nodes within the University of Oxford ARC HPC cluster.

5.1.2.3 Data analysis: Machine learning classification of AAA patient groups

The clinical parameters of interest were the presence or absence of an aneurysm, aneurysm size, and patient systemic response to aneurysm repair surgery. From the OxAAA study, the data sets listed were available to test the ability to classify according to each of these criteria:

1. **Presence or absence of aneurysm:**
    - Small aneurysm surveillance stream patients compared to healthy volunteers (S1 TP-S1 vs HV)
    - Large aneurysm patients' pre-surgical measurement compared to healthy volunteers (S2 TP-3 vs HV)
2. **Aneurysm size:**
    - Large aneurysm patients pre-surgical measurement compared to small aneurysm (S2 TP-3 vs S1 TP-S1)
3. **Patient systemic response to AAA repair surgery:**
    - Healthy volunteers compared to large aneurysm patients' post-surgical measurement (S2 TP-5 vs HV)
    - Large aneurysm patients' post-surgical measurement compared to small aneurysm (S2 TP-5 vs S1 TP-S1)
    - Large aneurysm patients' post-surgical measurement compared to large aneurysm patients' pre-surgical measurement (S2 TP-5 vs S2 TP-3)

Classification of patients into treatment groups was conducted using ML, similar to that conducted previously for the STEMI patients. The number of patients available for each comparison are listed in Table 5.1 Patient groups were unmatched for patient identity due to large amounts of missing data (samples were not available for patients across multiple treatment groups).

Table 5.1 Classification of AAA patient treatment groups.

Treatment group 1	Treatment group 2	n patients per group
S1 TP-S1	HV	20
S2 TP3	HV	20
S2 TP3	S1 TP-S1	89
S2 TP5	HV	20
S2 TP5	S1 TP-S1	44
S2 TP5	S2 TP3	44

The five ML methods described in Chapter 3 were employed to classify the mass spectral data according to these patient groups. Optimisation of model hyperparameters was performed as discussed previously in Chapter 4. The available mass spectra were undersampled where necessary in order to maintain class balance; the number of patients per class is shown in Table 5.1. Note that only a small number of samples were available for the HV group, and so class sizes are small where this class is featured. We recognise that this may lead to some overfitting. Cross-validation partitioning with 200 partitions was used to prevent overfitting with a holdout of 20%, as applied in previous analysis.

5.1.2.4 Data analysis: Feature-reduction and identification of significant mass peaks

Feature-reduction (FR) was conducted through the identification of MS peaks that correlated significantly with changes in the clinical data. This was conducted using the three FR methods described in Section 3.2.4, and used previously for STEMI data (see Chapter 4), namely statistical tests, analysis of peak overlap integrals, and ML feature ranking.

5.1.3 Results and Discussion

5.1.3.1 Spectral differences between patient groups

The spectra collected for the five different patient groups with the largest patient numbers were averaged to generate a spectrum representative of each patient group (Figure 5.2). Visual differences could be observed between the averaged spectra for each group. In particular, a group of peaks between m/z 250-310 is observed to be present at higher intensities in the AAA patient groups (Figures 5.2(b) and 5.2(c)) than in the healthy group. This group of peaks is also seen to be diminished in the post-surgery group (S2 TP-5, Figure 5.2(d)). The post-surgery spectrum is visually the most different compared to the other patient groups, which is likely due to the presence of metabolites associated with inflammation and healing after the trauma of AAA repair surgery.

While the spectra for small and large aneurysm patients share many features, there are also differences, as shown by their difference spectrum. This is calculated by subtracting one normalised spectrum from a second normalised spectrum, and the result filtered to show only the peaks with a significant difference ($P<0.001$) in mean intensity between patient groups according to an independent t-test. These difference spectra are shown in Figure 5.3. The 'small AAA'-HV difference spectrum and the 'large AAA'-HV spectrum are visually very similar (Figures 5.3(a) and 5.3(b)), suggesting that there are metabolites present due to AAA even when the AAA is small. Differences between the small and large aneurysm groups can be seen in the 250 - 310 m/z region, showing that the metabolite profile is affected by aneurysm size.

5.1.3.2 Machine learning classification results overview

The classification of patients into the groups defined previously was highly successful, with kappa scores of over 0.74, equivalent to 87% accuracy, achieved for all cases,

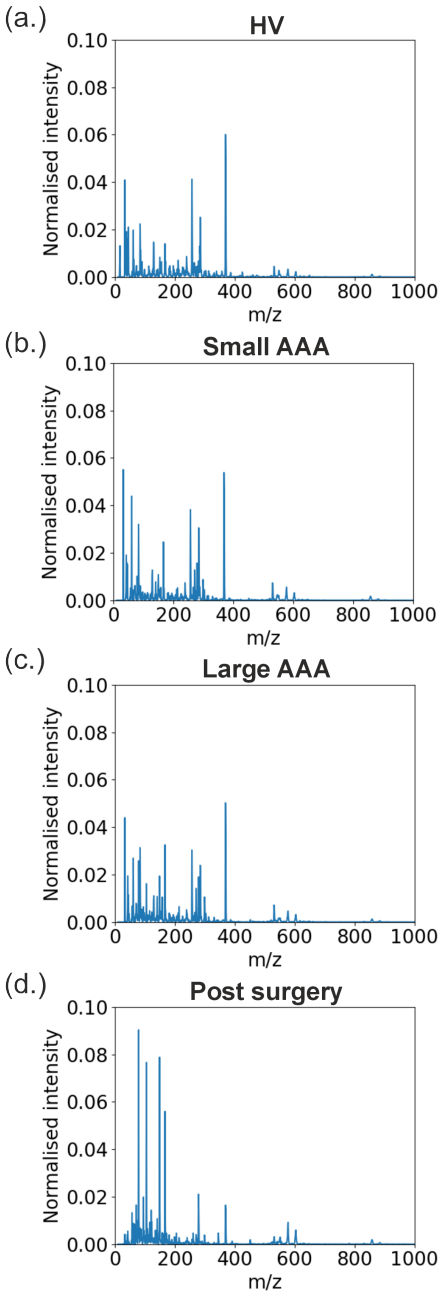

Figure 5.1

Figure 5.2 The averaged spectra for patient groups: (a) healthy volunteer, (b) small AAA, (c) large AAA, and (d) post-surgery. Spectra are displayed as normalised intensity plotted against *m/z*. Averaged spectra are calculated as the average of all patient spectra in each class.

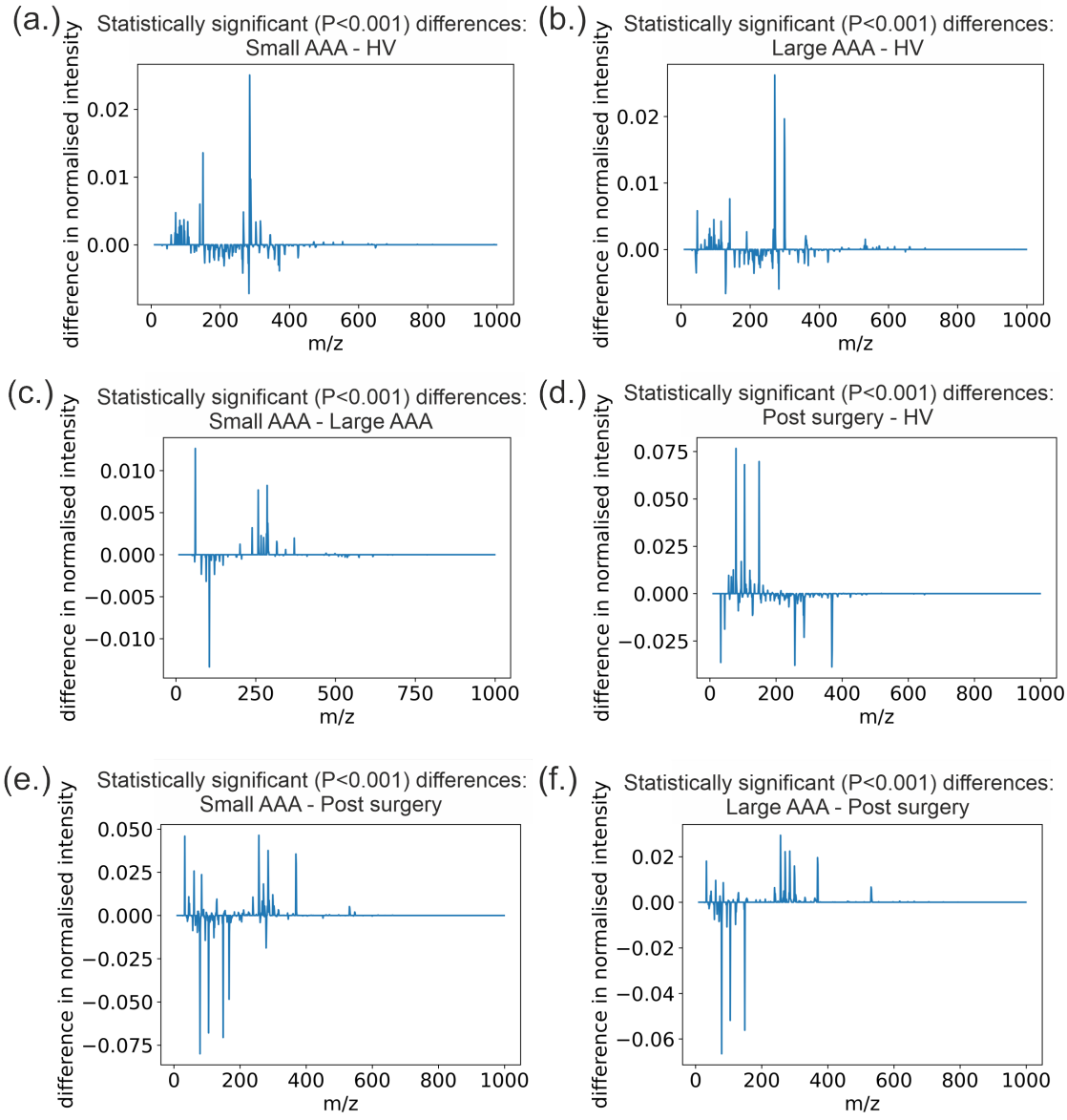

Figure 5.3 Difference spectra for patient classes showing only the peaks that show a significant ($p < 0.001$) difference between patient classes by an independent t-test. Spectra shown are S1-TPS1 - HV for small AAA - healthy volunteer patients (a), S2-TP3 - HV for large AAA - healthy volunteer patients (b), S2-TP3 - S1-TPS1 for large AAA - small AAA patients (c), S2-TP5 - HV for post-surgery - healthy volunteer patients (d), S1-TPS1 - S2-TP5 for small AAA patients - post-surgery (e), and S2-TP3 - S2-TP5 for large AAA - post-surgery patients (f).

Table 5.2 The best results obtained for each patient classification using range of machine learning (ML) models and feature reduction (FR) methods.

Clinical variable	**Highest κ score**	**κ Error**	**Equiv. accuracy /%**	**ML method**	**FR method**
Small AAA vs HV	0.964	0.151	98.2	RFC	FR
Large AAA vs HV	0.990	0.070	99.5	LDA	OI
Large AAA vs small AAA	0.865	0.090	93.2	SVM	None
Post-surgery vs HV	0.923	0.136	96.1	KNN	FR
Post-surgery vs small AAA	0.961	0.060	98.0	LDA	None
Post-surgery vs large AAA	0.740	0.163	87.0	RFC	None

and some instances of κ scores of well over 0.9 achieved. The best results for each classification are shown in Table 5.2, with the full results presented in Figure 5.4.

In most cases, the five ML models performed similarly to each other across the different classifications, with the exception of the NBC model performing significantly less well than the others in the 'large AAA vs small AAA' classification. The NBC model assumes variable (in this case m/z peak) independence, which is likely to be a poor assumption given the complexity of interlinked metabolic processes.

The use of feature selection methods did not largely effect the accuracy of the results, showing that only a small number of the mass peaks in the data are required for successful classification. It was found that no particular combination of ML model and feature selection method achieved consistently higher results in each case, in common with previous work of plasma samples from STEMI patients.

The uncertainties in the κ scores are low, implying that relationships found by ML models were robust and generalised across all patients, and were not the result of particular patient selections within the cross-validation partitioning or overfitting. The full results tables for this analysis are given in the Appendix C.1.

All classifications separating AAA patients from healthy volunteers were accurate to over 90%, supporting the use of AIMS technology to detect changes in circulating metabolites in venous blood plasma as a result of AAA. However, the number of healthy volunteer samples available used in this study was very low, so over-fitting and patient-selection biases may be enhancing these results. Nevertheless, given the

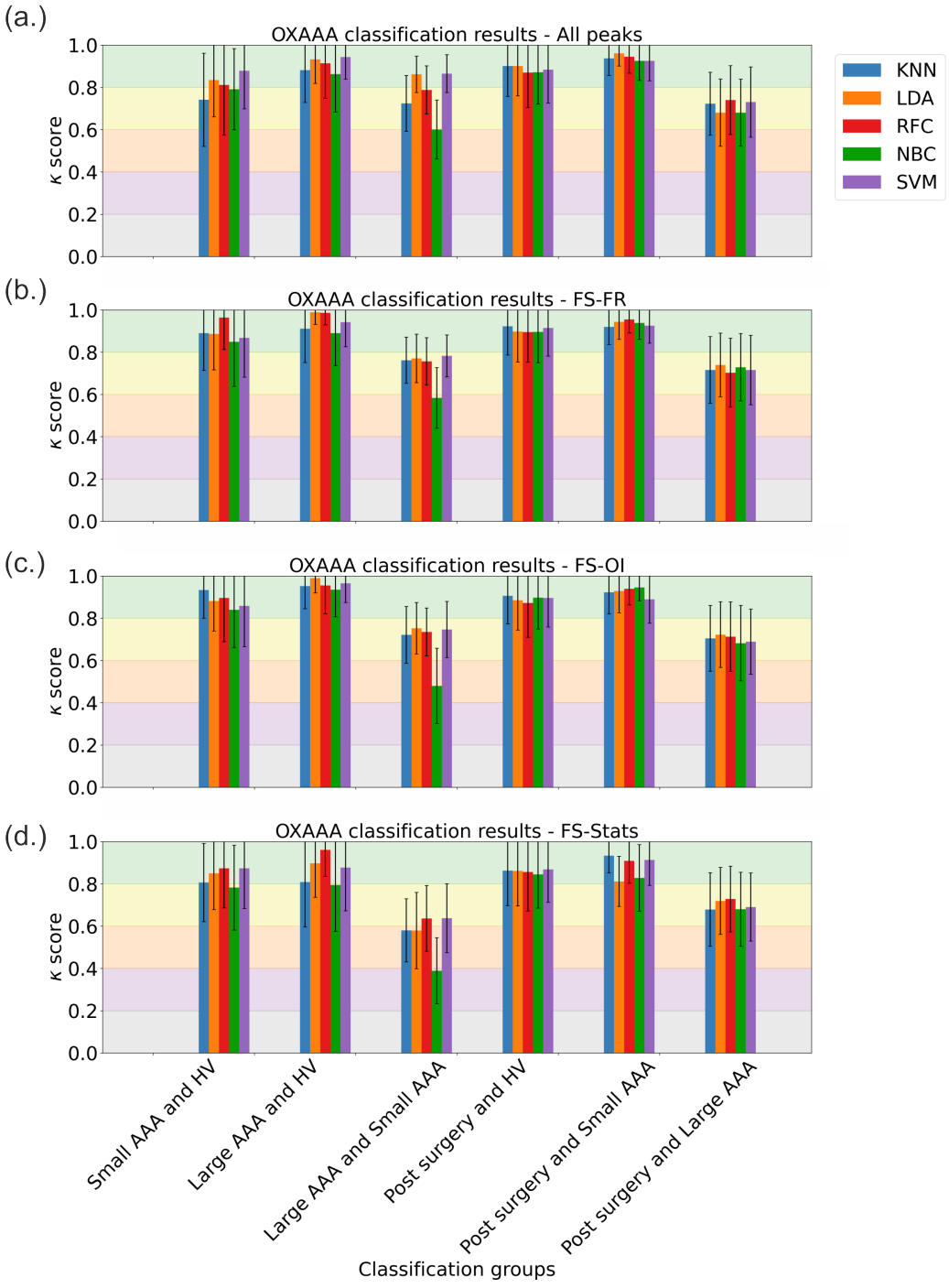

Figure 5.4 The κ obtained from the ML classification analysis of six different OxAAA group comparisons. Verrtical bars show the κ obtained for each ML model for that clinical variable, horizontal coloured bars show the level of agreement with *no agreement* in grey, *low agreement* in purple, *fair agreement* in light red, *moderate agreement* in yellow, and *strong agreement* in green. Results are shown for when all peaks were used (a), the FR FS method (b), the OI FS method (c), and stats FS method (d).

differences seen in the average spectra for the various AAA patient groups, it is expected that ML models would be able to separate these patient classes, so long as the in-class variation was low. Due to sample availability issues, the measurements on healthy volunteer samples were unfortunately conducted in a separate measurement batch from the other measurements, which were conducted in mixed class batches. Therefore we cannot rule out the presence of batch effects in the comparisons involving these samples. The study should be repeated with a larger HV data set and making measurements on mixed AAA patient HV batches.

The successful classification of small AAA patients from large AAA patients confirms that metabolic changes occur as the AAA size increases to > 5 cm, and these changes can be detected in venous blood plasma by rapid ASAP-MS measurement. The lowest accuracy obtained, and with the highest error, was for the classification of post-surgery vs large AAA patients. This may suggest that metabolic changes to the patient are still present after the AAA has been repaired, either due to lasting effects of the initial pathology, continued production of metabolically active molecules by abdominal aortic tissue, or continued presence of a systemic disease related to the AAA. As the AAA tissue is still present after surgical repair, the tissue may still be releasing metabolically active compounds into the circulation.

5.1.3.3 Analysis of peaks found to be key to successful patient classifications

When the peaks used in each FS analysis were analysed, a number of peaks were found to be highly important to five of the six classifications, and were identified in over 75% of the cross-validation partitions. The peaks found are shown in Table 5.3, and the distributions between patient groups are shown in box plots in Figure 5.5. The statistical difference between these distributions was calculated using an independent t-test (the t-statistic and p value for each peak is shown in the Appendix C.2).

In many cases, the same peaks were found to be important in all of the classifications. In several cases, consecutive peaks were found, which are likely to result from related molecules or fragmentations, isotopologues, or broad peaks being attributed to several different masses. The main trends identified between analyses are summarised as follows:

- **AAA vs healthy volunteers:** Three of the peaks found to be important in the 'large AAA vs HV' classification were also found to be important for the classification of 'small AAA vs HV', indicating these peaks are related to AAA pathology. The differences observed in the 'small AAA vs HV' (shown in Figure 5.5(a-b)) appear to be larger in the 'large AAA vs HV' comparison, which would correlate with the AAA development.

- **Post surgery vs AAA patients:** A large number of peaks were found to be important for the classification of post-surgery patients and AAA patients, and six out of eight of those were found both for classification involving both small and large AAA patients. This confirms that the post-surgery patient metabolic profile is quite different form pre-surgery AAA patients, as observed in the difference spectra. All the peaks identified showed a significant decrease in post-surgery patients compared to AAA patients (shown in Figures 5.5(d) and 5.5(e)), suggesting the peaks identified are biomarkers elevated by the presence of the AAA.

- **Post surgery vs healthy:** Two of the peaks found to be important for the 'post-surgery vs HV' classification were also found in the AAA patient to HV classifications. This is consistent with the previous suggestion that some systemic changes corresponding to AAA pathology are still present post-surgery.

- **Small AAA vs large AAA:** No peaks were found to be consistently used by FS methods across 75% of the partitions for the small vs large AAA patient

comparison. However, as the classification was successful, this suggests that patterns used for classification are not limited to a small number of peaks. Two peaks were found to be consistently used in over 50% of the FS models, shown in 5.5(c). There is more overlap between intensity distributions of these peaks compared to the analyses shown in the figure, but the confidence level is still high ($p < 0.001$). An explanation for this may be due to AAA size being a continuous variable being forced into a binary classification by the surgical threshold, but the variation in metabolite levels with AAA size will be continuous, patient specific, and subject to other sources of variation. The fact that this has been successfully classified is an exciting sign that metabolic markers linked with AAA size could be detected in venous plasma.

Table 5.3 The m/z ion peaks identified as being important for classification by the three FS methods for each pair of compared patient groups.

Clinical variable	**ID of m/z ion peaks found with $>$ 75% occurrence across 3 FS methods**
Small AAA vs HV	172, 206, 338, 353, 355, 357, 358, 425, 426, 444
Large AAA vs HV	281, 357, 358, 425, 459
Large AAA vs small AAA	None*
Post-surgery vs HV	227, 284, 357, 426,
Post-surgery vs small AAA	45, 239, 257, 258, 267, 284, 285, 286
Post-surgery vs large AAA	45, 256, 257, 258, 267, 268, 284, 285

Utilising the HMDB, and literature sources citing metabolomic studies relating to AAAs, enabled the tentative identification of three of the significant peaks identified in the above analysis at m/z 316, 426, and 357. These have molecular weights that may correspond to ions of the acyl carnitines decanoylcarnitine and oleoylcarnitine, and a derivative of the bile acid sterol cholanoic acid, respectively (see Table 5.4). These metabolites have been linked to the progression of AAAs in a study by Ciborowski *et al.*, in which researchers found that the concentration of these biomarkers in blood plasma was lowered in AAA patients compared to healthy volunteers [324]. This same trend was observed in our dataset, as observed in box-

plots 5.5(a) and 5.5(f), as well as a potentially related peak at m/z 425 shown in 5.5(b) for large AAA vs HV. Acylcarnitines are linked to altered fatty acid metabolism, and have been shown to be linked to AAA development in a number of studies [84, 327, 328], as well as more generally to other cardiovascular diseases [329]. The role of bile acids in AAAs is unclear, but studies have also found links between bile acids and cardiovascular disease in the 'gut-heart axis', which is also in agreement with the presence of systemic disease [330]. It is encouraging that the ASAP-MS results, in a small and fairly rapid analysis, can corroborate the findings of larger studies utilising high sensitivity and resolution instrumentation.

Table 5.4 Trends in tentatively assigned metabolite peaks compared to findings by M. Ciborowski *et al.* [324] for small AAA (S. AAA) and large AAA (L. AAA) compared to each other and to healthy volunteers.

	Literature data [%](p-value)			**ASAP-MS data [%](p-value)**		
Biomolecule	**S. AAA vs HV**	**L. AAA VS HV**	**S. AAA to L. AAA**	**S. AAA vs HV**	**L. AAA VS HV**	**S. AAA to L. AAA**
Decanoylcarnitine $m/z = 316$	-36% (0.05)	-60% (< 0.001)	-37% (0.1)	109% (0.078)	-64.9% (0.028)	-68.8% (< 0.001)
Cholanoic acid $m/z = 426$	+235% (0.2)	-76% (0.03)	-94% (< 0.001)	-96.9% (< 0.001)	-96.1% (0.001)	-86.8% (0.012)
Oleoylcarnitine $m/z = 357$	-48% (0.009)	-60% (< 0.001)	-22% (0.2)	-91.4% (< 0.001)	-82.5% (< 0.001)	-9.4% (0.012)

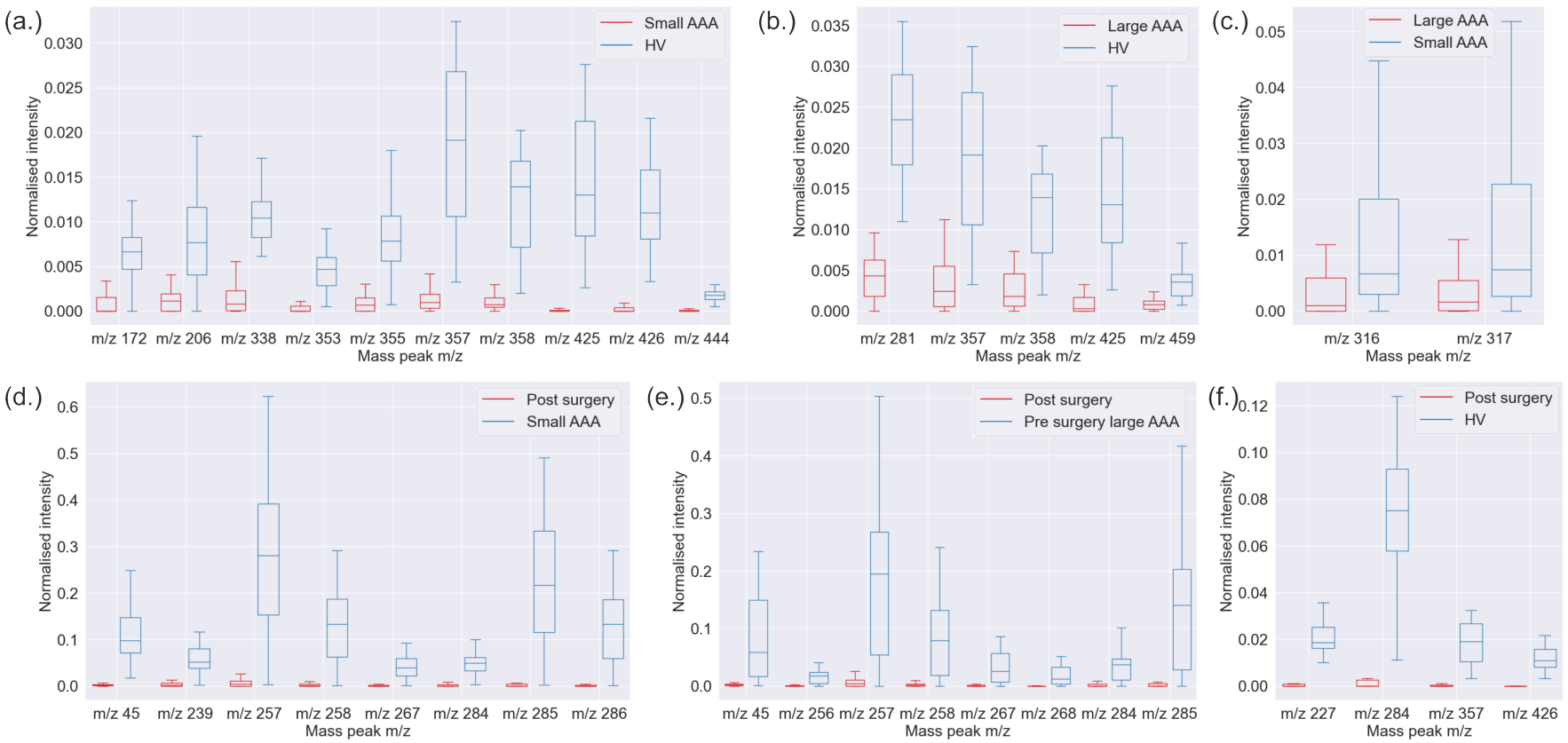

Figure 5.5 Boxplots showing the distribution of normalised peak intensities for peaks found to be important in over 75% of models for the classification of each of the classifications of AAA patient groups: small AAA and HV (a), large AAA and HV (b), post surgery and small AAA (d), post surgery and large AAA (e), and post surgery and HV (f). The peaks in (c) for large AAA vs small AAA were identified with a lower threshold, in over 50% of models. The central line represents the data median, with the inner box representing the inter-quartile range between the first and third quartiles. The whiskers extend from the box to the farthest data point lying within 1.5x the inter-quartile range from the box. All peaks were statistically different at $p < 0.001$.

5.2 The classification of human tissue samples from AAA patients

5.2.1 Introduction

In the first part of this chapter, it was demonstrated that blood plasma metabolites correlate with the presence and size of an AAA. The metabolic fingerprint of tissues in close proximity to the AAA may also hold information relevant to understanding AAA development and pathology. A number of tissue biopsy samples were available from patients enrolled in the OxAAA study who underwent AAA repair surgery, complementing the plasma samples discussed previously. As there were only a subset of patients for whom the AAA reached the surgical threshold of 5 cm, the number of tissue samples available was much lower than plasma samples. Metabolic fingerprints vary significantly between different tissue types, but changes related to the presence and progression of an AAA are likely to be more subtle, and may be difficult to distinguish from natural human variation, instrument error, and noise. The aims of this study were to understand the ability of ASAP-MS to determine the metabolic fingerprint of different tissue types (for example, fat and thrombus) and subtypes (for example omental fat and preperitoneal fat) and to see whether any additional differences could be related to the AAA or measured clinical information.

5.2.2 Methods

5.2.2.1 Study patients and criteria

This analysis used OxAAA tissue samples, as discussed in Section 2.3, which consisted of three sample types: fat, artery wall, and thrombus. The samples within these groups were then further categorised into eight tissue subtypes: omental, preperitoneal, and subcutaneous fat; distal, maximal, and proximal artery wall; and luminal and abluminal thrombus. The number of samples available for each

tissue type and subtype are listed in Table 5.5.

Table 5.5 The type and distribution of tissue samples used in this study.

Tissue type	Tissue subtype	n samples measured	Total samples by type
Fat	Pre-peritoneal	26	
Fat	Omental	32	85
Fat	Subcutaneous	27	
Thrombus	Abluminal	26	52
Thrombus	Luminal	26	
Wall	Maximal	30	
Wall	Distal	24	83
Wall	Proximal	29	

5.2.2.2 Materials and Instrumentation

Mass spectrometry analysis of tissue biopsy samples was conducted using the optimised tissue analysis method described in Section 2.5.5. Data analysis was conducted using custom software written in MATLAB R2022a and Python 3.7. Machine learning analysis was conducted on CPU nodes within the University of Oxford ARC. The practical work in this section was conducted with the help of Lloyd Smith, a Part II student in the group in 2023-2024. The data extraction, background subtraction and normalisation was conducted using the same methods as reported previously for plasma samples.

5.2.2.3 Data analysis: Machine learning classification of AAA patient groups

Classification of tissue samples was conducted using ML algorithms applied to the processed mass spectra, in a process similar to that described in the previous sections. Tissue mass spectra were classified into one of three overall tissue types: fat, thrombus, and aneurysm wall. Classification was conducted between binary pairs of tissue types, which produced three tissue type classifications: fat vs thrombus, fat vs wall, and wall vs thrombus. The tissue samples were then further classified by subtype, by comparing pairs of subtypes within the same overall tissue type.

This produced seven binary classifications: preperitoneal vs omental, preperitoneal vs subcutaneous, and omental vs subcutaneous for fat tissues; abluminal vs luminal for thrombus tissues; and maximal vs distal, maximal vs proximal, and distal vs proximal for aneurysm wall samples.

The five ML methods described in Chapter 3, and used previously for plasma, were employed for the classification of tissue samples. Optimisation of model hyperparameters was performed as conducted in previous analysis. The data were undersampled to maintain class balance, the number of patients per class is shown in Table 5.5. We recognise that this may lead to some overfitting due to the limited number of samples available and reduce the reliability of the results, particularly for the tissue subtype analysis, for which the number of samples per class are fewer than 30. However, in a proof-of-concept application, the ability to predict tissue type and subtype with some accuracy may suggest that with higher sample numbers, these predictions could be improved.

Cross-validation partitioning over 200 unique partitions was applied to reduce overfitting. Multiple tissue samples were taken from the same patients, with a minimum of one tissue subtype and a maximum of eight tissue subtypes available per patient. To prevent contamination of the test data with patient specific information, the cross-validation protocol was set-up to ensure that all samples relating to an individual patient would be used in the same cross-validation group, either all in training or all in testing.

Due to the different class sizes and the variation in sample numbers available for each patient, the training:testing ratio could not be maintained exactly throughout each analysis and each cross-validation. Instead of the commonly used 80:20 K-fold cross-validation method for training and test data, a leave-n-out cross-validation method was used, where a set number of unique patients were used for the test dataset in every partition ($n = 4$, with unique 2 patients per class). If samples were measured for that patient in both classes, all the samples available for that patient

will be held in the test data. An example of how this translated to class sizes is shown in Table 5.6. Using this LNO-CV method combines the reduced overfitting of K-fold cross-validation, and the suitability for small datasets of the leave-one-out cross-validation method. [a]

Table 5.6 Distribution of sample numbers within a typical partition.

Classification	n **spectra per class**	n **unique patients in classification**	n **spectra: training (averaged)**	n **spectra: testing (averaged)**
Fat vs thrombus	52	37	93	19
Fat vs wall	83	38	154	16
Wall vs thrombus	52	36	90	16
Subcutaneous vs omental	27	35	47	9
Subcutaneous vs preperitoneal	26	31	46	8
Preperitoneal vs omental	26	35	47	7
Abluminal vs luminal	26	32	46	8
Maximal vs proximal	29	35	53	7
Maximal vs distal	24	35	43	7
Proximal vs distal	24	34	41	9

Across this project, the impact of training data size was frequently questioned. Using this OxAAA tissue data, an investigation into the impact of training data size was undertaken. This study investigated the use of repeat sample measurements as a method of increasing training data availability, as well as changing the number of patient samples available. The results of this study are described in Appendix Sections C.4 and C.5. Overall, this investigation showed that it is the number of patients included in the data, and not the repeats used, that impact the accuracy of the results, and that the number of patients used in this analysis was sufficient for the investigation of patient classification with ML. This is in agreement with

[a] With a hold out of 2 patients per class used, equivalent to a leave-4-out cross-validation, the number of combinations of patients that could be chosen for the training smallest class available, which in this case is the distal wall class with 24 unique patients, the number of unique combinations of patient pairs than can be chosen is 276 ($_{35}C_4$). This is greater than the number of partitions chosen for use in the ML analysis, making 200 partitions appropriately sized for this analysis.

a common ML observation that a small, good-quality dataset, that spans a large breadth of the data area, can be sufficient for ML models. The impact of test data size was also investigated, and is shown in Appendix C.5.1.

5.2.2.4 Data analysis: Feature reduction and identification of significant mass peaks

The ML models were applied to all the m/z features in the dataset, as well as to selected features found to correlate with the tissue type using FR algorithms. Six different FR algorithms were used for this analysis, the three FR algorithms described previously (statistical feature reduction, OI, and χ^2), as well as three further methods: MRMR ML feature selection, PCA, and RelF ML feature selection. These methods were described in Section 3.2.4.

FR was applied to the training data within each data partition prior to ML, as described in previous chapters. For use of PCA as an FR method, the first 30 principal components were used as inputs to the machine learning algorithms. For the remaining four FR methods, the best 30 m/z variables determined by the relevant metric were used.

The recorded mass spectra variables were determined to be non-parametric in over 95% of variables using a Shapiro Wilkes test, and so non-parametric tests were used for statistical analysis. For the comparison of differences between grouped data, Mann-Whitney U tests were used. For the correlation of variables with the class variable, Spearman's rank correlation coefficient tests were used. In all tests, significance was determined when $p < 0.001$, unless stated otherwise.

5.2.3 Results and Discussion

5.2.3.1 Spectral differences between tissue types

The spectra collected for the different tissue types were averaged to generate representative average spectra. Visual differences were found between the resulting aver-

aged spectra for three main tissue types, shown in Figure 5.6. Differences between the fat spectrum and the other tissue types are visible in the intense peaks observed from m/z 500 to 1000. The peaks from 500 to 650 mostly arise from phospholipid diglycerides, particularly phosphatidylcholines, and the peaks from 800 to 1000 are mostly phospholipid triglycerides. The most intense peak in the aneurysm wall and thrombus spectra is cholesterol, observed at m/z 369 ([chol-H_2O]+H^+), however this was not the case in the fat spectrum in which the most intense peaks are observed in the phospholipid region.

The spectral intensity is normalised to unit AUC, and so the intensity shown will be relative to the largest observed peaks. The heatmaps show how the un-normalised spectral intensities vary over repeat measurements and across samples, and show the patterns observed in the averaged spectra to be well generalised. The heatmaps enable comparison of measured absolute metabolite ion signals, rather than values that may be suppressed by the normalisation process, particularly when intense peaks are present. Some variation is present between individual measurements due to variation in the amount of sample introduced to the instrument. The fat samples yield only low levels of signal from small molecule metabolites in the low-mass region of m/z 0 to 400, shown by the comparative lack of red and orange colours in the left of the heatmap. The intensity of the small molecule metabolite region is seen to increase from the fat to the wall samples, and is highest in the thrombus samples. The opposite trend is observed for high-mass lipid peaks.

5.2.3.2 Statistical analysis of tissue type spectra differences

Based on statistical analysis using a Mann-Whitney U test, a large number of m/z variables were determined to be significantly different between the tissue types. A summary of these results is shown in Table 5.7. As the statistical tests used were nonparametric, the statistical power of these results is lower than if parametric assumptions could be made (described previously in Chapter 3) [331]. The number

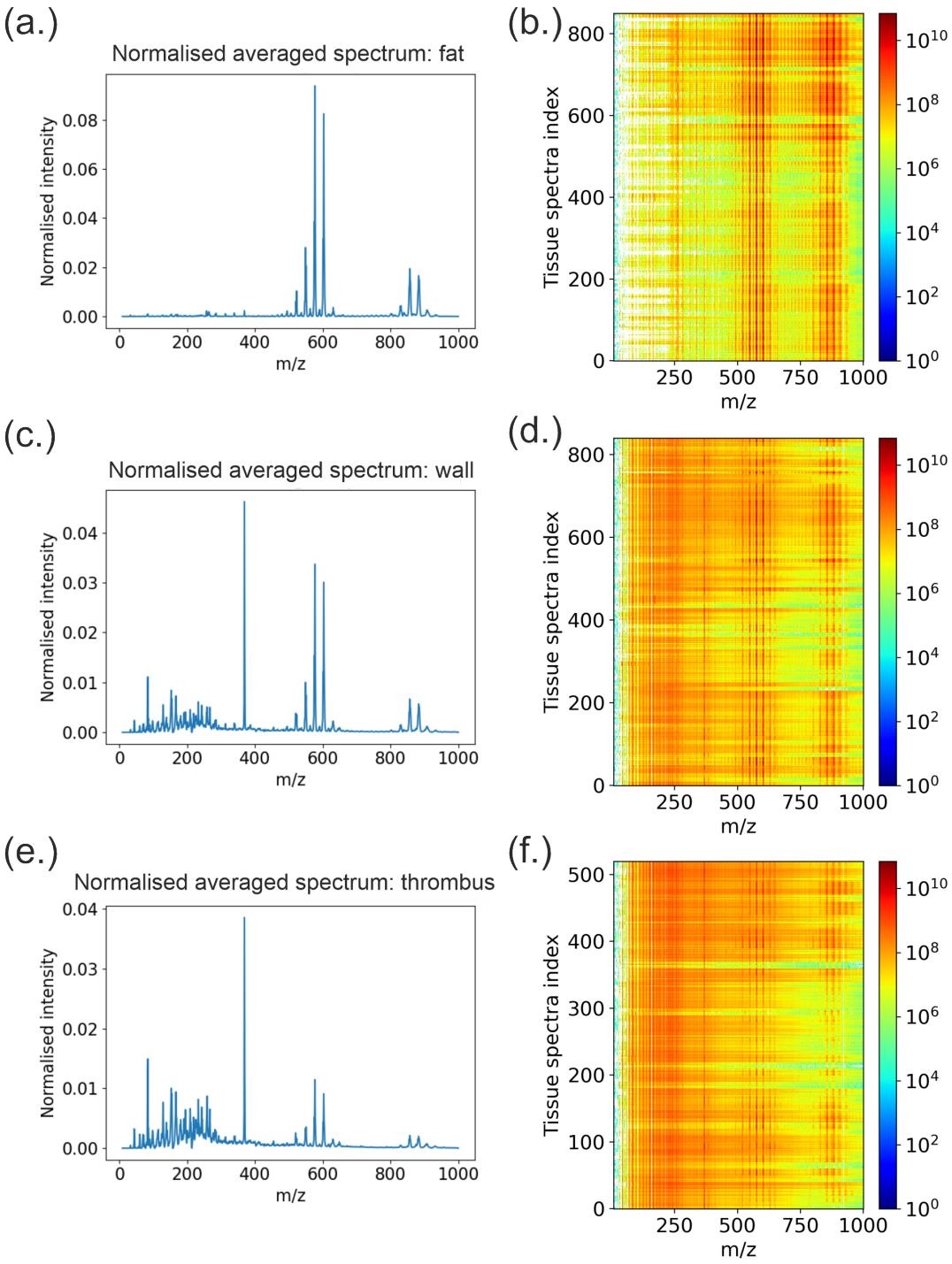

Figure 5.6 Comparison of the spectra for the averaged tissue types, and a heatmap representation of each repeat for each tissue type. The mass spectra on the left show the normalised intensity averaged over each repeat and each sample for each tissue type: fat (a), wall (c), and thrombus (e). The heatmaps shown the normalised intensity for each repeat measurement of each sample, plotted as *m/z* value (x), sample and repeat number (y), and measured intensity (shown by colour bar). The samples are plotted by type, showing fat (b), wall (d), and thrombus (f).

of significant mass peaks is therefore likely inflated. However, it is clear from these results that the differences between tissue types can be observed in a large number of metabolite peaks. In fact, over 80% of the peak intensities in each spectrum were significantly different in each comparison made.

The PCA results also showed that fat tissue clustered separately from other tissue types using just the first two or three principal components, with the wall vs thrombus comparison having much less distinctive clusters. Figure 5.7 shows that in both comparisons involving the fat samples, the clusters were well defined along the first principle component axis, which accounted for over 80% of variance in the data for both instances, as shown by the explained variance plots in Figure 5.7(b) and (d). The wall and thrombus clusters were less clearly defined, and the variance in the data was distributed more over PC2 and PC3 than for the previous comparisons. The AAA intraluminal thrombus structure is similar to that of the artery wall, as during the progression of AAA the thrombus thickens and takes on characteristics of the arterial wall, whilst the wall itself begins to weaken and lose structural integrity. Delineation of these two structures can be highly challenging in a clinical setting [332]. These two connected structures are expected to have similar metabolic profiles due to their proximity and function, and so the reduction of clustering by PCA is consistent with the physiology.

Table 5.7 Variables found to be significantly different ($p < 0.001$) between spectra grouped by tissue type.

Tissue type classification	**Number of m/z variables identified as significantly ($p < 0.001$) different between classes**	**Percentage of all variables**
Fat vs thrombus	920	92.7
Fat vs wall	917	92.4
Wall vs thrombus	811	81.8

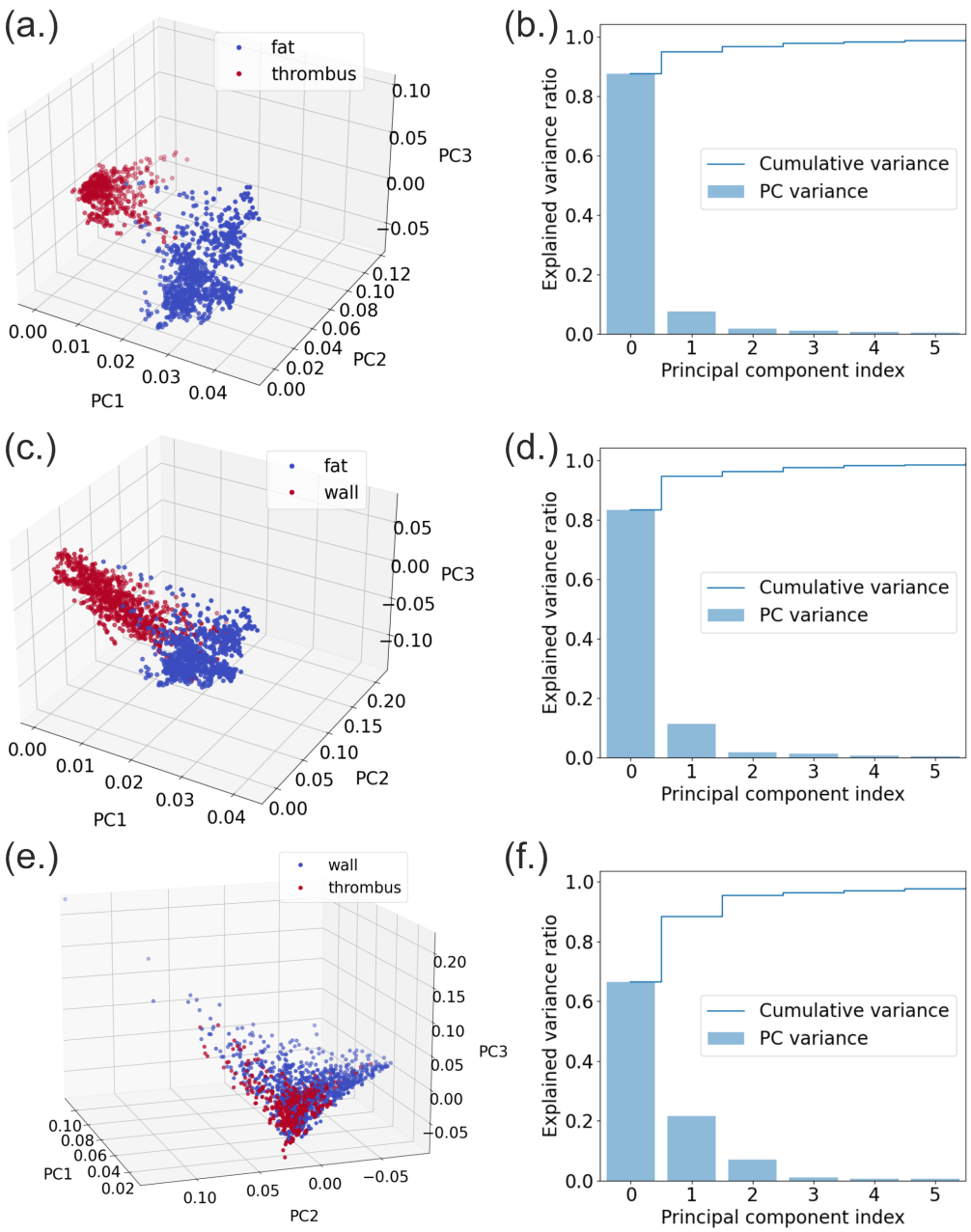

Figure 5.7 PCA analysis of tissue sample spectra. PCA cluster plots are shown on the left, with the first three principal components plotted, and the data coloured by class. The figures on the right show the distribution of data variance with principal components. By row, the analysis shown is for fat vs thrombus (a and b), fat vs wall data (c and d), and wall vs thrombus data (e and f). PCA clusters and variance for tissue type analysis.

5.2.3.3 Machine learning analysis of tissue type spectra differences

The ML clustering algorithms were able to successfully classify the tissue samples by tissue type with high levels of accuracy. The full results are shown in Figure 5.8. The majority of κ scores were above 0.6, which is equivalent to over 80% accuracy and 'moderate to strong' agreement between the predicted and true class. As shown by the previous PCA clustering analysis, the ML classification results show that fat is easily distinguished from both wall and thrombus with very high accuracy, with many results of $\kappa > 0.9$, indicating almost perfect accuracy at $> 95\%$. The classification of wall from thrombus was less successful than that of comparisons to fat tissue, with lower kappa scores and high error; but κ scores were still commonly within the 'moderate' or 'strong' agreement brackets.

The low error in classification across all analyses indicates a minimal reliance on the type of FR or ML methods employed. The only exception to this was for the use of PCA as an FR method, which gave lower accuracies than all other methods, and had much higher errors for all the tissue pairs. One explanation for this is that the way in which PCA translates the data away from the peak-based variables to a variance-based variable set is not optimal for this type of data, so the mass-dependent variance related to tissue type may be obscured. The PCA results were removed from the dataset for calculating the mean and range of accuracy across the methods due to the large difference in these results.

The low standard deviation in κ scores implies that differences in peak intensities correlating with tissue type are strongly represented in the data, so that each ML method is able to learn a pattern within the dataset even with the relatively small number of samples used for training. No FR method outperformed the others, with the best combinations of ML model and FR method (abbreviated to ML model-FR method) were: KNN/SVM-MRMR for fat:thrombus at 0.992 ± 0.04 κ, RFC-OI for fat:wall at 0.997 ± 0.04 κ, and LDA-noFR for wall:thrombus at 0.8 ± 0.19 κ.

FR method	ML model	Fat : Thrombus κ	Fat : Wall κ	Wall : Thrombus κ
All	KNN	0.97 ± 0.08	0.93 ± 0.08	0.55 ± 0.25
	SVM	0.97 ± 0.08	0.96 ± 0.07	0.73 ± 0.21
	LDA	0.97 ± 0.08	0.92 ± 0.08	0.8 ± 0.19
	NBC	0.96 ± 0.11	0.94 ± 0.1	0.53 ± 0.22
	RFC	0.98 ± 0.08	0.97 ± 0.06	0.7 ± 0.2
Stats	KNN	0.96 ± 0.09	0.99 ± 0.04	0.7 ± 0.21
	SVM	0.96 ± 0.1	0.99 ± 0.03	0.76 ± 0.2
	LDA	0.97 ± 0.08	0.94 ± 0.08	0.71 ± 0.21
	NBC	0.96 ± 0.09	0.98 ± 0.04	0.69 ± 0.2
	RFC	0.99 ± 0.05	0.97 ± 0.06	0.72 ± 0.2
OI	KNN	0.97 ± 0.08	0.98 ± 0.05	0.67 ± 0.23
	SVM	0.98 ± 0.06	0.99 ± 0.04	0.74 ± 0.2
	LDA	0.95 ± 0.11	0.92 ± 0.08	0.74 ± 0.19
	NBC	0.96 ± 0.09	0.96 ± 0.07	0.7 ± 0.2
	RFC	0.96 ± 0.08	1.0 ± 0.02	0.69 ± 0.21
Chi^2	KNN	0.97 ± 0.09	0.97 ± 0.07	0.66 ± 0.22
	SVM	0.97 ± 0.08	0.98 ± 0.05	0.73 ± 0.21
	LDA	0.96 ± 0.1	0.93 ± 0.08	0.7 ± 0.22
	NBC	0.96 ± 0.09	0.95 ± 0.08	0.7 ± 0.21
	RFC	0.98 ± 0.09	0.97 ± 0.06	0.73 ± 0.21
MRMR	KNN	0.99 ± 0.04	0.97 ± 0.06	0.66 ± 0.22
	SVM	0.99 ± 0.04	0.97 ± 0.06	0.7 ± 0.23
	LDA	0.97 ± 0.07	0.95 ± 0.07	0.66 ± 0.22
	NBC	0.99 ± 0.05	0.96 ± 0.07	0.65 ± 0.22
	RFC	0.98 ± 0.06	0.98 ± 0.08	0.69 ± 0.2
ReIF	KNN	0.98 ± 0.06	0.94 ± 0.08	0.69 ± 0.22
	SVM	0.97 ± 0.07	0.95 ± 0.09	0.78 ± 0.2
	LDA	0.95 ± 0.15	0.92 ± 0.08	0.78 ± 0.18
	NBC	0.97 ± 0.08	0.95 ± 0.09	0.6 ± 0.22
	RFC	0.97 ± 0.08	0.93 ± 0.09	0.73 ± 0.18
PCA	KNN	0.59 ± 0.41	0.46 ± 0.39	0.31 ± 0.29
	SVM	0.77 ± 0.21	0.33 ± 0.18	0.14 ± 0.24
	LDA	0.96 ± 0.09	0.39 ± 0.18	0.02 ± 0.07
	NBC	0.94 ± 0.14	0.18 ± 0.19	0.0 ± 0.05
	RFC	0.4 ± 0.44	0.27 ± 0.39	0.05 ± 0.13

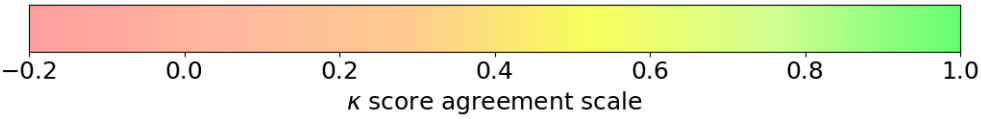

Figure 5.8 The κ scores achieved for each analysis of the tissue type classification by combinations of five ML models and seven FR methods applied to the mass spectra, averaged for each sample. The errors shown are the standard deviation in κ score achieved across each partition. The colourmap indicates the accuracy and level of agreement of the results by κ score.

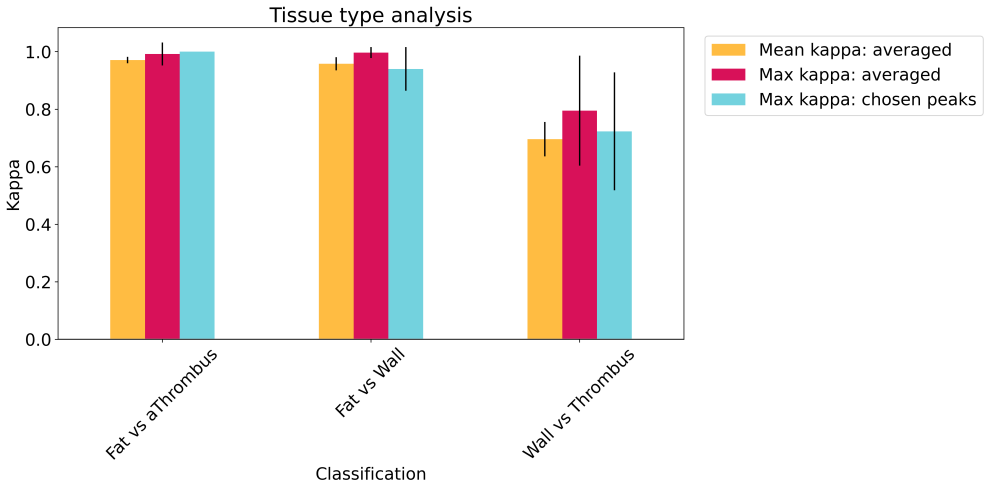

Figure 5.9 The κ scores achieved in ML classification of tissue by tissue type. Bars show the κ score from the average ML algorithm results, error bars represent standard deviations. Results are plotted as the mean κ score achieved across all ML algorithm results (yellow), or the maximum κ score achieved with a single ML algorithm and FR method (red). The analysis was repeated using peaks found to be significant by the FR analysis, and the maximum κ score determined for the data shown (blue).

5.2.3.4 AAA tissue subtype classification

The classification of tissue subtypes was more challenging, and very little success was achieved using ML algorithms. In addition, the low sample numbers significantly reduced the reliability of these results, and strong conclusions cannot be made using them. The results of tissue subtype classification are found in Appendix C.3.

5.2.3.5 Analysis of peaks found through feature reduction for tissue type classification

A number of peaks were found to be important for classifying mass spectra according to the different tissue types by analysis of the FR methods. When only the peaks identified in Table 5.8 were used as the features for each classification respectively, the accuracies achieved were high. The results for this analysis are shown in Table 5.9, and the top κ score achieved in each case shown in the blue bars in Figure 5.9.

The peaks found were all in the low to medium mass molecule range at less than 400 m/z, indicating that it is mainly the small molecule metabolites, rather

Table 5.8 Peaks found to be used by FR models with high occurrence ($> 75\%$).

Tissue type classification	**Identified m/z peaks that were reported with high occurrence across the FR methods**
Fat:Wall	204, 217, 274, 287, 305, 368, 369, 370, 372, 385, 387, 388
Fat:Thrombus	204, 212
Wall:Thrombus	204, 230, 258, 260, 280, 298, 310, 308, 330, 331, 347, 400

than lipids or small peptides, that the ML algorithms use to distinguish these tissue types.

Table 5.9 The κ scores obtained when only the peaks found with high occurrence were used as input variables for ML for the tissue type classification.

Analysis setup	**ML model**	**Fat:thrombus**	**Fat:wall**	**Wall:thrombus**
Identified peaks - averaged used	KNN	1.00±0.028	0.92±0.08	0.69±0.21
	SVM	1.00±0.0	0.94±0.08	0.639±0.21
	LDA	0.98±0.07	0.88±0.10	0.701±0.20
	NBC	0.99±0.03	0.93±0.08	0.723±0.21
	RF	0.98±0.06	0.94±0.08	0.672±0.21
Identified peaks - repeats used	KNN	0.99±0.01	0.91±0.06	0.65±0.17
	SVM	1.00±0.01	0.90±0.06	0.66±0.16
	LDA	0.99±0.03	0.86±0.07	0.63±0.16
	NBC	1.00±0.01	0.91±0.06	0.66±0.16
	RF	0.99±0.03	0.90±0.06	0.69±0.15

Analysis of the distribution of peak used by FR method showed a region of high FR-selection frequency observed at approximately 350-400 m/z in the χ^2, stats and OI FR methods across all the classifications. The region from m/z 350-400 appears to be important, and has been tentatively assigned to the sterol compounds, such as cholesterol (m/z 369 for cholesterol-H_2O or m/z 386 for parent molecule), lanosterol (m/z 386), and lumisterol-3 (m/z 385). These peaks are abundant in the spectra recorded for all biological tissue, as the molecules are highly abundant in tissue samples and ionise very efficiently in the ASAP source. They also are known to play important roles in human pathology, especially cardiovascular diseases [333].

5.3 The prediction of AAA growth rate, size, and flow-mediated dilatation

5.3.1 Introduction

One of the aims of the OxAAA study is to determine if indicators of AAA growth rate and size can be identified and measured. This would allow patients with rapidly growing AAAs to be monitored more frequently and potentially considered for earlier surgical intervention, thereby reducing the risk of rupture. Any mass spectrometric markers found to be associated with AAA growth can also be studied further in order to improve our understanding of disease progression. Prediction of AAA patient groups, and importantly the AAA size, was shown to be possible in Section 5.1, but these classifications were quite broad. As a next step we would like to explore whether AAA size can be determined with more precision than just classification as 'small' or 'large', as well as determining the ability to predict the AAA growth rate and FMD value from the ASAP-MS plasma measurements that were acquired previously. In this section, we describe a number of different analyses that were used to investigate these questions.

5.3.2 Materials and methods

The measurements used in this study were those described previously in Sections 5.1 and 5.2. For the present analysis, subsets of patients were selected for whom growth rate, AAA size, and FMD measurements were available. Growth rate (% annual growth) was available for 76 patients from the 'small AAA' patient group. Aneurysm size (mm) is available for 107 'small AAA' patients and 20 'large AAA' patients. FMD was available for 85 'small AAA' patients and 32 'large AAA' patients. The population distributions of these variables are shown in the histograms in Figure 5.10.

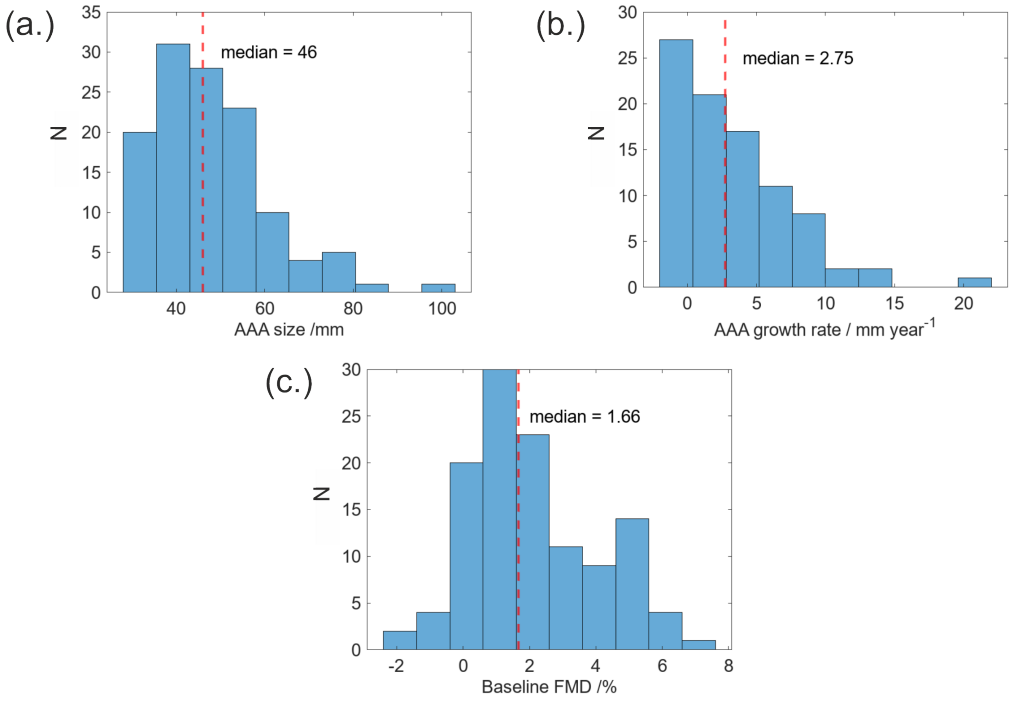

Figure 5.10 Histogram plots of the patient population distributions across each of the clinical parameters of interest: (a) AAA size in mm; (b) AAA growth rate as determined by ultrasound measurement in mm year^{-1}; and (c) FMD score baseline measurement %. The median value for each distribution is shown by the red dotted line.

The statistical methods described in previous sections for the OxAAA and OxAMI studies, were used to determine patterns in the ASAP-MS data that may be used for the prediction of the clinical measurements. Details of these methods are described in Section 5.1.2. Correlation analysis between the MS data and clinical measurements was conducted using both regression analysis for the continuous variables, and classification on the binary encoded versions of the three clinical variables. The encoding was conducted at above and below the median value in each dataset (shown in the histograms in Figure 5.10), as well as into extreme groupings of $n = 10, 20, 30$ as done for the OxAMI study in Section 4. The Shapiro-Wilk test for normality showed that the data were not normally distributed between the S1 TP-1 and S2-TP-3 classes used for analysis of these variables ($P > 0.05$), hence the non-parametric statistical tests, Wilcoxon rank-sum test for classification, and Spearman's correlation coefficient test for regression, were used.

5.3.2.1 Analysis of overlapping significant peaks

In an attempt to use this data to predict FMD or growth rate, the method put forward by R. Lee *et al.* in [201] to identify protein biomolecules within the context of the OxAAA study was utilised. The protocol discussed in this paper ([201]) shows how metabolic markers for a disease may be identified by finding the biomolecules that show significant differences between multiple overlapping studies. In the present case, we looked for m/z peaks that had significantly different intensities in the pre-vs-post surgery comparison, the fast-vs-slow growth comparison, and the omental fat vs thrombus comparison. These comparisons were selected as it was suggested that they would identify metabolite markers most related to the AAA in the pre-post surgery and omental fat vs thrombus comparisons. The intensity of any peak that was found to be significant in all of three classifications was then plotted against growth rate and FMD to explore the correlation with these two parameters. The idea behind this analysis is that it may help to distinguish peaks that are truly significant from those that are the result of false positives, the incidence of which is often high in this type of analysis. In R. Lee's study, the confidence level used was low, at $p < 0.05$. A confidence level of 5% means that there is only a 5% chance that the identified peak is a false positive.[b] A higher confidence level may be more appropriate for use with ASAP-MS data in order to further reduce any false positives.

To investigate growth rate, the 'extreme' groups of patients were used, in order to maximise the chance of identifying significant peaks. The pre vs post surgery, and omental fat vs thrombus classifications were conducted with matched samples (patient identities matched across the classes), and so a Wilcoxon signed-rank test was used. For the growth rate and large vs small AAA analyses, matched samples were not available, so the Wilcoxon rank-sum test was used, as previously.

[b]When there are a large number of peaks in the mass spectrum (approx. 1000) the number of false positives can be significant, 50 in this example. However, the overlap of false positives from multiple different statistical analyses is still likely to be small.

5.3.2.2 Symbolic regression analysis

The symbolic regression protocol that was developed using the OxAMI data and discussed in Section 4.2.1 was applied to the prediction of AAA size, FMD and growth rate. The PySR model was applied to this data using the optimised protocol discussed previously. The averaged and normalised ASAP-MS spectrum was collected for each patient. The three approaches of data input used previously in Section 4.2.2.3 were also applied to this dataset: approach (i) using all peaks in the spectrum (labelled all peaks), approach (ii) using a 10 component PCA feature reduction to train the PySR model (labelled PCA), and approach (iii) using statistically significant peaks found using a Pearson correlation coefficient at $p < 0.05$ (labelled stats).

The same hyperparameters that were used in Section 4.2.1 were employed for the OxAAA dataset, as they were shown to be appropriate for the analysis of ASAP-MS plasma data. The PySR analysis was run in triplicate for each clinical variable and input data set respectively, each time with a new training-test data partition.

The symbolic regression method assessment protocols used in Section 4.2.1 were applied. The accuracy of the resulting models were assessed by evaluation of the test data fit by Pearson's r correlation, and error described using RMSE. Where the algorithm generated multiple equations of differing complexity, the equation with the highest r^2 value when applied to the test data is reported as the output equation.

5.3.3 Results

5.3.3.1 Statistical analysis of continuous variables

Statistical analysis of peaks that varied significantly with clinical variables showed that a large number of peaks varied with AAA size, but very few peaks varied significantly with FMD or AAA growth rate. Table 5.10 shows how many significant peaks were found within both regression and classification tests for the three clinical

Table 5.10 The number of statistically significant peaks found for each clinical variable at four different significance levels.

Analysis type	Confidence	N significant peaks identified		
		AAA size	AAA Growth	FMD
Classification (WSR)	p<0.05	94	59	39
	p<0.01	20	4	6
	p<0.005	10	0	4
	p<0.001	2	0	1
Regression (Spearman)	p<0.05	144	42	42
	p<0.01	57	2	6
	p<0.005	39	1	2
	p<0.001	14	0	1
Overlap between regression and classification	p<0.05	70	22	23
	p<0.01	15	1	4
	p<0.005	8	0	2
	p<0.001	2	0	0

variables at four different significance levels. At the highest confidence levels of $p < 0.05$ and $p < 0.01$, very few peaks were identified for either FMD or AAA growth. The peaks found at the lower confidence levels are more likely to be due to chance as a result of the large number of MS peaks that were being assessed.[c]

5.3.3.2 ML classification results

Prediction of clinical variables by ML classification was successful for AAA size, echoing our results from Chapter 5.1, but no accuracy above $\kappa = 0.2$ was achieved for either AAA growth rate or FMD. The κ scores for the best ML models and relevant results tables for each of the three clinical variables are shown in Appendix C.6.

5.3.3.3 Analysis of overlapping significant peaks

The overlap of significant peaks identified across the three analysis of growth rate, pre vs post surgery, and omental fat vs thrombus was assessed at each confidence

[c]There are almost 1000 m/z peaks being assessed for their significance with each variable. If using a low confidence interval of $p < 0.01$, this gives a 1% chance that a significant result is due to chance. Therefore just by chance, it is likely that 10 peaks will be found to vary significantly. This is a type I error in statistics.

Table 5.11 The statistically significant peaks ($p < 0.05$) found to overlap in all three clinical variable classifications at different levels of extreme grouping for the growth rate variable.

Extreme growth rate class size	m/z peaks that fully overlap with all three analyses
N = 10	59, 73, 319, 414
N = 20	319, 403
N = 30	103, 146, 232, 255, 295, 301, 303, 304, 319, 331, 386, 387, 388, 403, 790

level, and for each extreme grouping. A small number of overlapping peaks were identified at $p < 0.05$, shown by the Venn diagrams in Figure 5.11, the identities of which are displayed in Table 5.11. No overlapping peaks were identified at higher confidence levels. Analysis of the correlation of the intensity of the peak at m/z=319 (the highest correlation identified) and growth rate showed a very weak negative correlation to growth rate by Spearman r correlation of $r = -0.356$ at $P = 0.002$, shown in Figure 5.12. The peaks identified were also assessed for correlation with FMD, with only one very weak correlation found at m/z=103 which is also shown in that Figure. This is shown next to the best, but also very low, correlation found in the original work by R. Lee in [201] for attractin, a molecule involved in immune response and inflammation. Low correlation values are common in medical settings, as discussed previously in Chapter 3, and so it could be interpreted that these molecules may be worth investigation for clinical value. However, these plots highlight the need for a much larger dataset of patients with growth rate measurements, particularly to populate the sparse high growth rate regions at $> 10\%$ per year. A further comparison at this confidence level between all four analysis (growth rate, AAA size, pre vs post surgery, and fat vs thrombus) showed very few peaks of interest, none of which were truly correlated at $r > 0.2$ on assessment of the correlation plots. The Venn diagrams of this analysis are shown in Appendix C.7.

Poor correlation with growth rate may well be due to the structure of the Ox-

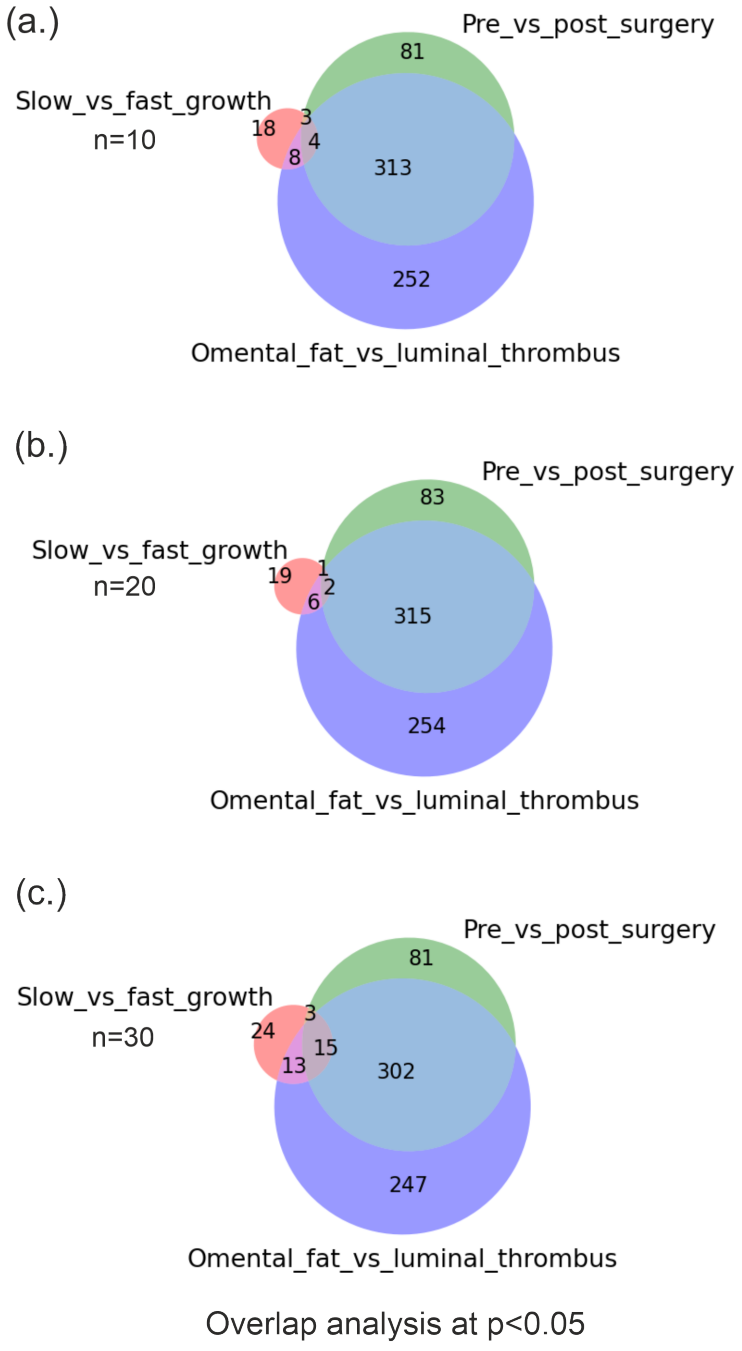

Figure 5.11 The significant peaks at the lowest confidence interval used (p<0.05) found in across three different patient group comparisons: pre vs post surgery (green), mental fat vs abluminal thrombus (blue), and slow vs fast growth (red), where the slow vs fast growth analysis has been conducted on extreme grouped patients at $n = 10$ per class (a), $n = 20$ per class (b), and $n = 30$ per class (c).

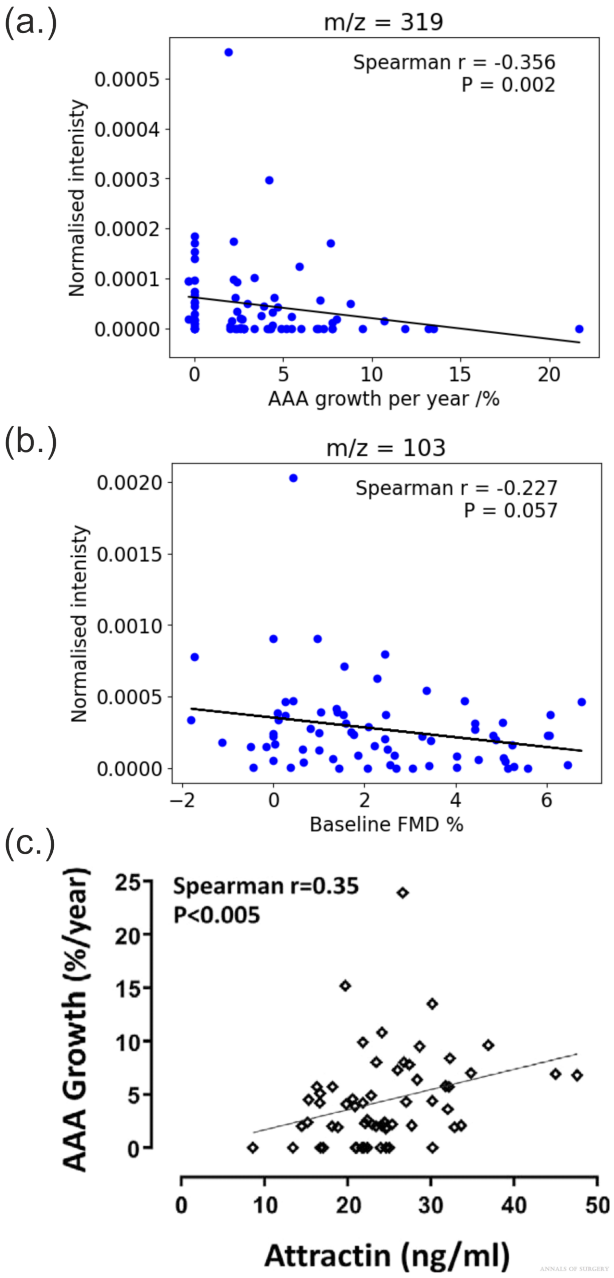

Figure 5.12 The change in normalised intensity with AAA growth rate for the most significant correlation found (p<0.05) at m/z = 319 (a), and the most significant correlation found from the overlap analysis peaks shown in Table 5.11 plotted against FMD (b). The correlation plot in (c) is taken from [201] and shows the correlation of the molecule attractin with AAA growth rate. Note the poor correlation in all three cases.

AAA study itself. The growth rate measurements in this study are calculated by assessing the difference between two ultrasound measurements, and hence only show an average growth for that patient over that time period. However, as it has been shown that growth rate is unlikely to be constant or linear for a single patient over time, and an aneurysm may well go through periods of growth and stabilisation over time, this may not be a good representation of the AAA growth rate for that patient. An improvement to this would be to take many measurements of growth rate over a period of years, and take the corresponding blood samples at each of those measurement times. Unfortunately this would require a prohibitive amount of time and resources to collect.

5.3.3.4 Symbolic regression of AAA clinical variables

Symbolic regression provided some encouraging results for the prediction of AAA variables, particularly for FMD and growth rate, the prediction of which had been previously particularly difficult. Figure 5.13 shows the best results achieved in each iteration of the analysis, across the three input data sets. It can be observed that the 'stats' data set proved favourable across all three variables, the peaks identified with which are described in Table 5.12. As previously, the 'PCA' data set resulted in very poor prediction accuracy across all variables.

Table 5.12 The statistically significant peaks identified by Pearson's correlation coefficient that were used as inputs in the PySR analysis.

Clinical variable	n stat sig peaks	Peak ids
FMD	11	344, 397, 414, 462, 471, 485, 829, 874, 929, 931, 935
Growth rate	7	145, 146, 232, 264, 319, 387, 535
AAA size	8	105, 466, 663, 753, 800, 907, 919, 935

FMD was predicted very well in the stats input methods, with one iteration generating an r^2 of 0.7. The plots in Figure 5.15 show that the equations result in a

very good visual fit for both the training and test data, albeit with a relatively large RMSE. This implies that there may well be metabolic indicators of FMD within the blood plasma, but that the relationship between FMD and peak intensities is complex. The relationship between complexity and accuracy for FMD was very right skewed, showing an increase in test data accuracy with increasing complexity. The complexity of the best equation was reported as 48, which is almost the maximum complexity that the PySR algorithm is capable of generating. The equation for the best fit for FMD was:

$$FMD_p = \left(\frac{I_{397} + I_{462} - \frac{I_{829}}{1.5675}}{0.0002} + 1.6346 + \frac{0.0952 \times I_{414} + 0.3719 \times I_{485} + I_{929}}{0.2925 \times I_{462} - I_{829} - \left(\frac{I_{397}}{-0.4403} + I_{485}\right)} \right) \times \exp\left(\frac{(I_{462}^2 + I_{929})^9}{(-(I_{485} + I_{931}) - 2.34104 \times 10^{-10})^9} \right) \quad (5.1)$$

which features 7 of the 11 statistically significant (mostly thought to be lipid) peaks found for FMD ($p < 0.05$). Investigation into the ionisation and fragmentation behaviour by ASAP-MS of biologically relevant lipids is ongoing.

Using a more sophisticated algorithm that is capable of generating equations with even higher complexity, the accuracy may potentially be improved further. However, given the number of samples in this dataset, the high complexity of this equation may imply that it is overfitting. Lower complexity equations also provided accurate predictions of FMD, for example, the following equation with a complexity of 23 generated a fit with high correlation shown in Figure 5.14:

$$FMD_p = 0.4542 \times \frac{I_{475}}{-I_{819} + 0.2456 \times (-I_{925} + I_{452}) - \frac{I_{387}}{-0.4326 + I_{475}}} + \frac{I_{387} + I_{452}}{0.0002 + 0.624} \quad (5.2)$$

The lower complexity of this equation makes it less likely to be the result of overfitting.

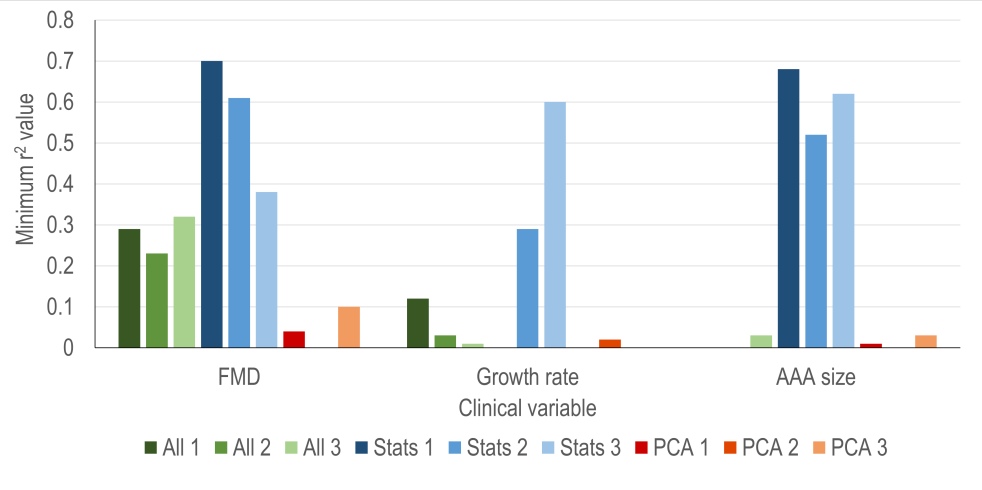

Figure 5.13 The accuracy obtained for each PySR output for the three clinical variables: FMD, Growth rate, and AAA size. The accuracy is shown as the r^2 fit between the predicted and true values for each variable, with the lower of either the test or training fit shown. Colours represent the data input method, with no feature reduction (green), statistical feature reduction (blue), and PCA feature reduction (red). Colour hues show the repeat results over multiple partitions.

The prediction of growth rate was less successful, but two potential equations were generated that appear to show a weak but positive fit to the data in the fit analysis, shown in 5.15(c) and (d). The test data is relatively well predicted in both examples. A clear skew to low growth rate values was observed for growth rate data,

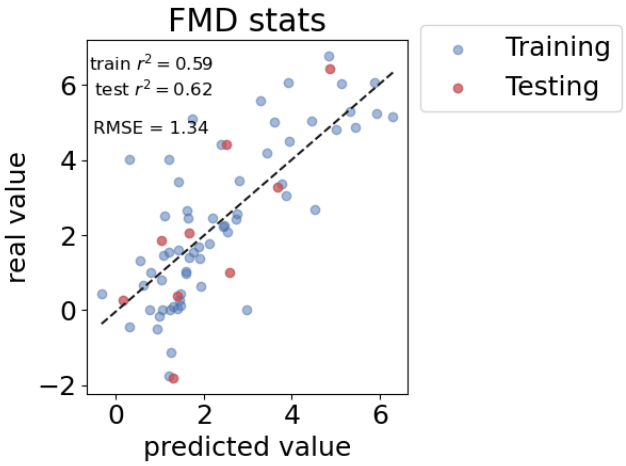

Figure 5.14 The fit of Equation 5.2 for FMD using the stats approach and the first partition.

with very little data present at the high growth rate range, which makes conducting

regression analysis more challenging. This is the same problem that was observed in Section 4.2.1 for MVO. With a larger and more evenly distributed dataset of growth rate data, this accuracy may be improved. The equation generated for growth rate was:

$$\begin{aligned} Growth\ rate = & \frac{I_{232}}{I_{232} + 0.4708 \times I_{319} - 3.1156 \times 10^{-5}} + \\ & \exp\left(\frac{I_{535}}{-I_{145} + 0.6049 \times I_{535} + 3.7112 \times 10^{-5}}\right) \\ & + \frac{1}{(-I_{145} + 0.8659 \times I_{264}) \times \frac{I_{232}^2 + I_{319}}{1.3773 \times 10^{-9}} + 0.2832 - \frac{3.6622 \times 10^{-9}}{(I_{535} - (I_{145} + I_{232}))^2}} \end{aligned} \quad (5.3)$$

and features 5 of the 7 significant peaks found previously for growth rate ($p < 0.05$). The complexity of this equation was also high, at 46.

The prediction of AAA size with PySR gave high accuracy scores, but visual analysis of the fit showed that these figures are misleading. The large vertical spread in the results relative to the trend line reveals that the PySR algorithm generates very similar predicted values across a range of true values. Interestingly, this vertical spread appears to separate patients in a similar way to the 'small' and 'large' AAA groupings, with two groups visible in the figures at $30 < x < 45$ mm and $45 < x < 55$ mm. As these binary groupings were well predicted with other classification methods, this may indicate that there is a metabolic change that occurs when an AAA reaches a certain size, in this case appropriately 45 mm rather than the 50 mm surgical threshold, and that change can be observed in the plasma. The best-fit equation for AAA size was:

$$\begin{aligned} size_p\ \ mm = & -\frac{I_{919}}{0.0704\,(I_{105} - (I_{466} + I_{800}))} - 10.451 \times \\ & \exp\Bigg(- 2.5676 \times 10^{40} \times I_{907}^2 \times (I_{466} - I_{753})^2 \times (I_{753} - I_{800})^2 \times \\ & (-0.1203 \times I_{800} + I_{935})^2 - I_{919} \times \left(-4595.36 + \frac{-2.5797}{I_{105} - 0.0039}\right)\Bigg) + 46.922 \end{aligned} \quad (5.4)$$

This was again very complex, at a complexity of 46, and used 6 of the 8 statistically significant peaks identified previously ($p < 0.05$). The use of these complex equations across the three clinical variables does raise concerns about overfitting, and should be tested further using larger datasets.

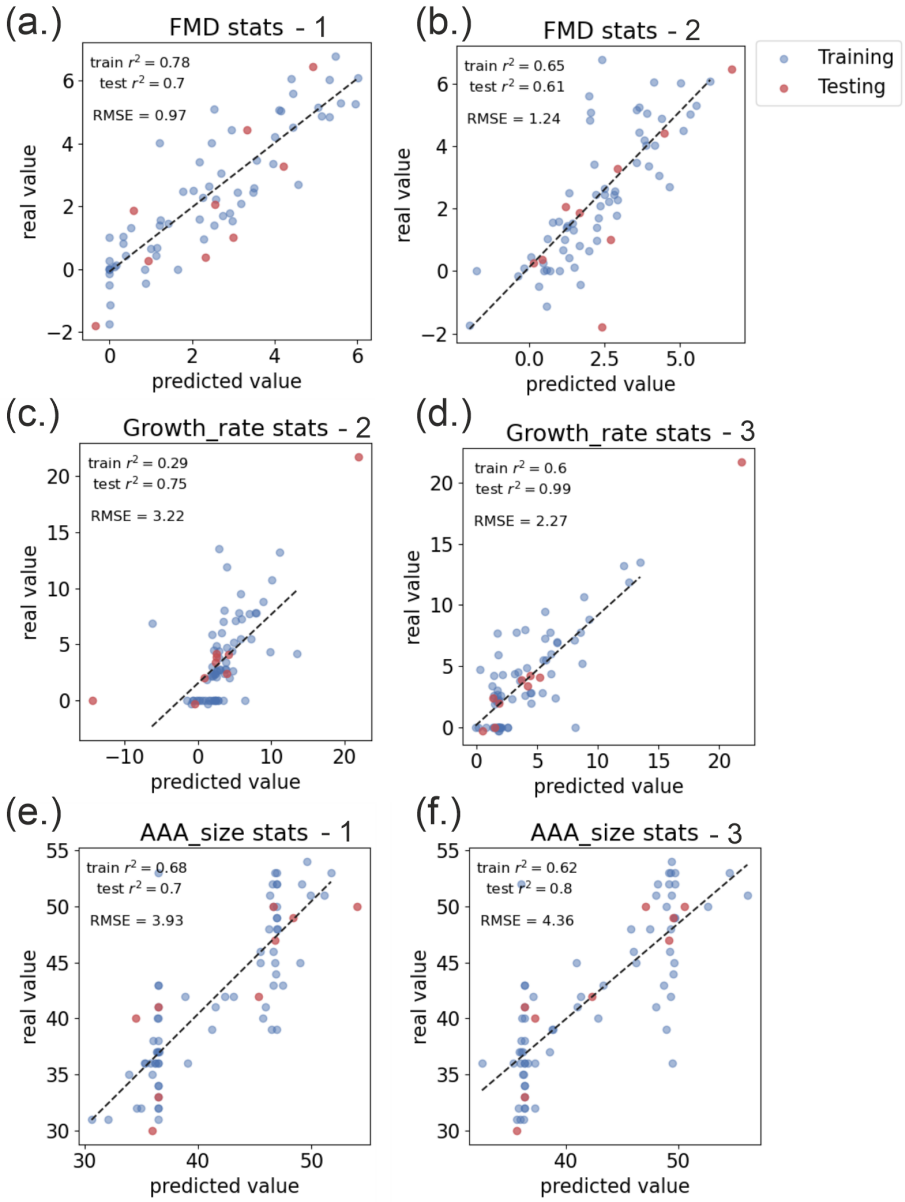

Figure 5.15 Plots of the fit of the best two equations generated using PySR for FMD (upper), growth rate (centre), and AAA size (lower). Training data is shown in blue and test data in red. The partition version is represented by $-n$ at the top of each plot.

5.4 Conclusions and outlook

The analysis described in this chapter has shown that ASAP-MS analysis of samples from AAA patients can be used to classify patients. In the first section, it was shown that plasma samples can be used to accurately classify AAA patients based on disease progression and pre/post-surgical status. By classifying AAA patients from healthy volunteers, this study shows that AAA pathology can be detected in the systemic circulation, supporting the consensus that AAA is the result of a systemic disease. Post-surgery AAA patients have been found to have a biochemical profile different from both AAA patients and healthy individuals, most likely due to the presence of markers for post-operative inflammation. However, metabolite peaks found to be important for the classification of both pre and post-surgical patients vs healthy individuals provide further evidence for the presence of systemic vascular abnormalities that persist after repair surgery. Patients with small and large AAAs were successfully distinguished, and two mass peaks were tentatively found to be important for classification. However, identification of significant mass peaks was challenging for this classification, probably due to the small dataset size.

We have further demonstrated that ASAP-MS can distinguish between broad tissue types. The distinct tissue types showed visibly different mass spectra and statistical differences in m/z peak intensities. ML classification models were able to distinguish these broad tissue types by the spectra alone with very high accuracy. The impact of FR methods was minimal on the final accuracy, but showed that a relatively small number of m/z peaks were driving the classification. The classification of tissue subtypes was far less successful. Although not directly useful for the OxAAA study, there are many other potential applications in clinical analysis where identifying a tissue type rapidly is useful. Examples include the identification of metastasised tumour origins and tumour boundary delineation, which is the focus of other work within our group.

A study of the impact of training data size (described in the Appendix) showed that the number of repeat measurements used has very little impact on the accuracy of the results, but that the accuracy increases with the number of patients used. This has implications for future study designs, in which the number of repeats may potentially be reduced to save time and resources.

The prediction of continuous clinical variables for AAA size, growth rate and FMD presented further challenges. However, using a range of methods, we have identified some promising results. Prediction of AAA size was relatively successful, with accurate ML predictions made using data grouped using new threshold sizes, extreme groups, and feature selection. A number of statistically significant biomolecule peaks were also identified, which could be explored further in future work. The ability to predict AAA size may assist with identifying biomarkers for growth in further studies, as large AAAs typically grow at a faster rate. The prediction of FMD and growth rate was less successful until the application of symbolic regression analysis, in which a number of complex equations were identified that predict these variables with relatively high accuracy based on the intensities of a small number of peaks.

Several mass peaks were found in this analysis that correlated with biomarkers found in the literature to be associated with AAA progression. This is an interesting finding, and confirms that ASAP-MS is capable of measurement of clinically relevant changes in mass peaks. However, we note that the identification of molecular peaks is not the aim of this analysis, and this should not be necessary for showing that AAA progression can be classified using ASAP-MS data.

The results from all of these tests have highlighted the need for a larger AAA dataset, particularly with more growth rate data, but these initial tests may enable analysis of such a dataset to be focused on areas of interest. We can conclude that further investigation into markers associated with AAA size and growth is warranted.

6

Investigation of different ASAP-MS instruments for clinical sample analysis

6.1 Introduction

As we are looking for subtle changes in metabolite patterns within the mass spectra, reducing any technical variation and noise will give the ML techniques the best chance of success. The measurements on clinical samples reported in previous chapters have been conducted using the Advion CMS-S expression ASAP-MS instrument. We have developed optimised protocols for the measurement of a range of biological sample types on this instrument, and have developed an understanding of the technical variation and systematic errors associated with this technique [180]. Other ASAP ionisation source instruments are available, each with their own novel features and advantages. In the following chapter, we discuss a series of analyses conducted using three different benchtop ASAP mass spectrometers:

- The Advion expression® compact mass spectrometer short m/z range (CMS-S) ASAP-MS instrument (Advion Ltd, Harlow, United Kingdom)[148].
- The Advion expression® compact mass spectrometer long m/z range (CMS-L) ASAP-MS instrument (Advion Ltd, Harlow, United Kingdom)[148].

- Waters RADIAN ASAP Direct Mass Detector (Waters, Wilmslow, United Kingdom).

A fourth, larger tandem instrument, the Waters Xevo G2-XS QTof MS equipped with an ASAP probe (ASAP-Q-Tof-MS, Waters, Wilmslow, United Kingdom) was also investigated. The aim of these measurements was to understand how the instruments compare across the following criteria:

1. Technical differences (ease of use);
2. Data quality and reproducibility for biological samples;
3. Costs and running requirements, such as source gas, instrument footprint, and consumables.

6.2 Materials and Methods

6.2.1 Biological samples

The samples used for this analysis were blood plasma and tissue samples from patients enrolled in the OxAAA study, as described in Section 2.3. These were chosen due to their availability and the large amount of data already collected using the Advion CMS-S instrument. No patient information relating to the OxAAA study was used for this analysis; the biological samples were used with permission from the study leads simply to represent the types of complex samples that we regularly measure. [a]

[a] A large number of experiments have been performed on the samples employed in this study. We recognise that changes to these samples will have been induced through the effect of freeze-thaw cycles and spending time at room temperature. There may, therefore, be features of the spectra recorded for these samples that are not representative of fresh samples. However, we believe that the samples are still adequate for method development. Samples were held on ice to help slow any chemical changes whilst measurements were taking place.

6.2.2 Measurement protocol for Advion CMS instruments

The plasma measurement protocol employed for the CMS-S and CMS-L instruments was the optimised protocol as described in Section 2.4.4 and [180]. The tissue measurement protocol used was the optimised tissue protocol described in 2.5.5. The practical work in this section was conducted with the help of Part II student Lloyd Smith (2023-2024), and DPhil student Liwen Song (2021-2025).

6.2.3 Measurement protocol for the Radian instrument

A measurement protocol for the Radian instrument was developed based on the Advion CMS-S protocol for blood plasma. Optimisation of the settings for source temperature, cone voltage, corona discharge current, scan time, and number of repeat measurements are discussed in Section 6.4. The Radian instrument was used on a short trial from Waters Corporation and Uvison instruments. The ASAP-MS instrument is set up in positive ion mode with a source temperature of 500 °C, a cone voltage of 12 V and corona discharge current of 3 μA. Capillaries are baked in an oven in advance at 300 °C for 30 mins. Prior to use, each capillary is cleaned using the 'bakeout' function. The mass range is selected to be 50-1200 m/z, and the scan time is set to 1 Hz. After starting the acquisition run, the clean capillary is inserted into the ion source housing for 30 s in order to record a background or blank spectrum. The probe is then removed from the spectrometer, cleaned and cooled with methanol, and dried on a lens tissue.

The plasma samples to be measured are thawed and vortex mixed. The probe is then used to wipe the inside surface of the sample container approximately 5 mm above the sample surface. A clean lens tissue is used to wipe off the excess sample, leaving a small amount of residue on the tip of the capillary. Wiping off excess sample was found to be required due to the sensitivity of the Radian instrument. Without this step, the spectra were seen to be saturated, with large amounts of carryover after the probe had been removed.

For tissue samples, the probe is used to touch the surface of the sample, and a piece of clean lens tissue is used to wipe off the excess sample from the capillary tip. The probe is then reinserted into the mass spectrometer for a period of 25 s, during which a series of mass spectra are recorded for the sample. The probe is then removed from the mass spectrometer and cleaned with methanol and lens tissue ready for the next repeat measurement. Five repeat measurements are made with the same capillary. After this the complete set of five measurements, including background measurement, is repeated with a new clean capillary, for a total of ten measurements per sample.

6.2.4 Measurement protocol for the ASAP-QTof-MS instrument

The measurement protocol for the ASAP-QTof-MS instrument was determined as that as similar as possible to the Radian instrument. The ion sources in the Radian and the ASAP-QTof-MS have the same settings and adjustments, but the structure, dimensions, and assembly of the ion sources are different, due to the different requirements of the inlets for these instruments. The ASAP-QTof-MS instrument is set up in positive ion mode in a continuum analysis mode, with a a corona discharge current of 3 μA, a sampling cone current of 30 A, a source offset of 80 A, a source temperature of 120 °C, a probe temperature of 450 °C, a cone gas flow rate of 50 Lh^{-1}, and a desolvation gas flow rate of 700 Lh^{-1}. Capillaries are baked in an oven in advance at 300 °C for 30 mins. Prior to use, each capillary is cleaned by changing the probe temperature to 600 °C for 2 mins, before returning to 450 °C. The mass range is selected to be 50-1200 m/z, and the scan time is set to 1 Hz. The ASAP ionisation source is not an optimised feature of the Waters QToF instrument, and consequently much of the method set-up is manual. The plasma sample to be measured is thawed and vortex mixed. The probe is then removed from the spectrometer, cleaned and cooled with methanol, and dried on a lens tissue. The probe

is then used to wipe the inside surface of the sample container approximately 5 mm above the sample surface. A clean lens tissue is used to wipe off the excess sample, leaving a small amount of residue on the tip of the capillary. The acquisition run was then started, with an acquisition time of 3 mins. The probe is then reinserted into the mass spectrometer for the 3 min acquisition time, during which a series of mass spectra are recorded for the sample. The acquisition stops automatically after the 3 min acquisition time. The probe is then removed from the mass spectrometer and cleaned with methanol and lens tissue, ready for the next repeat measurement. Three repeat measurements are made with the same capillary. Tissue samples were not measured using the ASAP-QToF-MS instrument.

6.3 Results

6.3.1 Subjective comparison of technical features and user observations

Prior to discussing the numerical data, a summary of some user observations and subjective assessments on using the Advion CMS and Waters Radian instruments is shown in Tables 6.1 and 6.2. The ASAP-QTof is not compared in this section as this is not a benchtop instrument, meaning it may not be appropriate for use in a clinical setting in the same manner. Overall, it was found that the Radian hardware was preferable for conducting experiments, in particular due to the shorter setup time and internal calibration system (eg Table 6.2 index 1, 2, 4, 11 -Advantages). The CMS instrument requires frequent cleaning and calibration, as noted in Table 6.1, which in addition to the longer setup time required for pumping down, makes the daily experimental procedure a time-consuming process. However, some features of the software for the Advion CMS instrument (e.g. the visual hardware read-backs, method setup page, and data output) were found to be optimal for data collection and analysis, with many user-friendly features making the instrument easy to use.

Table 6.1 User reported advantages and disadvantages of the Advion CMS ASAP instrument

Index	Advion CMS advantages	Disadvantages
1	Accessing the ion source housing is straightforward which makes cleaning convenient.	The unstable background and the need for frequent instrument cleaning and calibration reduces data reproducibility and slows down sampling
2	The glass window on the ion source housing allows the user to see the capillary, cone, and needle. This has been useful in identifying when the instrument needs cleaning, and troubleshooting when arcing between the cone and needle was observed.	The running and setup of the instrument is slow. The Advion CMS instrument requires approx. 1 hour to clean, calibrate and settle when used from standby, and 30 mins to pump down, 5 mins to calibrate, then 30 mins – 1 hour for the background to settle before the instrument can be used when starting from off and vented.
3	The probe is ergonomic, meaning sampling can be conducted with one hand. This facilitates working with biological samples.	Expensive capillaries that need cleaning prior to use.
4	The data recording software and user interface is user-friendly, with visual hardware readbacks that are easy to track.	Preset temperatures, and temperature range is restrictive
5	Setting the MS methods is straightforward (however this is may simply reflect user familiarity with the software).	The instrument is very sensitive to the sample consistency and sampling technique, producing poor chromatogram signals when the appropriate care is not taken.
6	The data output is in a format that is easy to integrate into analytical software and programs. It also includes the MS settings used for data collection, which is very useful when reviewing and archiving data.	
7	The total run time can be extended while the acquisition is running.	
8	The scan speed range of 200 – 2000 ms is appropriate for biological samples, and there are no fixed presets that must be used within this range, giving the user freedom to adapt the method to their requirements.	
9	Low nitrogen pressure requirements (approx. 3 bar) enable use of the in-house nitrogen supply.	

Table 6.2 User reported advantages and disadvantages of the Waters Radian ASAP instrument

Index	Radian advantages	Disadvantages
1	This system is very fast to set up to run from standby mode (1-2 mins), or from completely off and vented (10-15 mins).	The user interface is not user-friendly. The ion source settings are difficult to find and change, navigating between pages is challenging, and many different pages need to be open simultaneously to track the signal and settings together.
2	The detector is very sensitive, so very little sample is needed (although in the case of biological samples, this does require a wiping step to prevent sample carry-over after the capillary is removed).	The pressure required to run the instrument is higher than can be provided by the in-house nitrogen line, and so a nitrogen cylinder or nitrogen generator is required, which adds to the footprint and practical requirements.
3	The background signal is very stable, meaning runs rarely need to be aborted due to signal fluctuations.	Water's data output is challenging to convert to a useable format
4	The instrument requires only occasional cleaning, likely due to the high temperature range and the burn-off protocol providing effective cleaning of the source housing.	The acquisition time cannot be extended during an experiment, meaning if something is delayed the acquisition needs to be restarted.
5	A large temperature range of up to 600 °C with the facility for the user to set the temperature arbitrarily (to any point within the set range).	The pre-set scan speed settings in Hz are very restrictive.
6	Potential for running multiple experimental methods within a single acquisition.	
7	Stable and sensitive detector that is robust to intense signals and saturation, and needs replacing less frequently than in other instruments.	
8	Not needing to calibrate daily is another large time saving advantage.	
9	Capillary burn-off protocol in place provides excellent capillary cleaning and ion source cleaning.	
10	Small footprint, with the ability for the computer and gas supply to be integrated into the footprint. This may be beneficial for a clinical setting.	
11	The instrument runs the vacuum pump automatically, giving less room for user error or damage caused by incorrect use.	
12	Low cost consumables that are easy to acquire (standard capillary tubes).	

6.4 Data acquisition methods - quantitative comparisons

The optimisation protocols defined in Chapter 2 for plasma and tissue were adapted for use across the different instruments. The main focus of these experiments was to compare the performance of the CMS-S and Radian instruments, and a small number of experiments were also conducted on the CMS-L and ASAP-Q-Tof-MS.

6.4.1 Inter-instrument comparison 1: Optimum scan time

Scan time was defined previously as a measure of the time taken to acquire one scan over the scan range (seconds). The scan speed range for the Advion CMS instrument is 0.2 – 2 seconds, with an allowed interval change of 0.1 s. The scan speed range for the Waters Radian is defined by the scan frequency value, which is defined as how many scans are acquired per second, and has an allowed range of 1 – 20 Hz. This has fewer allowed preset options of 1, 2, 5, 10, 20 Hz (1 s, 0.5 s, 0.2 s, 0.1 s, 0.05 s) only.

The effect of scan time was investigated by recording spectra for each type of sample using a range of scan speeds, employing the protocols described previously in Sections 2.4.2.3 and 6.2.3 for the CMS and Radian respectively. Previous measurements on plasma using the Advion CMS showed that 900 ms was the optimal scan time, through comparison of signal-to-noise ratio achieved and number of unique m/z peaks identified (see Sections 2.4.2.3 and 2.5.3.2). Figure 6.1 shows the resulting scan time analysis averaged across the tissues for the Advion and Radian instruments. It was found that defining a clear optimum scan speed was challenging for either instrument using the tissue samples. For the Radian instrument, the data was more variable than previously observed for the Advion instrument, the signal-to-noise (blue) was on average slightly lower, but the number of peaks identified (red) was higher. The results for this analysis split by tissue type are shown in Appendix

D.1. While no clear trend was observed, the highest values for both variables were found at a scan frequency of 1 Hz. The scan speed of 1 Hz was therefore chosen for future measurements, in order to maintain similar settings between instruments where possible. We note that the low number of scan speeds available for the Radian instrument made this comparison challenging.

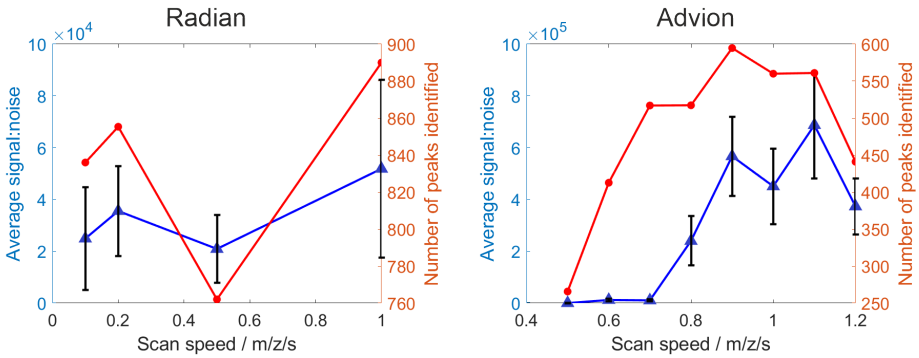

Figure 6.1 Analysis of scan speed averaged across the tissue samples for patient 77. Average signal to noise is plotted in blue on the left y axis, and number of peaks identified above the threshold intensity is plotted in red on the right y axis.

6.4.2 Inter-instrument comparison 2: Optimum acquisition time

To assess the optimum acquisition time, experiments were conducted in which the probe was left in the instrument for a 'long' acquisition time of 60 s. The method described previously in 2.4.2.5 was used for the CMS-S instrument, and an equivalent modification of method 6.2.3 was used for the Radian instrument, with an acquisition time of 60 s. The number of repeat measurements made was 5. The data were subjected to an Allan variance analysis (see Section 2.4.2.5). The Allan variance plots shown in Figure 6.2 reveal a very clear match in the optimum acquisition time between the Advion and Radian instruments for the same tissue samples. The minimum in the Allan variance plots for the maximal wall and preperitoneal fat was equivalent to an acquisition time of approximately 25 seconds. The abluminal thrombus sample did not show a typical time dependence in the plot, as observed

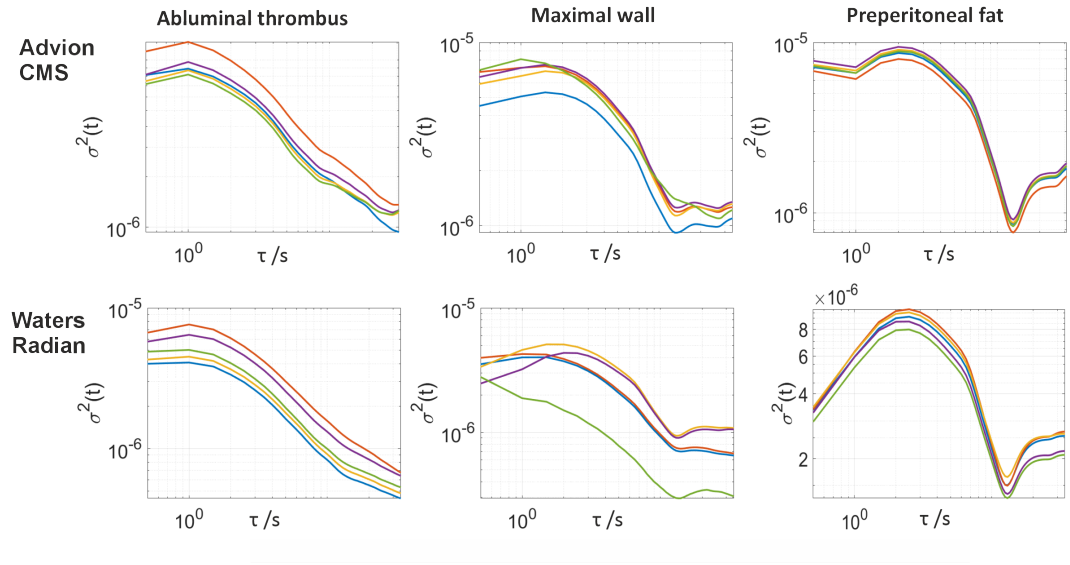

Figure 6.2 Allan variance plotted against τ in seconds. The plots shown are for abluminal thrombus, maximal wall, and preperitoneal fat from left to right, and for the Advion CMS-S and Radian instruments in the upper and lower plots respectively.

previously during method development for tissue samples. This may be due to the presence of whole blood in the thrombus samples.

6.4.3 Inter-instrument comparison 3: Optimum number of repeat measurements

The number of repeat measurements to be averaged over to generate the spectrum for each sample was assessed for the different instruments. Sets of 50 repeat measurements were taken of each different sample type for three different patients using the methods defined in 2.4.2.6 and an equivalent modification of method 6.2.3 used for the Radian instrument. The spectra were then averaged over increasingly large numbers of random selections of individual repeats to determine ΔI_N, which describes how the average spectra changes with N, as defined previously by Equation 2.3. Figure 6.3 compares how the results of this analysis for the two instruments. The Radian instrument displayed both a reduced change in the mean and lower error over random selections, showing that this instrument has a lower intra-sample variation in peak intensity over tissue samples. Both instruments showed a low ΔI_N

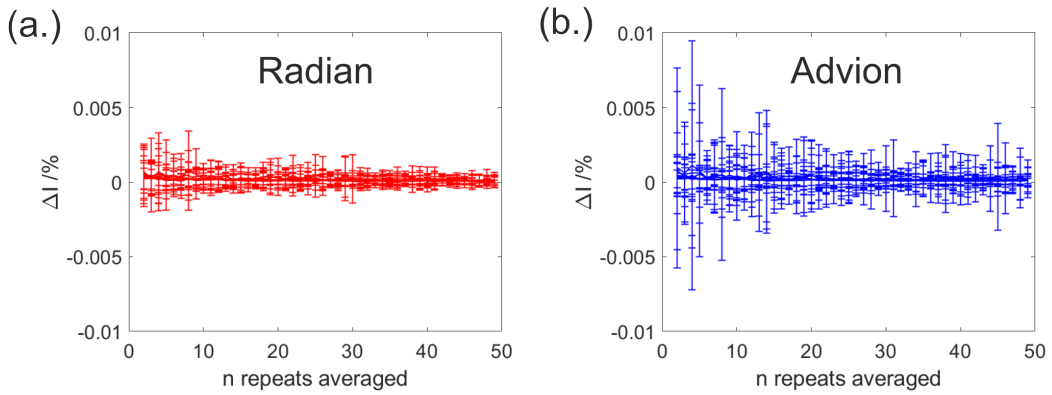

Figure 6.3 Analysis of the number of repeat measurements averaged to generate a patient spectrum. A set of 8 tissue samples were taken from patient 77. These were run for 50 repeat acquisitions on both the (a) Radian and (b) CMS-S instruments. The figures show the error in the mean change, error bars show the standard deviation in ΔI.

after approximately 10 repeats, which is consistent with data recorded previously for plasma samples.

6.4.4 Inter-instrument comparison 4: Optimum cone voltage

The cone voltage employed within an ASAP-MS instrument effects the extent to which molecules are ionised prior to entering the vacuum stages of the instrument. A higher cone voltage will result in the formation of more ions; however, it will also result in increased fragmentation of those molecules. As many biological molecules and metabolites decompose to common fragments, fragmentation reduces the available information on the metabolite composition of biological samples. This effect can be observed in the spectra as an increase in the abundance of low mass peaks as cone voltage is increased. On the CMS-S instrument, the cone voltage is set via the eight settings options available for the ion source, defined previously in Section 2.4.2.2 Table 2.3. The user can also define their own settings, but this involves further calibration and optimising, which is not a trivial process. The effect of the ion source settings on mass spectra recorded for blood plasma analysed with the Advion CMS-S instrument have been described previously in Section 2.4.2.2.

The effect of cone voltage was investigated for the Radian instrument by re-

cording mass spectra whilst changing the cone voltage from 3 V to 21 V, at 3 V increments. The cone voltage can be set much higher than this, up to 100 V, and given more time, further experiments would have been conducted at higher cone voltages. However, as it is known that high cone voltages cause increased fragmentation, which is unfavourable for these samples, only the low cone voltage analysis was conducted in order to observe how much of an impact this has on the data collected. No notable changes in the spectra were observed across the cone voltages used for these samples. The spectra recorded for various cone voltages are shown in the Appendix D.2.

6.4.5 Inter-instrument comparison 5: Optimum capillary temperature

Temperature affects the desorption of molecules from a surface, with behaviour that can be described by the van't Hoff Equation. As the temperature of the glass capillary tip of the ASAP probe increases, a larger number of molecules are desorbed from the surface of the capillary, and so more ions will be produced. For the Advion instrument, this behaviour is shown in Figure 2.7, referring to the settings shown previously in Section 2 Table 2.3. As the temperature is increased, mass peaks are observed across a greater number of m/z values, showing that a larger number of different molecules are desorbing from the surface. The relevant temperature for the CMS-S instrument is the capillary temperature. A capillary temperature of 250 °C is achieved using a source gas temperature of 400°C, which is the maximum temperature for this instrument (HT). For the Radian instrument, the maximum source gas temperature is 600 °C, which is higher than the maximum source gas temperature of the CMS-S, and so likely gives rise to a higher capillary temperature. Unfortunately, the capillary surface temperature is not reported for the Radian. The effect of source gas temperature on recorded spectra for blood plasma is shown in Figure 6.4. All spectra shown in the following have had the signal from cholesterol

(dehydrated cholesterol peak at m/z 369) removed. This is the largest peak in almost all spectra and has a highly variable intensity both within and between acquisitions. By excluding this from the spectra, the lower mass peaks of interest can be seen more clearly. At low temperatures, very little desorption occurs from the capillary surface. Between 400 °C and 500 °C, the spectra comprise many more peaks across a wide range of masses. From 500 °C to 600 °C, large signals at low masses are observed, with a very high number and intensity of low-mass molecules. This would need to be corrected for in data processing, and should be avoided if possible. From these observations, a temperature of 500 °C was determined to be optimal for the biological samples due to the large amount of high-mass information that is available at this temperature, and in combination with acceptable levels of background signal.

6.4.6 Inter-instrument comparison 6: Optimum corona discharge current

During APCI, a high voltage is applied to the corona discharge pin, creating a strong electric field at the tip of the pin which induces ionisation of the surrounding gas molecules. In positive mode, this ionisation process generates positive ions and free electrons. The electrons are accelerated away from the pin, colliding with further neutral gas molecules. The positive ions move towards the capillary inlet. This movement of ions generates a small but measurable corona discharge current. The user can adjust the corona discharge current, resulting in a change in the degree of ionisation that occurs in the ion source housing. The potential difference between the discharge electrode (the needle) and the counter electrode (the capillary inlet) is adjusted in order to tune the discharge current, and therefore the extent of ionisation. At a higher potential difference, an increase in the production of reactive ion species from the carrier and background gases occurs, facilitating the ionisation of the gaseous analytes. This can lead to stronger signal intensities and improved sensitivity. However, there is usually an optimal range. Beyond this range, fur-

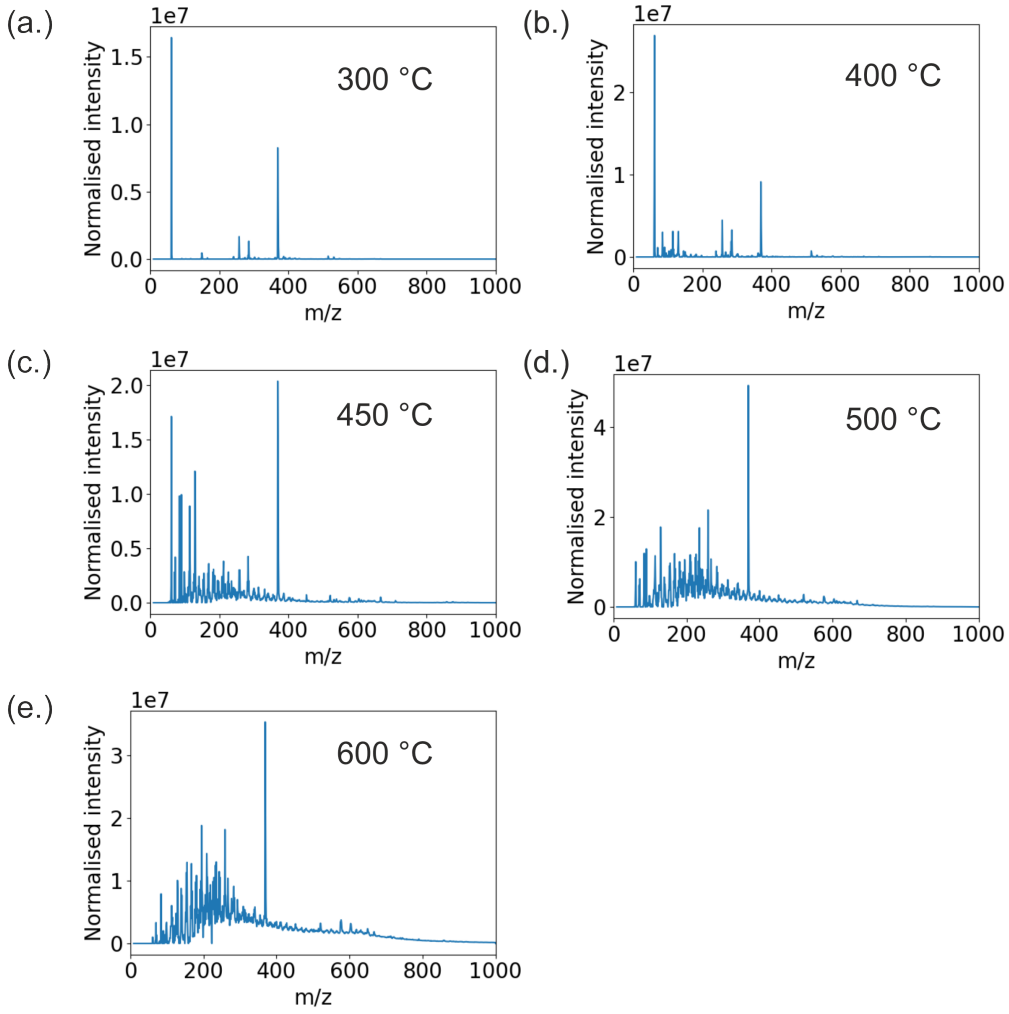

Figure 6.4 Averaged spectra of human blood plasma recorded at increasing source gas temperature from 300-600 $°C$ using the Waters Radian ASAP-MS. The spectra are normalised to unit area under the curve.

ther increases may not proportionally improve the signal, and can cause increased fragmentation of the analyte ion. Space-charge repulsion in the ion source may also reduce the ability to efficiently steer ions into the capillary inlet, and further increases can even induce arcing in the source.

The effect of tuning the corona discharge current was investigated for both the CMS-S and Radian instruments. For the CMS-S instrument, changing the corona discharge current from 1 to 5 μA resulted in no observable change in the spectra recorded. For the Radian instrument, changes were observed, as shown in Figure 6.5. The figure displays an increase in the intensity of the low mass metabolite

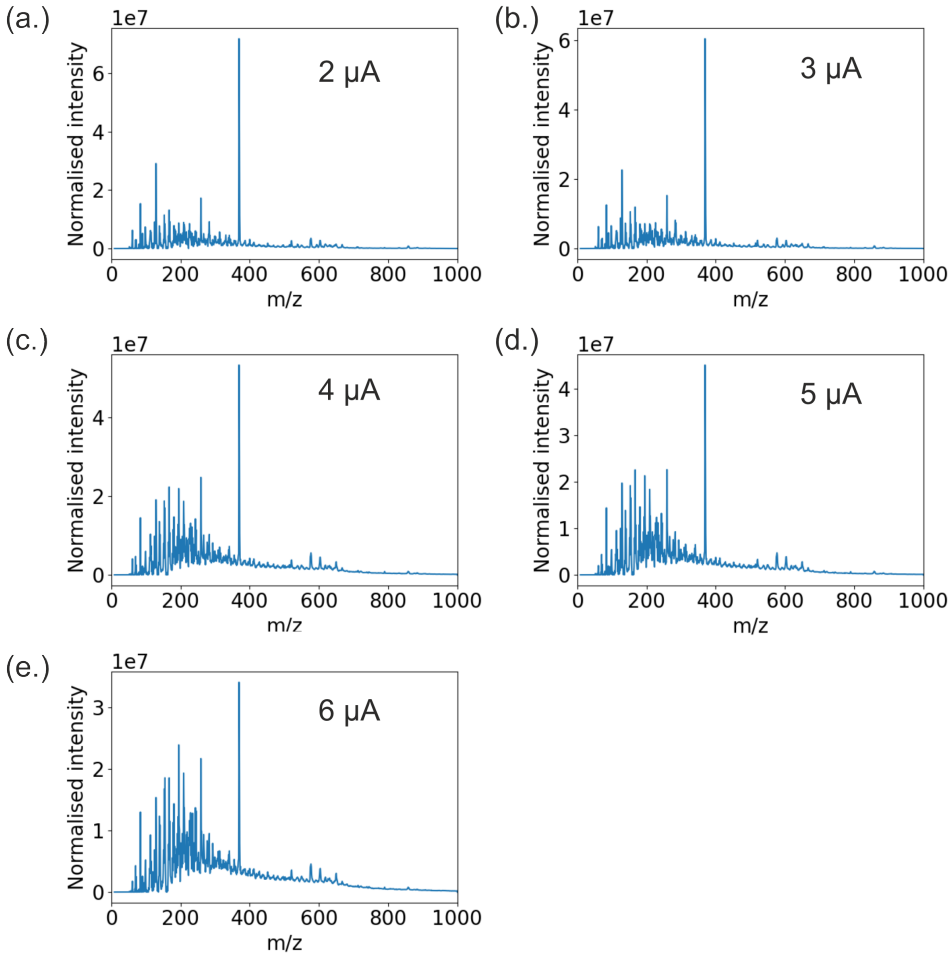

Figure 6.5 Averaged spectra for human blood plasma recorded at increasing corona discharge currents from 2 – 6 μA using the Waters Radian ASAP-MS. The spectra are normalised to unit area under the curve.

region as the corona discharge current increases.

An apparent increase in the baseline signal, caused by an increase in signal peak intensity and peak width, is also observed as the corona discharge increases, most evident at 6 μA. These features observed at the high discharge current are highly unfavourable. The increased current corresponds to more electrons, so a higher degree of electron ionisation, which eventually leads to the formation of the observed ions via proton transfer and other reactions. The increase in low-mass ion signals is presumably a result of increased fragmentation caused by increased collision energies. A corona discharge current of 3 μA was used in all further experiments on the Radian instrument to reduce the impact of this increased baseline signal. The lack of signal

change in the Advion system suggests that the ion signal is limited by a factor other than the discharge current in this instrument. Driving the current above 5 μA was not attempted on the Advion instrument due to a high risk of arcing.

6.5 Comparison of spectra between instruments

6.5.1 Comparison of tissue spectra: Radian and CMS-S

The spectra obtained for AAA tissue samples on the Radian and CMS-S instruments revealed differences in the intensity distributions of metabolite peaks. Example spectra are shown for the thrombus samples in Figure 6.6, and further tissue samples in Appendix D.3-D.5. The most notable difference is an increased intensity of higher mass peaks in the Radian data relative to the CMS-S spectra across all tissue types, particularly within the thrombus samples. Three metrics were used to compare the similarity of tissue samples recorded on the Radian and CMS-S instruments:

1. Percentage difference: The percentage difference between two instruments finds the percentage difference in normalised intensity between each peak, i, in spectrum a and spectrum b as a fraction of the intensity of each peak in spectrum a. This is then averaged over all n peaks, where n is the total number of m/z peaks in each spectrum.

$$\%D_{a,b} = \sum_{i=1}^{n} \frac{(a_i - b_i)}{a_i} \times 100\% \tag{6.1}$$

   The percentage difference was calculated for each peak, and then averaged over all peaks in the spectrum.

2. Pearson correlation (r): For a single sample measured across the two instruments, each pair of repeats available across the two measurements was also compared using Pearson correlation (r, defined previously in Section 3.2.4.1). In this instance, each repeat measurement (x_i or y_i) is subtracted from the

average intensity across all repeats recorded on that instrument ($\bar{x}$ or $\bar{y}$).

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{6.2}$$

This value varies between -1 and 1, where -1 shows a perfect inverse correlation, 0 shows a lack of correlation, and 1 shows a perfect positive correlation between the spectra.

3. Cosine similarity score: The cosine similarity measures the similarity between two sets of variables, x and y, when arranged as vectors projected in variable space. This is calculated through the cosine of the angle between the vectors, or the dot product of the vectors divided by the product of their lengths [258, 334]. Each repeat measurement recorded on instrument x (x_i) and instrument y (y_i) are compared iteratively. The complement to one of the cosine angle is defined as the cosine similarity.

$$\text{cosine angle} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}} \tag{6.3}$$

$$\text{cosine similarity} = 1 - \text{cosine angle} \tag{6.4}$$

Each analysis was conducted for tissue (below), and plasma (in the following section), in a method similar to that conducted by Zhvansky *et al.* [258].

The calculation of inter-instrument spectra metrics produced some interesting results, tabulated in Table 6.3. The Pearson's r correlation and Cosine angle were high in most cases, and produced very similar values between the two metrics. This is expected, as the Pearson's r coefficient becomes equal to the cosine angle when the mean is equal to zero, and since the average values of many of the m/z intensities in a spectrum are close to zero in a normalised spectrum, the two metrics produce similar outputs. In a number of cases the metrics did not correlate with the visual appearance of the spectra. For example, the thrombus data was observed to vary

significantly between instruments, but all three metrics showed high correlation. Bar chart representations of this data are shown in Appendix figure D.6.

The same analysis was conducted for a set of ten plasma samples, which were analysed using all four ASAP-MS instruments. Each sample was used to record ten repeats on each instrument. The average intra-sample instrument error was calculated as the standard deviation of peak intensities for a sample across the repeat measurements recorded on an instrument. The mass spectra of the samples were compared between instruments using Pearson's r correlation.

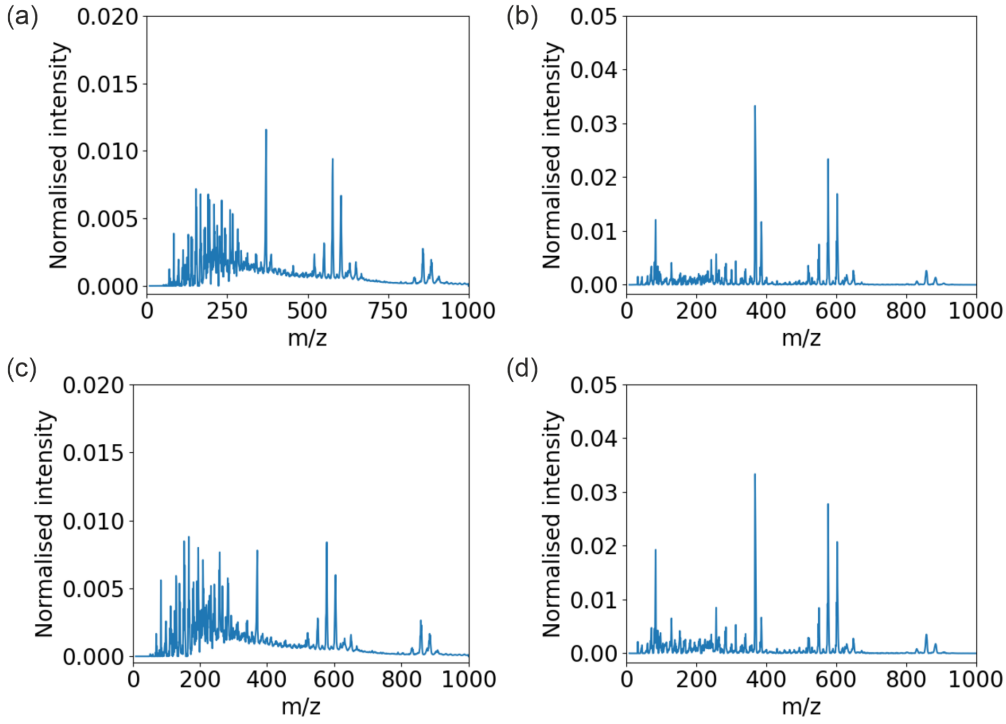

Figure 6.6 Mass spectra for Patient 77: thrombus recorded on the Waters Radian (left, a,c) and the Advion CMS-S (right, b,d) representing the different thrombus tissue subtypes of abluminal (top), and luminal (lower) thrombus.

An example of the results for a single sample, Sample 1, are shown in Figure 6.7. Along the diagonal are the intra-instrument results, showing very high values in the r correlation and cosine angle metrics. This reflects good reproducibility of measurements made with any individual instrument. The off-diagonal squares show the results of the inter-instrument analyses, and reveal lower values for both metrics. The most similarity was found between the spectra recorded using the CMS-

S and CMS-L instruments, which is not too surprising since they share identical ion sources and differ only in the quadrupole mass filter. The correlation between spectra recorded using the Radian and the CMS-L instrument was higher than that observed between the CMS-S instrument and the Radian.

Table 6.3 The instrument comparison metrics between the Radian and CMS-S instruments for a range of different tissue samples.

Patient n	**Tissue type**	**Percent Difference spec /%**	**Pearson's r**	**Cosine angle**	**n repeats**
77	Ab. thrombus	100 ± 321	0.88 ± 0.02	0.85 ± 0.03	50
77	Lum. thrombus	129 ± 112	0.8 ± 0.07	0.77 ± 0.06	50
77	Om. fat	262 ± 237	0.91 ± 0.02	0.91 ± 0.02	50
77	Pre. fat	574 ± 852	0.77 ± 0.07	0.78 ± 0.07	50
77	Sub. fat	510 ± 433	0.78 ± 0.03	0.79 ± 0.03	50
77	Dist.wall	336 ± 313	0.75 ± 0.12	0.76 ± 0.12	50
77	Max. wall	210 ± 268	0.79 ± 0.10	0.80 ± 0.1	50
77	Prox. wall	141 ± 164	0.83 ± 0.05	0.83 ± 0.05	50
75	Dist. wall	77 ± 29	0.73 ± 0.14	0.75 ± 0.13	10
76	Prox. wall	58 ± 11	0.81 ± 0.09	0.83 ± 0.08	10
78	Max. wall	84 ± 49	0.84 ± 0.06	0.85 ± 0.06	10
80	Prox. wall	1908 ± 1051	0.45 ± 0.35	0.51 ± 0.331	10
83	Dist. wall	776 ± 329	0.91 ± 0.03	0.90 ± 0.03	10

This can be attributed to the higher sensitivity at larger m/z values on both the Radian and CMS-L instruments. The correlation between spectra recorded using the ASAP-QTOF-MS instrument and with the lower-resolution benchtop instruments was very low. There are many differences between the ion sources, quadrupoles, and detectors employed in the high-resolution and lower resolution instruments, as well as the addition of a ToF stage, and so the spectra for samples recorded on this instrument have a significantly different distribution of masses and fragments. The fragmentation patterns observed are also different. The results shown for Sample 1 were found to be reproducible across the other five samples measured, the plots for

which are shown in Appendix D.7.

CV was calculated for the measurements made on each instrument, giving an indication of the uncertainty and reproducibility for the plasma samples. Figure 6.8 shows that the ASAP-QTOF instrument has the highest reproducibility, with an average CV of 11.98%. This is in agreement with the high specifications reported by the instrument manufacturers. The Radian instrument had a lower CV of 23.28%, which is still very good for measurements on blood plasma. The two CMS instruments showed very similar CV% at 35.74% and 38.26% for the CMS-S and CMS-L instruments, respectively.

6.6 Converting data between instrument formats

As we have shown, the mass spectra recorded on each instrument have characteristic features that distinguish it from data recorded on other instruments. For clinical applications it is vital that data recorded on different instrument types can be employed within the same analysis, eg. in building large clinical data sets with which to train ML algorithms. Inter-converting the mass spectra recorded on different instruments was attempted using instrument functions. The steps for generating an instrument function to transform spectra between two instruments (A and B) were: (i) to divide each spectrum for sample x in dataset A by each spectrum of sample x in dataset B; and then (ii) to average the resulting ratio over all of the repeat measurements made for a variety of different samples (see Appendix D.8 for further details). A PCA of the data allowed similarities and differences between the data sets to be visualised. A 'test' dataset of spectra that were not used in the generation of the function were then multiplied by the instrument function to convert from one instrument to the other in order to assess the success of the approach.

The initial training data recorded on two instruments was well separated within the two or three PC space in each case, and can be seen for the CMS-S to CMS-L and Radian to CMS-S examples in Figure 6.9 as the yellow and dark blue clustered points

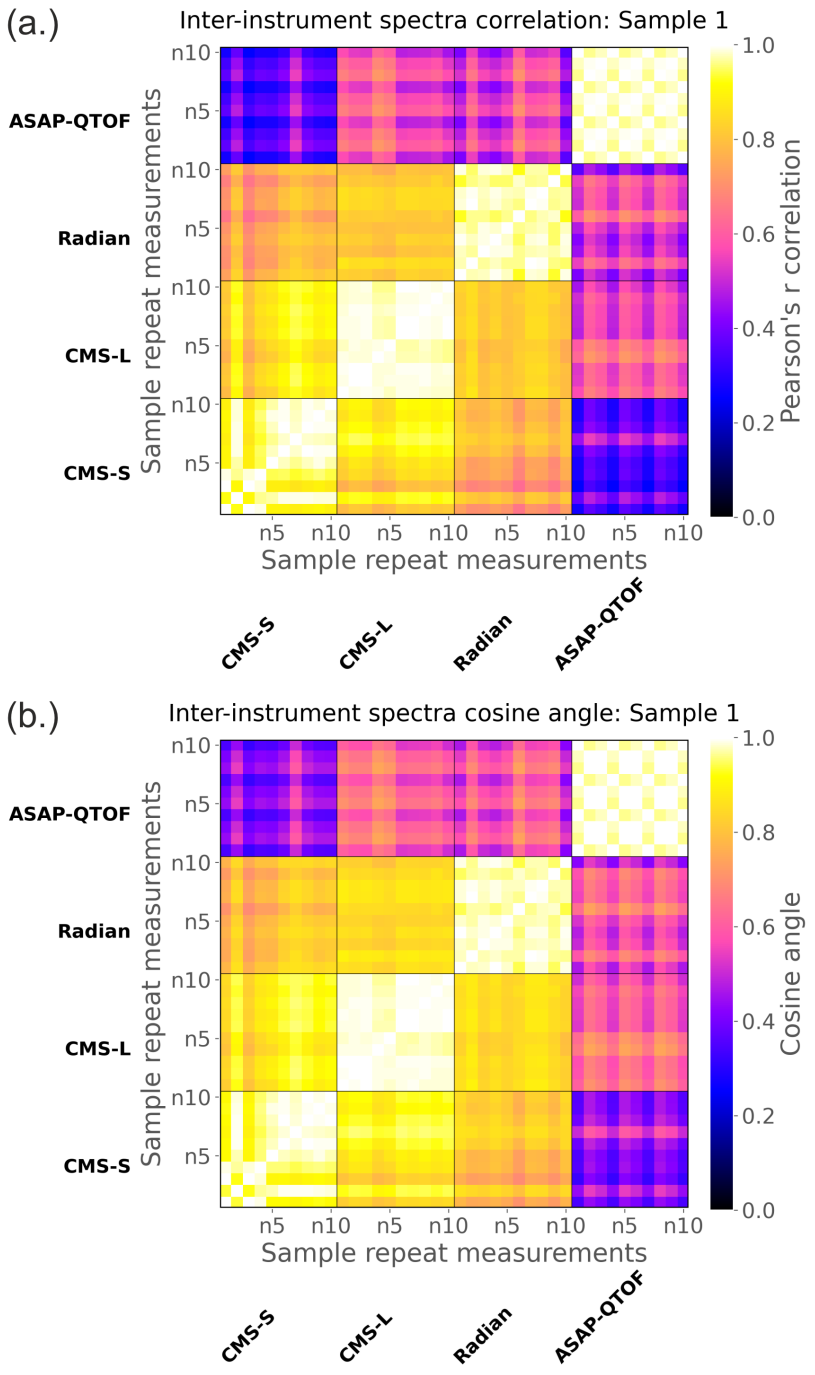

Figure 6.7 The Pearson's Correlation matrix (left) and Cosine angle matrix (right) showing inter- and intra-instrument sample variation for Sample 1, recorded on the four different ASAP-MS instruments. Darker colours represent lower correlation values or larger cosine angles plotted as 1-cosine angle. The x and y axes are split into the four instrument datasets. Each instrument dataset contains 10 repeat measurements, each represented by a pixel square in the diagram.

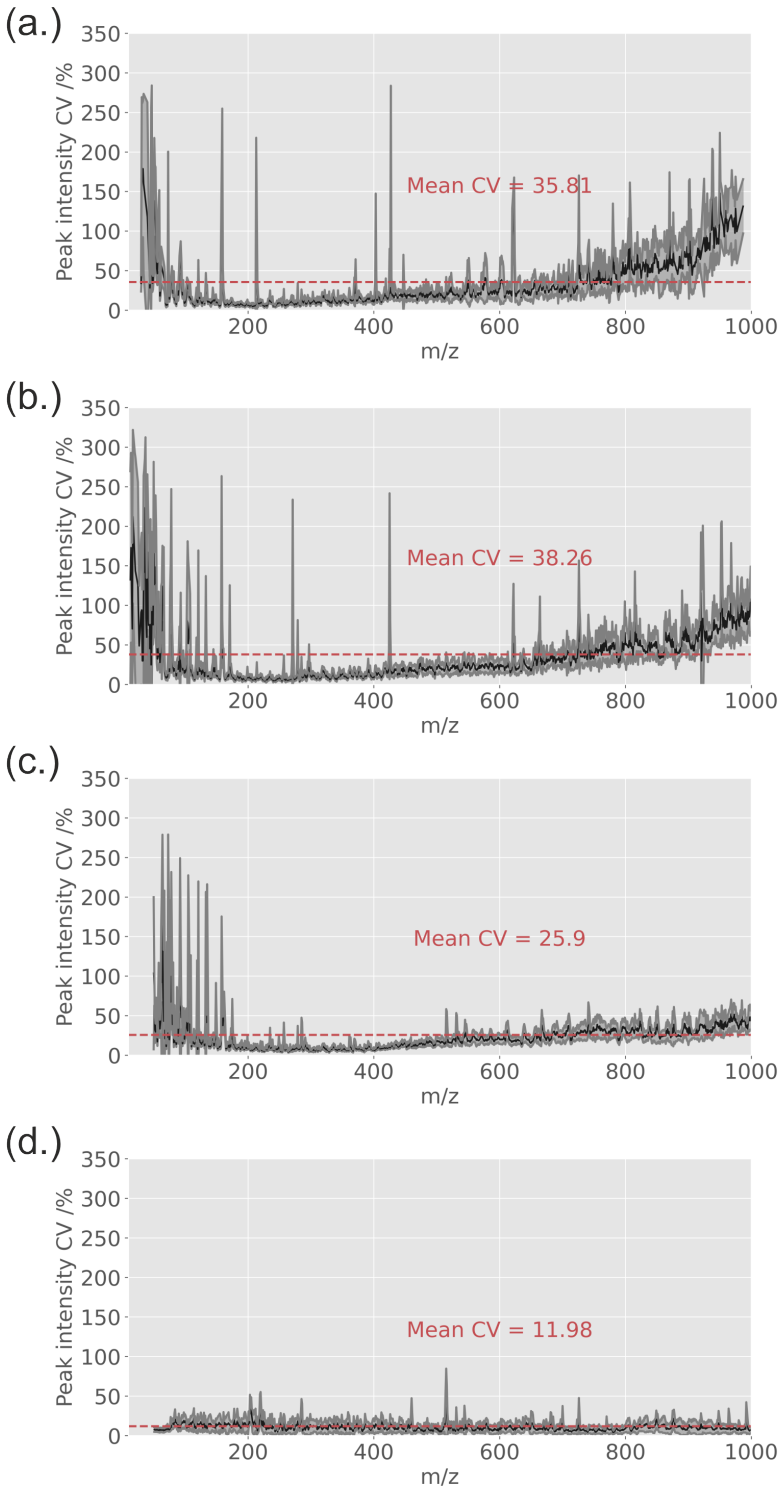

Figure 6.8 The CV values for each instrument (CMS-S in (a), CMS-L in (b), Radian in (c), QTOF in (d)) averaged over the recorded plasma data. CV is shown for each *m/z* value as a spectrum in black, with the standard deviation in CV added as grey shading to $\pm 1\sigma$. The average CV over the spectrum is shown as the red dashed line.

for instruments CMS-L and CMS-S, respectively. Prior to transformation, the data recorded using the CMS-S and CMS-L instruments form distinct clusters on both a 2D and a 3D PCA plot. After transforming the test data recorded on the CMS-L instrument to match that recorded on the CMS-S instrument, the transformed data points (turquoise in the Figure) cluster well with the CMS-S data. The same was observed for the conversion of Radian data to CMS-S data. The ideal position of the turquoise points would be overlapping with the green points, which show the same sample recorded on the target instrument.

The instrument functions themselves for these two conversions are visualised as a function of m/z in Figure 6.10. As expected, the largest functions are applied to the high mass peaks for the CMS-S to CMS-L conversion, which reflects the larger quadrupole and increased mass range (and subsequently higher sensitivity for larger masses) in the CMS-L instrument. The Radian was found to be less sensitive over the low mass range than the CMS-S instrument; hence the largest values for the instrument transformations functions are found in the low mass region of 10 - 200 m/z. The full instrument function visualisations are shown in Appendix D.12. Instrument functions were found to be highly mass-dependent, with a very non-smooth appearance. This suggests that the function was fitting some of the noise in the data sets, and so smoothing functions were utilised to see if this could be reduced.

Two common smoothing functions, the Gaussian filter, and a Savitzky-Golay filter, were applied to the raw instrument functions to assess the impact this would have on the conversion of the test data. Details of this study are described in Appendix sections D.9-D.12. The use of smoothing functions significantly decreased the accuracy of the data conversion, even with minimal smoothing. This suggests that there is a more complex mass dependence of the ionisation, fragmentation, quadrupole transmission, and detector probabilities, and their instrument dependence, and associated errors than can be represented by this sort of function.

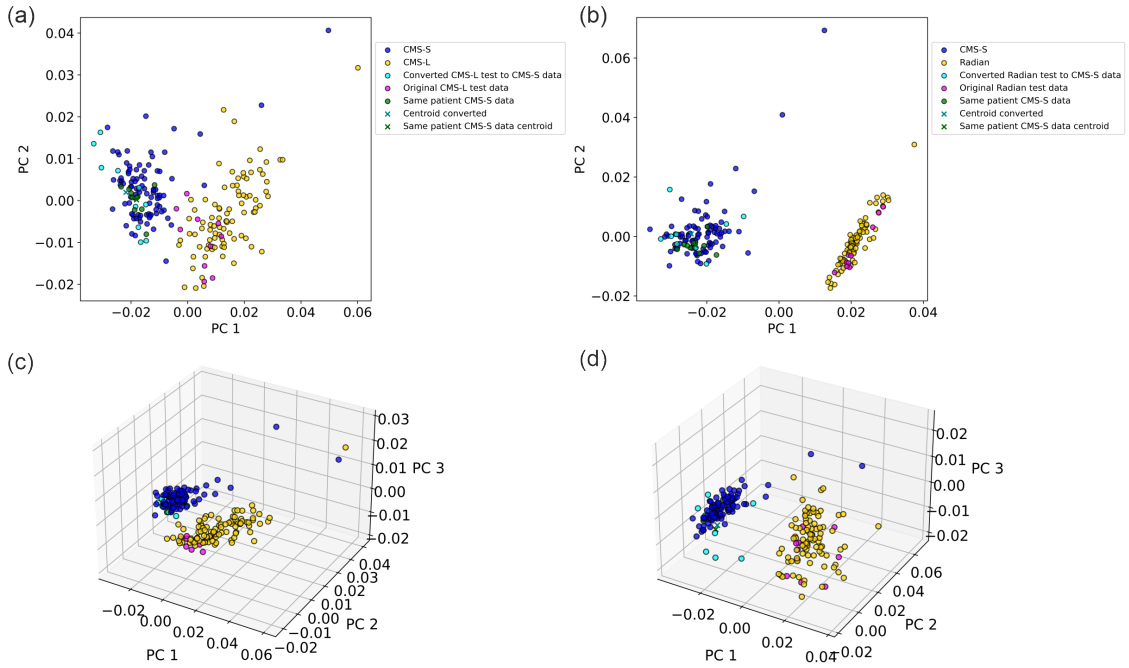

Figure 6.9 Plots of instrument spectra transformed using PCA, showing the 2D and 3D projections of PCs to determine clustering behaviour for the conversion of plasma data from one instrument (B) to another instrument (A). Plots show 2D projections of CMS-L to CMS-S converted data in (a), and Radian to CMS-S converted data in (b), and 3D projections of the same conversions in (c) and (d), respectively. Data shown is the training data recorded on instrument A (blue), training data recorded on instrument B (yellow), test data recorded on instrument A (green), test data recorded on instrument A (magenta), and the test data recorded on instrument B converted to instrument A (cyan).

The conversion of spectra recorded on the benchtop instruments to and from the QToF spectra was generally unsuccessful. The complexity of this instrument is greater than that of the simpler single-quadrupole benchtop instruments, and these differences are not well accounted for in a simple function. Also, the high-resolution spectra recorded with the QTof instrument were converted to a 'summed unit m/z' format so that they could be compared with spectra from the other instruments, which further complicates the transformation (and may fail in the same way as the smoothing functions described previously).

This concept was explored in a pilot study, utilising a neural network optimised using Bayesian optimisation to determine if high-resolution data can be converted to low-resolution data (called down-scaling), and the opposite (called super-resolution).

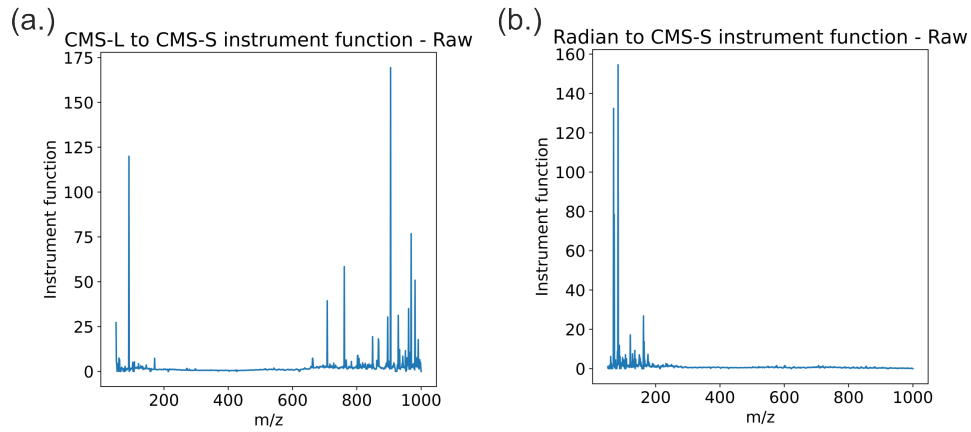

Figure 6.10 Plots show the instrument functions plotted as mass spectra. The CMS-L to CMS-S instrument function in is shown in (a), Radian to CMS-S instrument function in (b). Note the significant mass dependence of the function, which will be a result of differences in ionisation, fragmentation, and transmission of ions through the quadrupole.

Plasma spectra were generated on both the ASAP-TOF-MS and the CMS-S instrument (n=100), and this was used to generate artificial training data to build the neural network. A detailed description of this study and the results can be found in Appendix D.13. During the course of this analysis, no neural network (NN) was generated in which the accuracy of the NN on the test data was greater than 0%. It is likely that much more experimental data across a much wider range of samples would be required to generate a model that was capable of performing this transformation accurately. It was hoped that finding a method of converting between these instrument types would enable us to more accurately and confidently identify molecular peak identities in our CMS-S data using a high-resolution measurement. Unfortunately, by ASAP-TOF-MS there is still a high degree of fragmentation and many differences in the instrument setup and conditions that result in differences in the mass spectra. Tentative chemical formula assignments can be made for peaks, but without using internal standards, or correlation to the lower-resolution spectra, we cannot be certain that the same molecules are contributing to the same peaks in each case.

6.7 Conclusions and outlook

The results from this study have further given an interesting insight into the ability of different ASAP-MS instruments to effectively record representative spectra of biological samples. The overall analysis points to the Radian instrument generating more reproducible results than the CMS-S instrument, and the much faster instrument setup gives potential advantages for clinical applications, or when analysing large numbers of samples in conducting high-throughput analysis. The Advion instrument software was thought to be more user-friendly, but the relatively larger variability in the CMS-S data will increase the errors experienced during data analysis. Converting data between instrument formats using an instrument function was shown to be successful when analysed using PCA for some conversions between spectra recorded on bench-top instruments. These results offer the promise for data to be recorded on different instruments and combined to generate a single dataset, which offers the potential to expand the quantity of data available for model training considerably. More work is needed to generate an instrument function that is generalised across sample types and is robust and accurate enough to use in ML applications; this could be conducted by collecting larger datasets of patients on two or more instruments and training classification algorithms on a mixed dataset. The QTOF instrument was shown to have very low error when compared to the simpler ASAP-MS instruments, but the instrument functions were not able to convert data between low and high-resolution instrument types. A more complex function may be able to complete such a conversion, but would require a larger dataset. The potential to convert low-resolution data to the higher-resolution instruments may allow for ion identities to be found.

7

Investigation of the effect of normalisation, scaling and transformation methods

7.1 Introduction

Ambient ionisation mass spectrometry data sets are prone to high error in the mass spectral intensity, which is mainly due to the variable amount of sample introduced into the instrument on the capillary tip each time. This makes absolute quantitation between repeat measurements unfeasible without employing the use of standards or normalisation techniques. Spiking samples with a standard is a possible method of introducing an internal concentration standard, but this approach is not suitable for some samples, such as solid samples, and makes application of the method within a surgical environment challenging. Post-measurement methods applied in the data processing stage that are worth investigating include normalisation methods, scaling methods, and transformation methods. A range of normalisation methods are available for reducing the impact of intra and inter-sample variation on a data set. For MS based -omics research, normalisation methods have been investigated for improving the biological information content and reducing the impact of error and

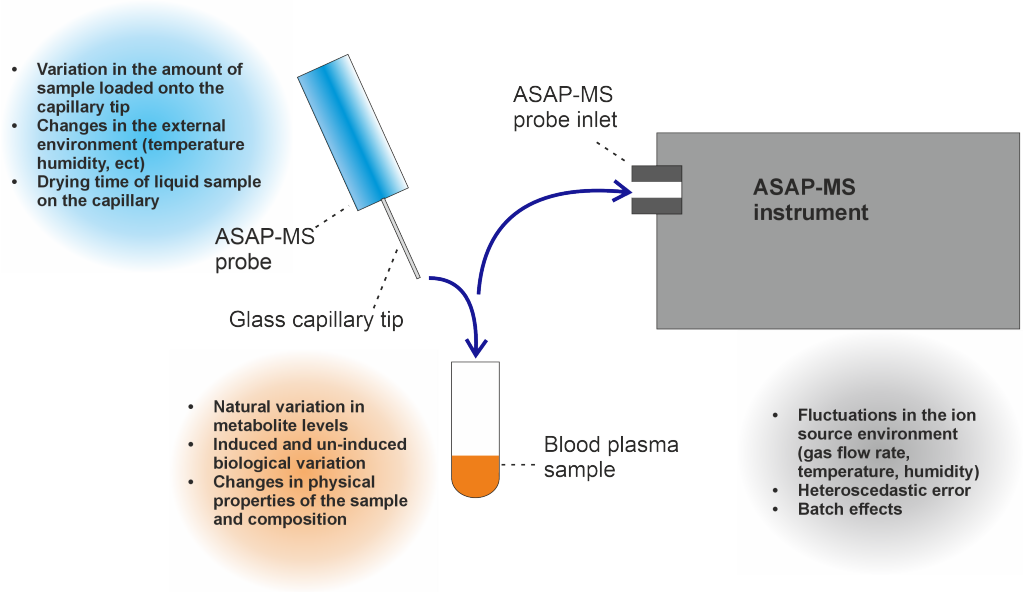

Figure 7.1 Identified sources of error that may occur in the ASAP-MS measurement of human blood samples, grouped into biological sample-related error (orange), sample loading-related error (blue), and ASAP instrument-related error (grey).

noise in a data set [208]. A summary of these sources of error, grouped into biological sample related error (orange), sample loading related error (blue), and ASAP instrument related error (grey), is shown in Figure 7.1.

The approaches to be investigated can be classified as either **normalisation methods**, **scaling methods**, or **transformation methods**. The definitions of these groups are ambiguous and often overlap. In the present work, they will be defined as follows.

- **Normalisation methods** – *methods that alter intra-sample variation by changing the range of the data obtained across all features in a single sample.* In this application, this corrects for the different amount of sample introduced on each measurement by scaling the intensity values within a mass spectrum for a sample by a suitable scaling factor determined from that mass spectrum (e.g. total intensity, mean or median peak height), such that different spectra can be compared or averaged. Normalisation methods treat each mass spectrum in isolation, and the absolute peak intensities within a mass spectrum are

changed, but relative intensities are preserved (see Section 7.3.1 for details).

- **Scaling methods** – *methods that impact inter-sample variation by changing the range of the data obtained for a single feature across different samples.* This is similar to normalisation, but uses a scaling factor determined from multiple mass spectra (see Section 7.3.2 for details). Individual mass spectra are not related in isolation but scaled as a data set, and as in the normalisation methods, absolute peak intensities within a mass spectrum are changed, but relative intensities are preserved.

- **Transform methods** – *methods that impact the shape of the whole data set to produce features that are more normally distributed.* This is done by applying a function to the peak intensities in each mass spectrum, changing both the absolute and relative intensities of the peaks (see Section 7.3.3 for details). Transformation methods also treat each mass spectrum in isolation.

A literature review of this topic found that there is no 'gold standard' normalisation method, and recommended that researchers should assess a range of methods to determine which best suits their data [210, 211, 335–337]. Conducting a literature review using the *Web of Science* search tool (Clarivate Web of Science, Copyright Clarivate 2023) of papers that featured 'atmospheric solids analysis probe mass spectrometry' from 2013 - 2023 yielded 33 papers described as related to the fields of analytical chemistry or biomedical research, 19 of which featured some form of data normalisation. Of these, 13 used relative abundance or unit mass normalisation methods (e.g. [338]), two used internal calibration peaks (e.g. [339]), three used a combination of Pareto scaling and relative abundance (e.g. [340]), and one used a combination of quantile normalisation, log transformation, and Pareto scaling ([341]). These named methods will be defined in the next section. This showed that even within the small but growing field of ASAP-MS, there is a wide range of accepted normalisation methods.

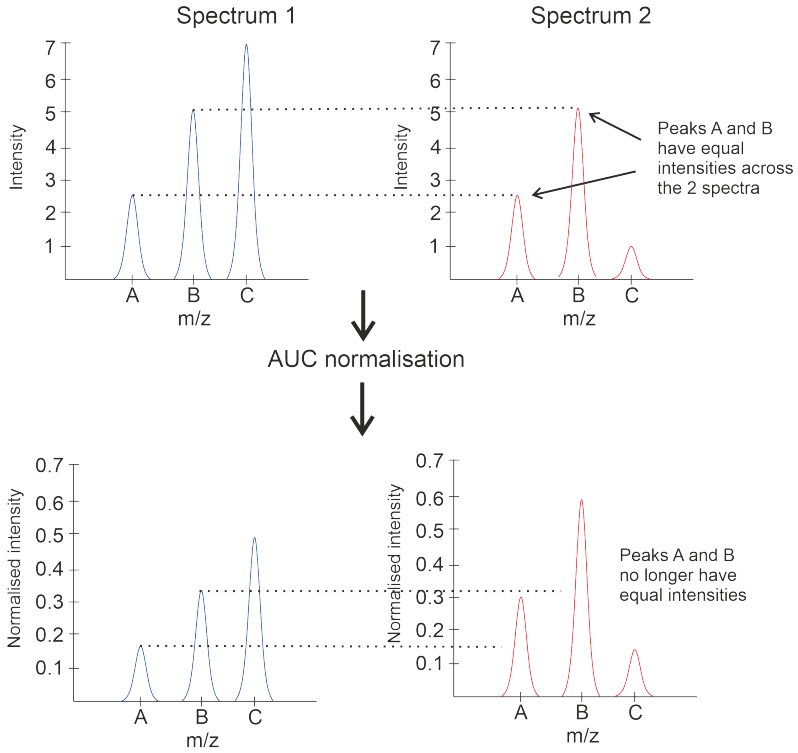

Figure 7.2 An illustration of the impact of AUC normalisation on the apparent and relative intensities of three mass peaks. The raw intensities are shown for three peaks (A, B and C) in two spectra at the top, and after the application of AUC normalisation underneath. The raw intensity of peak C changes drastically. Note the resulting change in normalised intensity of peaks A and B, even though no change in raw intensity was recorded.

One of the challenges presented by -omics data is that metabolites and biomolecules do not self-average, meaning that a concentration increase in one group of metabolites is not generally balanced by a decrease in another group [335]. Some approaches to normalisation, such as unit AUC and mean/median methods, do not account for this behaviour, and may lead to artifacts in the normalised data. An example of this is shown in Figure 7.2. The raw intensity of peak C changes from being the most intense peak in spectrum 1 to a much lower intensity in spectrum 2 in Figure 7.2(a). When AUC normalisation is applied, this results in an artificial increase in the normalised intensity of peaks A and B in spectrum 2 in Figure 7.2(b) relative to spectrum 1, even though the raw intensities do not change between the two spectra.

Normalisation, scaling and transformation methods can be combined in different

ways to produce composite methods that reduces the impact of variation in a data set in an optimal way, which may change depending on the analysis to be conducted. Open-source software packages such as Metaboanalyst provide a platform on which users can compare the effects of various normalisation methods on data sets, before performing multivariate statistics analysis with the goal of conducting data classification or regression analysis [268, 342]. However, the there is not currently a way to quantify the impact of changing methods, and the user must rely on visual changes to the results in a trail-and-error approach. The optimum normalisation approach for a particular data set can be assessed using a number of methods, including statistical methods, unsupervised clustering analysis such as PCA, and supervised ML methods such as KNN. Using a number of ASAP-MS data sets collected for complex biological samples, we have compared the effect of a large number of normalisation methods by assessing the impact of each method on the separation of classes.

7.2 Sources of error and uncertainty in ASAP-MS data

Normalisation, scaling and transformation methods are commonly employed to reduce the impact of measurement uncertainty on a biological data set. There are a large number of sources of error and uncertainty in the ASAP-MS data for biological samples, which should be understood before error reduction is employed in order to ensure that appropriate techniques are being used at the experimental design stage, the experimental measurement stage, and finally the data processing stages. A review paper written by van den Berg *et al.* describes this process well for biological MS data variation [208].

Sources of variation in a typical data set recorded for blood plasma using ASAP-MS were identified. Firstly, natural variation in the concentrations of common, abundant metabolites can be very large, and the most abundant metabolites are not

necessarily of interest to the study of a particular disease [208]. This is a common problem in untargeted metabolomics studies, as abundant molecules can obscure the measurement of less concentrated molecules. Next, there are induced and uninduced sources of biological variation. Induced biological variation describes how the biological context of each metabolite determines how much variation is observed naturally [343]. For example, primary metabolites linked to central metabolic processes will be relatively constant, but metabolites related to peripheral and secondary processes may show much larger changes [344]. Uninduced biological variation describes how a metabolite present at the same concentrations in different samples may not give the same signal response in a measurement, and may show large variations even under the same experimental conditions. The causes of this are still unclear.

Then there is technical variation in the analytical measurement, to which ambient ionisation techniques are particularly prone. Some sources of this technical variation for ASAP-MS can come from the varying amount of sample loaded onto the capillary, and from differences in the physical and chemical properties of the biological matrix such as viscosity of the blood plasma, lipid content, and water content. Many other physical factors, such as fluctuations in the carrier gas pressure and temperature, the environment of the lab, the capillary drying time, and humidity, can also add to this variation [345]. As some of these physical factors can vary on a daily basis, some of this variation may be expressed within batch effects, whereby signals can be observed to change between batches of measurements, introducing unpredictable batch-dependent structure into the data [346, 347]. Batch effects are a large part of AIMS methods, and need to be carefully considered in both experimental setup, and data analysis. The use of batch correction algorithms is a potential method of reducing the impact of batch effects, but they can also obscure patterns in the dataset. Ensuring samples are measured randomly over a large number of batches and using standard normalisation techniques, can be sufficient for this task in many cases [346–348]. A detailed study of ASAP-MS technical error, with a focus on

batch effects, was conducted within our group by Liwen Song *et al.*, and further information on this can be found in our publication on this subject [285].

Finally, there is error due to heteroscedasticity. It is often assumed that error caused by sources of variation will cause symmetric or normally centred errors, and that these will be constant across the observations, but this is rarely true. This makes it easy to introduce additional unwanted structure ('hereroscedastic error'), to the data during data processing, as many processing and normalisation methods assume normal distributions and constant error structure [208, 349].

7.3 Normalisation, scaling and transformation methods

A total of 24 normalisation, scaling, and transformation methods, listed in Table 7.1, were investigated for use with ASAP-MS data. Each method is defined further in the next section. Each method type was compared with a 'None' option, in which the data remains unchanged. These methods were applied sequentially in the order (i)

Table 7.1 Normalisation, scaling and transformation methods

Normalisation	**Scaling**	**Transformation**
None	None	None
Sum	Mean centre scaling	Log
AUC	Median centre scaling	Exponential
Mean	Auto scaling	Square
Median	Pareto scaling	Square root
Quantile	Range scaling	Cube
Vector		Cube root
Minimised vector		
Histogram bin position		
Histogram intensity		
Histogram function		

normalisation, (ii) scaling and (iii) transformation. This was done to first reduce any intra-sample variation between repeats, then to reduce inter-sample variation within features, before transforming the data distributions. Conducting scaling first may

enhance variation caused by intra-sample variation. This generated 420 different method combinations, including a blank method in which no normalisation, scaling or transformation was conducted. In the equations described below, $I_{m,S}$ is the intensity of the peak at mass m/z in spectrum S, and $(N_{m,S})$ is the normalised intensity of that same peak.

7.3.1 Normalisation methods

Sum normalisation - For a sample spectrum, calculate the sum of all the peak intensities $(\sum (I_S))$, and divide each mass peak $(I_{m,S})$ by the sum.

$$N_{m,s} = \frac{I_{m,S}}{\sum (I_S)} \tag{7.1}$$

AUC normalisation - For a sample spectrum, calculate the area under the spectrum $(\int I_S)$, and divide each mass peak $(I_{m,S})$ by the area. When the mass interval is 1 Da and the peak shape is assumed constant across masses, the AUC and Sum normalisation methods are equivalent.

$$N_{m,s} = \frac{I_{m,S}}{\int (I_S)} \tag{7.2}$$

Mean normalisation - For a sample spectrum, calculate the mean intensity signal for that sample across all the mass peaks $(\bar{I}_S)$, and divide each mass peak $(I_{m,S})$ by that mean.

$$N_{m,s} = \frac{I_{m,S}}{\bar{I}_S} \tag{7.3}$$

Median normalisation - For a sample spectrum, calculate the median intensity signal for that sample across all the mass peaks $(\tilde{I}_S)$, and divide each mass peak $(I_{m,S})$ by the median.

$$N_{m,s} = \frac{I_{m,S}}{\tilde{I}_S} \tag{7.4}$$

Quantile normalisation - Normalise by quantile distribution. The method

to complete this is as follows: Tabulate the recorded data, so that the samples are in columns, and the rows are m/z values (Table A). Rank the peak intensities for each sample. Then create a new table equivalent to Table A but with the rank reported instead of intensity (Table B). Take Table B and order the rank values in each Sample column from smallest to largest, irrespective of the m/z value (Table C). For each row in Table C, calculate an average rank, and rank the averages to generate a 'rank dictionary'. Finally, take Table B and reassign the rank values by finding the appropriate rank average from the rank dictionary [350, 351].

Vector normalisation - For each sample spectrum, calculate the square root of the sum of the squared peak intensities. Divide the intensity of each mass peak in the spectrum by this value.

$$N_{m,s} = \frac{I_{m,S}}{\sqrt{\sum (I_S)^2}} \tag{7.5}$$

Minimised vector normalisation - For each sample spectrum, S, minimise the Euclidean distance, $f(k)$, between the spectrum vector and the mean centroid vector (of all spectra regardless of label) by optimising a spectrum multiplication factor k_S. The minimisation factor is then applied to the spectrum.

$$\begin{aligned} N_{m,s} &= k_S I_{m,S} \\ f(k) &= \sqrt{\sum_{S'=1}^{n} (\bar{I}'_S - k_S I_S)^2} \\ \text{solve } \frac{df(k_S)}{dk} &= 0 \text{ to find } k_S \end{aligned} \tag{7.6}$$

Histogram distribution normalisation - The histogram normalisation methods were developed specifically for the purposes of the present study. The idea behind this was to normalise the spectra to the area under the curve of a peak height distribution. By doing this, the normalisation constant would be minimally affected by a few peaks with wildly varying intensities, but would instead be determined

mostly by the intensities of peaks with fairly reproducible intensities.

The initial set-up is to generate a histogram of the peak intensity distribution for each spectrum. The structure of the histogram was divided into segments based on percentile boundaries, and the width of each bin was then optimised (see below). Equally sized intensity bins were not used as it was observed that the peak distribution was highly skewed towards the low mass region, and many highly variable peaks were observed in the low mass region, and the higher mass region appeared very sparsely populated. It was proposed that dividing the histogram into 5 sections of variable bin numbers would represent the distribution of peaks found in a spectrum. A diagram representing the structure of the histogram setup is shown in Figure 7.3. A single bin from 0 to the lower limit percentile (P_1), initially set at 20%, represents peaks with 0 values and very low intensities and background noise. The second set of bins from P_1 to P_2 is given a flexible number of bins (n_1) initially set at 30. This represents peaks with low intensity values. The third and fourth sets of bins from $P2$ to $P3$ and P_3 to P_4 also have flexible bin numbers (n_2 and n_3) initially set at 30 each. The final region of the distribution from P_4 to 100% is represented by one bin and represents the most intense peaks that often vary considerably. The optimum number of bin segments, B, and their positions, $P\%$, and number of bins, n, were determined though a genetic algorithm that optimised for maximum in-space cluster efficiency, which is described in the next section by Equation 7.14. Three different methods were used to represent the spectra using the final optimised histogram.

i Histogram bin - Exchange each mass peak intensity with the bin number that the peak was placed into, where bin number ranges from 1 (first bin from 0% to $P1$), to Pn.

ii Histogram intensity - Represent a spectrum as a reduced dimension spectrum where the m/z axis is replaced with the bin number, and the intensity axis is replaced with the bin frequency.

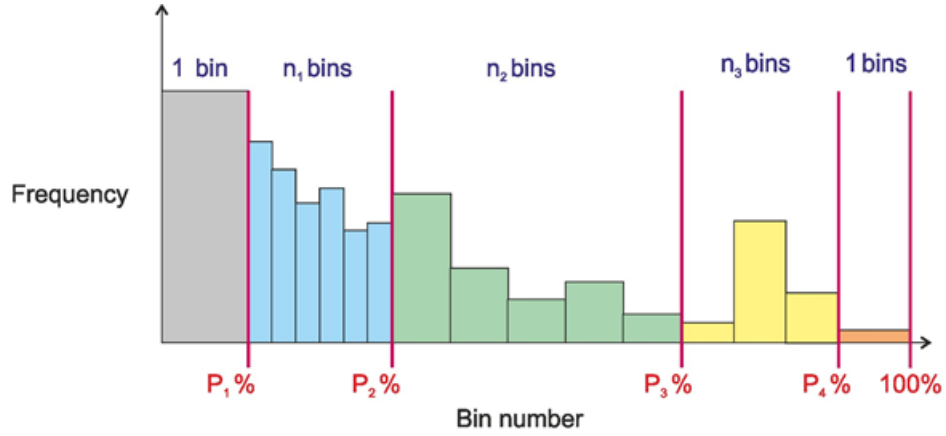

Figure 7.3 A diagram showing the histogram normalisation function, showing the distribution of histogram bins. Bin segments (B) are shown separated by the red lines at locations $P\%$, and the number of bins in each segment, n, shown by the coloured bars.

iii Histogram function - The normalised spectrum $N_{m,s}$ is represented as the multiplication factor for each peak that converts the intensity value ($I_{m,S}$) to the bin number for that peak $bin[I_{m,S}]$.

$$N_{m,s} = \frac{bin[I_{m,S}]}{I_{m,S}} \tag{7.7}$$

7.3.2 Scaling methods

Mean centre scaling - For each mass peak, calculate the mean intensity across all samples, $\bar{I_m}$, and divide each peak of that mass in each spectrum, $I_{m,S}$, by the mean.

$$N_{m,s} = \frac{I_{m,S}}{\bar{I}_m} \tag{7.8}$$

Median centre scaling - For each mass peak, calculate the median intensity across all samples, $\tilde{I_m}$, and divide each peak of that mass in each spectrum, $I_{m,S}$, by the median.

$$N_{m,s} = \frac{I_{m,S}}{\tilde{I}_m} \tag{7.9}$$

Auto scaling - As for mean centre scaling, calculate the mean intensity across

all samples, $\bar{I_m}$, and divide each peak of that mass in each spectrum, $I_{m,S}$, by the mean. The mean centre scaled value is then divided by the standard deviation of intensity for that mass across the spectra. The aim of this method is to decrease the importance of highly variable peaks relative to the stable peaks.

$$N_{m,s} = \frac{I_{m,S}/\bar{I}_m}{\sigma I_m} \tag{7.10}$$

Pareto scaling - The same method as for auto scaling, but the mean centre scaled value is divided by the square root of the standard deviation of intensity for that mass across the spectra.

$$N_{m,s} = \frac{I_{m,S}/\bar{I}_m}{\sqrt{\sigma I_m}} \tag{7.11}$$

Range scaling - For each mass peak, calculate the intensity range across all samples, $\max(I_m) - \min(I_m)$, and divide the mean centred intensity for each mass, $I_{m,S}$, by the range.

$$N_{m,s} = \frac{I_{m,S}/\bar{I}_m}{\max(I_m) - \min(I_m)} \tag{7.12}$$

7.3.3 Transformation methods

For every intensity value, $I_{m,s}$, in the data set, apply one of the six functions,

$$\begin{aligned} f(I_{m,s}) &= log(I_{m,s}), exp(I_{m,s}), (I_{m,s})^2, (I_{m,s})^3, \sqrt{(}I_{m,s}), (or)\, \sqrt[3]{(}I_{m,s}) \\ N_{m,s} &= f(I_{m,S}) \end{aligned} \tag{7.13}$$

7.4 Material and methods

The aim of this study was to determine how changing normalisation, scaling and transformation methods applied to an ASAP-MS plasma data set impacted the resulting data analysis conducted on that data. In this instance, the final analysis

to be used was either ML classification, or clustering efficiency (defined below).

ASAP-MS measurements taken for the OxAMI and OxAAA plasma samples, described in Sections 4 and 5 were used to study the normalisation methods described. The data sets used were chosen as it was known that two clear data classes should be present, for which ML classification could be conducted. The OxAMI data set has been shown previously to be classified with weak to moderate agreement, and so improvements to ML accuracy may be found using optimised normalisation methods. The OxAAA data set was shown to classify into patient treatment groups with moderate to strong agreement, and so could act as a good comparison data set.

7.4.1 Method evaluation

Two methods for evaluating the effect of the normalisation, scaling and transformation methods were used. A clustering efficiency method, and a machine learning method.

Cluster efficiency evaluation - The normalisation, scaling and transformation methods described were assessed on how well the resulting processed data sets clustered into two classes in N dimensional space. The extent of clustering was assessed by calculating the cluster size to centroid separation ratio, C_R,

$$C_R = \frac{\sqrt{\sum_{n=m/z}^{i=1} (|c_0 - c_1|)^2}}{\sigma d_1 + \sigma d_0} \quad \text{where} \quad d_n = \sqrt{\sum (|c_n - S_n|)^2} \tag{7.14}$$

where c_n is the centroid of the data points of the samples with label n (c_0 = centroid of class 0, c_1 = centroid of class 1), d_n is the average distance between each sample in class n (S_n) and the centroid of class n cluster (c_n), and σ is the standard deviation of d_n, as described by Equation 7.14. The C_R value is therefore described by the Euclidean distance (described previously in Section 3.2.2.1) between

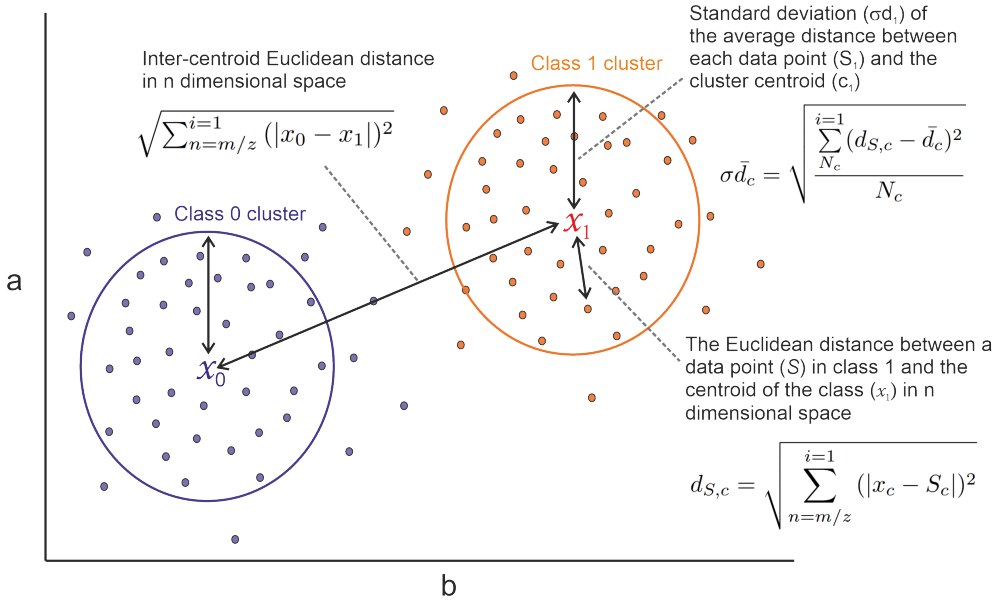

Figure 7.4 A 2d graphical representation of the cluster ratio analysis. Samples from two classes: class 0 in blue, and class 1 in orange. The two cluster centroids are shown by x_0 and x_1 respectively. Data is shown with respect to two axis, a and b, which represent two *m/z* values, but in reality the dimensional space will be much larger.

the two clusters divided by the cluster sizes represented by the standard deviation of sample to centroid distances. A larger C_R value represents a data set within which individuals within the same group are tightly clustered together, and there is a clear separation between the two clusters of the two data groups. A graphical representation of this is shown in 2D in Figure 7.4.

Machine learning evaluation - As improving ML classification accuracy is the primary aim of the investigation into normalisation, scaling and transformation methods, the effect of each method on the accuracy obtained from ML models trained on the processed data was evaluated. The normalisation, scaling and transformation methods described were assessed based on the extent of successful classification of patient variables by a supervised linear discriminant analysis (LDA) ML classification model. The accuracy of each ML model was assessed using Cohen's kappa statistic, as described previously in 3.12.

The LDA model was chosen as it was found to give reproducible and accurate predictions of clinical variables in previous analysis. Furthermore, this ML model uses

Bayes' Theorem to classify data, assuming normally distributed and homoscedastic data classes. It was proposed that this assumption would improve in correlation with increasing C_R. Four other supervised learning ML methods were investigated, the results for which were very similar and are described in Appendix E.1.

The ML anlayis was repeated over 50 different 80:20 partitions of the data, and the accuracy obtained from the summed confusion matrix across these partitions. Repeating this over multiple partitions prevents overfitting from particular selections of patients. Error was reported as the standard deviation in κ score over the partitions.

7.5 Results

7.5.1 The effect of normalisation, scaling and transformation methods on clustering efficiency

A set of AAA data was assessed for the extent of clustering, C_R, into healthy volunteers (HV) and small aneurysm patient (S1) groups after each of the normalisation methods had been applied. The colourmap table in Figure 7.6 shows the full set of results for all of the combined methods. The maximum C_R value was obtained using the Histogram function normalisation method, combined with pareto scaling and a cubic transform. The histogram function method generated high C_R values with all combinations of scaling and transformation methods. With no normalisation methods, both range scaling and autoscaling also performed well with all combinations of transform method. The AAA data used for this analysis is known to classify well by ML methods with very high accuracy ($> 95\%$) achieved. To determine if a higher C_R score is indicative of increased classification accuracy by ML methods, a comparative study was conducted using data that is more challenging to cluster.

The normalisation, scaling and transform methods were applied to each of a set of 9 clinical variables from the OXAMI data set. Method evaluation was conducted

Table 7.2 Best normalisation, scaling and transform methods for OXAMI variables assessed by cluster efficiency

	Best normalisation method		
Clinical variable	Normalisation	Transform	Scaling
Creatinine	Histogram bin	None	None
Mortality	Histogram function	log	Median centre
HF	AUC	log	Auto
IMR	Histogram bin	None	None
MDRD	Histogram intensity	cube	Range scaling
MVO	Histogram bin	None	None
ITime	Histogram bin	None	None
TScore	Histogram bin	None	None
Troponin	Histogram bin	None	None

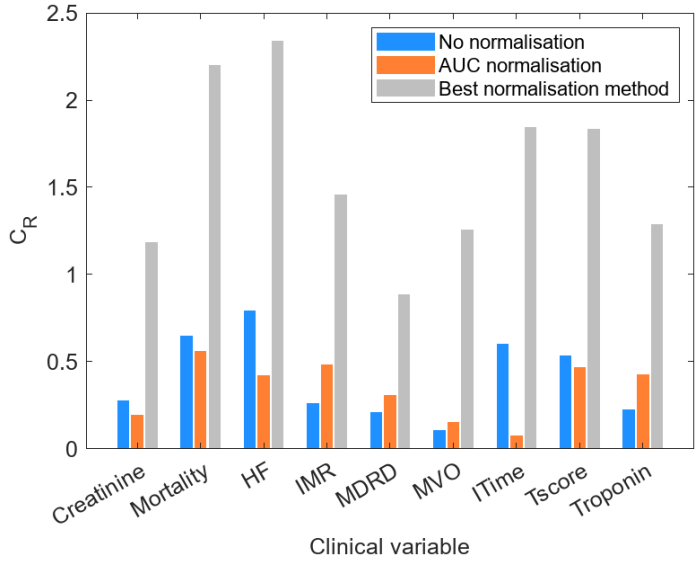

Figure 7.5 Comparison of normalisation, scaling and transform methods for OXAMI clinical variables evaluated by cluster efficiency. Cluster efficiency when no normalisation, scaling and transformation is used is shown in blue, where AUC normalisation with no other scaling or transformations is shown in orange, and the best normalisation, scaling and transformation combined method is shown in grey

using the cluster efficiency measurement. The methods that gave the highest C_R values are highlighted in Table 7.2. The most common high performing methods combination was Histogram bin normalisation, no transformation and no scaling. A comparison of no methods, the previously used AUC method, and the best performing method, is shown for each clinical variable in Figure 7.5. In all cases, clustering was greatly improved in the best performing method.

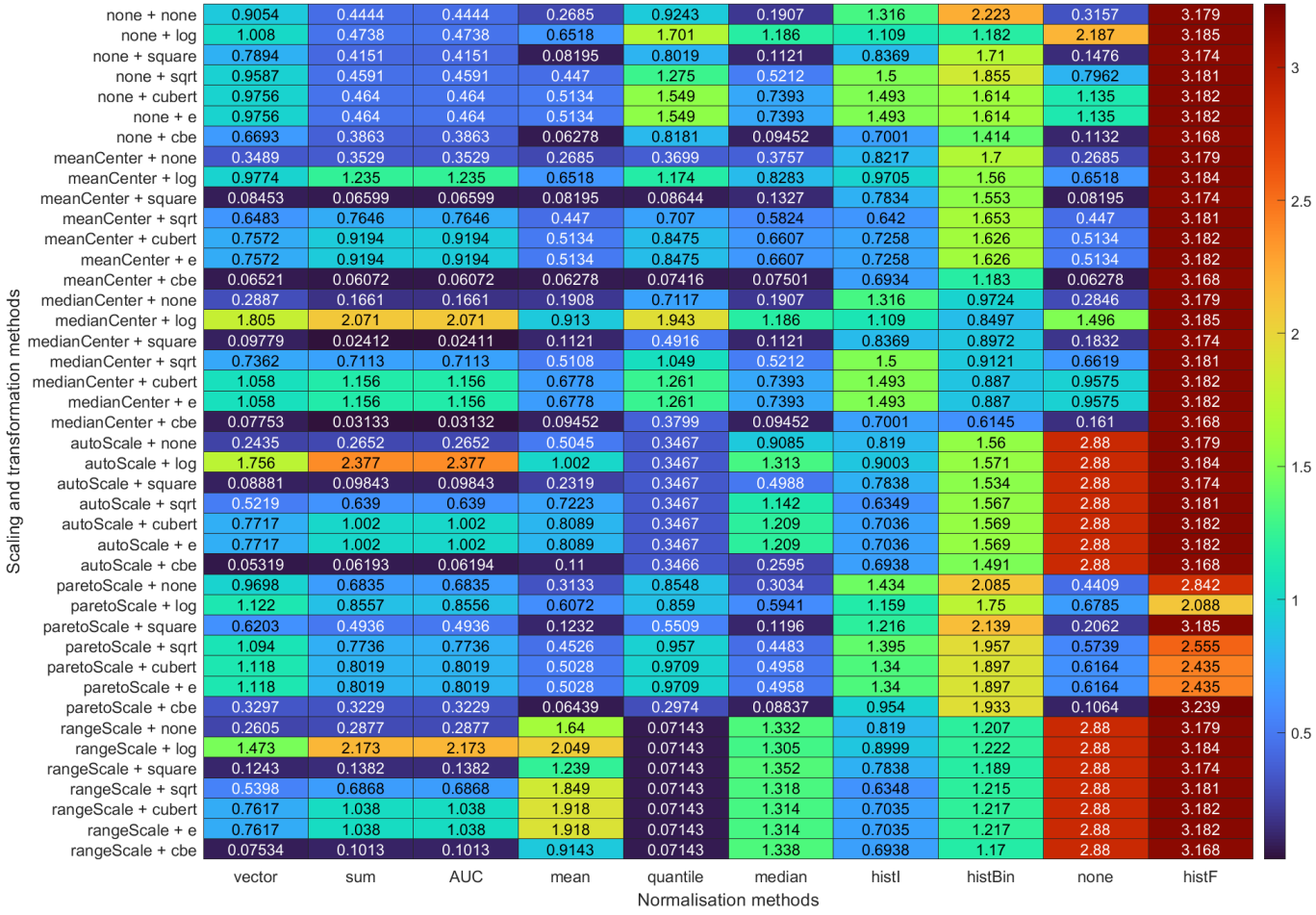

Figure 7.6 The cluster size to centroid ratio values for the OxAAA data clustered into HV and S1 patient groups. The colourmap represents the C_R value from low (blue) to red (high).

7.5.2 The effect of normalisation, scaling and transformation methods on machine learning classification

The classification of AAA patients into small aneurysm and HV groups, and STEMI patients into the nine chosen clinical variables, was conducted using the normalisation methods described. Figure 7.7 shows the κ scores obtained for each of the five ML algorithms used to classify the OxAAA data into 'healthy volunteer' and 'small aneurysm' categories. The figure shows the results obtained with no data normalisation, with AUC normalisation, and using the combined normalisation, scaling, and transformation method that led to the best performance of the relevant ML algorithm. The best-performing methods for each ML algorithm are listed in Table 7.3.

The results show that AUC normalisation leads to better classification than no normalisation, and that optimisation of the normalisation, scaling, and transformation methods leads to further improvement in the classification performance of all ML models. With this optimisation, the RFC classifier had the lowest (but still very good) accuracy of 93% (κ 0.9), while all other ML methods yielded accuracies of 96% (κ 0.925) or higher. No clear single method was found to yield the best performance over multiple ML methods, though we note that there is more variation amongst the best-peforming transformation methods than the best-performing normalisation and scaling methods.

Table 7.4 shows the model used to obtain the highest accuracy model for each of the STEMI clinical variables. Normalisation methods that performed well for this data were mean, median, vector, and AUC normalisation. The scaling methods that performed well were mean centre scaling, median centre scaling, and auto scaling. Many transformation methods were found to be effective, but were highly data set-specific. Unlike for the clustering evaluation method, no single combination of methods performed the best consistently.

The comparison of no normalisation, AUC normalisation and the best normalisa-

Table 7.3 The best-performing normalisation, scaling, and transformation methods for ML classification of OxAAA spectra into small aneurysm and HV categories.

ML method	Normalisation	Scaling	Transformation	Accuracy
KNN	None	None	Cube root	96%
SVM	AUC	None	Square root	98%
LDA	AUC	Mean centre	Cube	98%
NBC	Quantile	None	Exp	97%
RFC	AUC	None	Square	93%

Table 7.4 The best performing normalisation, scaling and transform methods for OXAMI variables assessed by LDA ML classification accuracy

Clinical variable	Best normalisation method			% Accuracy range
	Normalisation	Scaling	Transform	
Creatinine	Mean	None	Log	<60%
Mortality	Vector	Mean centre	None	80-90%
HF	Min vector	Mean centre	Log	80-90%
IMR	AUC	Mean centre	Square-root	70-80%
MDRD	Median	Median centre	Cube-root	60-70%
MVO	Mean	Auto	Cube-root	<60%
ITime	Vector	Mean centre	None	70-80%
TScore	None	None	None	70-80%
Troponin	Median	None	Cube-root	60-70%

tion method combination for each clinical variable is shown in Figure 7.8. Improvements in ML accuracy were obtained for each variable, with the greatest improvements observed for HF and MVO when compared to the previous AUC method. Methods that previously showed a high clustering efficiency, such as the histogram-based methods, did not correspond to high ML clustering accuracy. The change in κ score from 'no normalisation' to the best performing method is low, from 0.1-0.2 κ in most cases, which would be within a single κ interpretation band, as described by McHugh in [267]. In addition, for the Thrombus score variable, the no-normalisation method produced the highest accuracy model. The impact of normalisation methods on ML can therefore be said to be relatively low. The results of ML analysis using the four other ML algorithms are described in Appendix E.1.

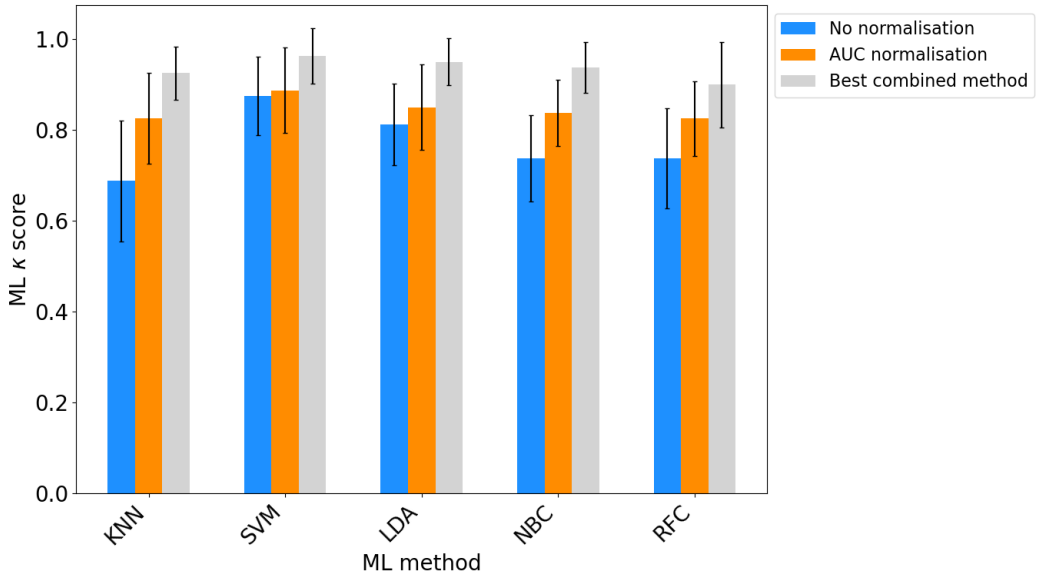

Figure 7.7 Comparison of normalisation, scaling and transform methods for AAA S1 vs HV classification evaluated by κ score. The ML clustering κ score using the five different ML models when no normalisation, scaling and transformation is used is shown in blue, where AUC normalisation with no other scaling or transformations is shown in orange, and the best normalisation, scaling and transformation combined method is shown in grey. Error bars show the standard deviation in κ over 50 partitions.

7.6 Comparison of normalisation assessment methods

The two different methods used to evaluate the various processed data sets were compared to determine whether maximising cluster efficiency is an appropriate method of improving ML classification accuracy. The ML and C_R results obtained for the classification of patient groups was compared. Scatter plots showing this comparison for S1-HV AAA patient classification and patient mortality is shown in Figure 7.9. No correlation was observed between the C_R values determined in the clustering evaluation and kappa values determined from the ML evaluation in any of the clinical classifications attempted across both studies. This suggests that the overall location of a point in vector space is not a defining part of the ML classification success for this data set. As shown by the high ML κ score, and in previous results for

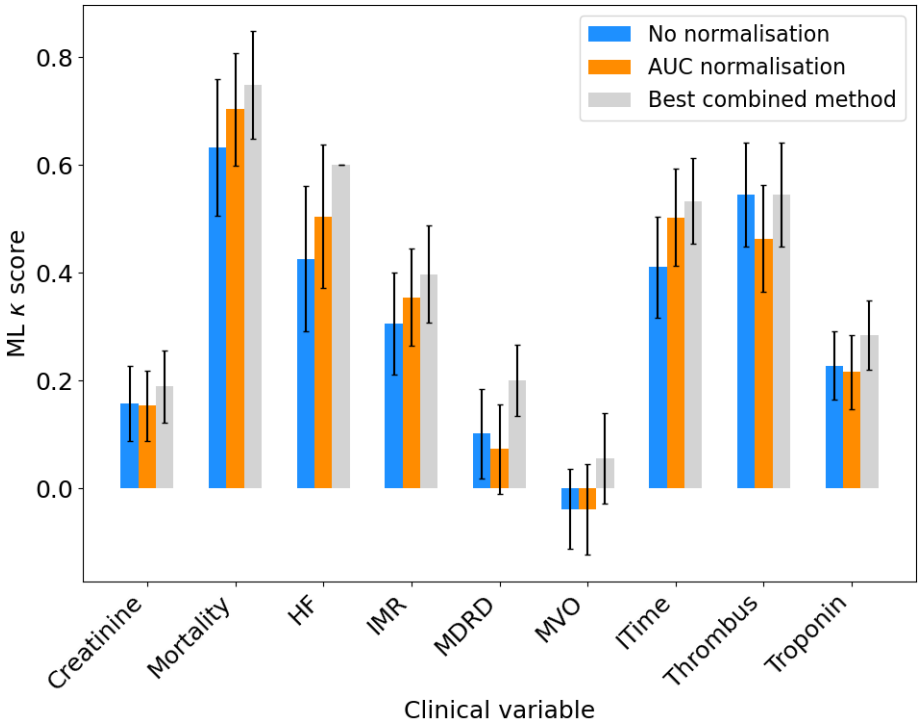

Figure 7.8 Comparison of normalisation, scaling and transform methods for OXAMI clinical variables evaluated by κ score. The ML clustering κ score using an LDA model when no normalisation, scaling and transformation is used is shown in blue, where AUC normalisation with no other scaling or transformations is shown in orange, and the best normalisation, scaling and transformation combined method is shown in grey. Error bars show the standard deviation in κ over 50 partitions.

these data, the ML vector-based models are able to successfully classify this data. The spatial clustering of this data is therefore shown to be peak-specific, rather than dependent on the entire vector space. This is in agreement with the success of previous feature reduction analysis.

7.6.1 Batch effect correction

As discussed previously, batch effects may be present in the dataset that could impact the results of clustering and classification analysis. A number of methods of batch effect correction were also investigated, in line with work reported by L. Song *et al.* [285]. Two commonly available batch correction algorithms were applied to a data set before classification was attempted: the ComBat and ICA algorithms

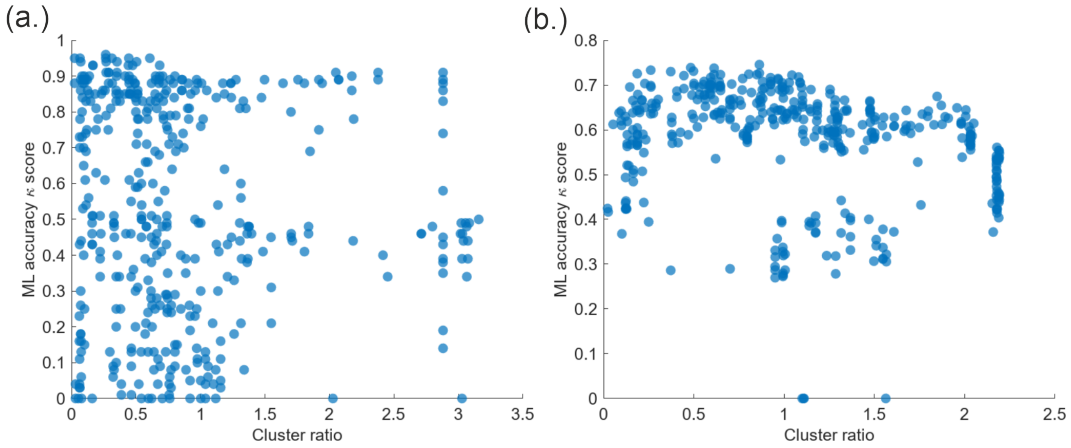

Figure 7.9 Comparison of normalisation evaluation methods for AAA S1 and HV patient classification (a) and patient mortality (b). The cluster ratio is plotted against ML accuracy as κ score. No correlation was observed in either case.

[352, 353]. The ComBat algorithm uses an empirical Bayes framework to adjust for batch-to-batch variability, specifically by estimating and removing batch-specific effects such as shifts in mean and scale by modelling them as covariates, then applying Bayesian shrinkage to stabilise the estimates [352, 354]. The ICA algorithm tries to decompose the data into statistically independent components. Assuming that batch effects manifest as distinct, independent signals within the data, components associated with batch effects are identified by correlation with batch label and removed or adjusted [353]. These methods were applied to both the OxAMI and OxAAA plasma data sets prior to classification by ML methods. Unfortunately, classification accuracy decreased substantially in all cases on application of these batch correction algorithms, indicating that the algorithms had removed or masked features of the measured data required to classify the biological data. As the plasma samples were measured randomly distributed across many batches (with the exception of some HV samples described in Chapter 5), and statistically significant differences in intensities were found to disappear after batch correction, it was determined that these batch correction methods were too severe for these data and that batch-related error was potentially too low to negatively impact these results. To reinforce this hypothesis, when each data set was visualised in PCA space

coloured by batch, no significant batch clustering or separation was observed. For higher throughput analysis, this may have a much larger effect.

7.7 Conclusions

Having investigated a large number of approaches to data normalisation, the impact of changing normalisation methods was found to be minimal, and it was notable that the ML models performed with high accuracy even on un-normalised data. This highlights the strong underlying patterns within the data sets. Additionally, no correlation was found between clustering efficiency ratio and ML accuracy scores: high clustering efficiency did not indicate a higher kappa score for a given ML method. This was unexpected, as a greater inter-centroid distance was assumed to enhance ML accuracy. However, this likely reflects the complexity of the data, where not all peaks contribute equally to ML performance—an aspect that clustering efficiency does not capture. No single combination of normalisation, scaling, and transformation methods proved universally optimal, with the most effective approach varying by data set. AUC normalisation performed well in most cases based on ML kappa scores, and was deemed suitable for standardising ASAP-MS data across data sets. Furthermore, data transformations were found to be beneficial in many instances, suggesting their potential for use in fine-tuning ML algorithms to enhance classification accuracy for specific analyses once some initial analyses have been completed.

8

Conclusions, evaluation and future work

The results of the studies presented in the preceding chapters have shown that ASAP-MS can be used to conduct untargeted metabolomic fingerprinting of clinical samples, and that machine learning methods can be employed to predict clinically relevant information when trained on the ASAP-MS data sets. Proof of concept was established with respect to both clinical study groups analysed.

For STEMI patients enrolled in the OxAMI study, we were able to predict six different clinical metrics related to patient risk with over 76% accuracy, with patient mortality and MVO predicted with over 80% accuracy. Given the relatively high uncertainty associated with both the ASAP-MS measurements and some of the more subjective or qualitative clinical measurements, improvements may be possible. We also note the low patient numbers and uncertainty over the quality and reproducibility of the coronary aspirate samples employed in the proof-of-concept work. Next steps for this study will focus on expanding the patient cohort and exploring different sample types, including plasma from different locations in the heart, venous plasma, and potentially whole blood.

Symbolic regression analysis was useful for predicting relationships between a subset of mass peaks and clinical variables, and would be interesting to explore further. Although there were some challenges in using symbolic regression analysis, particularly with the tendency of models to overfit, evaluating the results of this method was achievable by strict analysis of the performance of both training and

test data via a 'human-in-the-loop' approach.

For AAA patients, the prediction of the various patient groups based on aneurysm presence and size, and pre/post-surgical status, was extremely accurate, with a particular highlight being the separation of 'small AAA' and 'large AAA' groups with over 93% accuracy. This result implies a strong relationship between aneurysm size and detectable circulating biological markers. The prediction of AAA growth rate and FMD was more challenging, but the use of symbolic regression proved useful in this application, with the identification of some weak correlations with a subset of mass peaks in the data set that could be explored further.

A number of potentially interesting mass peaks were identified within the analysis of the AAA data set, including metabolite peaks that correspond to those identified in literature reports as being involved in AAA progression. Identifying these molecules, and their relationship to the overall disease pathology, will require access to further samples and to instrumentation capable of conducting high-resolution metabolomics analysis. However, peak identification is not a feature of the ASAP-MS method, and should not prohibit patient metabolic fingerprints from being classified.

In the work described in Chapter 6, several commercially available ASAP-MS instruments were evaluated to assess their relative suitability for human plasma analysis. Certain features of each instrument were found to enhance the usability and improve the reproducibility in plasma measurements, particularly in the case of the Waters Radian ASAP. However, due to the short loan period for this instrument and clinical study constraints, we were unable to record sufficient data to evaluate how these improvements influence the prediction of clinical outcomes, which could be addressed in future research. Data conversion between instruments appears feasible through the use of instrument functions, provided the configurations of the instruments are sufficiently similar (e.g. the CMS instruments and Radian). The Waters ASAP-Q-TOF instrument, despite offering higher mass resolution and improved reproducibility, did not present a clear advantage over more compact systems due to

its high cost and difficulty of use. The inability of this instrument to resolve peak identities without the use of internal standards and the substantial increase in data volume over the quadrupole-based instruments, given the limited sample size, may hinder its practicality for machine learning applications in this context.

In Chapter 7, a large number of methods for normalisation, scaling and transformation of the mass spectra were also investigated, and assessed via their impact on ML classification accuracy. No method, or combination of methods, was found to give significantly better results than other methods, in agreement with previous findings reported in the literature. It is worth noting that in many cases, the unnormalised data was also well classified.

We believe that this work has shown that ASAP-MS has a place in conducting rapid fingerprinting of clinical samples related to different diseases, and that machine learning techniques provide a rapid and interpretable method of analysing the resulting data in a way that will be useful in a clinical setting. We have investigated different methods by which this analysis can be optimised, and recognise that considerable further work is required before this can be translated to a clinical setting.

8.1 Method evaluation

In the following, we explore various ways in which our approach can be improved in the future, challenges and opportunities for further development, as well as considering alternative measurement techniques worth exploring.

8.1.1 Challenges and opportunities

In common with all forms of MS, ASAP-MS signals are affected by matrix effects (changes in analyte signal due to the presence of other compounds in a particular sample matrix), ion suppression (reduction in signal of an analyte due to the presence of interfering co-eluting compounds), and selective ionisation (favouring

the ionisation of molecules with particular physical properties) [355]. As a result, the mass spectra of biological samples are only a partial representation, or 'fingerprint', of the whole metabolic content of the sample, rather than a fully quantitative measurement.

Choosing to classify patients based on a semi-qualitative fingerprint is currently not a uniformly accepted technique by the medical field outside of research. Many medical tests have strict limits of detection and acceptable ranges for measurements of specific, identifiable biomarkers. This is due in part to the historic, and current, dependence on LC-MS, and has provided a safe and standardised way to develop and conduct clinical testing for several decades. New techniques based on pattern recognition and clustering, commonly employed by AIMS methods to classify samples with very high accuracy, often do not adhere to these requirements. Therefore, without metabolite identifications, many AIMS clinical tools are failing to be approved for use in UK, EU and US markets. This is less of a problem in less highly regulated fields, such as detection of explosives, detection of drugs and counterfeit drugs, detecting pollutants in the environment, and food science, where untargeted classification techniques that use AIMS methods are being developed and approved [356]. New protocols and regulations that cover the use of untargeted and qualitative clustering rather than quantitative measurement will need to be developed [355]. This topic was discussed in some detail by the AIMS community, led by Zoltan Takats, at the British Mass Spectrometry Society Ambient Ionisation Special Interest Group (BMSS AI SIG) 20th anniversary celebration in 2024 [3]. The development of new regulations will hopefully also coincide with further improvements to AIMS methods, quantification, and reproducibility. The use of software and AI in medical devices has its own separate (and rapidly evolving) regulations, which further complicates the development of these tools [357, 358].

Other challenges and opportunities have been identified that are more specific to ASAP-MS.

Sample volumes - At present, there is low reproducibility in the sample volume introduced to the MS on the capillary. A potential method around this would be to use open capillaries injected with a known volume of sample, but this adds a further preparation step, and has not yet been tested for biological samples. Work in this area is being advanced by scientists at Astra Zeneca, but is yet to be published [359].

Automation - ASAP-MS requires manual introduction of samples, which makes conducting larger scale studies very time and labour intensive. In the studies discussed here, we identified that it would be advantageous to expand the data sets to include samples from a larger number of patients, which would be very challenging using the current setup. Automated methods have been developed for other MS based clinical tools [360–363]. An automated method involving a robotic sample delivery system, either using ASAP-MS or a different AIMS technology (e.g.. DESI), and automated interaction with the instrument software would be beneficial.

Ion fragmentation and peak identification - During ASAP-MS ionisation, as in most other forms of ionisation, fragmentation of ions can occur. Without separation techniques such as chromatography, or tandem MS approaches to allow ion selection, this complicates the observed spectra, and makes ion peak identification a further challenge. This was demonstrated for aspirin in [170], and in our own work, where we have rarely seen a single molecule produce only a single MS peak. When trying to identify ion peaks and assign them to individual metabolites, this needs to be borne in mind: one metabolite may give rise to multiple ion peaks, which may or may not include a parent ion, and a single ion peak may be generated in multiple processes by different parent ions. The single-quadrupole benchtop ASAP-MS instruments have low resolution, which amplifies the problem of identifying metabolite peaks. In this study, data is binned to unit m/z or 'mass window', as done in most other AIMS studies. This will reduce the data represented in the spectra [364].

Analysis of lipids - Another challenge for conducting untargeted metabolomics analysis with ASAP-MS is the analysis of lipids. Lipids have a large amount of

structural isomerism, and are oligomeric, resulting in different lipids frequently dissociating into common fragments. This makes identifying lipids by ASAP-MS alone unfeasible, yet lipids are known to be highly important in many disease states, and many of the important molecules identified as being significant for classification appear in regions of the spectra that are populated by lipid peaks. Even when a specific parent precursor ion can be selected for, techniques such as ion mobility MS and labelled lipid standards are usually utilised to distinguish between these highly isomeric and isobaric lipid species. Without these more complex and high-resolution techniques, only the molecular formula can be distinguished, without being able to access other levels of the lipid annotation hierarchy [337, 365]. Analysis into the effect of ASAP-MS ionisation of lipid molecules is ongoing, but was not discussed in this thesis.

Quality control samples - The data collected for the STEMI and AAA studies did not feature control samples for use in comparing samples, normalisation, or mitigating uncertainty and batch effects. Control samples are commonly used in analytical measurements to normalise the data and correct for batch-dependent variation in the data. Introduction of quality control (QC) samples and assessment of their utility in ASAP-MS measurements will be an important step for future experiments. The majority of ASAP-MS publications do not discuss the use of QC samples, and so this could be beneficial for the wider ASAP-MS research community [366]. Pooled human plasma samples obtained from collections such as biobanks would be suitable QC samples for most of the work described in this thesis. A set of experiments to investigate this are discussed in Section 8.3. Also of interest would be biological samples from pathology-matched controls, which are patients with similar disease states to that being studied. For example, samples from patients with aneurysms which are not located in the abdominal aorta. This would allow fingerprints to be identified highlighting metabolites related to the specific pathology, and potentially reduce false positive rates.

8.2 Comparison with alternative analytical techniques

A number of other AIMS based techniques have already been developed for use in clinical settings, and have been of interest to the AIMS community for some time [355]. For example, a number of mass-spectrometry based surgical probes such as the iKnife (using REIMS) [367], MassSpec pen (using DESI) [368], and SpiderMass (using RIR-LA REIMS) [362], are being evaluated for use as tissue-analysis methods that can be used surgically to perform tasks such as cancer diagnosis and tumour margin identification. These techniques utilise multivariate statistics and machine learning within software packages that give the surgical team a rapid result, usually using a colour coded traffic light system. Robotic systems have also been integrated into the workflows, allowing highly accurate and rapid screening, and in some cases even imaging, of a tissue area [360–362, 369]. The ASAP-MS method differs from these techniques as it does not involve the use of tissue electrocautery (as used in the iKnife), which will impact the molecules found in the samples. DESI based techniques (such as the MassSpec pen) require the use of a cumbersome probe which requires a constant supply of solvent (water) and is not as easily applied to both solid and liquid samples. The ASAP-MS method requires a simpler handheld probe which can be autoclaved, or even made single-use, which may be a potential improvement in clinical settings, and so could be used in parallel.

Paper-spray ionisation is another alternative AIMS technique that is being developed for use with liquid biological samples [363, 370, 371]. Paper-spray ionisation uses ESI, which will again produce a very different spectrum to that produced by ASAP. A limitation of this method is that the paper used to introduce the sample can impact the spectrum observed, and is potentially more challenging to regulate and sterilise than the borosilicate glass used in ASAP. This may lead to increased error and variation in the spectra observed. However, it has been shown that this technique can be automated, and given the similarities of this technique to ASAP

(the sample is introduced on the tip of a substrate), this may provide useful direction for the future automation of ASAP techniques [363]. An improvement on the paper-spray method is coated blade spray ionisation, in which metal probes coated with a sorbent coating are used to transfer samples for electrospray ionisation [372, 373]. Again, this utilises a different method of ionisation to ASAP-MS, which will result in differences in the observed spectrum.

A large number of other ambient ionisation techniques have been developed recently, summaries of which can be found in review articles [2, 138, 139]. Many of these methods are being investigated for biological applications, with each method bringing a range of advantages and disadvantages when compared with other techniques. It is likely that no single method will be the best suited for all clinical applications. Information gained from the study of one method may well aid the development of a better tool or protocol using another technique. Therefore, it is not argued here that ASAP-MS is necessarily the best technique to use for a surgical probe. However, it has been shown to be a simple, rapid, method that can detect pathologically related metabolic changes in patients with cardiovascular diseases. This information may help to advance the field of applied AIMS.

8.3 Future experiments

A number of further experiments have been identified throughout this thesis that would advance the conclusions drawn from this work. A key requirement that was identified repeatedly is the need for increased numbers of both clinical samples and control samples.

8.3.1 Measurement of healthy human blood

We aim to acquire a large number ($n > 1000$) of frozen standard human venous plasma from healthy volunteers of mixed age and gender, likely from a human biobank. Analysis of these samples would be completed using ASAP-MS, in par-

allel with metabolomics and lipidomics assays by UHPLC-MS/MS, and GC-MS if possible. This data would provide us with data that could be used to explore the following questions:

1. Can we identify molecular peaks in ASAP-MS spectra by correlating this data to other high-resolution analytical measurements? Could we use machine learning techniques to assist with this?

2. Can we determine what variation occurs in ASAP-MS spectra due to non-pathological human variation (eg. age, gender) in human plasma? Can we quantify this variation?

3. Can we use healthy plasma as a control sample during ASAP-MS experiments? How many controls need to be run during an experiment, and how does this affect the data collected? How can we use these data to normalise with respect to human variation and measurement batches in an automated way?

4. Do we need to spike samples or introduce standards in order to improve the ability to identify molecular peaks or conduct normalisation?

This is probably the most important project that needs to occur before further ASAP clinical applications can be explored.

8.3.2 Measurement of whole and dried blood by ASAP-MS

In a surgical setting, it is likely that whole blood would be used instead of blood plasma. Whole blood analysis presents new challenges for ASAP ionisation, as this matrix is less homogenous and contains larger structures such as red blood cells that may disrupt the collection of clean mass spectra. We have also observed that tissue samples contaminated with whole blood produce complex spectra and tend to burn on the capillary tip of the ASAP probe, leading to poor chromatogram signals. A study should be conducted into whether whole blood can be used in

the place of blood plasma for ASAP-MS clinical studies. Potentially, whole blood could be collected from finger-prick blood samples, instead of invasive blood draws. This may make the ethics of approving such a study more feasible, and improve patient retention through using less invasive techniques, but would require significant method development and optimisation. Another option would be to use dried blood spots, also taken from fingerpick samples, similar to that described by Orlowsky et al [179]. Again, this would require the development of a protocol for the analysis of whole blood by ASAP-MS.

8.3.3 Measurement of venous blood samples from OxAMI patients

As the study using ASAP-MS on coronary aspirate blood plasma for STEMI patients has yielded some interesting conclusions, repeating this study on the same patients using blood samples from different regions of the circulation may help us to understand why this analysis was successful, and potentially improve the results. Samples of venous blood would allow us to determine whether metabolic changes can be detected systemically, allowing blood to be used from a standard venous blood draw. Analysis of paired coronary artery and coronary sinus blood (i.e. before and after passage through the affected artery) would allow us to understand chemical changes to the blood that occurred within the region of the heart that is ischaemic and impacted by the infarct.

8.3.4 Measurement of samples from a larger AAA patient cohort with increased growth rate monitoring data

The successful prediction of AAA growth rate is a highly sought-after capability. Using ASAP-MS data and symbolic regression analysis, we were able to predict this with a weak correlation. To test these results and potentially improve them, a much larger set of AAA patient samples would be needed, with an improved distribution of

growth rate data compared to the highly skewed set that was analysed in the present study. It would also be beneficial to have multiple time points over which growth rate was measured for each patient, given that it is known that growth rate is not constant over time in most patients. The OxAAA study has recently collaborated with multiple other AAA studies around the world, and access to thousands of plasma samples for AAA patients, with the accompanying clinical data, may be possible.

8.3.5 Measurement automation

The greatest bottleneck in our current method is the manual measurement of samples. Automating the measurement protocol will be necessary for the analysis of the increased patient data sets that have been alluded to in the previous sections. For ASAP ionisation, this will prove challenging, given that the instrument is designed around a handheld probe, and that this is one of the key features that makes it potentially ideal for a surgical setting. Utilising other AIMS techniques that have already been automated, such as DESI, may be an easier way of achieving measurements at a greater scale. Given that we have shown that ASAP-MS data can be converted to allow mixed-instrument measurement batches, it may be possible to develop models that can accept both ASAP and DESI biological measurements for classification.

8.3.6 Development of a single measure of patient risk for use in a clinical setting

The final stage of development in relation to a given disease would be to create a tool that is suitable for use in a clinical setting. Taking the prediction of STEMI patient risk as an example, there are still a number of considerations to address in relation to developing a suitable indicator of patient risk. A risk score could come from a composite of some of the variables predicted within this study, but would need some

clinical input to determine. Publication of a model trained to complete the data analysis on new samples to predict this variable would be required, within a software package suitable for use by non-specialists. The instrument handling, cleaning, and operation would need to be automated or simplified to make it straightforward for clinical staff or technicians to complete quickly; this would be easier to address with the Waters Radian instrument than the current Advion CMS-S, as discussed in Chapter 6. This process will likely take some time, and the ASAP-MS instrument may not be the most suitable instrument. However, it is clear that there is much interest in using AIMS devices in clinical settings, and given the current rapid pace of progress, we predict that it will not be long until surgical use of mass spectrometry becomes commonplace.

References

(1) World Health Organisation, *Cardiovascular diseases*, 2025, https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1.

(2) S. Rankin-Turner, P. Sears and L. M. Heaney, *Analytical Science Advances*, 2023, **4**, 133–153.

(3) Z. Takats, 'Ambient ionisation 20 years on – Big breakthroughs and remaining challenges', Conference, BMSS Ambient Ionisation SIG Meeting 2024 - 20 Years of Ambient ionisation - A celebrationBirmingham, 2024.

(4) H. Akbar, C. Foth, R. A. Kahloon and S. Mountfort, *StatPearls*, 2024.

(5) A. Singh, A. S. Museedi and S. A. Grossman, *StatPearls*, 2023.

(6) R. Baruah and A. Hartley, *ST-elevation myocardial infarction. BMJ Best Practice.* 2023, https://bestpractice.bmj.com/topics/en-gb/3000103.

(7) K. J. Schulte and H. N. Mayrovitz, *Cureus*, 2023, **15**, e37522.

(8) A. D. Hosseiny, S. Moloi, J. Chandrasekhar and A. Farshid, *Open Heart*, 2016, **3**, e000405.

(9) S. F. Carville, R. Henderson and H. Gray, *Clinical Medicine*, 2015, **15**, 362.

(10) C. Weston, S. Perwaiz, J. Wang, J. Kerr, A. Dana, M. de Belder, P. Ludman, M. Mamas, C. Gale, A. Wragg, C. Miles, B. Alkofer, A. Keys, C. Ighofose, T. Quinn and J. Goodfellow, *Management of Heart Attack: analyses from the Myocardial Ischaemia National Audit Project (MINAP) and the National Audit of Percutaneous Coronary Intervention (NAPCI)*, 2023, https://www.nicor.org.uk/myocardial-ischaemia-minap-heart-attack-audit/.

(11) H. Basit, A. Malik and M. R. Huecker, *StatPearls*, 2023.

(12) NICE, *MI - secondary prevention and Risk Factors - NICE*, https://cks.nice.org.uk/topics/mi-secondary-prevention/background-information/risk-factors/.

(13) A. Surendran, N. Atefi, H. Zhang, M. Aliani and A. Ravandi, *Metabolites 2021, Vol. 11, Page 685*, 2021, **11**, 685.

(14) C. Li, C. Deng, B. Shi and R. Zhao, *International Journal of Cardiology*, 2024, **405**, 131965.

(15) K. Thygesen, J. S. Alpert, A. S. Jaffe, B. R. Chaitman, J. J. Bax, D. A. Morrow, H. D. White, S. Corbett, M. Chettibi, H. Hayrapetyan, F. X. Roithinger, F. Aliyev, V. Sujayeva, M. J. Claeys, E. Smajić, P. Kala, K. K. Iversen, E. E. Hefny, T. Marandi, P. Porela, S. Antov, M. Gilard, S. Blankenberg, P. Davlouros, T. Gudnason, R. Alcalai, F. Colivicchi, S. Elezi, G. Baitova, I. Zakke, O. Gustiene, J. Beissel, P. Dingli, A. Grosu, P. Damman, V. Juliebo, J. Legutko, J. Morais, G. Tatu-Chitoiu, A. Yakovlev, M. Zavatta, M. Nedeljkovic, P. Radsel, A. Sionis, T. Jemberg, C. Müller, L. Abid, A. Abaci and A. Parkhomenko, *Circulation*, 2018, **138**, e618–e651.

(16) J. L. Björkegren and A. J. Lusis, *Cell*, 2022, **185**, 1630–1645.

(17) L. Alonso-Herranz, J. Albarrán-Juárez and J. F. Bentzon, *Frontiers in Cardiovascular Medicine*, 2023, **10**, 1254114.

(18) A. Undas, K. Szułdrzyński, K. E. Brummel-Ziedins, W. Tracz, K. Zmudka and K. G. Mann, *Blood*, 2008, **113**, 2070.

(19) G. Pocock and C. Richards, in *Human Physiology : the basics of medicine*, 1999, ch. 13.

(20) G. S. Bodor, *EJIFCC*, 2016, **27**, 95.

(21) X. Peng, J. Du and Y. Wang, *Biomedicine & Pharmacotherapy*, 2024, **170**, 116079.

(22) G. Pocock and C. Richards, in *Human physiology : the basis of medicine*, ed. C. D. Richards, Oxford University Press, 1999, ch. 15.

(23) J. I. Okada, K. Fujiu, K. Yoneda, T. Iwamura, T. Washio, I. Komuro, T. Hisada and S. Sugiura, *Journal of Physiological Sciences*, 2020, **70**, 1–14.

(24) P. Bamford, T. D. Henry, W. W. O'Neill and C. L. Grines, *Journal of the Society for Cardiovascular Angiography and Interventions*, 2024, **3**, 102395.

(25) L. H. Mughal and S. Sastry, *JRSM Cardiovascular Disease*, 2022, **11**, DOI: 10.1177/20480040221075519.

(26) M. Harker, S. Carville, R. Henderson and H. Gray, *Heart*, 2014, **100**, 536–543.

(27) C. Krittanawong, B. Yue, D. Mahtta, B. Narasimhan, A. Kumar, Z. Wang, S. K. Sharma, J. E. Tamis-Holland, S. S. Brar, R. Mehran, M. Alam, H. Jneid and S. S. Virani, *The American Journal of Cardiology*, 2022, **173**, 25–32.

(28) C. S. Kwok, A. Narain, H. M. Pacha, T. S. Lo, E. W. Holroyd, M. C. Alraies, J. Nolan and M. A. Mamas, *Cardiovascular Revascularization Medicine*, 2020, **21**, 375–391.

(29) J. A. Rymer, A. Y. Chen, L. Thomas, G. C. Fonarow, E. D. Peterson and T. Y. Wang, *Journal of the American Heart Association*, 2019, **8**, e012059.

(30) H. M. Bolooki and A. Askari, in *Current Clinical Medicine: Expert Consult*, W.B. Saunders, 2010, ch. 2, 65–71.e1.

(31) D. J. F. Lassen, H. E. Bøtker, C. J. Terkelsen, J. F. Lassen, J. F. Lassen, H. E. Bøtker and C. J. Terkelsen, *Nature Reviews Cardiology 2012 10:1*, 2012, **10**, 41–48.

(32) M. Ahmad, P. Mehta, A. K. R. Reddivari and S. Mungee, in *StatPearls*, StatPearls Publishing, 2023.

(33) BCIS, *The PCI Procedure - British Cardiovascular Intervention Society*, https://www.bcis.org.uk/public-information/the-pci-procedure/.

(34) P. Widimsky, *Primary PCI in the treatment of acute ST elevation myocardial infarction*, 2006, https://www.escardio.org/Journals/E-Journal-of-Cardiology-Practice/Volume-4/vol4n18-Title-Primary-PCI-in-the-treatment-of-acute-ST-elevation-myocardial-i.

(35) G. Sardella and R. E. Stio, *World Journal of Cardiology*, 2014, **6**, 924.

(36) S. S. Jolly, S. James, V. Džavík, J. A. Cairns, K. D. Mahmoud, F. Zijlstra, S. Yusuf, G. K. Olivecrona, H. Renlund, P. Gao, B. Lagerqvist, A. Alazzoni, S. Kedev, G. Stankovic, B. Meeks and O. Frøbert, *Circulation*, 2017, **135**, 143–152.

(37) I. B. Wilkinson, T. Raine, K. Wiles, P. Hateley, D. Kelly and I. McGurgan, in *Oxford Handbook of Clinical Medicine*, Oxford University PressOxford, 2024, ch. 3, pp. 84–151.

(38) R. Sweis and A. Jivan, *Acute Myocardial Infarction (MI) - Cardiovascular Disorders - MSD Manual Professional Edition*, 2024, https://www.msdmanuals.com/professional/cardiovascular-disorders/coronary-artery-disease/acute-myocardial-infarction-mi.

(39) C. L. Hvas and J. B. Larsen, *International Journal of Molecular Sciences*, 2023, **24**, 14179.

(40) H. P. Rang, in *Rang and Dale's Pharmacology*, ed. J. Ritter, R. J. (J. Flower, G. (Henderson and H. P. Rang, Elsevier Churchill Livingstone, Eighth edition., 2016, ch. 21.

(41) R. Partow-Navid, N. Prasitlumkum, A. Mukherjee, P. Varadarajan and R. G. Pai, *The International Journal of Angiology : Official Publication of the International College of Angiology, Inc*, 2021, **30**, 67.

(42) R. Rossi, A. Bagnacani, F. Sgura, D. E. Monopoli, F. Coppi, M. Talarico, C. Rolando and G. Boriani, *Coronary Artery Disease*, 2020, **31**, 348–353.

(43) W. Y. Shi and J. A. Smith, in *Primary Angioplasty*, Springer, 2018, ch. 16, pp. 211–221.

(44) NHS, *Liverpool Heart and Chest Hospital | Coronary Artery Bypass Grafting*, https://www.lhch.nhs.uk/coronary-artery-bypass-grafting.

(45) R. McGeoch, S. Watkins, C. Berry, T. Steedman, A. Davie, J. Byrne, S. Hillis, M. Lindsay, S. Robb, H. Dargie and K. Oldroyd, *JACC: Cardiovascular Interventions*, 2010, **3**, 715–722.

(46) G. S. Yoon, S. G. Ahn, S. I. Woo, M. H. Yoon, M. J. Lee, S. H. Choi, J. Y. Seo, S. W. Kwon, S. D. Park and K. W. Seo, *Journal of Clinical Medicine*, 2021, **10**, 4752.

(47) W. F. Fearon, A. F. Low, A. S. Yong, R. McGeoch, C. Berry, M. G. Shah, M. Y. Ho, H.-S. Kim, J. P. Loh and K. G. Oldroyd, *Circulation*, 2013, **127**, 2436–2441.

(48) H. Duman, M. Çetin, M. E. Durakoğlugil, H. Değirmenci, H. Hamur, M. Bostan, Z. Karadağ and Y. Çiçek, *Medical Science Monitor : International Medical Journal of Experimental and Clinical Research*, 2015, **21**, 3540.

(49) C. T. O'Connor, A. Ibrahim, A. Buckley, C. Maguire, R. Kumar, J. Kumar, S. Arnous and T. J. Kiernan, *The British Journal of Cardiology*, 2022, **29**, 17.

(50) M. Ghobrial, B. Bawamia, T. Cartlidge, I. Spyridopoulos, V. Kunadian, A. Zaman, M. Egred, A. McDiarmid, M. Williams, M. Farag and M. Alkhalil, *Journal of Clinical Medicine*, 2023, **12**, 5934.

(51) I. Y. Elgendy and H. Jneid, *Journal of Thoracic Disease*, 2018, **10**, 1343.

(52) C. Rios-Navarro, V. Marcos-Garces, A. Bayes-Genis, O. Husser, J. Nuñez and V. Bodi, *Journal of Clinical Medicine*, 2019, **8**, 1805.

(53) S. Kasim, P. N. F. A. Rudin, S. Malek, K. S. Ibrahim, W. A. W. Ahmad, A. Y. Y. Fong, W. Y. Lin, F. Aziz and N. Ibrahim, *Scientific Reports 2024 14:1*, 2024, **14**, 1–16.

(54) K. I. Shakhgeldyan, N. S. Kuksin, I. G. Domzhalov, V. Y. Rublev and B. I. Geltser, *Computers in Biology and Medicine*, 2024, **170**, 107953.

(55) P. Chen, B. Wang, L. Zhao, S. Ma, Y. Wang, Y. Zhu, X. Zeng, Z. Bai and B. Shi, *BMC Cardiovascular Disorders*, 2023, **23**, 1–9.

(56) S. Kasim, S. Malek, M. F. Aziz and K. S. Ibrahim, *European Heart Journal*, 2022, **43**, DOI: 10.1093/EURHEARTJ/EHAB849.176.

(57) H. P. Sritharan, H. Nguyen, J. Ciofani, R. Bhindi and U. K. Allahwala, *Frontiers in Cardiovascular Medicine*, 2024, **11**, DOI: 10.3389/fcvm.2024.1454321.

(58) A. Spirito, A. Quagliana, M. Coiro, G. D. Melaku, S. Vandenberghe, G. Leibundgut, J. Häner, M. Moccetti, M. Araco, H. M. Garcia-Garcia and M. Valgimigli, *EuroIntervention*, 2022, **18**, 242.

(59) G. L. D. Maria, H. M. Garcia-Garcia, R. Scarsini, A. Finn, Y. Sato, R. Virmani, R. Bhindi, J. L. Ciofani, J. Nuche, H. B. Ribeiro, W. Mathias, C. Yerasi, T. A. Fischell, L. Otterspoor, F. Ribichini, B. Ibañez, N. H. Pijls, R. S. Schwartz, N. K. Kapur, G. W. Stone and A. P. Banning, *European Heart Journal. Acute Cardiovascular Care*, 2021, **10**, 687–697.

(60) R. A. Kotronias, F. Marin, M. Emfietzoglou, J. P. Langrish, A. J. Lucking, K. M. Channon, A. P. Banning and G. L. D. Maria, *Cardiovascular Revascularization Medicine*, 2023, **52**, 75–85.

(61) I. Merinopoulos, T. Gunawardena, N. Corballis, U. Bhalraam, J. Reinhold, U. Wickramarachchi, C. Maart, T. Gilbert, P. Richardson, S. Sulfi, T. Sarev, C. Sawh, T. Wistow, A. Ryding, M. O. Mohamed, A. Perperoglou, M. A. Mamas, V. S. Vassiliou and S. C. Eccleshall, *Cardiovascular Interventions*, 2023, **16**, 771–779.

(62) B. Beska, D. E. Raharjo and V. Kunadian, *Revista Espanola de Cardiologia*, 2021, **74**, 1004–1006.

(63) A. Albaeni, Y. SHAN, R. A. Thakker, R. Saxena, S. Priyadarshni, D. Gaalema, Y.-F. Kuo and H. Jneid, *Journal of the American College of Cardiology*, 2024, **83**, 1335.

(64) S. V. Popov, L. N. Maslov, A. V. Mukhomedzyanov, B. K. Kurbatov, A. S. Gorbunov, M. Kilin, V. N. Azev, M. S. Khlestkina and G. Z. Sufianova, *Pharmaceutics 2023, Vol. 15, Page 1029*, 2023, **15**, 1029.

(65) L. C. Hsiao, Y. N. Lin, W. C. Shyu, M. Ho, C. R. Lu, S. S. Chang, Y. C. Wang, J. Y. Chen, S. Y. Lu, M. Y. Wu, K. Y. Li, Y. K. Lin, W. Y. I. Tseng, M. Y. Su, C. T. Hsu, C. K. Tsai, L. T. Chiu, C. L. Chen, C. L. Lin, K. C. Hu, D. Y. Cho, C. H. Tsai, K. C. Chang and L. B. Jeng, *Frontiers in Cardiovascular Medicine*, 2022, **9**, 961920.

(66) N. Sakalihasan, R. Limet and O. D. Defawe, *Lancet*, 2005, **365**, 1577–1589.

(67) T. L. Forbes, D. K. Lawlor, G. DeRose and K. A. Harris, *Ann Vasc Surg*, 2006, **20**, 564–8.

(68) X. Li, G. Zhao, J. Zhang, Z. Duan and S. Xin, *PLoS One*, 2013, **8**, e81260.

(69) I. M. Nordon, R. J. Hinchliffe, I. M. Loftus and M. M. Thompson, *Nature Reviews Cardiology*, 2011, **8**, 92–102.

(70) P. W. Stather, N. Dattani, M. J. Bown, J. J. Earnshaw and T. A. Lees, *European Journal of Vascular and Endovascular Surgery*, 2013, **45**, 231–234.

(71) M. Crowther, S. Goodall, J. L. Jones, P. R. F. Bell and M. M. Thompson, *Journal of Vascular Surgery*, 2000, **32**, 575–583.

(72) S. Goodall, M. Crowther, D. M. Hemingway, P. R. Bell and M. M. Thompson, *Circulation*, 2001, **104**, 304–309.

(73) S. Goodall, M. Crowther, P. R. Bell and M. M. Thompson, *Journal of Vascular Surgery*, 2002, **35**, 937–942.

(74) A. S. Ward, *Archives of Surgery*, 1992, **127**, 990–991.

(75) S. Makita, A. Ohira, R. Tachieda, S. Itoh, Y. Moriai, H. Niinuma, M. Nakamura and K. Hiramori, *American Heart Journal*, 2000, **140**, 297–302.

(76) I. Nordon, R. Brar, J. Taylor, R. Hinchliffe, I. M. Loftus and M. M. Thompson, *Journal of Vascular Surgery*, 2009, **50**, 171–176.e1.

(77) R. Lee, K. Bellamkonda, A. Jones, N. Killough, F. Woodgate, M. Williams, I. Cassimjee, A. Handa, A. Antonopoulos, C. Antoniades, K. M. Channon, R. Perera, K. Hurst, I. Milosevic, C. R. Darby, A. Halliday, L. J. Hands, P. Lintott, T. R. Magee, A. Northeast, J. Perkins and E. Sideso, *European Journal of Vascular and Endovascular Surgery*, 2017, **53**, 820–829.

(78) R. Lee, K. Bellamkonda, A. Jones, N. Killough, F. Woodgate, M. William, I. Cassimjee and A. Handa, *European Journal of Vascular and Endovascular Surgery*, 2019, **58**, e57–e58.

(79) Nuffield Department of Surgical Sciences John Radcliffe Hospital University of Oxford, *Oxford Abdominal Aortic Aneurysm (OxAAA) Study*, 2023, https://www.nds.ox.ac.uk/research/oxaaa.

(80) M. Niknejad, *Radiopaedia.org*, 2024, DOI: 10.53347/RID-185841.

(81) D. D'Souza, T. Walizai and M. T. Niknejad, *Radiopaedia.org*, 2008, DOI: 10.53347/RID-826.

(82) A. Piechota-Polanczyk, A. Jozkowicz, W. Nowak, W. Eilenberg, C. Neumayer, T. Malinski, I. Huk and C. Brostjan, *Front Cardiovasc Med*, 2015, **2**, 19.

(83) S. J. Cameron, H. M. Russell and A. P. Owens, *Blood*, 2018, **132**, 2619–2628.

(84) Q. Wang, G. Yesitayi, B. Liu, D. Siti, M. Ainiwan, A. Aizitiaili and X. Ma, *International Journal of Biological Sciences*, 2023, **19**, 3869.

(85) NHS, *Abdominal aortic aneurysm (AAA) screening - NHS*, https://www.nhs.uk/conditions/abdominal-aortic-aneurysm-screening/.

(86) H. Hong, Y. Yang, B. Liu and W. Cai, *Current vascular pharmacology*, 2010, **8**, 808.

(87) K. Knypl, *The abdominal aortic artery aneurysm and cardiovascular risk factors*, 2020, https://www.escardio.org/Journals/E-Journal-of-Cardiology-Practice/Volume-18/the-abdominal-aortic-artery-aneurysm-and-cardiovascular-risk-factors.

(88) E. Altobelli, L. Rapacchietta, V. F. Profeta and R. Fagnano, *International Journal of Environmental Research and Public Health 2018, Vol. 15, Page 2805*, 2018, **15**, 2805.

(89) B. Sonesson, T. Länne, F. Hansen and T. Sandgren, *European Journal of Vascular Surgery*, 1994, **8**, 89–95.

(90) Cambridge University Hositals NHS Foundation Trust, *Treatment of abdominal aortic aneurysms (AAA)*, Web Page, 2021, https://www.cuh.nhs.uk/patient-information/treatment-of-abdominal-aortic-aneurysms-aaa/#:~:text=The%20traditional%20operation%20involves%20cutting,the%20rest%20of%20your%20life..

(91) J. J. Reimerink, M. J. van der Laan, M. J. Koelemay, R. Balm and D. A. Legemate, *British Journal of Surgery*, 2013, **100**, 1405–1413.

(92) B. O. Patterson, P. J. Holt, R. Hinchliffe, I. M. Loftus and M. M. Thompson, *European Journal of Vascular and Endovascular Surgery*, 2008, **36**, 637–645.

(93) F. A. Lederle, S. E. Wilson, G. R. Johnson, D. B. Reinke, F. N. Littooy, C. W. Acher, D. J. Ballard, L. M. Messina, I. L. Gordon, E. P. Chute, W. C. Krupski, S. J. Busuttil, G. W. Barone, S. Sparks, L. M. Graham, J. H. Rapp, M. S. Makaroun, G. L. Moneta, R. A. Cambria, R. G. Makhoul, D. Eton, H. J. Ansel, J. A. Freischlag and D. Bandyk, *N Engl J Med*, 2002, **346**, 1437–44.

(94) G. S. Kim, H. J. Ahn, W. H. Kim, M. J. Kim and S. H. Lee, *Yonsei Medical Journal*, 2011, **52**, 339.

(95) W. K. Kieffer, S. Sonnenberg, R. A. Windhaber, N. Pal and R. M. Pemberton, *https://doi.org/10.1308/003588412X13171221501465*, 2015, **94**, 177–180.

(96) G. A. Sicard, R. M. Zwolak, A. N. Sidawy, R. A. White and F. S. Siami, *Journal of Vascular Surgery*, 2006, **44**, 229–236.

(97) M. Khashram, J. A. Williman, P. N. Hider, G. T. Jones and J. A. Roake, *European Journal of Vascular and Endovascular Surgery*, 2016, **51**, 203–215.

(98) NCBI, *Abdominal aortic aneurysm: Learn More – When is surgery recommended for the treatment of abdominal aortic aneurysm? - InformedHealth.org - NCBI Bookshelf*, 2006, `https://www.ncbi.nlm.nih.gov/books/NBK441584/`.

(99) T. Kudo, *Annals of Vascular Diseases*, 2019, **12**, 157.

(100) K. Berry, J. Gudgeon and J. Taylor, *BJA Education*, 2022, **22**, 208–215.

(101) J. J. V. Wyngaarden, M. D. Ross and B. R. Hando, *Journal of Orthopaedic and Sports Physical Therapy*, 2014, **44**, 500–507.

(102) S. M. Alabdullatif, M. H. Alajwad, M. F. Kareemah, I. A. Almaghasilah, A. M. Alsaedan, N. T. Alenazi, A. S. Alshaibani, M. K. Abusaleh, L. M. Maawadh, A. A. Alsolaiman and M. Alshammari, *Cureus*, 2021, **13**, e18587.

(103) F. A. Lederle, *Acta Chirurgica Belgica*, 2009, **109**, 7–12.

(104) A. Siika, M. Bogdanovic, M. L. Liljeqvist, T. C. Gasser, R. Hultgren and J. Roy, *Scientific Reports*, 2023, **13**, 9283.

(105) K. L. Kristensen, M. Dahl, L. M. Rasmussen and J. S. Lindholt, *Arteriosclerosis, thrombosis, and vascular biology*, 2017, **37**, 730–736.

(106) H. Kurvers, F. J. Veith, E. C. Lipsitz, T. Ohki, N. J. Gargiulo, N. S. Cayne, W. D. Suggs, C. H. Timaran, G. Y. Kwon, S. J. Rhee and C. Santiago, *Journal of the American College of Surgeons*, 2004, **199**, 709–715.

(107) M. Vega de Céniga, R. Gómez, L. Estallo, N. de la Fuente, B. Viviens and A. Barba, *Annals of Vascular Surgery*, 2008, **22**, 37–44.

(108) R. Ristl, J. Klopf, A. Scheuba, F. Wolf, M. Funovics, B. Gollackner, A. Wanhainen, C. Neumayer, M. Posch, C. Brostjan and W. Eilenberg, *British Journal of Surgery*, 2022, **109**, 211–219.

(109) S. A. Badger, C. Jones, J. McClements, L. L. Lau, I. S. Young and C. C. Patterson, *Vascular medicine*, 2011, **16**, 415–421.

(110) K. Vardulaki, T. Prevost, N. Walker, N. Day, A. Wilmink, C. Quick, H. Ashton and R. Scott, *Journal of British Surgery*, 1998, **85**, 1674–1680.

(111) M. V. De Ceniga, R. Gómez, L. Estallo, L. Rodríguez, M. Baquer and A. Barba, *European journal of vascular and endovascular surgery*, 2006, **31**, 231–236.

(112) S. Thompson, L. Brown, M. Sweeting, M. Bown, L. Kim, M. Glover, M. Buxton and J. Powell, *Health Technology Assessment (Winchester, England)*, 2013, **17**, 1–118.

(113) S. L. Olson, M. A. Wijesinha, A. M. Panthofer, W. C. Blackwelder, J. Upchurch, Gilbert R., M. L. Terrin, J. A. Curci, B. T. Baxter and J. S. Matsumura, *JAMA Surgery*, 2021, **156**, 363–370.

(114) A. Chandrashekar, A. Handa, P. Lapolla, N. Shivakumar, E. Ngetich, V. Grau and R. Lee, *Annals of Surgery*, 2020, **277**, e175.

(115) E. Akkoyun, S. T. Kwon, A. C. Acar, W. Lee and S. Baek, *Computers in Biology and Medicine*, 2020, **117**, 103620.

(116) S. S. Dhesi, P. Adusumilli, N. Ravikumar, M. A. Waduud, R. Frood, A. F. Frangi, G. McDermott, J. H. Rudd, Y. Huang, J. R. Boyle, M. Elkhawad, D. E. Newby, N. Joshi, J. Y. Kwan, P. Coughlin, M. A. Bailey and A. F. Scarsbrook, *Algorithms*, 2025, **18**, 86.

(117) NICE, *Abdominal aortic aneurysm: diagnosis and management NICE guideline*, 2020, `www.nice.org.uk/guidance/ng156`.

(118) M. D. Silverstein, S. R. Pitts, E. L. Chaikof and D. J. Ballard, *Baylor University Medical Center Proceedings*, 2005, **18**, 345–367.

(119) M. Kim, J. H. Han, D. H. Kim, M. Yoon and H. J. Jung, *Vascular Specialist International*, 2023, **39**, DOI: `10.5758/VSI.230075`.

(120) M. Wu and J. Xiong, *Journal of Endovascular Therapy*, 2025, DOI: `10.1177/15266028251316136`.

(121) A. Hosseini, T. Sahranavard, Ž. Reiner, T. Jamialahmadi, Y. A. Dhaheri, A. H. Eid and A. Sahebkar, *European Journal of Pharmaceutical Sciences*, 2022, **178**, DOI: `10.1016/J.EJPS.2022.106284`.

(122) L. M. Weaver, C. D. Loftin and C. G. Zhan, *Biomedicine and Pharmacotherapy*, 2022, **153**, DOI: `10.1016/J.BIOPHA.2022.113340`.

(123) P. Du, Y. Hou, C. Su, J. Gao, Y. Yang, J. Zhang, X. Cui and J. Tang, *Frontiers in Bioengineering and Biotechnology*, 2023, **11**, DOI: `10.3389/FBIOE.2023.1324406`.

(124) Y. Qi, H. Jiang, Y. Lun, Q. Gang, S. Shen, H. Zhang, M. Liu, Y. Wang and J. Zhang, *Journal of the American Heart Association*, 2025, **14**, DOI: `10.1161/JAHA.124.037802`.

(125) M. Mogi, *Hypertension Research*, 2023, **46**, 801–802.

(126) K. L. Summers, E. K. Kerut, F. To, C. M. Sheahan and M. G. Sheahan, *Journal of Vascular Surgery*, 2024, **79**, 1057–1067.e2.

(127) G. G. Salzler, E. J. Ryer, R. W. Abdu, A. Lanyado, T. Sagiv, E. N. Choman, A. A. Tariq, J. Urick, E. G. Mitchell, R. M. Maff, G. DeLong, S. L. Shriner and J. R. Elmore, *Journal of Vascular Surgery*, 2024, **79**, 776–783.

(128) T. K. Chung, P. H. Gueldner, O. U. Aloziem, N. L. Liang and D. A. Vorp, *Scientific Reports 2024 14:1*, 2024, **14**, 1–10.

(129) K. DiLosa, G. Brittenham, C. Pozolo, N. Hedayati, M. Kwong, S. Maximus and M. Humphries, *Journal of Vascular Surgery*, 2024, **80**, 107–113.

(130) J. O. Olukorode, C. N. Onwuzo, E. O. Otabor, N. O. Nwachukwu, R. Omiko, O. Omokore, H. Kristilere, Y. Oladipupo, R. Akin-Adewale, O. Kuku, J. O. Ugboke and T. Joseph-Erameh, *Cureus*, 2024, DOI: 10.7759/CUREUS.58673.

(131) A. F. Pérez, R. L. González, P. C. Lee, A. Vouyouka and Y. Erben, *Journal of Surgical Research*, 2025, **305**, 258–264.

(132) J. McCullagh and N. Oldham, in *Mass Spectrometry*, Oxford University Press, 2023, ch. 1.

(133) R. Ekman, J. Silberring, A. M. Westman-Brinkmalm, A. Kraj, D. M. Desiderio and N. M. Nibbering, *Mass Spectrometry : Instrumentation, Interpretation, and Applications*, John Wiley & Sons, Incorporated, 2008.

(134) E. de Hoffmann, *Mass Spectrometry : Principles and Applications.* Ed. V. Stroobant and E. D. Hoffmann, John Wiley & Sons, Incorporated, 3rd ed., 2007.

(135) J. H. Gross, *Mass Spectrometry*, Springer International Publishing, 3rd edn., 2017.

(136) *What is Mass Spectrometry | Scripps Research*, https://masspec.scripps.edu/learn/ms/.

(137) Z. Takáts, J. M. Wiseman, B. Gologan and R. G. Cooks, *Science (New York, N.Y.)*, 2004, **306**, 471–473.

(138) C. L. Feider, A. Krieger, R. J. Dehoog and L. S. Eberlin, *Analytical chemistry*, 2019, **91**, 4266.

(139) S. Rankin-Turner, J. C. Reynolds, M. A. Turner and L. M. Heaney, *Analytical Science Advances*, 2022, **3**, 67–89.

(140) T. H. Chen, H. Y. Hsu and S. P. Wu, *Forensic Science International*, 2016, **267**, 1–6.

(141) A. Kim, P. F. Kelly, M. A. Turner and J. C. Reynolds, *Rapid Communications in Mass Spectrometry*, 2023, **37**, e9422.

(142) M. J. Culzoni, P. Dwivedi, M. D. Green, P. N. Newton and F. M. Fernández, *MedChemComm*, 2013, **5**, 9–19.

(143) S. Mathias, M. Amerio-Cox, T. Jackson, D. Douce, B. McCullough, A. Sage, P. Luke, C. Crean and P. Sears, *Journal of the American Society for Mass Spectrometry*, 2024, DOI: 10.1021/JASMS.4C00277/ASSET/IMAGES/LARGE/JS4C00277_0003.JPEG.

(144) D. R. Ifa and L. S. Eberlin, *Clinical chemistry*, 2015, **62**, 111.

(145) Z. Takats, N. Strittmatter and J. S. McKenzie, *Advances in Cancer Research*, 2017, **134**, 231–256.

(146) M.-Z. Huang, C.-Y. Lee, K.-L. Chen, S.-Y. Lin, C.-M. Wu and J. Shiea, *Green Analytical Chemistry*, 2025, **13**, 100269.

(147) *Ambient Ionisation | BMSS*, https://www.bmss.org.uk/special-interest-groups/ambient-ionisation/.

(148) Advion Inc., *Advion expression CMS*, 2020, (visited on 28/11/2022).

(149) H. R. Tan, L. Y. Chan, H. H. Lee, Y. Q. Xu and W. Zhou, *Food Control*, 2022, **134**, 108736.

(150) T. Damiani, N. Dreolin, S. Stead and C. Dall'Asta, *Talanta*, 2021, **227**, 122116.

(151) Y. Wu, L. Huang, Y. Xu, Y. Zhang, L. Nie, S. Kang, F. Wei and S. Ma, *Food Chemistry*, 2025, **462**, 140965.

(152) E. Lee and M. Wood, *Analysis of drug-infused papers by ASAP-MS (poster)*, tech. rep., Waters Corporation UK, 2022.

(153) R. D. Soares, D. K. John, M. P. Thomé, P. S. Corrêa, K. S. Souza and M. F. Ferrão, *Drug Testing and Analysis*, 2025, **0**, 1–8.

(154) K. A. R. Monroy, R. Koerber and G. F. Verbeck, *Rapid Communications in Mass Spectrometry*, 2025, **39**, e9994.

(155) D. Burns, S. Mathias, B. J. McCullough, C. J. Hopley, D. Douce, N. Lumley, S. Bajic and P. Sears, *International Journal of Mass Spectrometry*, 2022, **471**, 116735.

(156) M. Jones, *Spectroscopy Europe*, 2021, **33**, 26–29.

(157) C. Henry and P. D. Rainville, *Demonstrating the Waters RADIAN ASAP Direct Ionization Detector as a Rapid Screening Tool for Discovery Reaction Monitoring*, tech. rep., Waters Corporation, 2021.

(158) Waters Corporation, *RADIAN ASAP Direct Mass Detector*, `https://www.waters.com/nextgen/gb/en/products/mass-spectrometry/mass-spectrometry-systems/radian-asap-direct-mass-detector.html`.

(159) Waters Corporation, *Atmospheric Solids Analysis Probe (ASAP)*, `https://www.waters.com/nextgen/gb/en/services/instrument-services/instrument-upgrades/mass-spectrometry-ms-upgrades/atmospheric-solids-analysis-probe.html`.

(160) Advion Inc., *Software for CMS*, 2020, (visited on 28/11/2022).

(161) Advion, Inc., *RevC-Expression CMS User's Manual*, Advion, Ltd., Harlow, UK, 2014.

(162) F. J. Andrade, J. T. Shelley, W. C. Wetzel, M. R. Webb, G. Gamez, S. J. Ray and G. M. Hieftje, *Analytical Chemistry*, 2008, **80**, 2654–2663.

(163) S. P. Pasilis and G. J. V. Berkel, in *Encyclopedia of Spectroscopy and Spectrometry (Third Edition)*, ed. J. C. Lindon, G. E. Tranter and D. W. Koppenaal, Academic Press, Third Edition, 2017, pp. 819–829.

(164) M. Schmid, G. S. Parkinson and U. Diebold, *ACS Physical Chemistry Au*, 2023, **3**, 44–62.

(165) J. McCullagh and N. Oldham, in *Mass Spectrometry*, Oxford University Press, 2023, ch. 2.

(166) W. C. Byrdwell, *Lipids*, 2001, **36**, 327–346.

(167) Creative Proteomics, *Atmospheric Pressure Chemical Ionization - Creative Proteomics*, `https://www.creative-proteomics.com/support/atmospheric-pressure-chemical-ionization.htm`.

(168) Waters Corporation, *Atmospheric Pressure Ionisation Sources: Their Use and Applicability (White paper)*, tech. rep., 2017.

(169) C. N. McEwen and B. S. Larsen, *Journal of the American Society for Mass Spectrometry*, 2009, **20**, 1518–1521.

(170) N. M. Bingham, J. S. Wright, S. C. Mathias, D. Douce and P. Sears, *Rapid Communications in Mass Spectrometry*, 2024, **38**, e9689.

(171) J. McCullagh and N. Oldham, in *Mass Spectrometry*, Oxford University Press, 2023, ch. 5.

(172) J. McCullagh and N. Oldham, in *Mass Spectrometry*, Oxford University Press, 2023, ch. 3.

(173) Shimadzu Corporation, *Introduction to mass analyzers : SHIMADZU (Shimadzu Corporation)*, `https://www.shimadzu.com/an/service-support/technical-support/analysis-basics/fundamental/mass_analyzers.html`.

(174) TRACES centre, *Quadrupoles: How do they work?*, tech. rep., University of Toronto, 2020.

(175) R. Pedder, 49th ASMS Conference on Mass Spectrometry and Allied Topics, 2001.

(176) E. de Hoffmann and V. Stroobant, in *Mass Spectrometry: Principles and Applications*, John Wiley and Sons, 3rd, 2007, ch. 3.

(177) F. Zydel, J. R. Smith, V. S. Pagnotti, R. J. Lawrence, C. N. Mcewen and B. R. Capacio, *Drug Testing and Analysis*, 2012, **4**, 308–311.

(178) D. Carrizo, I. Nerín, C. Domeño, P. Alfaro and C. Nerín, *Journal of Pharmaceutical and Biomedical Analysis*, 2016, **124**, 149–156.

(179) A. Orlowsky, 2021.

(180) A. S. Eardley-Brunt, A. Jones, T. Mills, L. Song, R. Kotronias, P. Lapolla, A. Handa, R. Lee, K. Channon, G. L. de Maria and C. Vallance, *International Journal of Mass Spectrometry*, 2025, **508**, 117386.

(181) L. Chen, F. Zhong and J. Zhu, *Metabolites*, 2020, **10**, 348.

(182) T. Cajka and O. Fiehn, *Analytical Chemistry*, 2016, **88**, 524–545.

(183) SAS UK, *Machine Learning: What it is and why it matters | SAS UK*, 2025, `https://www.sas.com/en_gb/insights/analytics/machine-learning.html`.

(184) I. H. Sarker, *SN Computer Science*, 2021, **2**, 1–21.

(185) T. Hastie, *The elements of statistical learning : data mining, inference, and prediction*, ed. R. Tibshirani and J. H. (H. Friedman, Springer, Second edition., 2017.

(186) J. A. Sidey-Gibbons and C. J. Sidey-Gibbons, *BMC Medical Research Methodology*, 2019, **19**, 1–18.

(187) Z. Ye, Y. Xu, L. Tang, M. Wu, B. Wu, T. Zhu and J. Wang, *Cardiovascular Diabetology*, 2023, **22**, 1–12.

(188) T. H. Sun, C. C. Wang, Y. L. Wu, K. C. Hsu and T. H. Lee, *Scientific Reports 2023 13:1*, 2023, **13**, 1–14.

(189) W. DeGroat, H. Abdelhalim, K. Patel, D. Mendhe, S. Zeeshan and Z. Ahmed, *Scientific Reports 2024 14:1*, 2024, **14**, 1–13.

(190) L. M. Petrick and N. Shomron, *Cell Rep Phys Sci*, 2022, **3**, DOI: 10.1016/j.xcrp.2022.100978.

(191) C. Sirocchi, F. Biancucci, M. Donati, A. Bogliolo, M. Magnani, M. Menotta and S. Montagna, *Computer Methods and Programs in Biomedicine*, 2024, **250**, 108163.

(192) Y. Deng, Y. Yao, Y. Wang, T. Yu, W. Cai, D. Zhou, F. Yin, W. Liu, Y. Liu, C. Xie, J. Guan, Y. Hu, P. Huang and W. Li, *Nature Communications 2024 15:1*, 2024, **15**, 1–17.

(193) A. G. S. James, L. Hand, T. Mills, L. Song, A. S. Brunt, P. E. B. Mann, A. F. Worrall, M. I. Stewart and C. Vallance, *Journal of Chemical Education*, 2023, **100**, 1343–1350.

(194) J. T. Watson and O. D. Sparkman, *Introduction to mass spectrometry : instrumentation, applications and strategies for data interpretation*, John Wiley and Sons, Chichester, England, 4th, 2007.

(195) E. d. Hoffmann and V. Stroobant, *Mass spectrometry : principles and applications*, Wiley, Chichester, Third, 2007.

(196) A. Borlotti, M. Jerosch-Herold, D. Liu, D. Viliani, A. Bracco, M. Alkhalil, G. L. D. Maria, K. M. Channon, A. P. Banning, R. P. Choudhury, S. Neubauer, R. K. Kharbanda and E. Dall'Armellina, *JACC: Cardiovascular Imaging*, 2019, **12**, 1783–1793.

(197) F. Cuculi, N. Herring, A. R. D. Caterina, A. P. Banning, B. D. Prendergast, J. C. Forfar, R. P. Choudhury, K. M. Channon and R. K. Kharbanda, *Heart*, 2013, **99**, 1198–1203.

(198) G. L. D. Maria, F. Cuculi, N. Patel, S. Dawkins, G. Fahrni, G. Kassimis, R. P. Choudhury, J. C. Forfar, B. D. Prendergast, K. M. Channon, R. K. Kharbanda and A. P. Banning, *European Heart Journal*, 2015, **36**, 3165–3177.

(199) Department of Cardiovascular Medicine University of Oxford, *Information for Health Professionals*, https://oxami.org.uk/information-for-health-professionals/.

(200) Oxford University Clinical Academic Graduate School, *The Oxford Abdominal Aortic Aneurysm Study (OxAAA) — OUCAGS*, 2017, https://www.oucags.ox.ac.uk/featured-projects/clinical-lectureships-1/the-oxford-abdominal-aortic-aneurysm-study-oxaaa.

(201) R. Lee, I. Cassimjee, H. Huang, P. Lapolla, E. Ngetich, A. Chandrashekar, P. Charles, B. Kessler, R. Fischer and A. Handa, *Annals of surgery*, 2022, **275**, 1206–1211.

(202) The MathWorks, Inc., *Inertial Sensor Noise Analysis Using Allan Variance*, Web Page, 2023, https://uk.mathworks.com/help/fusion/ug/inertial-sensor-noise-analysis-using-allan-variance.html.

(203) W. J. Riley, *National Institute of Standards and Technology*, 2007, **1065**, 1–123.

(204) W. J. Qian, T. Liu, V. A. Petyuk, M. A. Gritsenko, B. O. Petritis, A. D. Polpitiya, A. Kaushal, W. Xiao, C. C. Finnerty, M. G. Jeschke, N. Jaitly, M. E. Monroe, R. J. Moore, L. L. Moldawer, R. W. Davis, R. G. Tompkins, D. N. Herndon, D. G. Camp and R. D. Smith, *J Proteome Res*, 2009, **8**, 290–9.

(205) P. D. Piehowski, V. A. Petyuk, D. J. Orton, F. Xie, R. J. Moore, M. Ramirez-Restrepo, A. Engel, A. P. Lieberman, R. L. Albin, D. G. Camp, R. D. Smith and A. J. Myers, *J Proteome Res*, 2013, **12**, 2128–37.

(206) M. Woolman, A. Tata, E. Bluemke, D. Dara, H. J. Ginsberg and A. Zarrine-Afsar, *Journal of the American Society for Mass Spectrometry*, 2017, **28**, 145–153.

(207) N. Abbassi-Ghadi, E. A. Jones, K. A. Veselkov, J. Huang, S. Kumar, N. Strittmatter, O. Golf, H. Kudo, R. D. Goldin, G. B. Hanna and Z. Takats, *Analytical Methods*, 2014, **7**, 71–80.

(208) R. A. van den Berg, H. C. Hoefsloot, J. A. Westerhuis, A. K. Smilde and M. J. van der Werf, *BMC Genomics*, 2006, **7**, 142.

(209) T. Cook, Y. Ma and S. Gamagedara, *Journal of Pharmaceutical and Biomedical Analysis*, 2020, **177**, 112854.

(210) E. Dubois, A. N. Galindo, L. Dayon and O. Cominetti, *Biosystems*, 2022, **215-216**, 104661.

(211) B. B. Misra, *European Journal of Mass Spectrometry*, 2020, **26**, 165–174.

(212) T. Hastie, R. Tibshirani and J. Friedman, in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, Second Edition, 2009, ch. 14.

(213) E. Zhvansky, A. Sorokin, V. Shurkhay, D. Zavorotnyuk, D. Bormotov, S. Pekov, A. Potapov, E. Nikolaev and I. Popov, *Mass Spectrometry*, 2021, **10**, A0094.

(214) K. R. Moon, D. van Dijk, Z. Wang, S. Gigante, D. B. Burkhardt, W. S. Chen, K. Yim, A. van den Elzen, M. J. Hirn, R. R. Coifman, N. B. Ivanova, G. Wolf and S. Krishnaswamy, *Nature Biotechnology 2019 37:12*, 2019, **37**, 1482–1492.

(215) Y. Liu, *Knowledge-Based Systems*, 2012, **26**, 207–215.

(216) scikit-learn, *Unsupervised learning — scikit-learn 1.6.1 documentation*, https://scikit-learn.org/stable/unsupervised_learning.html.

(217) MathWorks, *Machine Learning with MATLAB*, Electronic Book, 2023, https://uk.mathworks.com/campaigns/offers/machine-learning-with-matlab.html.

(218) I. T. Jolliffe, *Principal component analysis*, Springer, 2nd ed., 2011.

(219) C. Aliferis, G. Simon, C. Aliferis and G. Simon, *Health Informatics*, 2024, 477–524.

(220) C. Maver, D. Dubinsky and S. Ahmad, *The Path to Machine Intelligence: Classic AI vs. Deep Learning vs. Biological Approach*, 2022, https://www.numenta.com/blog/2022/01/25/the-path-to-machine-intelligence/.

(221) J. Faouzi and O. Colliot, *Neuromethods*, 2023, **197**, 25–75.

(222) T. Hastie, in *The elements of statistical learning: data mining, inference, and prediction*, ed. R. Tibshirani and J. Friedman, Springer, 2017, ch. 13.

(223) Z.-H. Zhou, in Springer, 2021, ch. 6.

(224) Z.-H. Zhou, in Springer, 2021, ch. 7.

(225) Z.-H. Zhou, in Springer, 2021, ch. 3.

(226) G. James, D. Witten, T. Hastie and R. Tibshirani, 2021, DOI: 10.1007/978-1-0716-1418-1.

(227) Z.-H. Zhou, in Springer, 2021, ch. 4.

(228) P. Flach, in *Machine learning : the art and science of algorithms that make sense of data*, Cambridge University Press, 2012, ch. 8, pp. 231–261.

(229) G. Guo, H. Wang, D. Bell, Y. Bi and K. Greer, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2003, **2888**, 986–996.

(230) B. Boehmke and B. Greenwell, in *Hands-On Machine Learning with R*, CRC Press, 2020, ch. 8.

(231) IBM, *What is the k-nearest neighbors algorithm? | IBM*, https://www.ibm.com/think/topics/knn.

(232) C. Cortes, V. Vapnik and L. Saitta, *Machine Learning 1995 20:3*, 1995, **20**, 273–297.

(233) B. Boehmke and B. Greenwell, in *Hands-On Machine Learning with R*, CRC Press, 2020, ch. 14.

(234) R. Berwick, *An Idiot's guide to Support vector machines (SVMs)*, https://web.mit.edu/6.034/wwwbob/svm-notes-long-08.pdf.

(235) scikit-learn, *1.4. Support Vector Machines — scikit-learn 1.6.1 documentation*, https://scikit-learn.org/stable/modules/svm.html.

(236) P. Flach, in *Machine learning : the art and science of algorithms that make sense of data*, Cambridge University Press, 2012, ch. 9.

(237) *Naive Bayes Algorithm in ML: Simplifying Classification Problems*, https://www.turing.com/kb/an-introduction-to-naive-bayes-algorithm-for-beginners.

(238) N. Mohanty, A. L.-S. John, R. Manmatha and T. M. Rath, in *Handbook of Statistics*, ed. C. R. Rao and V. Govindaraju, Elsevier, 2013, vol. 31, ch. 10, pp. 249–267.

(239) S. Soaad, S. Yahaya, Y.-F. Lim, H. Ali and Z. Omar, 2016, DOI: 10.3844/jmssp.2016.312.316.

(240) scikit-learn, *Linear Discriminant Analysis — scikit-learn 1.6.1 documentation*, https://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis.LinearDiscriminantAnalysis.html.

(241) IBM, *What Is Linear Discriminant Analysis? | IBM*, https://www.ibm.com/think/topics/linear-discriminant-analysis.

(242) P. Flach, in *Machine learning : the art and science of algorithms that make sense of data*, Cambridge University Press, 2012, ch. 11.

(243) T. Hastie, R. Tibshirani and J. Friedman, in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, ed. T. Hastie, R. Tibshirani and J. Friedman, Springer New York, 2009, ch. 15, pp. 587–604.

(244) IBM, *What Is Random Forest? | IBM*, https://www.ibm.com/think/topics/random-forest.

(245) N. Makke and S. Chawla, *Artificial Intelligence Review*, 2024, **57**, 1–38.

(246) N. Makke and S. Chawla, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2023, **13830 LNCS**, 38–50.

(247) D. Ferrari, V. Guidetti, Y. Wang and V. Curcin, *AMIA Annual Symposium Proceedings*, 2023, **2022**, 442.

(248) D. Angelis, F. Sofos and T. E. Karakasidis, *Archives of Computational Methods in Engineering 2023 30:6*, 2023, **30**, 3845–3865.

(249) Y. A. Radwan, G. Kronberger and S. Winkler, 2024.

(250) S. Katoch, S. S. Chauhan and V. Kumar, *Multimedia Tools and Applications*, 2021, **80**, 8091–8126.

(251) M. Cranmer, 2023.

(252) N. J. Christensen, S. Demharter, M. MacHado, L. Pedersen, M. Salvatore, V. Stentoft-Hansen and M. T. Iglesias, *Bioinformatics*, 2022, **38**, 3749–3758.

(253) P. Flach, in *Machine learning : the art and science of algorithms that make sense of data*, Cambridge University Press, 2012, ch. 3.2.

(254) D. F. Hamilton, M. Ghert and A. H. Simpson, *Bone & Joint Research*, 2015, **4**, 152.

(255) A. M. Stefan and F. D. Schönbrodt, *Royal Society Open Science*, 2023, **10**, DOI: 10.1098/RSOS.220346.

(256) G. D. Smith and S. Ebrahim, *BMJ : British Medical Journal*, 2002, **325**, 1437.

(257) Laerd Statistics, *Testing for Normality using SPSS Statistics*, https://statistics.laerd.com/spss-tutorials/testing-for-normality-using-spss-statistics.php#:~:text=value%20of%20the%20Shapiro%2DWilk,enhanced%20testing%20for%20normality%20guide..

(258) E. S. Zhvansky, S. I. Pekov, A. A. Sorokin, V. A. Shurkhay, V. A. Eliferov, A. A. Potapov, E. N. Nikolaev and I. A. Popov, *Scientific Reports*, 2019, **9**, 914.

(259) B. R. Kirkwood, in *Essential medical statistics*, ed. J. A. C. Sterne, Blackwell Pub., 2nd ed., 2003, ch. 10.

(260) H. Liu and R. Setiono, *Proceedings of the International Conference on Tools with Artificial Intelligence*, 1995, 388–391.

(261) M. L. McHugh, *Biochemia Medica*, 2013, **23**, 143–149.

(262) MathWorks, *fsrmrmr*, `https://uk.mathworks.com/help/stats/fsrmrmr.html`.

(263) M. Radovic, M. Ghalwash, N. Filipovic and Z. Obradovic, *BMC Bioinformatics*, 2017, **18**, 1–14.

(264) MathWorks, *relieff*, `https://uk.mathworks.com/help/stats/relieff.html`.

(265) R. J. Urbanowicz, M. Meeker, W. L. Cava, R. S. Olson and J. H. Moore, *Journal of Biomedical Informatics*, 2018, **85**, 189–203.

(266) B. R. Kirkwood, in *Essential medical statistics*, ed. J. A. C. Sterne, Blackwell Pub., 2nd ed., 2003, ch. 36.

(267) M. L. McHugh, *Biochemia Medica*, 2012, 276–282.

(268) *MetaboAnalyst*, `https://www.metaboanalyst.ca/home.xhtml`.

(269) A. S. J. Eardley-Brunt, T. Mills, R. Kontronias, G. L. de Maria, K. Channon, T. O. A. M. I. Study and C. Vallance, *Analyst*, 2025, **24**, DOI: `10.1039/D5AN00565E`.

(270) C. H. Wang, H. T. Wang, K. H. Wu, F. J. Cheng, C. I. Cheng, C. T. Kung and F. C. Chen, *Emergency Medicine International*, 2022, **2022**, 5389072.

(271) C. Montalto, R. A. Kotronias, F. Marin, D. Terentes-Printzios, M. Shanmuganathan, M. Emfietzoglou, R. Scalamera, I. Porto, J. Langrish, A. Lucking, R. Choudhury, R. Kharbanda, K. M. Channon, G. L. D. Maria and A. Banning, *International journal of cardiology*, 2021, **339**, 1–6.

(272) R. Scarsini, M. Shanmuganathan, G. L. D. Maria, A. Borlotti, R. A. Kotronias, M. K. Burrage, D. Terentes-Printzios, J. Langrish, A. Lucking, G. Fahrni, F. Cuculi, F. Ribichini, R. P. Choudhury, R. Kharbanda, V. M. Ferreira, K. M. Channon and A. P. Banning, *JACC: Cardiovascular Imaging*, 2021, **14**, 1948–1959.

(273) G. A. Waard, G. Fahrni, D. D. Wit, H. Kitabata, R. Williams, N. Patel, P. F. Teunissen, P. M. V. D. Ven, S. Umman, P. Knaapen, D. Perera, T. Akasaka, M. Sezer, R. K. Kharbanda and N. V. Royen, *Heart (British Cardiac Society)*, 2018, **104**, 127–134.

(274) G. Fahrni, M. Wolfrum, G. L. D. Maria, F. Cuculi, S. Dawkins, M. Alkhalil, N. Patel, J. C. Forfar, B. D. Prendergast, R. P. Choudhury, K. M. Channon, A. P. Banning and R. K. Kharbanda, *J Am Heart Assoc*, 2017, **6**, DOI: `10.1161/jaha.116.005409`.

(275) N. Patel, R. Petraco, E. Dall'Armellina, G. Kassimis, G. L. D. Maria, S. Dawkins, R. Lee, B. D. Prendergast, R. P. Choudhury, J. C. Forfar, K. M. Channon, J. Davies, A. P. Banning and R. K. Kharbanda, *JACC: Cardiovascular Interventions*, 2015, **8**, 1410–1421.

(276) R. Scarsini, G. L. D. Maria, A. Borlotti, R. A. Kotronias, J. P. Langrish, A. J. Lucking, R. P. Choudhury, V. M. Ferreira, F. Ribichini, K. M. Channon, R. K. Kharbanda and A. P. Banning, *Cardiovascular revascularization medicine : including molecular interventions*, 2019, **20**, 1148–1155.

(277) G. L. D. Maria, M. Alkhalil, M. Wolfrum, G. Fahrni, A. Borlotti, L. Gaughran, S. Dawkins, J. P. Langrish, A. J. Lucking, R. P. Choudhury, I. Porto, F. Crea, E. Dall'Armellina, K. M. Channon, R. K. Kharbanda and A. P. Banning, *EuroIntervention : journal of EuroPCR in collaboration with the Working Group on Interventional Cardiology of the European Society of Cardiology*, 2017, **13**, 935–943.

(278) G. L. D. Maria, G. Fahrni, M. Alkhalil, F. Cuculi, S. Dawkins, M. Wolfrum, R. P. Choudhury, J. C. Forfar, B. D. Prendergast, T. Yetgin, R. J. V. Geuns, M. Tebaldi, K. M. Channon, R. K. Kharbanda, P. M. Rothwell, M. Valgimigli and A. P. Banning, *EuroIntervention : journal of EuroPCR in collaboration with the Working Group on Interventional Cardiology of the European Society of Cardiology*, 2016, **12**, 1223–1230.

(279) H. Tian, X. Zhao, Y. Zhang and Z. Xia, *Biomedicine and Pharmacotherapy*, 2023, **163**, 114827.

(280) A. G. Beck, M. Muhoberac, C. E. Randolph, C. H. Beveridge, P. R. Wijewardhane, H. I. Kenttämaa and G. Chopra, *ACS Measurement Science Au*, 2024, **4**, 233.

(281) N. Verbeeck, R. M. Caprioli and R. V. de Plas, *Mass Spectrometry Reviews*, 2020, **39**, 245–291.

(282) N. H. Tran, R. Qiao, L. Xin, X. Chen, C. Liu, X. Zhang, B. Shan, A. Ghodsi and M. Li, *Nature Methods 2018 16:1*, 2018, **16**, 63–66.

(283) W. M. Abdelmoula, S. A. Stopka, E. C. Randall, M. Regan, J. N. Agar, J. N. Sarkaria, W. M. Wells, T. Kapur and N. Y. Agar, *Bioinformatics*, 2022, **38**, 2015–2021.

(284) Division of Cardiovascular Medicine Radcliffe Department of Medicine University of Oxford, *OXAMI: Oxford Acute Myocardial Infarction Study*, 2023, `https://oxami.org.uk/`.

(285) L. Song, J. G. Reese, M. A. Platt, C. Lewis, A. S. J. Eardley-Brunt, B. Sun, O. Ansorge and C. Vallance, *Analyst*, 2025, DOI: `10.1039/D5AN00166H`.

(286) 'The effect of normalisation, scaling and transformation methods on machine learning analysis applied to ASAP-MS human blood plasma measurements'. A. S. Eardley-Brunt, L. Song, The Oxford Acute Myocardial Infarction Study, The Oxford Abdominal Aortic Aneurysm Study and C. Vallance, forthcoming.

(287) J. A. Rodrigues, K. Melleu, M. M. Schmidt, C. A. M. Gottschall, M. A. P. D. Moraes and A. S. D. Quadros, *Arquivos Brasileiros de Cardiologia*, 2018, DOI: 10.5935/abc.20180178.

(288) scikit-learn, *Importance of Feature Scaling — scikit-learn 1.6.1 documentation*, https://scikit-learn.org/stable/auto_examples/preprocessing/plot_scaling_importance.html.

(289) K. Kira and L. A. Rendell, *Proceedings of the 9th International Workshop on Machine Learning, ICML 1992*, 1992, 249–256.

(290) R. Spencer, F. Thabtah, N. Abdelhamid and M. Thompson, *Digital Health*, 2020, **6**, DOI: 10.1177/2055207620914777/ASSET/IMAGES/LARGE/10.1177_2055207620914777-FIG4.JPEG.

(291) A. Donner and M. Eliasziw, *Biometrics*, 1994, **50**, 550.

(292) J. Sim and C. C. Wright, *Physical Therapy*, 2005, **85**, 257–268.

(293) A. Abbas, G. H. Matthews, I. W. Brown, J. S. Shambrook, C. R. Peebles and S. P. Harden, *Br J Radiol*, 2015, **88**, 20140470.

(294) G. L. D. Maria, M. Alkhalil, M. Wolfrum, G. Fahrni, A. Borlotti, L. Gaughran, S. Dawkins, J. P. Langrish, A. J. Lucking, R. P. Choudhury, I. Porto, F. Crea, E. Dall'Armellina, K. M. Channon, R. K. Kharbanda and A. P. Banning, *JACC. Cardiovascular imaging*, 2019, **12**, 837–848.

(295) H. Bulluck and C. Berry, *JACC. Cardiovascular imaging*, 2019, **12**, 1593–1594.

(296) G. Sianos, M. I. Papafaklis, J. Daemen, S. Vaina, C. A. van Mieghem, R. T. van Domburg, L. K. Michalis and P. W. Serruys, *Journal of the American College of Cardiology*, 2007, **50**, 573–583.

(297) C. M. Gibson, J. A. D. Lemos, S. A. Murphy, S. J. Marble, C. H. McCabe, C. P. Cannon, E. M. Antman and E. Braunwald, *Circulation*, 2001, **103**, 2550–2554.

(298) D. S. Wishart, D. Tzur, C. Knox, R. Eisner, A. C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M.-A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D. D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G. E. Duggan, G. D. MacInnis, A. M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B. D. Sykes, H. J. Vogel and L. Querengesser, *Nucleic Acids Research*, 2007, **35**, DOI: 10.1093/nar/gkl923.

(299) D. S. Wishart, A. Guo, E. Oler, F. Wang, A. Anjum, H. Peters, R. Dizon, Z. Sayeeda, S. Tian, B. L. Lee, M. Berjanskii, R. Mah, M. Yamamoto, J. Jovel, C. Torres-Calzada, M. Hiebert-Giesbrecht, V. W. Lui, D. Varshavi, D. Varshavi, D. Allen, D. Arndt, N. Khetarpal, A. Sivakumaran, K. Harford, S. Sanford, K. Yee, X. Cao, Z. Budinski, J. Liigand, L. Zhang, J. Zheng, R. Mandal, N. Karu, M. Dambrova, H. B. Schiöth, R. Greiner and V. Gautam, *Nucleic Acids Research*, 2021, **50**, DOI: 10.1093/nar/gkab1062.

(300) T. Hahn, S. Yao, L. M. Dunford, J. Thomas, J. Lohr, P. Arora, M. Battiwalla, S. L. Smiley and P. L. McCarthy, *Biology of Blood and Marrow Transplantation*, 2009, **15**, 574–579.

(301) UK Kidney Association, *Measurement of kidney function | UK Kidney*, https://www.ukkidney.org/health-professionals/information-resources/uk-eckd-guide/measurement-kidney-function.

(302) *GitHub - MilesCranmer/PySR: High-Performance Symbolic Regression in Python and Julia*, https://github.com/MilesCranmer/PySR.

(303) A. Tonda, *Genetic Programming and Evolvable Machines 2024 26:1*, 2024, **26**, 1–4.

(304) *Feyn Documentation · Symbolic AI using the QLattice*, https://docs.abzu.ai/.

(305) *GitHub - abzu-ai/QLattice-clinical-omics*, https://github.com/abzu-ai/QLattice-clinical-omics?tab=readme-ov-file.

(306) *GitHub - hengzhe-zhang/EvolutionaryForest: An open source python library for automated feature engineering based on Genetic Programming*, https://github.com/hengzhe-zhang/EvolutionaryForest.

(307) H. Zhang, A. Zhou and H. Zhang, *IEEE Transactions on Evolutionary Computation*, 2022, **26**, 735–749.

(308) H. Zhang, A. Zhou, Q. Chen, B. Xue and M. Zhang, *IEEE Transactions on Evolutionary Computation*, 2023, DOI: 10.1109/TEVC.2023.3243172.

(309) *HeuristicLab*, https://dev.heuristiclab.com/trac.fcgi/.

(310) *GitHub - heal-research/HeuristicLab: HeuristicLab - An environment for heuristic and evolutionary optimization*, https://github.com/heal-research/HeuristicLab.

(311) *GitHub - marcovirgolin/GP-GOMEA: Genetic Programming version of GOMEA. Also includes standard tree-based GP, and Semantic Backpropagation-based GP*, https://github.com/marcovirgolin/GP-GOMEA.

(312) M. Virgolin, T. Alderliesten, C. Witteveen and P. A. Bosman, *Evolutionary Computation*, 2021, **29**, 211–237.

(313) *Differentiable Cartesian Genetic Programming — dCGP 1.6.1 documentation*, https://darioizzo.github.io/dcgp/.

(314) *GitHub - darioizzo/dcgp: Implementation of a differentiable CGP (Cartesian Genetic Programming)*, https://github.com/darioizzo/dcgp/.

(315) D. Izzo, F. Biscani and A. Mereta, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, **10196 LNCS**, 35–51.

(316) E. Wu, J. T. Ortiz, P. Tejedor, D. C. Lee, C. Bucciarelli-Ducci, P. Kansal, J. C. Carr, T. A. Holly, D. Lloyd-Jones, F. J. Klocke and R. O. Bonow, *Heart (British Cardiac Society)*, 2008, **94**, 730–736.

(317) G. W. Stone, H. P. Selker, H. Thiele, M. R. Patel, J. E. Udelson, E. M. Ohman, A. Maehara, I. Eitel, C. B. Granger, P. L. Jenkins, M. Nichols and O. Ben-Yehuda, *Journal of the American College of Cardiology*, 2016, **67**, 1674–1683.

(318) M. Wamil, A. Borlotti, D. Liu, A. B. e Gala, A. Bracco, M. Alkhalil, G. L. D. Maria, S. K. Piechnik, V. M. Ferreira, A. P. Banning, R. K. Kharbanda, S. Neubauer, R. P. Choudhury, K. M. Channon and E. Dall'Armellina, *International Journal of Cardiovascular Imaging*, 2019, **35**, 1297–1308.

(319) D. Venetsanos, J. Alfredsson, M. Segelmark, E. Swahn and S. S. Lawesson, *BMJ Open*, 2015, **5**, e007835.

(320) E. H. Bae, S. Y. Lim, K. H. Cho, J. S. Choi, C. S. Kim, J. W. Park, S. K. Ma, M. H. Jeong and S. W. Kim, *American Journal of Kidney Diseases*, 2012, **59**, 795–802.

(321) *API Reference - PySR*, `https://ai.damtp.cam.ac.uk/pysr/api/`.

(322) A. Gupta, T. S. Stead and L. Ganti, *Academic Medicine & Surgery*, 2024, DOI: `10.62186/001C.125154`.

(323) E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes-Bascarán and Á. Fernández-Leal, *Artificial Intelligence Review 2022 56:4*, 2022, **56**, 3005–3054.

(324) M. Ciborowski, J. Teul, J. L. Martin-Ventura, J. Egido and C. Barbas, *PLoS ONE*, 2012, **7**, e31982.

(325) Y. Guo, S. Wan, M. Han, Y. Zhao, C. Li, G. Cai, S. Zhang, Z. Sun, X. Hu, H. Cao and Z. Li, *Med Sci Monit*, 2020, **26**, e926766.

(326) G. Wang, L. Bi, G. Wang, F. Huang, M. Lu and K. Zhu, *Vascular*, 2018, **26**, 301–314.

(327) Q. Guo, X. Xu, X. Li, Y. Mao, S. Li, Y. Yao, X. Li, Y. Li, J. Feng, Y. Shu and X. Xu, *Frontiers in Pharmacology*, 2024, **15**, 1514293.

(328) Y. Hou, W. Guo, T. Fan, B. Li, W. Ge, R. Gao and J. Wang, *Frontiers in Cardiovascular Medicine*, 2021, **8**, 630269.

(329) O. Deda, E. Panteris, T. Meikopoulos, O. Begou, T. Mouskeftara, E. Karagiannidis, A. S. Papazoglou, G. Sianos, G. Theodoridis and H. Gika, *Biomolecules*, 2022, **12**, 354.

(330) Z. Zhang, T. Lv, X. Wang, M. Wu, R. Zhang, X. Yang, Y. Fu and Z. Liu, *Biomedicine & Pharmacotherapy*, 2024, **174**, 116567.

(331) M. T. Politi, J. C. Ferreira and C. M. Patino, *J Bras Pneumol*, 2021, **47**, 20210292.

(332) B. Rodriguez-Vila, J. Tarjuelo-Gutierrez, P. Sánchez-González, P. Verbrugghe, I. Fourneau, G. Maleux, P. Herijgers and E. J. Gomez, 2015, DOI: `10.1155/2015/202539`.

(333) E. Jung, S. Y. Kong, Y. S. Ro, H. H. Ryu and S. D. Shin, *International Journal of Environmental Research and Public Health*, 2022, **19**, 8272.

(334) V. Kotu and B. Deshpande, in ed. V. Kotu and B. Deshpande, Morgan Kaufmann, Second Edition, 2019, ch. 4, pp. 65–163.

(335) M. Sysi-Aho, M. Katajamaa, L. Yetukuri and M. Oresic, *BMC Bioinformatics*, 2007, **8**, 93.

(336) S. M. Kohl, M. S. Klein, J. Hochrein, P. J. Oefner, R. Spang and W. Gronwald, *Metabolomics*, 2011, **8**, 146.

(337) H. C. Köfeler, R. Ahrends, E. S. Baker, K. Ekroos, X. Han, N. Hoffmann, M. Holcapek, M. R. Wenk and G. Liebisch, *Journal of Lipid Research*, 2021, **62**, 100138.

(338) C. Vitali, H. G. Janssen, F. S. Ruggeri and M. W. Nielen, *Analytical Chemistry*, 2023, **95**, 1395–1401.

(339) X. Xiao, L. L. Miller, K. J. Parchert, D. Hayes and J. M. Hochrein, *Rapid Communications in Mass Spectrometry*, 2016, **30**, 1639–1646.

(340) H. R. Tan, L. Y. Chan, A. Ong, Y.-Q. Xu, X.-B. Zhang and W. Zhou, *Food Chemistry*, 2023, **408**, 135135.

(341) S. Ollivier, P. Jéhan, F. Lambert, D. Olivier-Jimenez, J. Boustie, F. L.-L. Dévéhat and N. L. Yondre, *Phytochemical Analysis*, 2022, **33**, 1028–1035.

(342) Z. Pang, Y. Lu, G. Zhou, F. Hui, L. Xu, C. Viau, A. F. Spigelman, P. E. Macdonald, D. S. Wishart, S. Li and J. Xia, *Nucleic Acids Research*, 2024, **52**, W398–W406.

(343) F. Traquete, J. Luz, C. Cordeiro, M. S. Silva and A. E. N. Ferreira, *Metabolites*, 2021, **11**, 788.

(344) D. M. Pott, S. Osorio and J. G. Vallarino, *Frontiers in Plant Science*, 2019, **10**, 835.

(345) K. Tamama, *Clinica Chimica Acta*, 2021, **514**, 40–47.

(346) B. Balluff, C. Hopf, T. P. Siegel, H. I. Grabsch and R. M. Heeren, *Journal of the American Society for Mass Spectrometry*, 2021, **32**, 628.

(347) S. X. Phua, K. P. Lim and W. W. B. Goh, *Computational and Structural Biotechnology Journal*, 2022, **20**, 4369–4375.

(348) Q. Liu, D. Walker, K. Uppal, Z. Liu, C. Ma, V. L. Tran, S. Li, D. P. Jones and T. Yu, *Scientific Reports 2020 10:1*, 2020, **10**, 1–13.

(349) O. M. Kvalheim, F. Brakstad and Y.-Z. Liang7, *Anal. Chem*, 1994, **66**, 43–51.

(350) S. C. Hicks and R. A. Irizarry, *bioRxiv*, 2014, 012203.

(351) Y. Zhao, L. Wong and W. W. B. Goh, *Scientific Reports 2020 10:1*, 2020, **10**, 1–11.

(352) A. Behdenna, M. Colange, J. Haziza, A. Gema, G. Appé, C. A. Azencott and A. Nordor, *BMC Bioinformatics*, 2023, **24**, 1–9.

(353) A. Hyvärinen and E. Oja, *Neural Networks*, 2000, **13**, 411–430.

(354) W. E. Johnson, C. Li and A. Rabinovic, *Biostatistics (Oxford, England)*, 2007, **8**, 118–127.

(355) C. R. Ferreira, K. E. Yannell, A. K. Jarmusch, V. Pirro, Z. Ouyang and R. G. Cooks, *Clinical chemistry*, 2015, **62**, 99.

(356) T. H. Kuo, E. P. Dutkiewicz, J. Pei and C. C. Hsu, *Analytical Chemistry*, 2020, **92**, 2353–2363.

(357) *Software and AI as a Medical Device Change Programme: Gov.UK Roadmap*, `https://www.gov.uk/government/publications/software-and-ai-as-a-medical-device-change-programme/software-and-ai-as-a-medical-device-change-programme-roadmap`.

(358) Office of Regulatory Policy Center for Drug Evaluation and Research, *Validation of Analytical Procedures Q2(R2)*, 2022, `https://www.ema.europa.eu/en/ich-q2r2-validation-analytical-procedures-scientific-guideline`.

(359) R. McGuire and C. Walsh, 'Investigating the functionality of RADIAN-ASAP for the assay and identification of a compound in pharmaceutical drug development', Conference, BMSS Ambient Ionisation SIG Meeting 2024 - 20 Years of Ambient ionisation - A celebration Birmingham, 2024.

(360) E. Manoli, J. Higginson, N. Tolley, A. Darzi, J. Kinross, B. Temelkuran and Z. Takats, *Scientific Reports 2024 14:1*, 2024, **14**, 1–11.

(361) V. Plekhova, L. V. Meulebroek, M. D. Graeve, A. Perdones-Montero, M. D. Spiegeleer, E. D. Paepe, E. V. de Walle, Z. Takats, S. J. Cameron and L. Vanhaecke, *Nature Protocols 2021 16:9*, 2021, **16**, 4327–4354.

(362) N. Ogrinc, A. Kruszewski, P. Chaillou, P. Saudemont, C. Lagadec, M. Salzet, C. Duriez and I. Fournier, *Analytical Chemistry*, 2021, **93**, 14383–14391.

(363) N. S. García-Rojas, H. Guillén-Alonso, S. MacKay, C. Torres-Calzada, L. D. Soto-Rodriguez, R. Winkler and D. S. Wishart, *HardwareX*, 2024, **19**, e00551.

(364) S. Mathias, D. Burns, T. Hambidge, B. J. McCullough, C. J. Hopley, D. Douce, A. Sage and P. Sears, *Drug Testing and Analysis*, 2024, **16**, 807–816.

(365) T. Züllig, M. Trötzmüller and H. C. Köfeler, *Analytical and Bioanalytical Chemistry*, 2020, **412**, 2191–2209.

(366) B. A. Neely, Y. Perez-Riverol and M. Palmblad, *Journal of Biomolecular Techniques*, 2024, **35**, DOI: `10.7171/3FC1F5FE.42308A9A`.

(367) M. Tzafetas, A. Mitra, M. Paraskevaidi, Z. Bodai, I. Kalliala, S. Bowden, K. Lathouras, F. Rosini, M. Szasz, A. Savage, J. Balog, J. McKenzie, D. Lyons, P. Bennett, D. MacIntyre, S. Ghaem-Maghami, Z. Takats and M. Kyrgiou, *Proceedings of the National Academy of Sciences of the United States of America*, 2020, **117**, 7338–7346.

(368) J. Zhang, J. Rector, J. Q. Lin, J. H. Young, M. Sans, N. Katta, N. Giese, W. Yu, C. Nagi, J. Suliburk, J. Liu, A. Bensussan, R. J. Dehoog, K. Y. Garza, B. Ludolph, A. G. Sorace, A. Syed, A. Zahedivash, T. E. Milner and L. S. Eberlin, *Science translational medicine*, 2017, **9**, eaan3968.

(369) D. Calligaris, I. Norton, D. R. Feldman, J. L. Ide, I. F. Dunn, L. S. Eberlin, R. G. Cooks, F. A. Jolesz, A. J. Golby, S. Santagata and N. Y. Agar, *Journal of mass spectrometry : JMS*, 2013, **48**, 1178.

(370) L. Xiong, S. Sun, X. Lu, X. Wang, Q. Yu and X. Qian, *Talanta*, 2024, **278**, 126476.

(371) Y. Ren, H. Wang, J. Liu, Z. Zhang, M. N. McLuckey and Z. Ouyang, *Chromatographia*, 2013, **76**, 1339.

(372) D. A. Rickert, G. A. Gómez-Ríos, E. Nazdrajić, M. Tascon, V. Kulasingam and J. B. Pawliszyn, *Analytical and Bioanalytical Chemistry*, 2020, **412**, 5067–5076.

(373) G. A. Gómez-Ríos, M. Tascon, N. Reyes-Garcés, E. Boyacl, J. Poole and J. Pawliszyn, *Scientific Reports 2017 7:1*, 2017, **7**, 1–7.

(374) S. S. Markin, E. A. Ponomarenko, Y. A. Romashova, T. O. Pleshakova, S. V. Ivanov, V. V. Beregovykh, S. L. Konstantinov, G. I. Stryabkova, Z. Y. Chefranova, Y. A. Lykov, I. M. Karamova, A. G. Koledinskii, K. M. Shestakova, P. A. Markin, N. E. Moskaleva and S. A. Appolonova, *Scientific Reports 2024 14:1*, 2024, **14**, 1–17.

(375) M. Kohlhauer, S. Dawkins, A. S. Costa, R. Lee, T. Young, V. R. Pell, R. P. Choudhury, A. P. Banning, R. K. Kharbanda, K. Saeb-Parsy, M. P. Murphy, C. Frezza, T. Krieg and K. M. Channon, *Journal of the American Heart Association*, 2018, **7**, DOI: 10.1161/JAHA.117.007546.

(376) M. Deidda, C. Piras, G. Binaghi, D. Congia, A. Pani, A. Boi, F. Sanna, A. Rossi, B. Loi, C. C. Dessalvi, L. Atzori, M. Porcu and G. Mercuro, *Scientific Reports 2019 9:1*, 2019, **9**, 1–5.

(377) G. Gundogdu, O. Senol, F. D. Miloglu, Y. Koza, F. Gundogdu, A. Hacımüftüoğlu and A. M. A. El-Aty, *Biomedical Chromatography*, 2020, **34**, e4738.

(378) M. M. Ibrahim, *Obesity Reviews*, 2010, **11**, 11–18.

(379) I. H. Sarker, *SN Computer Science*, 2021, **2**, 1–20.

(380) M. A. Al-Garadi, Y. C. Yang, H. Cai, Y. Ruan, K. O'Connor, G. H. Graciela, J. Perrone and A. Sarker, *BMC Medical Informatics and Decision Making*, 2021, **21**, 1–13.

(381) Y. Bouras and L. Li, *Buildings 2023, Vol. 13, Page 2252*, 2023, **13**, 2252.

(382) F. Es-sabery, A. Hair, J. Qadir, B. de Abajo, B. Garcia-Zapirain and I. la Torre Díez, *IEEE Access*, 2021, **PP**, 1.

(383) D. Crusius, F. Cipcigan and P. C. Biggin, *Faraday Discussions*, 2025, DOI: 10.1039/D4FD00091A.

(384) M. C. Chambers, B. MacLean, R. Burke, D. Amodei, D. L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, K. Hoff, D. Kessner, N. Tasman, N. Shulman, B. Frewen, T. A. Baker, M. Y. Brusniak, C. Paulse, D. Creasy, L. Flashner, K. Kani, C. Moulding, S. L. Seymour, L. M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E. W. Deutsch, R. L. Moritz, J. E. Katz, D. B. Agus, M. MacCoss, D. L. Tabb and P. Mallick, *Nature Biotechnology 2012 30:10*, 2012, **30**, 918–920.

(385) *ProteoWizard*, https://proteowizard.sourceforge.io/.

(386) I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, *Science Robotics*, 2014, **3**, 2672–2680.

(387) *Deep Convolutional Generative Adversarial Network*, `https://www.tensorflow.org/tutorials/generative/dcgan`.

(388) MathWorks, *Train Generative Adversarial Network (GAN)*, `https://uk.mathworks.com/help/deeplearning/ug/train-generative-adversarial-network.html`.

(389) *pygan PyPI*, `https://pypi.org/project/pygan/`.

(390) *NEAT Overview — NEAT-Python 0.92 documentation*, `https://neat-python.readthedocs.io/en/latest/neat_overview.html`.

(391) H. Alibrahim and S. A. Ludwig, *2021 IEEE Congress on Evolutionary Computation, CEC 2021 - Proceedings*, 2021, 1551–1559.

(392) N. Basha, G. Ibrahim, H. A. Choudhury, M. S. Challiwala, R. Fezai, B. Malluhi, H. Nounou, N. Elbashir and M. Nounou, *Gas Science and Engineering*, 2023, **113**, 204964.

(393) I. Bello, B. Zoph, V. Vasudevan and Q. V. Le, Proceedings of the 34th International Conference on Machine Learning, ed. D. Precup and Y. W. Teh, PMLR, 2017, vol. 70, pp. 459–468.

(394) Analytics Vidhya, *Tuning the Hyperparameters and Layers of Neural Network Deep Learning*, `https://www.analyticsvidhya.com/blog/2021/05/tuning-the-hyperparameters-and-layers-of-neural-network-deep-learning/`.

(395) A. Bulat, J. Yang and G. Tzimiropoulos, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, **11210 LNCS**, 187–202.

(396) M. Krasser, *Single image super-resolution with deep neural networks*, 2019, `https://krasserm.github.io/2019/09/04/super-resolution/`.

Appendices

A

Appendix: Method development for the analysis of biological samples by ASAP-MS

A.1 Optimisation of tissue acquisition time

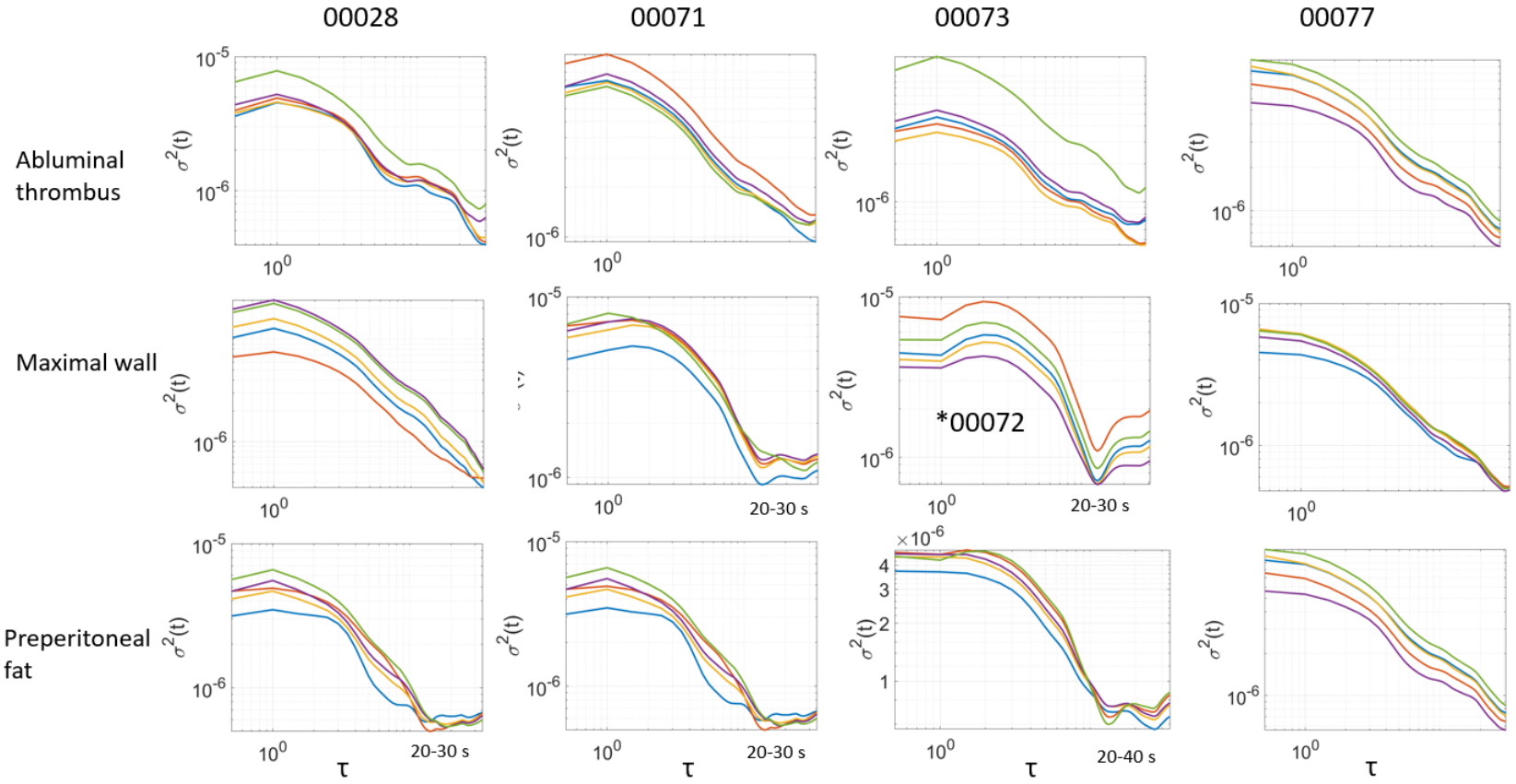

Figure A.1 The Allan variance analysis of three tissue types: abluminal thrombus (upper), maximal wall (centre), and preperitoneal fat (lower) across four patients. Each colour represents one of five individual repeat measurements.

A.2 Optimisation of tissue repeat measurements

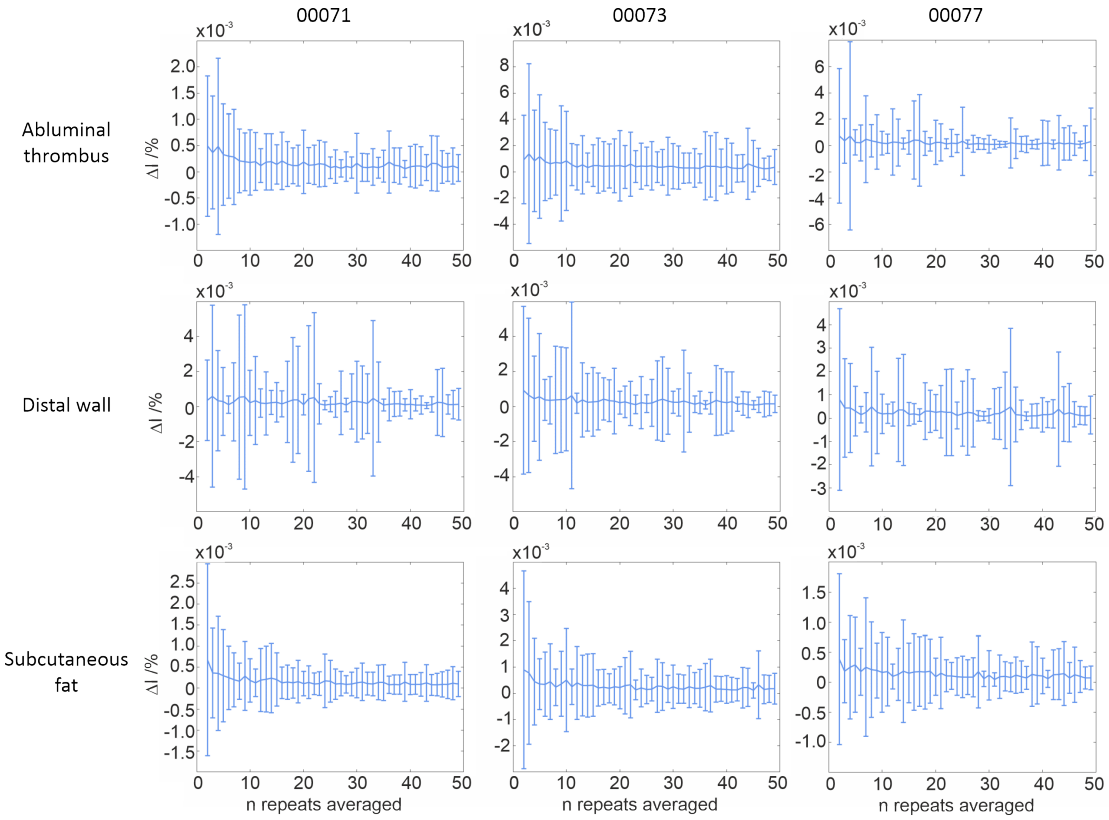

Figure A.2 The change in ΔI_N as a function of number of repeats averaged for three different tissue subtypes: abluminal thrombus (top), distal wall (centre), and subcutaneous fat (lower) across three different patients: patient 00071 (left), patient 00073 (centre) and patient 00077 (right).

B

Appendix: Prediction of STEMI patient outcomes using ASAP-MS and ML analysis

B.1 ML classification methods

The models were trained within MATLAB using the functions `fitcknn`, `fitcsvm`, `fitcdiscr`, `fitcnb`, and `fitcensemble`, for the KNN, SVM, LDA, NBC, and RFC algorithms, respectively. Hyperparameter optimisation was conducted for each model using the MATLAB argument of `OptimiseHyperparameters` inside each ML fitting function with the `'auto'` keyword, prior to completing the classification analysis. The parameters optimised are shown in Table B.1.

Table B.1 Hyperparameter optimisation for ML classification models.

Model	Hyperparameters optimised
KNN	Number of neighbours
	Distance metric
SVM	Box constraint
	Kernel Scale
LDA	Delta
	Gamma
NBC	Distribution names
	Width
RFC	Method
	Number of learning cycles
	Learn rate
	For tree method: minimum leaf size
	For discriminant method: Delta and gamma
	For KNN method: Distance and Number of neighbours

B.2 ML classification results - OxAMI 'All patients'

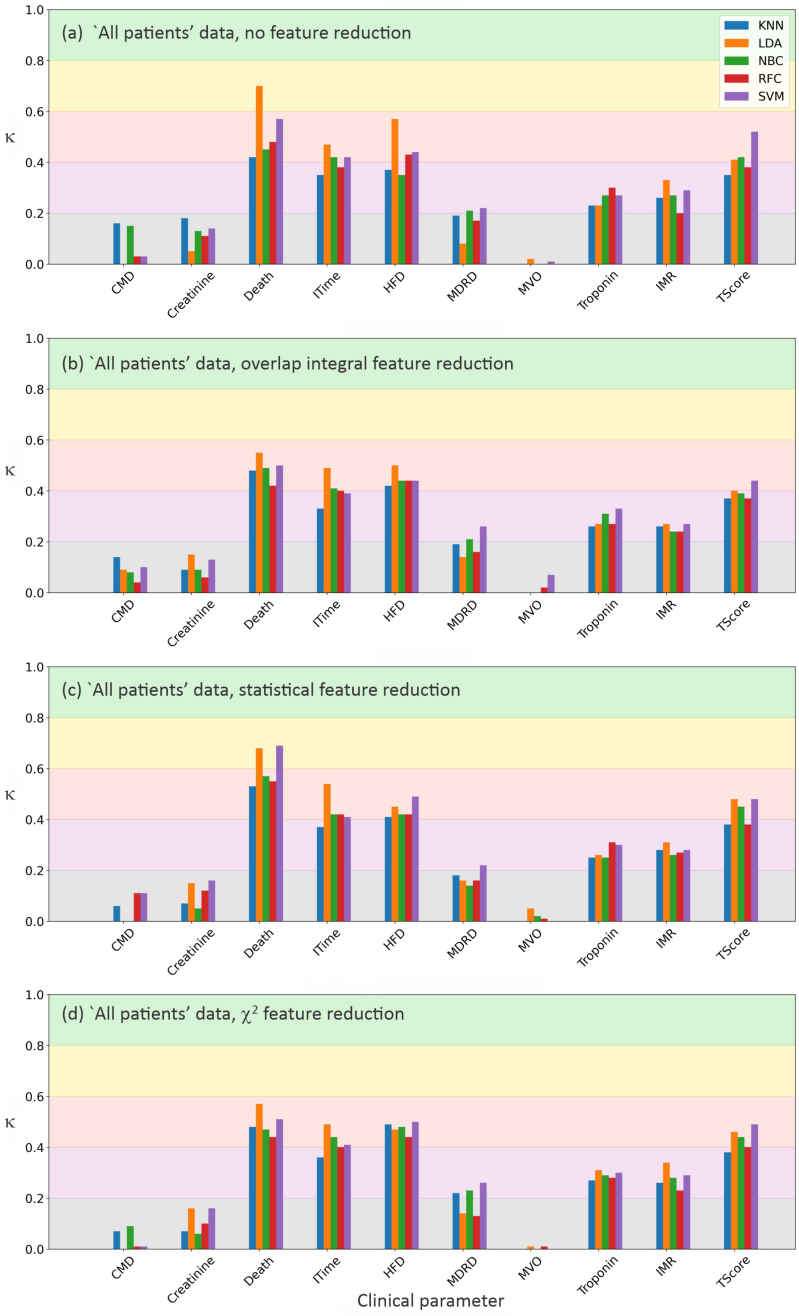

Figure B.1 The κ scores obtained from analysing 'All patients' data set with: (a) no feature reduction; (b) feature reduction with overlap integral method; (c) feature reduction with statistical method; and (d) feature reduction by χ^2 method. Bars show the κ score obtained for each ML model, shading shows the level of agreement, with *no agreement* in grey, *low agreement* in purple, *fair agreement* in light red, *moderate agreement* in yellow, and *strong agreement* in green.

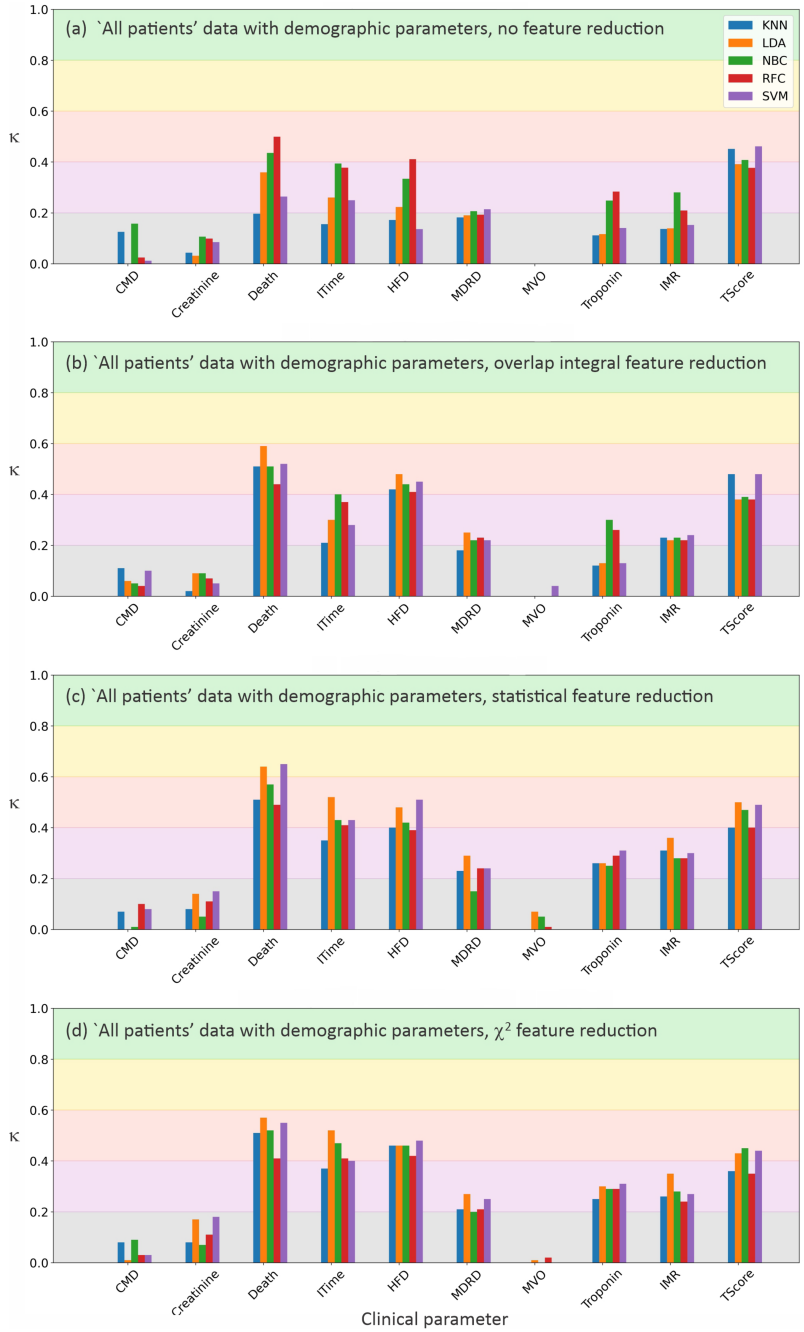

Figure B.2 The κ scores obtained from analysing 'All patients' data set with the inclusion of demographic data, and: (a) no feature reduction; (b) feature reduction with overlap integral method; (c) feature reduction with statistical method; and (d) feature reduction by χ^2 method. Bars show the κ score obtained for each ML model, shading shows the level of agreement, with *no agreement* in grey, *low agreement* in purple, *fair agreement* in light red, *moderate agreement* in yellow, and *strong agreement* in green.

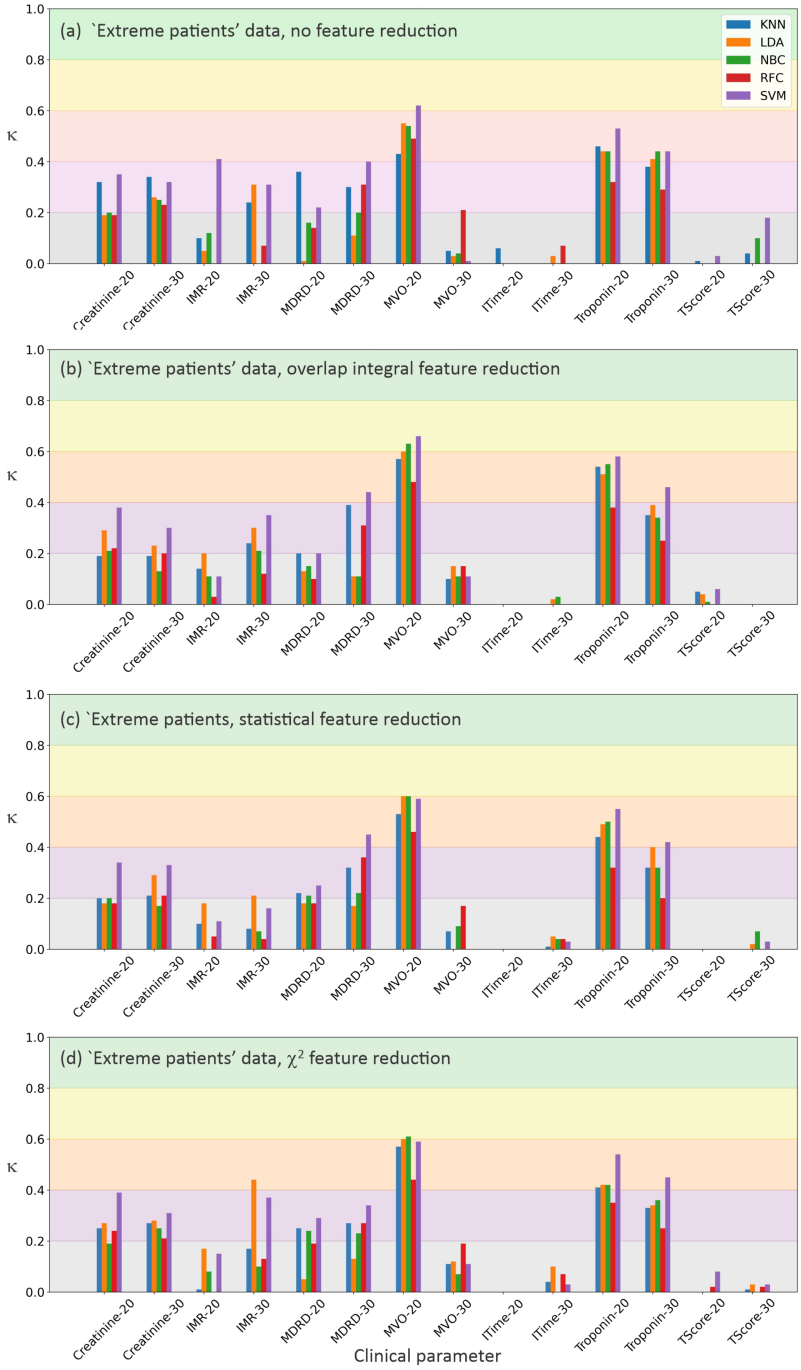

Figure B.3 The κ scores obtained from analysing 'Extreme patients' data set with: (a) no feature reduction; (b) feature reduction with overlap integral method; (c) feature reduction with statistical method; and (d) feature reduction by χ^2 method. Bars show the κ score obtained for each ML model, shading shows the level of agreement, with *no agreement* in grey, *low agreement* in purple, *fair agreement* in light red, *moderate agreement* in yellow, and *strong agreement* in green.

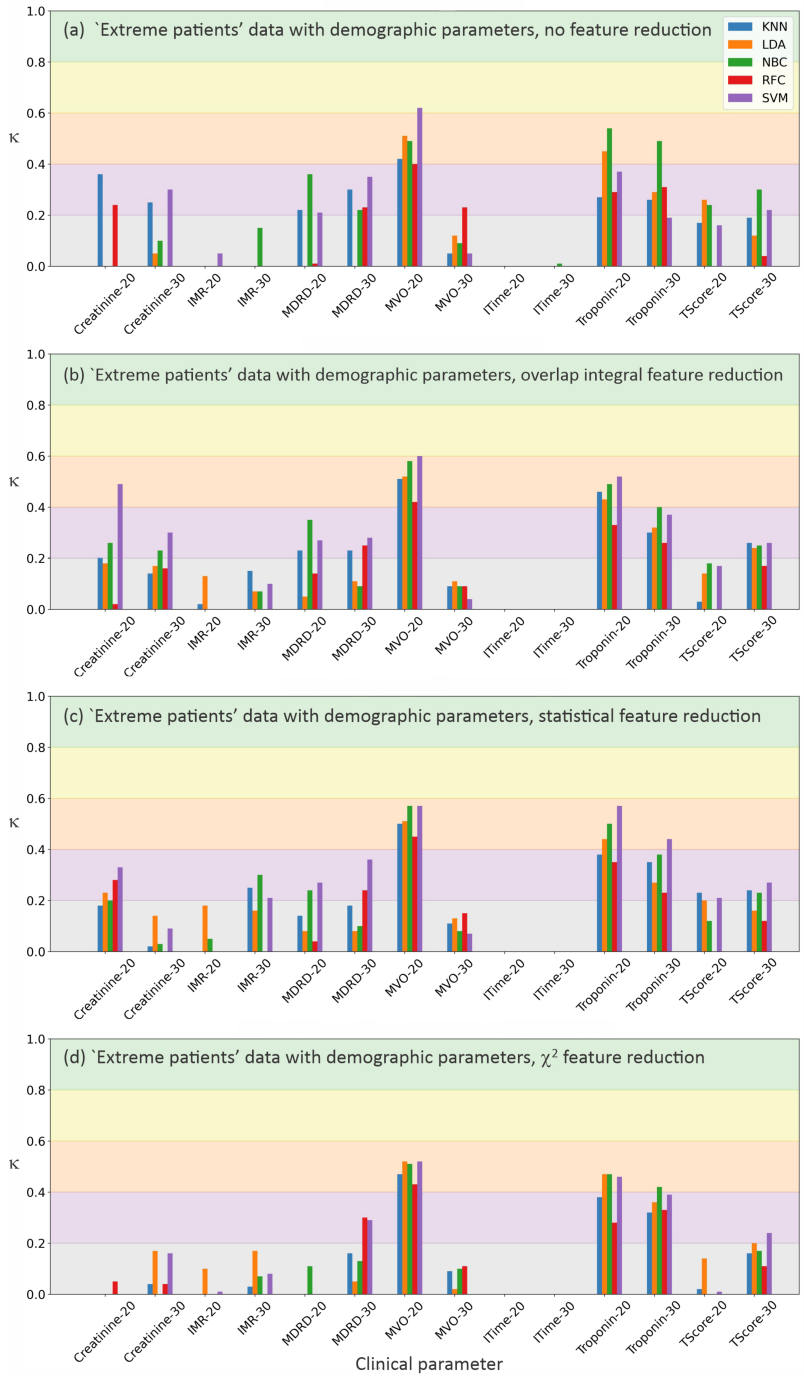

Figure B.4 The κ scores obtained from analysing 'Extreme patients' data sets with inclusion of clinical data and: (a) no feature reduction; (b) feature reduction with overlap integral method; (c) feature reduction with statistical method; and (d) feature reduction by χ^2 method. Bars show the κ score obtained for each ML model , shading shows the level of agreement, with *no agreement* in grey, *low agreement* in purple, *fair agreement* in light red, *moderate agreement* in yellow, and *strong agreement* in green.

B.3 ML full results for the classification of OxAMI data

Table B.2 The full ML results for the classification of OxAMI data by five ML classification algorithms.

Clinical variable	KNN kappa	SVM kappa	LDA kappa	NBC kappa	RFC kappa
Creatinine	0.158±0.134	0.153±0.113	0.053±0.128	0.121±0.128	0.089±0.14
Mortality	0.448±0.249	0.591±0.229	0.706±0.212	0.437±0.268	0.493±0.274
ITime	0.308±0.201	0.439±0.196	0.448±0.214	0.396±0.197	0.387±0.218
HF	0.4±0.264	0.405±0.247	0.544±0.257	0.347±0.263	0.393±0.278
MDRD	0.18±0.145	0.229±0.141	0.081±0.161	0.197±0.15	0.166±0.155
MVO	-0.041±0.183	0.004±0.168	0.008±0.169	-0.04±0.148	-0.001±0.184
Troponin	0.218±0.142	0.262±0.127	0.223±0.139	0.252±0.131	0.265±0.14
IMR	0.242±0.173	0.281±0.179	0.324±0.17	0.257±0.172	0.225±0.168
Tscore	0.337±0.201	0.54±0.184	0.382±0.189	0.411±0.193	0.383±0.203

Table B.3 The full ML results for the classification of OxAMI data with additional demographic information by five ML classification algorithms.

Clinical variable	KNN kappa	SVM kappa	LDA kappa	NBC kappa	RFC kappa
Creatinine	0.186±0.136	0.172±0.135	0±0	0.123±0.152	0.102±0.132
Mortality	0.384±0.273	0.565±0.269	-0.111±0	0.377±0.267	0.43±0.289
ITime	0.324±0.186	0.433±0.192	0±0	0.481±0.223	0.409±0.208
HF	0.41±0.281	0.397±0.272	0±0	0.291±0.243	0.417±0.266
MDRD	0.214±0.148	0.229±0.148	0±0	0.204±0.153	0.199±0.149
MVO	-0.026±0.18	0.031±0.165	0±0	-0.02±0.16	0.002±0.198
Troponin	0.221±0.142	0.279±0.134	0±0	0.248±0.115	0.275±0.125
IMR	0.23±0.176	0.305±0.17	-0.04±0	0.273±0.187	0.201±0.197
Tscore	0.358±0.194	0.498±0.175	0±0	0.444±0.201	0.381±0.194

Table B.4 The full ML results for the classification of OxAMI data by five ML classification algorithms with χ^2 feature reduction applied.

Clinical variable	KNN kappa	SVM kappa	LDA kappa	NBC kappa	RFC kappa
Creatinine	0.065±0.139	0.164±0.135	0.141±0.124	0.045±0.143	0.118±0.128
Mortality	0.456±0.269	0.496±0.286	0.552±0.265	0.487±0.274	0.415±0.278
ITime	0.38±0.19	0.413±0.202	0.506±0.179	0.443±0.183	0.395±0.216
HF	0.485±0.271	0.505±0.243	0.472±0.258	0.484±0.239	0.458±0.262
MDRD	0.242±0.147	0.266±0.149	0.137±0.145	0.228±0.153	0.144±0.165
MVO	-0.004±0.189	-0.042±0.194	0.007±0.177	-0.018±0.172	-0.014±0.189
Troponin	0.262±0.146	0.297±0.13	0.304±0.129	0.303±0.134	0.28±0.135
IMR	0.232±0.173	0.238±0.179	0.31±0.168	0.255±0.164	0.223±0.18
Tscore	0.388±0.202	0.49±0.192	0.458±0.193	0.456±0.195	0.379±0.204

Table B.5 The full ML results for the classification of OxAMI data with included additional demographic variables by five ML classification algorithms with χ^2 feature reduction applied.

Clinical variable	KNN kappa	SVM kappa	LDA kappa	NBC kappa	RFC kappa
Creatinine	0.083±0.129	0.182±0.137	0.176±0.135	0.053±0.132	0.109±0.127
Mortality	0.49±0.278	0.558±0.251	0.59±0.262	0.52±0.242	0.457±0.283
ITime	0.364±0.173	0.411±0.185	0.503±0.171	0.451±0.182	0.413±0.208
HF	0.456±0.253	0.458±0.24	0.487±0.25	0.45±0.224	0.428±0.251
MDRD	0.221±0.153	0.238±0.145	0.248±0.145	0.219±0.162	0.18±0.145
MVO	-0.002±0.191	-0.01±0.199	0.031±0.197	0.013±0.189	0.03±0.202
Troponin	0.248±0.154	0.288±0.143	0.29±0.138	0.273±0.145	0.288±0.14
IMR	0.248±0.191	0.246±0.186	0.334±0.171	0.261±0.171	0.255±0.188
Tscore	0.397±0.2	0.463±0.212	0.465±0.177	0.446±0.19	0.371±0.211

Table B.6 The full ML results for the classification of OxAMI data by five ML classification algorithms with OI feature reduction applied.

Clinical variable	KNN kappa	SVM kappa	LDA kappa	NBC kappa	RFC kappa
Creatinine	0.095±0.139	0.143±0.123	0.154±0.14	0.076±0.143	0.062±0.129
Mortality	0.444±0.282	0.477±0.28	0.536±0.296	0.454±0.284	0.38±0.293
ITime	0.332±0.188	0.417±0.199	0.487±0.196	0.416±0.191	0.397±0.202
HF	0.391±0.255	0.402±0.245	0.454±0.267	0.431±0.261	0.389±0.271
MDRD	0.196±0.167	0.263±0.153	0.156±0.156	0.219±0.149	0.17±0.151
MVO	-0.026±0.195	0.035±0.174	-0.023±0.187	-0.023±0.192	-0.008±0.199
Troponin	0.268±0.143	0.309±0.14	0.278±0.138	0.316±0.147	0.29±0.128
IMR	0.246±0.181	0.262±0.17	0.27±0.165	0.216±0.183	0.241±0.179
Tscore	0.387±0.212	0.446±0.202	0.397±0.186	0.385±0.194	0.348±0.217

Table B.7 The full ML results for the classification of OxAMI data with additional demographic data by five ML classification algorithms with OI feature reduction applied.

Clinical variable	KNN kappa	SVM kappa	LDA kappa	NBC kappa	RFC kappa
Creatinine	0.034±0.127	0.064±0.14	0.108±0.125	0.088±0.142	0.068±0.127
Mortality	0.47±0.3	0.49±0.285	0.576±0.289	0.499±0.266	0.439±0.276
ITime	0.22±0.184	0.287±0.166	0.321±0.186	0.409±0.199	0.392±0.206
HF	0.431±0.243	0.453±0.226	0.506±0.265	0.457±0.24	0.432±0.259
MDRD	0.164±0.163	0.217±0.146	0.257±0.134	0.218±0.158	0.223±0.158
MVO	-0.013±0.199	0.027±0.162	-0.03±0.184	-0.034±0.177	0.004±0.191
Troponin	0.124±0.124	0.141±0.135	0.135±0.129	0.303±0.122	0.28±0.143
IMR	0.2±0.151	0.229±0.155	0.216±0.168	0.243±0.173	0.231±0.18
Tscore	0.463±0.193	0.464±0.184	0.378±0.186	0.385±0.203	0.364±0.202

Table B.8 The full ML results for the classification of OxAMI data by five ML classification algorithms with statistical feature reduction applied.

Clinical variable	KNN kappa	SVM kappa	LDA kappa	NBC kappa	RFC kappa
Creatinine	0.069±0.116	0.154±0.132	0.134±0.117	0.049±0.125	0.111±0.133
Mortality	0.488±0.281	0.676±0.236	0.656±0.224	0.542±0.245	0.523±0.254
ITime	0.334±0.205	0.398±0.182	0.536±0.189	0.428±0.207	0.396±0.201
HF	0.392±0.267	0.484±0.251	0.51±0.257	0.439±0.257	0.469±0.268
MDRD	0.18±0.154	0.242±0.148	0.157±0.16	0.149±0.161	0.174±0.164
MVO	-0.001±0.195	-0.005±0.183	0.073±0.178	0.045±0.194	0.026±0.178
Troponin	0.256±0.137	0.319±0.128	0.251±0.131	0.266±0.112	0.305±0.129
IMR	0.333±0.188	0.308±0.164	0.322±0.178	0.278±0.178	0.268±0.182
Tscore	0.387±0.188	0.452±0.181	0.467±0.2	0.447±0.208	0.397±0.208

Table B.9 The full ML results for the classification of OxAMI data with additional demographic data included by five ML classification algorithms with statistical feature reduction applied.

Clinical variable	KNN kappa	SVM kappa	LDA kappa	NBC kappa	RFC kappa
Creatinine	0.068±0.125	0.158±0.143	0.136±0.123	0.043±0.13	0.108±0.131
Mortality	0.509±0.298	0.685±0.239	0.645±0.25	0.591±0.247	0.558±0.269
ITime	0.349±0.2	0.399±0.193	0.531±0.172	0.424±0.197	0.406±0.198
HF	0.392±0.248	0.489±0.267	0.493±0.237	0.406±0.242	0.397±0.273
MDRD	0.235±0.144	0.224±0.143	0.286±0.152	0.175±0.137	0.246±0.155
MVO	-0.013±0.194	-0.025±0.208	0.079±0.167	0.046±0.19	0.026±0.192
Troponin	0.246±0.14	0.305±0.127	0.242±0.124	0.244±0.127	0.294±0.13
IMR	0.29±0.195	0.272±0.179	0.346±0.155	0.269±0.175	0.275±0.171
Tscore	0.39±0.202	0.474±0.172	0.497±0.18	0.468±0.203	0.403±0.216

B.4 ML full results for classification of 'extreme' patients data

Table B.10 The full ML results for the classification of 'Extreme' OxAMI data by five ML classification algorithms.

Clinical variable	**KNN kappa**	**SVM kappa**	**LDA kappa**	**NBC kappa**	**RFC kappa**
Creatinine-20	0.334±0.302	0.373±0.294	0.26±0.326	0.23±0.31	0.217±0.329
Creatinine-30	0.339±0.254	0.35±0.239	0.271±0.262	0.213±0.262	0.183±0.294
IMR-20	0.125±0.342	0.388±0.314	0.063±0.322	0.1±0.291	0.014±0.354
IMR-30	0.263±0.336	0.305±0.291	0.333±0.296	-0.005±0.312	0.04±0.349
MDRD-20	0.366±0.319	0.263±0.318	0.037±0.28	0.148±0.29	0.168±0.379
MDRD-30	0.296±0.253	0.421±0.253	0.057±0.265	0.163±0.264	0.337±0.289
MVO-20	0.436±0.279	0.646±0.259	0.589±0.274	0.533±0.285	0.464±0.306
MVO-30	0.057±0.248	0.015±0.232	0.054±0.217	0.05±0.248	0.239±0.283
Pain-20	0.081±0.335	-0.111±0.31	-0.03±0.273	-0.064±0.278	-0.07±0.314
Pain-30	-0.007±0.272	-0.115±0.193	-0.03±0.22	-0.065±0.238	0.034±0.27
Troponin-20	0.47±0.317	0.51±0.28	0.438±0.291	0.454±0.275	0.318±0.309
Troponin-30	0.379±0.272	0.46±0.23	0.413±0.251	0.445±0.245	0.253±0.254
TScore-20	0.03±0.315	0.01±0.295	-0.063±0.239	-0.02±0.226	-0.065±0.319
TScore-30	-0.001±0.262	0.14±0.242	-0.011±0.209	0.079±0.222	-0.054±0.256

Table B.11 The full ML results for the classification of 'Extreme' OxAMI data with additional demographic data included by five ML classification algorithms.

Clinical variable	KNN kappa	SVM kappa	LDA kappa	NBC kappa	RFC kappa
Creatinine-20	0.37±0.629	-0.055±0.391	-0.035±0.272	-0.07±0.407	0.16±0.661
Creatinine-30	0.272±0.365	0.302±0.342	0.067±0.249	0.091±0.304	0.021±0.339
IMR-20	-0.016±0.318	0.029±0.324	-0.045±0.244	0.017±0.234	-0.19±0.332
IMR-30	0±0.301	-0.1±0.317	-0.035±0.307	0.085±0.344	-0.205±0.44
MDRD-20	0.154±0.489	0.214±0.488	0.068±0.373	0.322±0.481	0.067±0.478
MDRD-30	0.278±0.378	0.365±0.333	-0.052±0.259	0.297±0.329	0.215±0.437
MVO-20	0.393±0.324	0.602±0.294	0.546±0.279	0.501±0.305	0.418±0.303
MVO-30	0.078±0.233	0.088±0.256	0.119±0.262	0.065±0.255	0.194±0.306
Pain-20	-0.063±0.38	-0.254±0.219	-0.195±0.282	-0.238±0.281	-0.15±0.364
Pain-30	-0.127±0.307	-0.088±0.193	-0.108±0.247	0.006±0.277	0.022±0.369
Troponin-20	0.312±0.386	0.354±0.382	0.41±0.371	0.522±0.355	0.222±0.389
Troponin-30	0.268±0.305	0.227±0.298	0.254±0.299	0.491±0.307	0.339±0.322
TScore-20	0.153±0.52	0.187±0.506	0.201±0.499	0.229±0.489	-0.13±0.484
TScore-30	0.282±0.378	0.233±0.332	0.052±0.305	0.355±0.358	0.088±0.351

Table B.12 The full ML results for the classification of 'Extreme' OxAMI data by five ML classification algorithms with χ^2 feature reduction applied.

Clinical variable	KNN kappa	SVM kappa	LDA kappa	NBC kappa	RFC kappa
Creatinine-20	0.175±0.329	0.293±0.317	0.158±0.336	0.211±0.328	0.218±0.318
Creatinine-30	0.237±0.242	0.327±0.277	0.292±0.269	0.176±0.27	0.27±0.3
IMR-20	0.059±0.336	0.14±0.315	0.16±0.324	-0.003±0.298	0.071±0.332
IMR-30	0.066±0.315	0.156±0.364	0.198±0.323	0.124±0.335	0.023±0.324
MDRD-20	0.238±0.368	0.293±0.374	0.203±0.299	0.226±0.303	0.19±0.36
MDRD-30	0.295±0.308	0.44±0.26	0.18±0.288	0.228±0.265	0.372±0.274
MVO-20	0.502±0.298	0.554±0.302	0.562±0.27	0.574±0.278	0.478±0.31
MVO-30	0.058±0.261	-0.006±0.259	0.062±0.26	0.109±0.289	0.125±0.273
Pain-20	-0.045±0.305	-0.06±0.331	-0.078±0.327	-0.063±0.336	-0.141±0.304
Pain-30	0.026±0.26	0.051±0.286	0.047±0.281	0.049±0.267	0.104±0.283
Troponin-20	0.448±0.299	0.57±0.271	0.503±0.272	0.493±0.289	0.36±0.306
Troponin-30	0.323±0.25	0.394±0.224	0.39±0.231	0.277±0.275	0.24±0.231
TScore-20	-0.074±0.323	-0.063±0.34	-0.056±0.292	-0.015±0.338	-0.075±0.331
TScore-30	-0.022±0.273	0.018±0.29	0.007±0.279	0.078±0.286	-0.073±0.262

Table B.13 The full ML results for the classification of 'Extreme' OxAMI data with additional demographic data included by five ML classification algorithms with χ^2 feature reduction applied.

Clinical variable	KNN kappa	SVM kappa	LDA kappa	NBC kappa	RFC kappa
Creatinine-20	-0.235±0.584	-0.055±0.378	-0.13±0.494	0.005±0.526	0.06±0.685
Creatinine-30	0.034±0.362	0.162±0.363	0.1±0.39	0.04±0.359	0.079±0.378
IMR-20	-0.057±0.365	0.003±0.369	0.139±0.426	-0.039±0.373	-0.178±0.376
IMR-30	0.14±0.549	0.155±0.471	0.115±0.503	0.22±0.56	-0.105±0.474
MDRD-20	-0.035±0.459	-0.106±0.303	-0.119±0.406	0.18±0.476	-0.164±0.41
MDRD-30	0.108±0.408	0.282±0.411	0.097±0.385	0.165±0.385	0.313±0.409
MVO-20	0.458±0.279	0.524±0.29	0.423±0.286	0.484±0.287	0.378±0.327
MVO-30	0.06±0.265	0.009±0.279	0.009±0.277	0.094±0.315	0.131±0.267
Pain-20	-0.162±0.362	-0.176±0.289	-0.078±0.418	-0.168±0.383	-0.174±0.389
Pain-30	-0.144±0.323	-0.124±0.291	-0.102±0.335	-0.139±0.328	-0.088±0.35
Troponin-20	0.408±0.366	0.498±0.366	0.429±0.37	0.436±0.352	0.334±0.375
Troponin-30	0.308±0.351	0.365±0.362	0.33±0.344	0.395±0.32	0.344±0.364
TScore-20	0.016±0.484	0.036±0.556	0.071±0.562	-0.026±0.561	-0.158±0.469
TScore-30	0.142±0.38	0.192±0.354	0.147±0.351	0.14±0.376	0.088±0.338

Table B.14 The full ML results for the classification of 'Extreme' OxAMI data by five ML classification algorithms with OI feature reduction applied.

Clinical variable	KNN kappa	SVM kappa	LDA kappa	NBC kappa	RFC kappa
Creatinine-20	0.253±0.37	0.412±0.326	0.324±0.376	0.139±0.332	0.273±0.372
Creatinine-30	0.27±0.276	0.33±0.255	0.293±0.273	0.232±0.256	0.185±0.282
IMR-20	0.038±0.316	0.173±0.35	0.18±0.322	0.103±0.311	-0.056±0.311
IMR-30	0.199±0.306	0.335±0.349	0.404±0.337	0.06±0.347	0.143±0.375
MDRD-20	0.257±0.397	0.37±0.356	0.129±0.338	0.208±0.325	0.23±0.364
MDRD-30	0.265±0.284	0.403±0.328	0.133±0.288	0.278±0.24	0.306±0.294
MVO-20	0.571±0.283	0.61±0.276	0.631±0.265	0.654±0.274	0.493±0.293
MVO-30	0.088±0.297	0.146±0.257	0.122±0.257	0.069±0.269	0.194±0.274
Pain-20	-0.048±0.307	-0.166±0.318	-0.105±0.311	-0.14±0.294	-0.089±0.345
Pain-30	0.002±0.28	0.042±0.28	0.09±0.278	-0.005±0.285	0.071±0.274
Troponin-20	0.438±0.313	0.579±0.279	0.408±0.302	0.439±0.302	0.361±0.301
Troponin-30	0.341±0.263	0.443±0.234	0.324±0.265	0.379±0.253	0.226±0.248
TScore-20	-0.026±0.339	0.011±0.353	-0.088±0.325	-0.125±0.303	-0.056±0.316
TScore-30	0.032±0.278	0.05±0.253	0.028±0.241	-0.029±0.233	-0.003±0.256

Table B.15 The full ML results for the classification of 'Extreme' OxAMI data with additional demographic data included by five ML classification algorithms with OI feature reduction applied.

Clinical variable	KNN kappa	SVM kappa	LDA kappa	NBC kappa	RFC kappa
Creatinine-20	0.08±0.637	0.295±0.547	0.135±0.631	0.065±0.666	0.175±0.638
Creatinine-30	0.043±0.353	0.095±0.337	0.101±0.341	0.003±0.401	-0.005±0.313
IMR-20	-0.079±0.345	-0.076±0.361	0.133±0.442	0.023±0.383	-0.178±0.373
IMR-30	0.215±0.557	0.095±0.487	0.195±0.591	0.26±0.494	-0.14±0.437
MDRD-20	0.127±0.483	0.266±0.558	0.108±0.502	0.314±0.54	-0.025±0.481
MDRD-30	0.245±0.432	0.415±0.377	0.078±0.388	0.163±0.372	0.238±0.362
MVO-20	0.503±0.301	0.562±0.301	0.536±0.271	0.537±0.283	0.414±0.292
MVO-30	0.129±0.27	0.073±0.315	0.115±0.318	0.076±0.308	0.175±0.313
Pain-20	-0.18±0.378	-0.231±0.299	-0.239±0.361	-0.222±0.343	-0.216±0.401
Pain-30	-0.101±0.35	-0.071±0.268	-0.084±0.333	-0.212±0.296	-0.075±0.348
Troponin-20	0.48±0.398	0.57±0.397	0.492±0.363	0.522±0.35	0.388±0.396
Troponin-30	0.338±0.35	0.45±0.339	0.288±0.332	0.371±0.332	0.228±0.323
TScore-20	0.172±0.534	0.246±0.453	0.176±0.581	0.139±0.486	-0.085±0.493
TScore-30	0.34±0.368	0.328±0.338	0.237±0.374	0.25±0.363	0.1±0.38

Table B.16 The full ML results for the classification of 'Extreme' OxAMI data by five ML classification algorithms with statistical feature reduction applied.

Clinical variable	KNN kappa	SVM kappa	LDA kappa	NBC kappa	RFC kappa
Creatinine-20	0.246±0.329	0.383±0.323	0.308±0.35	0.194±0.343	0.183±0.329
Creatinine-30	0.234±0.262	0.333±0.241	0.243±0.259	0.145±0.245	0.17±0.262
IMR-20	0.111±0.331	0.078±0.356	0.191±0.329	0.08±0.372	0.021±0.341
IMR-30	0.228±0.351	0.324±0.331	0.278±0.312	0.175±0.338	0.189±0.332
MDRD-20	0.166±0.39	0.189±0.393	0.084±0.325	0.152±0.298	0.056±0.352
MDRD-30	0.341±0.298	0.43±0.254	0.134±0.249	0.123±0.291	0.317±0.257
MVO-20	0.562±0.27	0.663±0.27	0.576±0.276	0.64±0.269	0.48±0.297
MVO-30	0.14±0.262	0.123±0.265	0.133±0.261	0.069±0.249	0.152±0.27
Pain-20	-0.038±0.298	-0.078±0.299	-0.076±0.298	-0.035±0.291	-0.079±0.322
Pain-30	-0.027±0.276	0.035±0.266	0.022±0.28	0.01±0.27	0.009±0.267
Troponin-20	0.521±0.311	0.561±0.284	0.483±0.28	0.55±0.282	0.351±0.288
Troponin-30	0.376±0.27	0.457±0.242	0.372±0.257	0.349±0.275	0.248±0.268
TScore-20	-0.006±0.318	0.03±0.353	0.008±0.305	-0.013±0.313	-0.023±0.325
TScore-30	-0.055±0.229	-0.064±0.213	-0.104±0.225	-0.067±0.23	-0.1±0.283

Table B.17 The full ML results for the classification of 'Extreme' OxAMI data with additional demographic data included by five ML classification algorithms with statistical feature reduction applied.

Clinical variable	**KNN kappa**	**SVM kappa**	**LDA kappa**	**NBC kappa**	**RFC kappa**
Creatinine-20	0.255±0.73	0.44±0.623	0.2±0.73	0.31±0.69	-0.015±0.526
Creatinine-30	0.111±0.377	0.219±0.332	0.146±0.344	0.214±0.339	0.136±0.377
IMR-20	0.019±0.391	-0.018±0.431	0.101±0.44	0.002±0.441	0.054±0.467
IMR-30	0.165±0.509	0.085±0.547	0.08±0.637	0.09±0.513	-0.22±0.472
MDRD-20	0.325±0.573	0.26±0.579	0.083±0.504	0.394±0.503	0.052±0.541
MDRD-30	0.178±0.426	0.237±0.399	0.103±0.368	0.075±0.368	0.18±0.41
MVO-20	0.503±0.311	0.585±0.296	0.5±0.278	0.592±0.258	0.44±0.301
MVO-30	0.132±0.269	0.082±0.276	0.117±0.274	0.105±0.279	0.112±0.282
Pain-20	-0.175±0.371	-0.136±0.329	-0.214±0.36	-0.139±0.358	-0.18±0.378
Pain-30	-0.08±0.322	-0.05±0.329	-0.064±0.369	-0.034±0.324	-0.059±0.335
Troponin-20	0.476±0.41	0.54±0.385	0.492±0.367	0.54±0.351	0.36±0.387
Troponin-30	0.335±0.339	0.399±0.348	0.353±0.299	0.37±0.318	0.311±0.349
TScore-20	0.02±0.442	0.219±0.467	0.143±0.465	0.106±0.501	-0.106±0.453
TScore-30	0.245±0.345	0.288±0.353	0.215±0.364	0.208±0.373	0.163±0.398

B.5 Significant m/z peaks determined through FR for STEMI group classification

The m/z ion peaks identified as being important for classification of patients according to each clinical parameter by all of the three feature reduction methods employed, with the associated t statistic and p score shown. These data were generated from the 'All patients' data set, with the exception of the data for MVO, which was generated from the MVO-20 data set.

Clinical variable	m/z peaks	t statistic	p score
Mortality	129	4.27	1.02×10^{-4}
	169	4.82	1.75×10^{-5}
	156	5.3	3.62×10^{-6}
	197	4.74	2.26×10^{-5}
HFD	208	4.73	2.00×10^{-5}
	219	2.79	7.59×10^{-3}
	532	5.45	1.73×10^{-6}
	189	-3.9	2.99×10^{-4}
MVO	355	4.77	2.71×10^{-5}
	387	-5.08	1.03×10^{-5}
	531	-4.06	2.37×10^{-4}
	532	-4.41	8.34×10^{-5}
	533	-3.97	3.04×10^{-4}
IMR	45	5.47	2.32×10^{-7}
	46	5.6	1.30×10^{-7}
	538	-2.94	3.89×10^{-3}
ITime	45	5.11	1.72×10^{-6}
	46	-5.72	1.29×10^{-7}
	387	4.65	1.11×10^{-4}
TScore	45	-5.64	1.89×10^{-7}
	46	-5.87	6.92×10^{-8}
	115	2.50	1.42×10^{-2}
MVO	45	-2.65	8.63×10^{-3}
	46	4.38	1.82×10^{-5}
	49	3.38	8.71×10^{-4}
	370	-5.92	1.23×10^{-8}
	371	-5.72	3.50×10^{-8}
	576	4.84	2.50×10^{-6}
Creatinine	114	-2.91	3.91×10^{-3}
	62	-4.48	1.14×10^{-5}

B.6 Identification of ion peaks using ASAP-Q-TOF-MS analysis of blood plasma

Peaks found within the OxAMI dataset were matched to ion peaks identified in the same ASAP-Q-TOF data. We note that these peaks may not be associated with the same ions across both instruments due to differences in the mass spectrometers. In all but 1 case, the CMS-S peak corresponded to more than one peak in the QTOF data.

The references investigated for matching molecules: [374], [375], [376], [377].

Table B.18 The peaks identified using ASAP-QTOF-MS in the same m/z bins as for the CMS-S data for significant peaks identified as significant for OxAMI data classification. The error in ppm between the found and reference masses is shown in column 5.

CMS-S peak mass	TOF peak mass	Potential molecular formula (assuming H^+)	Potential biomolecules	Error PPM for protonated version	Source
114	114.036 114.066 114.092	C_5H_7NS $C_4H_7N_3O$ $C_6H_{11}NO$	Creatinine	-10.5 -1.65 5.78	[375]
115	115.053 115.086	C_9H_6 $C_5H_{10}N_2O$	Proline	-10.66 -5.12	[374]
129	128.896 129.067 129.092 129.1	N/A $C_5H_8N_2O_2$ $C_7H_{12}O_2$ $C_4H_{10}N_5$	Dihydrothymine	8.88 7.7 -6.95	[375]
156	156.078 156.136	$C_6H_9N_3O_2$ $C_7H_{15}N_4$	Histidine	7.99 -6.07	[375]

CMS-S peak mass	TOF peak mass	Potential molecular formula (assuming H^+)	Potential biomolecules	Error PPM for protonated version	Source
	156.176	$C_{10}H_{21}N$		8.48	
169	168.975	$C_2H_3NO_6P$		-12.28	
	169.063	$C_5H_{13}O_4P$		3.42	
	169.097	$C_8H_{12}N_2O_2$		-0.91	
	169.126	N/A			
	169.16	$C_{11}H_{20}O$		7.73	
	169.196	$C_{12}H_{24}$		5.45	
189	189.074	$C_6H_{10}N_3O_4$	Glycine-3 tripeptide	-2.15	
	189.101	$C_5H_{13}O_4P$		3.42	
	189.127	$C_{13}H_{16}O$		-2.07	
	189.164	$C_{14}H_{20}$		1.18	
197	197.107	$C_{13}H_{12}N_2$		-1.65	
	197.129	$C_{10}H_{16}N_2O_2$		2.77	
	197.153	$C_{12}H_{20}O_2$		-3.08	
	197.191	$C_{13}H_{24}O$		5.11	
208	208.036	$C_{10}H_7O_5$		-3	
	208.097	$C_{11}H_{13}NO_3$		0.87	
	208.144	$C_{11}H_{17}N_3O$		-2.11	
	208.18	$C_{12}H_{21}N_3$		-3.96	
	208.215	N/A			
219	219.047	$C_6H_8N_3O_6$		-7.24	
	219.174	$C_{15}H_{22}O$		-1.56	
	219.211	$C_{16}H_{26}$		1.24	
355	355.072	$C_{21}H_{10}N_2O_4$		1.88	
	355.291	$C_{20}H_{39}N_2OP$		10.48	

CMS-S peak mass	TOF peak mass	Potential molecular formula (assuming H^+)	Potential bio-molecules	Error PPM for protonated version	Source
	355.33	$C_{21}H_{42}N_2O_2$		-5.36	
	355.358	$C_{23}H_{46}O_2$		2.65	
	355.476	N/A			
	369.543	N/A			
	369.588	N/A			
	369.654	N/A			
370	369.705	N/A			
	369.806	N/A			
	369.895	N/A			
	370.358	$C_{27}H_{45}$		-3.79	
	370.507	N/A			
	370.548	N/A			
	370.593	N/A			
	370.655	N/A			
	370.714	N/A			
	370.765	N/A			
371	370.789	N/A			
	370.811	N/A			
	370.852	N/A			
	370.898	N/A			
	370.94	N/A			
	371.232	$C_{22}H_{30}N_2O_3$		-2.48	
	371.362	$C_{22}H_{46}N_2O_2$		-3.25	
378	378.363	$C_{24}H_{45}N_2O$		6.7	
531	530.539	$C_{33}H_{56}NO_2P$		5.39	
	531.415	$C_{30}H_{59}O_5P$		-4.31	

CMS-S peak mass	TOF peak mass	Potential molecular formula (assuming H^+)	Potential biomolecules	Error PPM for protonated version	Source
352	531.53 532.421	N/A $C_{29}H_{57}NO_7$		 0.41	
533	532.521 533.421 533.473	$C_{36}H_{67}O_2$ $C_{33}H_{56}O_5$ $C_{31}H_{65}O_4P$	 Phosphoric acid tri-C8-16-alkyl esters	-0.72 0.95 6.89	
534	533.533	$C_{36}H_{68}O_2$		7.11	
537	537.378 537.497	$C_{31}H_{52}O_7$ $C_{33}H_{64}N_2O_3$		-1.08 -3.67	
538	538.38 538.5	$C_{31}H_{54}O_5P$ $C_{34}H_{65}O_4$		3.41 8.24	
576	575.514 576.516 576.582	$C_{34}H_{71}O_4P$ $C_{37}H_{67}O_4$ $C_{39}H_{75}O_2$		-3.95 8 8.3	

B.7 PySR hyperparameters

Hyperparameter	Set value
`Iterations`	1000
`Maxsize`	50
`Ncycles_per_iteration`	10000
`Population_size`	1000
`Binary_operators`	"+", "*","-","/"
`Unary_operators`	"exp", "square", "cube", "inv(x) = 1/x"
`Weight_randomize`	0.2
`Extra_sympy_mappings`	"inv": lambda x: 1 / x
`Elementwise_loss`	$loss(prediction, target) = (prediction - target)^2$

B.8 Symbolic regression hyperparameter optimisation

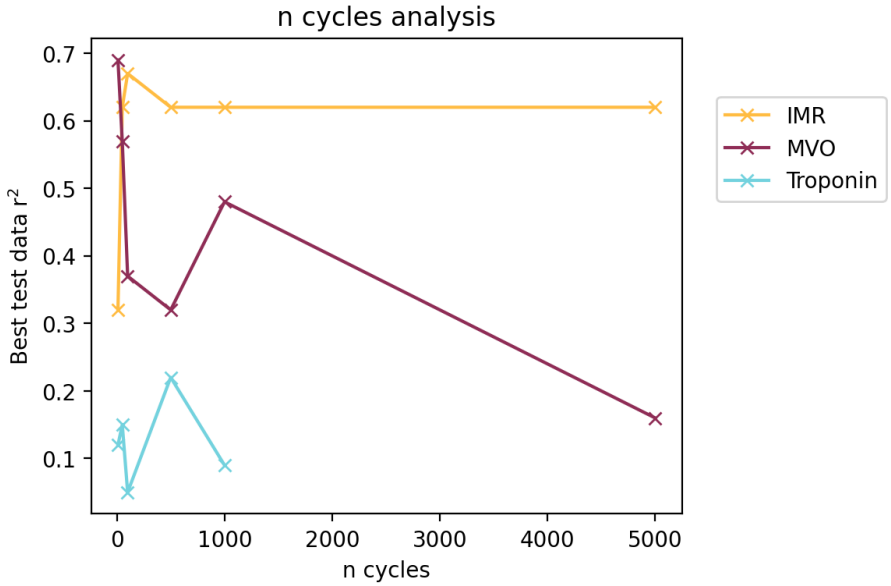

Figure B.5 Assessment of the impact of increasing `n_cycles` on the accuracy of PySR models for predicting IMR, Troponin and MVO. Troponin was not assessed above 1000 cycles as the troponin dataset was much larger resulting in prohibitively long run times.

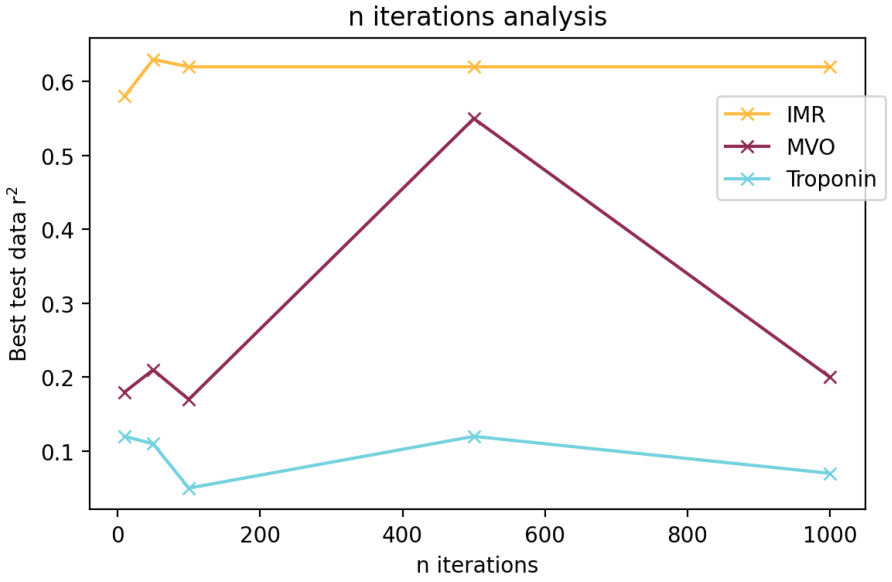

Figure B.6 Assessment of the impact of increasing `n_iterations` on the accuracy of PySR models for predicting IMR, Troponin and MVO.

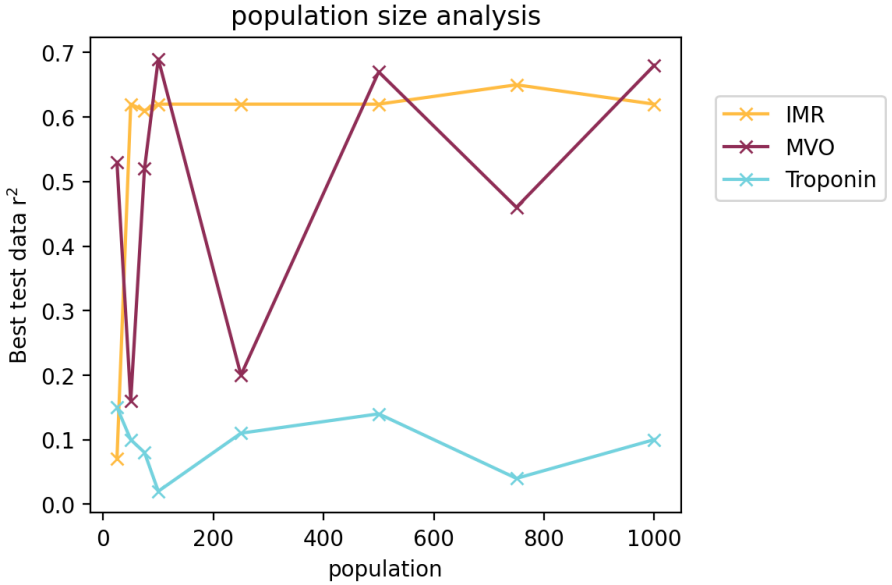

Figure B.7 Assessment of the impact of increasing `population_size` on the accuracy of PySR models for predicting IMR, Troponin and MVO.

B.9 PySR results equations

Table B.19 The PySR generated equations that produced the most accurate prediction of each OxAMI clinical parameter.

Parameter	PySR generated equation
AIS	$y = \frac{0.00092929037}{(-I_{56}/(I_{44}/0.207251 - 0.3582202) + I_{63} - I_{70} + I_{73})} + \frac{0.002034326}{(I_{73} - 0.0028416575)} + \frac{1}{(I_{44} + 0.031837903))}$
Creatinine	$y = \frac{(I_{61})^2}{(I_{128} + I_{150} - 0.0058631115)^2} + 69.11897$
FIS	$y = \frac{0.5738643}{(I_{44} + ((I_{164})^2/(I_{239})^2) + I_{239} + 0.017887488)} + \frac{2 \times I_{60}}{I_{370} - 0.089101024}$
IMR	$y = 49450.184 \times I_{522} + 49.094116 + \frac{0.21932559}{I_{45} - 0.12894352} - \frac{0.0034090339}{0.0070158155 - I_{239}}$
MDRD	$y = -0.5960459 \times \frac{PC_5}{(PC_4 - 2.0908031 \times PC_9} + \frac{PC_6}{(0.022744797 - PC_3} - \frac{258.28058 + 1.4503913}{PC_1 - PC_7 + 0.0071767517} \times \frac{(PC_4 - PC_5) + 87.524635}{\exp(PC_2)} + \frac{0.013850129}{PC_2 - (-0.017110461)}$
MSI	$y = PC_1 + PC_3 - PC_5 + (PC_2 + PC_9 \times (-1.688105 + \frac{1}{0.2753508 - (PC_7)^2/PC_{10})}) \times (-0.2753508 \times \frac{PC_6}{PC_4} + \frac{PC_6 - 0.0130959}{PC_3} + 0.6322356$
MVO	$y = -(I_{18} + I_{533}) \times \frac{-1}{(I_{332}/(-0.515032) + I_{387} - (I_{337} + I_{532})} + \frac{(-0.061694)}{(-I_{337} - I_{357} + 0.000187)} + \frac{(I_{516}/(0.017642 - 2.88622 \times I_{60}) + I_{532})}{(16.03 \times (I_{44} + I_{337} - I_{259}) + 0.001744)} + 1.343616$
Pain	$y = 367.40747 + \frac{1}{(PC_5/0.05972878 - 0.08666947)} + \frac{0.07434914}{PC_3 - 0.00210751}$
Troponin	$y = PC_5 \times (-1640.1641 + \frac{-3.3397906}{PC_2 - PC_3}) - -126.04444$

C

Appendix: Prediction of AAA patient groups using ASAP-MS and ML analysis

C.1 ML full results for classification of clinical data

Table C.1 ML accuracies for the classification of AAA patients based on clinical groups using all ASAP-MS variables available.

Clinical variable	KNN kappa	SVM kappa	LDA kappa	NBC kappa	RFC kappa
Small AAA vs HV	0.741±0.22	0.879±0.181	0.835±0.174	0.791±0.192	0.811±0.237
Large AAA vs HV	0.881±0.152	0.944±0.105	0.933±0.114	0.863±0.179	0.914±0.165
Large AAA vs Small AAA	0.724±0.132	0.865±0.09	0.862±0.087	0.601±0.14	0.788±0.113
Post surgery vs HV	0.901±0.144	0.884±0.158	0.901±0.142	0.871±0.15	0.87±0.166
Post surgery vs Small AAA	0.937±0.081	0.925±0.094	0.961±0.06	0.925±0.092	0.945±0.078
Post surgery vs Pre surgery	0.723±0.149	0.73±0.167	0.68±0.159	0.68±0.158	0.74±0.163

Table C.2 ML accuracies for the classification of AAA patients based on clinical groups using χ^2 FR selected ASAP-MS variables.

Clinical variable	KNN kappa	SVM kappa	LDA kappa	NBC kappa	RFC kappa
Small AAA vs HV	0.89±0.177	0.868±0.186	0.886±0.17	0.849±0.211	0.964±0.151
Large AAA vs HV	0.911±0.16	0.943±0.117	0.989±0.058	0.89±0.154	0.986±0.057
Large AAA vs Small AAA	0.762±0.109	0.783±0.099	0.77±0.115	0.584±0.144	0.757±0.112
Post surgery vs HV	0.923±0.136	0.915±0.134	0.898±0.144	0.896±0.147	0.895±0.143
Post surgery vs Small AAA	0.92±0.084	0.926±0.083	0.944±0.083	0.938±0.078	0.955±0.064
Post surgery vs Pre surgery	0.716±0.158	0.715±0.165	0.739±0.151	0.728±0.159	0.703±0.164

Table C.3 ML accuracies for the classification of AAA patients based on clinical groups using OI FR selected ASAP-MS variables.

Clinical variable	KNN kappa	SVM kappa	LDA kappa	NBC kappa	RFC kappa
Small AAA vs HV	0.934±0.134	0.859±0.193	0.881±0.142	0.84±0.179	0.896±0.207
Large AAA vs HV	0.953±0.107	0.965±0.091	0.99±0.07	0.935±0.128	0.955±0.134
Large AAA vs Small AAA	0.721±0.134	0.747±0.133	0.753±0.122	0.48±0.178	0.735±0.113
Post surgery vs HV	0.906±0.134	0.896±0.138	0.885±0.141	0.898±0.149	0.873±0.162
Post surgery vs Small AAA	0.922±0.102	0.889±0.113	0.929±0.104	0.946±0.063	0.939±0.077
Post surgery vs Pre surgery	0.705±0.156	0.689±0.154	0.722±0.155	0.682±0.179	0.712±0.165

Table C.4 ML accuracies for the classification of AAA patients based on clinical groups using stats FR selected ASAP-MS variables.

Clinical variable	KNN kappa	SVM kappa	LDA kappa	NBC kappa	RFC kappa
Small AAA vs HV	0.806±0.185	0.874±0.191	0.85±0.172	0.783±0.201	0.874±0.188
Large AAA vs HV	0.809±0.213	0.878±0.205	0.898±0.161	0.795±0.22	0.961±0.126
Large AAA vs Small AAA	0.58±0.149	0.638±0.163	0.579±0.181	0.388±0.157	0.637±0.155
Post surgery vs HV	0.863±0.166	0.869±0.156	0.861±0.166	0.845±0.16	0.856±0.185
Post surgery vs Small AAA	0.934±0.081	0.913±0.12	0.812±0.119	0.828±0.157	0.909±0.106
Post surgery vs Pre surgery	0.679±0.174	0.691±0.161	0.719±0.158	0.68±0.175	0.728±0.155

C.2 Significant peaks determined through RF for AAA plasma classification

The m/z ion peaks identified as being important for classification of patients according to each patient group within the OxAAA study by all three feature reduction methods employed in 75% of the analysis, with the associated t statistic and p score (Table C.5). The results show the peaks found for the *Large AAA vs Small AAA* analysis in 50% of feature reduction analysis as no peaks were found in 75% of analysis.

Table C.5 The significant *m/z* peaks identified for each AAA patient group classification and the relevant statistics.

Patient groups	m/z peaks	t statistic	p score
Small AAA vs HV	172	-6.99	2.53×10^{-8}
	206	-5.95	6.77×10^{-7}
	338	-8.36	3.84×10^{-10}
	353	-7.45	6.12×10^{-9}
	355	-6.84	4.09×10^{-8}
	357	-8.36	3.84×10^{-10}
	358	-8.55	2.17×10^{-10}
	425	-8.52	2.38×10^{-10}
	426	-8.64	1.66×10^{-10}
	444	-7.16	1.50×10^{-8}
Large AAA vs HV	281	-7.73	2.59×10^{-9}
	357	-7.0	2.49×10^{-8}
	358	-6.94	2.99×10^{-8}
	425	-7.75	2.46×10^{-9}
	459	-4.83	2.23×10^{-5}
Post surgery vs HV	227	-8.89	7.99×10^{-11}
	284	-10.37	1.24×10^{-12}
	357	-8.88	8.30×10^{-11}
	426	-8.80	1.05×10^{-10}
Post surgery vs Small AAA	45	-6.77	1.48×10^{-9}
	239	-6.63	2.83×10^{-9}
	257	-7.13	2.98×10^{-10}
	258	-6.9	8.16×10^{-10}
	267	-7.06	4.04×10^{-10}
	284	-6.72	1.89×10^{-9}
	285	-7.49	5.70×10^{-11}
	286	-6.99	5.49×10^{-10}
Post surgery vs Large AAA pre surgery	45	-4.56	1.65×10^{-5}
	256	-10.62	2.57×10^{-17}
	257	-10.42	6.51×10^{-17}
	258	-10.42	6.58×10^{-17}
	267	-10.12	2.58×10^{-16}
	268	-11.12	2.55×10^{-18}
	284	-10.20	1.82×10^{-16}
	285	-9.8	1.17×10^{-15}
Large AAA vs Small AAA	316	-2.75	7.21×10^{-3}
	317	-3.34	1.26×10^{-3}

C.3 Classification of AAA tissue subtypes

C.3.1 Statistical analysis of tissue subtype spectral differences

The statistical analysis results for the peak distributions across the tissue subtypes data are shown in Table C.6, with a much lower number of significantly different peak distributions between the subtypes than were observed in the tissue type analysis. The most different tissue subtypes appeared to be subcutaneous:omental fat, preperitoneal:omental fat, and abluminal:luminal thrombus. The differences between fat subtypes can be justified biologically, as omental fat is functionally and structurally distinct from non-visceral fats. Omental fat is more metabolically active, excreting higher levels of hormones and adipokines, pro-inflammatory cytokines, free fatty acids, and contains a higher number of large adipocytes [378]. Preperitoneal fat and subcutaneous fat are much more similar. Both are non-visceral fats, differing mainly by location in the body rather than as distinct tissue types. The differences observed between abluminal and luminal thrombus is interesting, as the delineation of these parts of the thrombus can be subjective. Not all patients have a thrombus present, and the width of the thrombus can vary in volume by a large amount. Very few peaks were found to be statistically difference between the wall

Table C.6 Percentage of peaks identified as significantly different between groups.

Tissue subtype classification	**Tissue type**	**% of m/z variables identified as significantly ($P < 0.001$) different between classes**	**% of all variables**
Subcut.:omental	Fat	456	46.9
Subcut.:Preperit.	Fat	64	6.5
Preperit.:omental	Fat	359	36.2
Abluminal:luminal	Thrombus	384	38.7
Maximal:proximal	Wall	29	2.9
Maximal:distal	Wall	14	1.4
Proximal:distal	Wall	13	1.3

subtypes. This suggests it may be difficult for the ML algorithms to distinguish between these tissue samples, as metabolic changes that may be present related to their location with respect to the AAA are not strongly expressed, or easily observed using this technique. PCA clustering analysis was not able to separate any of the tissue subtypes using the first three principle components.

C.3.2 Machine learning analysis of tissue subtypes

The ML classification of tissue subtypes reflected the results obtained in the statistical analysis. The full results table is shown in Figure C.1.

No successful analysis with a κ score above the lower threshold of 0.2 was achieved for classification of the wall subtypes. These tissue types are not distinct enough to be classified using these methods, and so metabolic changes due to AAA progression cannot be investigated using these methods within the AAA wall.

Classification of abluminal and luminal thrombus was achieved with moderate success, with a maximum κ score of 0.48 using the averaged-NBC-stats combination of methods.

Classification of subcutaneous:omental fat was the most successful subtype classification, with a κ score of 0.512 achieved using the repeats-RelF-SVM combination of methods.

The classification of subcutaneous:preperitoneal and preperitoneal:omental were less successful, with only weak agreement (κ 0.2-0.4, 60-70% accuracy) obtained for these classifications. Spatially, these tissue types are closer in proximity to each other. As the subcutaneous:omental classification was the most successful, and these fat subtypes are both the most distinct and the furthest apart spatially, this suggests that the spatially similar fat subtypes share a large proportion of their mass spectral features.

FR method	ML model	Subc. : Om. κ	Sub. : Prep. κ	Om : Prep. κ	Ablum. : Lum. κ	Max. : Prox. κ	Max. : Dist. κ	Prox. : Dist. κ
All	KNN	0.2 ± 0.31	-0.06 ± 0.33	0.23 ± 0.31	0.26 ± 0.28	0.03 ± 0.28	-0.09 ± 0.34	-0.03 ± 0.37
	SVM	0.3 ± 0.29	0.07 ± 0.25	0.07 ± 0.3	0.35 ± 0.29	0.05 ± 0.27	-0.12 ± 0.29	-0.08 ± 0.25
	LDA	0.3 ± 0.3	0.03 ± 0.3	0.17 ± 0.33	0.41 ± 0.29	-0.08 ± 0.17	-0.07 ± 0.26	-0.12 ± 0.14
	NBC	0.36 ± 0.29	-0.11 ± 0.19	-0.02 ± 0.26	0.28 ± 0.28	-0.05 ± 0.19	-0.09 ± 0.27	-0.01 ± 0.29
	RFC	0.36 ± 0.32	0.34 ± 0.33	0.29 ± 0.33	0.34 ± 0.32	-0.02 ± 0.34	0.02 ± 0.37	0.03 ± 0.4
Stats	KNN	0.34 ± 0.32	-0.12 ± 0.29	0.29 ± 0.33	0.46 ± 0.28	-0.03 ± 0.28	-0.06 ± 0.3	-0.01 ± 0.36
	SVM	0.37 ± 0.31	0.02 ± 0.27	0.22 ± 0.29	0.46 ± 0.31	-0.02 ± 0.29	-0.12 ± 0.29	-0.11 ± 0.23
	LDA	0.38 ± 0.32	0.05 ± 0.3	0.22 ± 0.34	0.37 ± 0.28	-0.12 ± 0.19	-0.11 ± 0.31	-0.13 ± 0.16
	NBC	0.28 ± 0.29	-0.09 ± 0.22	0.28 ± 0.33	0.48 ± 0.27	-0.08 ± 0.23	-0.14 ± 0.25	-0.04 ± 0.25
	RFC	0.38 ± 0.32	0.3 ± 0.36	0.28 ± 0.35	0.37 ± 0.3	-0.03 ± 0.28	-0.02 ± 0.34	0.08 ± 0.39
OI	KNN	0.26 ± 0.29	0.02 ± 0.31	0.13 ± 0.33	0.48 ± 0.31	0.01 ± 0.32	-0.03 ± 0.35	-0.01 ± 0.38
	SVM	0.43 ± 0.28	0.15 ± 0.28	0.13 ± 0.31	0.47 ± 0.31	-0.1 ± 0.26	-0.08 ± 0.3	-0.09 ± 0.25
	LDA	0.42 ± 0.29	0.23 ± 0.31	0.22 ± 0.34	0.39 ± 0.29	-0.1 ± 0.18	-0.05 ± 0.28	-0.15 ± 0.23
	NBC	0.25 ± 0.3	-0.04 ± 0.34	0.12 ± 0.35	0.46 ± 0.3	-0.05 ± 0.32	-0.07 ± 0.38	-0.05 ± 0.34
	RFC	0.41 ± 0.28	0.29 ± 0.32	0.15 ± 0.35	0.38 ± 0.28	0.04 ± 0.33	0.03 ± 0.39	-0.02 ± 0.35
Chi^2	KNN	0.3 ± 0.34	-0.02 ± 0.33	0.21 ± 0.33	0.35 ± 0.29	0.04 ± 0.34	0.03 ± 0.37	-0.02 ± 0.38
	SVM	0.4 ± 0.34	0.1 ± 0.29	0.15 ± 0.3	0.45 ± 0.3	-0.1 ± 0.24	-0.11 ± 0.28	-0.09 ± 0.27
	LDA	0.37 ± 0.33	0.15 ± 0.34	0.19 ± 0.33	0.35 ± 0.34	-0.1 ± 0.23	-0.09 ± 0.28	-0.14 ± 0.27
	NBC	0.26 ± 0.32	-0.06 ± 0.32	0.15 ± 0.38	0.42 ± 0.29	-0.04 ± 0.34	-0.03 ± 0.34	-0.0 ± 0.36
	RFC	0.4 ± 0.33	0.21 ± 0.32	0.17 ± 0.33	0.38 ± 0.31	0.01 ± 0.36	0.01 ± 0.35	-0.01 ± 0.37
MRMR	KNN	0.35 ± 0.3	0.19 ± 0.38	0.25 ± 0.32	0.47 ± 0.29	-0.06 ± 0.32	-0.08 ± 0.35	-0.03 ± 0.37
	SVM	0.4 ± 0.28	0.21 ± 0.32	0.18 ± 0.29	0.47 ± 0.26	-0.02 ± 0.24	-0.15 ± 0.2	-0.15 ± 0.22
	LDA	0.38 ± 0.29	0.15 ± 0.32	0.16 ± 0.33	0.39 ± 0.26	-0.06 ± 0.29	0.14 ± 0.34	0.08 ± 0.37
	NBC	0.32 ± 0.3	0.13 ± 0.34	0.26 ± 0.32	0.47 ± 0.28	-0.06 ± 0.28	-0.16 ± 0.29	-0.12 ± 0.24
	RFC	0.38 ± 0.31	0.27 ± 0.33	0.23 ± 0.32	0.41 ± 0.3	0.02 ± 0.38	-0.04 ± 0.38	0.17 ± 0.4
ReliF	KNN	0.3 ± 0.31	0.17 ± 0.32	0.14 ± 0.33	0.39 ± 0.32	0.12 ± 0.34	-0.02 ± 0.4	-0.05 ± 0.36
	SVM	0.43 ± 0.29	0.27 ± 0.27	0.13 ± 0.3	0.41 ± 0.28	-0.02 ± 0.26	-0.1 ± 0.3	-0.05 ± 0.26
	LDA	0.38 ± 0.3	0.24 ± 0.33	0.09 ± 0.32	0.36 ± 0.26	0.08 ± 0.32	0.03 ± 0.33	-0.09 ± 0.33
	NBC	0.32 ± 0.3	0.19 ± 0.32	0.16 ± 0.33	0.38 ± 0.3	0.07 ± 0.32	-0.05 ± 0.36	0.04 ± 0.36
	RFC	0.4 ± 0.32	0.27 ± 0.36	0.16 ± 0.34	0.35 ± 0.31	0.17 ± 0.34	0.06 ± 0.38	0.12 ± 0.38
PCA	KNN	0.01 ± 0.2	0.01 ± 0.17	0.0 ± 0.15	0.06 ± 0.18	0.0 ± 0.11	-0.03 ± 0.19	-0.02 ± 0.16
	SVM	0.01 ± 0.15	-0.03 ± 0.13	-0.06 ± 0.13	0.0 ± 0.08	-0.06 ± 0.06	-0.08 ± 0.09	-0.09 ± 0.1
	LDA	-0.07 ± 0.12	-0.03 ± 0.15	-0.09 ± 0.11	0.0 ± 0.0	-0.06 ± 0.0	-0.07 ± 0.02	-0.07 ± 0.06
	NBC	-0.04 ± 0.18	-0.04 ± 0.17	-0.06 ± 0.16	-0.03 ± 0.14	-0.06 ± 0.19	-0.03 ± 0.15	-0.01 ± 0.15
	RFC	0.01 ± 0.04	-0.01 ± 0.06	0.01 ± 0.05	0.02 ± 0.05	0.0 ± 0.03	-0.02 ± 0.03	-0.02 ± 0.01

Figure C.1 The κ scores achieved for each analysis of the tissue subtype classification by combinations of five ML models and seven FR methods applied to the mass spectra, with the spectra averaged across each sample. The errors shown are the standard deviation in κ score achieved across each partition. The colourmap indicates the accuracy and level of agreement of the results by κ score. Tissue subtypes are abbreviated to Subcut. (subcutaneous), Om. (Omental), Prep. (Preperitoneal), Lum. (Luminal), Ablum. (Abluminal), Dist. (Distal), Max. (Maximal), and Prox. (Proximal).

C.3.3 Analysis of peaks found through feature reduction for tissue subtype classification

Analysis of the peaks used by the FR methods for the tissue subtype classifications found that the same peaks were only used consistently for the most successful classifications. The peaks found are shown in Table C.7. For the two fat classifications in which m/z peaks were used by FR methods in over 75% of the time, all of the peaks found were in the high mass molecule and lipid region at $m/z > 600$. This confirms that it is lipid peaks being used to distinguish fat samples, rather than small molecule metabolites. When only the peaks found by FR methods were used as the input features for ML, the accuracy was similar to that of the best performing ML algorithms that used all the peaks for the fat and thrombus analyses (shown in Table C.8).

Table C.7 Peaks found to be used by FR models with high occurrence ($> 75\%$).

Tissue subtype classification	**Tissue type**	**Identified m/z peaks that were reported with high occurrence across the FR methods**
Subcut.:Omental	Fat	689, 659, 943, 941, 633, 634, 635, 657
Subcut.:Preperit.	Fat	659, 941, 632, 634, 942
Preperit.:Omental	Fat	None
Abluminal:Luminal	Thrombus	455, 267, 191, 219, 215, 190, 415, 233, 268, 220, 232, 218
Maximal:Proximal	Wall	None
Maximal:Distal	Wall	None
Proximal:Distal	Wall	None

The ability to classify abluminal and luminal thrombus was an interesting achievement, as these tissue types are structurally and metabolically similar, and are difficult to delineate clinically. This may be of interest for the analysis of AAA progression through monitoring metabolic changes.

As stated previously, the number of samples used in the subtype analysis was very low, and so the conclusions made here are purely speculative. A much larger study

Table C.8 The κ scores obtained when only the peaks found with high occurance were used as input variables for ML for the tissue type classification.

Analysis setup	**ML model**	**Subcut.: Omental**	**Subcut.: Preperit.**	**Abluminal: Luminal**
Identified peaks - averaged used	KNN	0.51±0.31	0.42±0.32	0.25±0.31
	SVM	0.43±0.32	0.35±0.29	0.15±0.29
	LDA	0.43±0.29	0.40±0.33	0.20±0.30
	NBC	0.48±0.27	0.41±0.3	0.06±0.32
	RFC	0.51±0.31	0.41±0.32	0.26±0.35
Identified peaks - repeats used	KNN	0.36±0.20	0.28±0.2	0.32±0.23
	SVM	0.43±0.24	0.38±0.23	0.43±0.28
	LDA	0.40±0.25	0.36±0.26	0.41±0.31
	NBC	0.52±0.24	0.43±0.25	0.41±0.27
	RFC	0.43±0.21	0.27±0.20	0.41±0.24

would need to be conducted to improve the confidence of these results. However, as the tissue subtype classification was poor, particularly for the classification of wall samples, it is unlikely that this analysis will be helpful for investigating AAA development.

C.4 Tissue classification: Using repeat samples to increase the training data size

Using all 10 repeat spectra for a patient as opposed to the averaged data was assessed as a method of increasing the size of the data available for model training. All samples and repeats from a single patient were kept together in either just the training or test data to prevent data leakage. The use of this data resulted in no significant increase in accuracy across either the tissue type classifications or tissue subtype classifications conducted ($P > 0.05$ by comparison of means by paired sample t-test). The colour maps in Figures C.4 and C.4, and the bar charts in Figures C.4 and C.5, show the impact of using the repeats or averaged datasets on the accuracy obtained. Using the repeat data did not improve the accuracy or error substantially, despite increasing the number of training data points to greater than the 10 times limit recommended for ML analysis. Therefore, this may not be an appropriate way of increasing training data size for this analysis. This may show that a factor of 10 increase in samples is not enough to reduce error in the analysis from false classification. Alternatively, the error in a variation in a single patient repeat measurement may not be a good representation of inter-patient variability which is needed to train generalised ML models. Finally, it may more simply show that these results are the best possible outcomes using simple ML classification models on this particular dataset. A further investigation into the impact of training data size on the ML model accuracy through changing both repeat numbers and patient numbers was then undertaken (see next section).

FR method	ML model	Fat : Thrombus κ	Fat : Wall κ	Wall : Thrombus κ
All	KNN	0.98 ± 0.04	0.9 ± 0.09	0.61 ± 0.16
	SVM	0.99 ± 0.02	0.93 ± 0.08	0.77 ± 0.14
	LDA	0.97 ± 0.06	0.93 ± 0.09	0.8 ± 0.13
	NBC	0.97 ± 0.05	0.9 ± 0.08	0.55 ± 0.18
	RFC	0.98 ± 0.04	0.96 ± 0.05	0.81 ± 0.13
Stats	KNN	0.97 ± 0.05	0.94 ± 0.05	0.67 ± 0.14
	SVM	0.96 ± 0.06	0.95 ± 0.05	0.71 ± 0.14
	LDA	0.96 ± 0.07	0.89 ± 0.07	0.71 ± 0.14
	NBC	0.96 ± 0.08	0.93 ± 0.06	0.6 ± 0.16
	RFC	0.97 ± 0.05	0.95 ± 0.05	0.7 ± 0.14
OI	KNN	0.99 ± 0.04	0.95 ± 0.05	0.63 ± 0.16
	SVM	0.99 ± 0.04	0.96 ± 0.04	0.69 ± 0.17
	LDA	0.98 ± 0.05	0.9 ± 0.07	0.71 ± 0.16
	NBC	0.98 ± 0.05	0.93 ± 0.05	0.6 ± 0.17
	RFC	0.98 ± 0.04	0.95 ± 0.04	0.68 ± 0.17
Chi^2	KNN	0.99 ± 0.04	0.94 ± 0.05	0.62 ± 0.15
	SVM	0.98 ± 0.05	0.95 ± 0.05	0.69 ± 0.15
	LDA	0.98 ± 0.05	0.89 ± 0.07	0.7 ± 0.14
	NBC	0.98 ± 0.04	0.92 ± 0.06	0.6 ± 0.17
	RFC	0.98 ± 0.04	0.94 ± 0.05	0.69 ± 0.15
MRMR	KNN	0.99 ± 0.03	0.94 ± 0.06	0.64 ± 0.16
	SVM	0.99 ± 0.03	0.96 ± 0.05	0.7 ± 0.17
	LDA	0.98 ± 0.05	0.93 ± 0.06	0.69 ± 0.16
	NBC	0.99 ± 0.03	0.92 ± 0.07	0.62 ± 0.17
	RFC	0.98 ± 0.03	0.95 ± 0.05	0.71 ± 0.16
ReliF	KNN	0.98 ± 0.04	0.91 ± 0.09	0.65 ± 0.16
	SVM	0.98 ± 0.04	0.93 ± 0.07	0.76 ± 0.14
	LDA	0.98 ± 0.03	0.9 ± 0.07	0.75 ± 0.14
	NBC	0.98 ± 0.05	0.91 ± 0.07	0.55 ± 0.18
	RFC	0.97 ± 0.09	0.94 ± 0.06	0.75 ± 0.14
PCA	KNN	0.81 ± 0.22	0.6 ± 0.29	0.26 ± 0.25
	SVM	0.78 ± 0.18	0.49 ± 0.24	0.25 ± 0.22
	LDA	0.52 ± 0.2	0.03 ± 0.05	0.05 ± 0.15
	NBC	0.82 ± 0.19	0.19 ± 0.25	0.13 ± 0.18
	RFC	0.9 ± 0.19	0.92 ± 0.06	0.08 ± 0.16

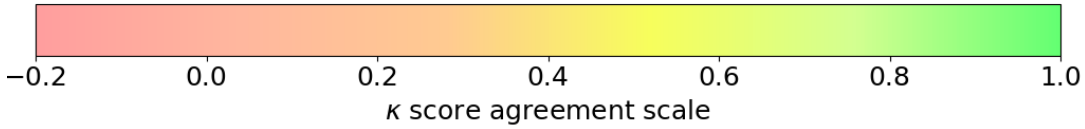

Figure C.2 The κ scores achieved for each analysis of the tissue type classification by combinations of five ML models and seven FR methods applied to the mass spectra, with all repeats of each sample used. The errors shown are the standard deviation in κ score achieved across each partition. The colourmap indicates the accuracy and level of agreement of the results by κ score.

FR method	ML model	Subc. : Om. κ	Sub. : Prep. κ	Om : Prep. κ	Ablum. : Lum. κ	Max. : Prox. κ	Max. : Dist. κ	Prox. : Dist. κ
All	KNN	0.14 ± 0.24	-0.04 ± 0.23	0.05 ± 0.27	0.19 ± 0.23	-0.04 ± 0.17	-0.04 ± 0.21	-0.13 ± 0.22
	SVM	0.43 ± 0.24	0.18 ± 0.22	0.16 ± 0.22	0.32 ± 0.26	0.03 ± 0.23	0.01 ± 0.23	-0.2 ± 0.2
	LDA	0.45 ± 0.25	0.11 ± 0.19	0.24 ± 0.23	0.34 ± 0.25	0.02 ± 0.22	0.02 ± 0.22	-0.18 ± 0.22
	NBC	0.17 ± 0.2	-0.07 ± 0.24	0.15 ± 0.22	0.24 ± 0.22	-0.12 ± 0.22	-0.11 ± 0.21	-0.16 ± 0.2
	RFC	0.41 ± 0.25	0.08 ± 0.23	0.08 ± 0.23	0.37 ± 0.26	0.01 ± 0.22	-0.0 ± 0.23	-0.13 ± 0.22
Stats	KNN	0.32 ± 0.22	0.05 ± 0.21	0.06 ± 0.19	0.26 ± 0.21	0.01 ± 0.19	-0.01 ± 0.23	-0.09 ± 0.22
	SVM	0.39 ± 0.22	0.27 ± 0.2	0.18 ± 0.2	0.37 ± 0.23	0.04 ± 0.23	-0.0 ± 0.25	-0.13 ± 0.24
	LDA	0.34 ± 0.23	0.25 ± 0.23	0.18 ± 0.22	0.38 ± 0.23	0.01 ± 0.2	0.02 ± 0.24	-0.15 ± 0.24
	NBC	0.37 ± 0.26	0.26 ± 0.26	0.26 ± 0.2	0.36 ± 0.23	-0.02 ± 0.17	-0.06 ± 0.24	-0.15 ± 0.22
	RFC	0.43 ± 0.21	0.2 ± 0.22	0.11 ± 0.2	0.35 ± 0.23	0.02 ± 0.19	0.02 ± 0.24	-0.08 ± 0.23
OI	KNN	0.21 ± 0.27	0.03 ± 0.24	0.1 ± 0.26	0.24 ± 0.22	0.01 ± 0.18	-0.01 ± 0.21	-0.04 ± 0.22
	SVM	0.39 ± 0.24	0.19 ± 0.21	0.08 ± 0.23	0.4 ± 0.22	-0.0 ± 0.18	-0.02 ± 0.23	-0.05 ± 0.23
	LDA	0.35 ± 0.23	0.19 ± 0.21	0.1 ± 0.23	0.42 ± 0.22	-0.01 ± 0.2	-0.0 ± 0.22	-0.07 ± 0.23
	NBC	0.25 ± 0.29	0.1 ± 0.25	0.02 ± 0.22	0.32 ± 0.22	-0.04 ± 0.17	-0.01 ± 0.22	-0.07 ± 0.22
	RFC	0.34 ± 0.29	0.12 ± 0.25	0.03 ± 0.25	0.37 ± 0.21	0.01 ± 0.19	0.03 ± 0.24	-0.04 ± 0.24
Chi^2	KNN	0.31 ± 0.23	0.09 ± 0.26	0.1 ± 0.25	0.28 ± 0.2	-0.05 ± 0.19	0.05 ± 0.24	-0.09 ± 0.21
	SVM	0.38 ± 0.23	0.26 ± 0.24	0.13 ± 0.22	0.39 ± 0.23	-0.1 ± 0.2	0.02 ± 0.25	-0.04 ± 0.22
	LDA	0.35 ± 0.23	0.22 ± 0.24	0.12 ± 0.22	0.4 ± 0.23	-0.09 ± 0.19	0.01 ± 0.24	-0.03 ± 0.22
	NBC	0.31 ± 0.25	0.13 ± 0.25	0.17 ± 0.22	0.35 ± 0.22	-0.12 ± 0.21	0.02 ± 0.22	-0.04 ± 0.22
	RFC	0.4 ± 0.22	0.14 ± 0.25	0.05 ± 0.22	0.34 ± 0.25	-0.02 ± 0.2	0.02 ± 0.23	-0.05 ± 0.23
MRMR	KNN	0.3 ± 0.26	0.13 ± 0.25	0.09 ± 0.22	0.29 ± 0.21	-0.04 ± 0.19	0.03 ± 0.24	-0.09 ± 0.21
	SVM	0.43 ± 0.25	0.18 ± 0.22	0.12 ± 0.22	0.42 ± 0.24	-0.01 ± 0.2	-0.01 ± 0.25	-0.11 ± 0.25
	LDA	0.39 ± 0.24	0.17 ± 0.24	0.12 ± 0.21	0.43 ± 0.25	-0.03 ± 0.2	0.0 ± 0.26	-0.11 ± 0.27
	NBC	0.35 ± 0.27	0.2 ± 0.26	0.17 ± 0.23	0.37 ± 0.24	-0.08 ± 0.19	-0.1 ± 0.21	-0.07 ± 0.25
	RFC	0.4 ± 0.25	0.13 ± 0.24	0.07 ± 0.23	0.41 ± 0.24	-0.02 ± 0.2	0.01 ± 0.24	-0.11 ± 0.24
ReliF	KNN	0.3 ± 0.26	0.13 ± 0.22	0.1 ± 0.26	0.21 ± 0.19	-0.04 ± 0.17	0.03 ± 0.23	-0.05 ± 0.26
	SVM	0.51 ± 0.21	0.27 ± 0.24	0.14 ± 0.23	0.33 ± 0.23	-0.02 ± 0.22	0.03 ± 0.24	-0.09 ± 0.24
	LDA	0.51 ± 0.21	0.26 ± 0.25	0.15 ± 0.22	0.37 ± 0.22	-0.01 ± 0.22	0.03 ± 0.23	-0.12 ± 0.25
	NBC	0.38 ± 0.25	0.17 ± 0.23	0.08 ± 0.24	0.33 ± 0.23	-0.06 ± 0.16	-0.04 ± 0.23	-0.06 ± 0.23
	RFC	0.42 ± 0.25	0.23 ± 0.22	0.03 ± 0.27	0.37 ± 0.21	0.02 ± 0.21	-0.0 ± 0.24	-0.0 ± 0.24
PCA	KNN	0.01 ± 0.08	0.01 ± 0.1	0.0 ± 0.11	0.16 ± 0.21	0.01 ± 0.16	-0.01 ± 0.22	0.0 ± 0.16
	SVM	0.0 ± 0.02	-0.0 ± 0.05	-0.01 ± 0.02	-0.01 ± 0.14	-0.01 ± 0.09	-0.06 ± 0.12	-0.02 ± 0.15
	LDA	-0.0 ± 0.0	0.0 ± 0.01	-0.0 ± 0.05	-0.01 ± 0.11	-0.01 ± 0.06	-0.06 ± 0.12	-0.01 ± 0.13
	NBC	-0.06 ± 0.05	-0.03 ± 0.13	-0.07 ± 0.09	0.25 ± 0.2	-0.02 ± 0.14	0.01 ± 0.22	0.03 ± 0.18
	RFC	-0.01 ± 0.12	-0.01 ± 0.1	-0.01 ± 0.05	0.14 ± 0.2	-0.0 ± 0.05	-0.04 ± 0.11	-0.0 ± 0.1

Figure C.3 The κ scores achieved for each analysis of the tissue subtype classification by combinations of five ML models and seven FR methods applied to the mass spectra, with all repeats of each sample used. The errors shown are the standard deviation in κ score achieved across each partition. The colourmap indicates the accuracy and level of agreement of the results by κ score. Tissue subtypes are abbreviated to Subcut. (subcutaneous), Om. (Omental), Prep. (Preperitoneal), Lum. (Luminal), Ablum. (Abluminal), Dist. (Distal), Max. (Maximal), and Prox. (Proximal).

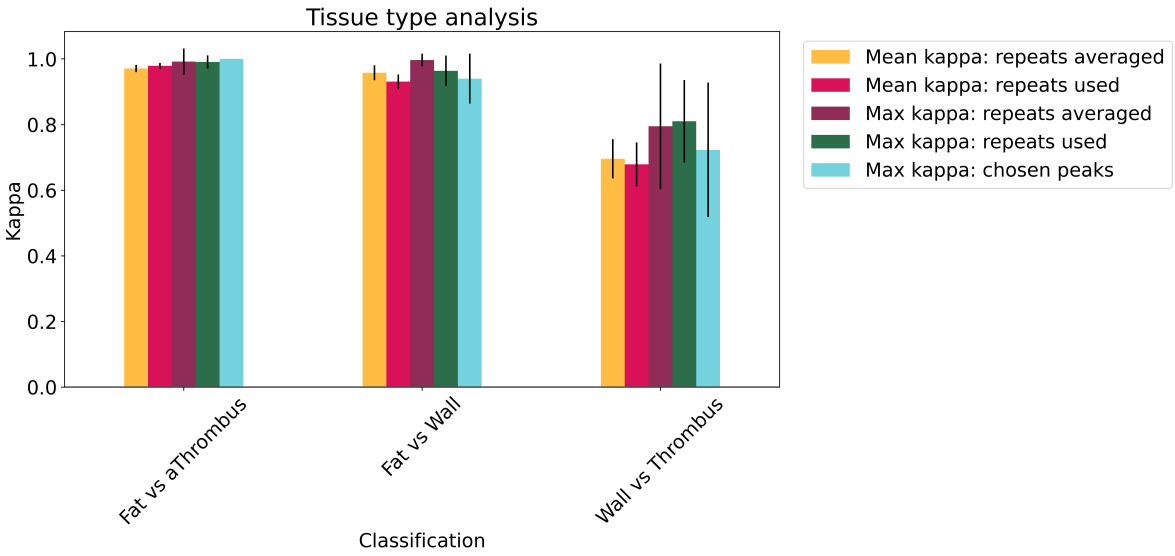

Figure C.4 The κ scores achieved in ML classification of tissue by tissue type. Bars show the κ score from the average ML algorithm results, error bars represent standard deviations. Results are shown for analysis where repeats were averaged to one spectrum per patient (yellow), or repeats were used as separate samples (red) are plotted as the mean and standard deviations of all ML algorithm results respectively. The best κ score achieved with a single ML algorithm and feature reduction method was plotted for both the repeat averaged (purple) and repeats used (green). Where appropriate, the analysis was repeated using peaks found to be significant by the FR analysis, and the maximum κ score achieved for the data shown (blue)

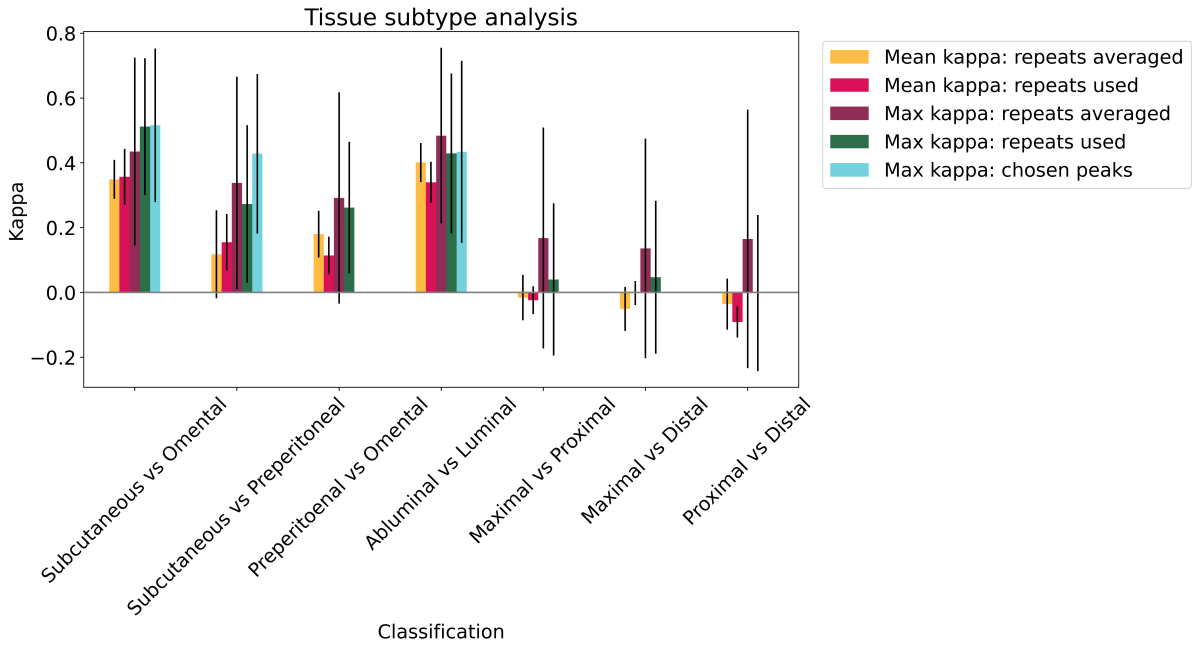

Figure C.5 The κ scores achieved in ML classification of tissue by tissue subtype. Bars show the κ score from the average ML algorithm results, error bars represent standard deviations. Results are shown for analysis where repeats were averaged to one spectrum per patient (yellow), or repeats were used as separate samples (red) are plotted as the mean and standard deviations of all ML algorithm results respectively. The best κ score achieved with a single ML algorithm and feature reduction method was plotted for both the repeat averaged (purple) and repeats used (green). Where appropriate, the analysis was repeated using peaks found to be significant by the FR analysis, and the maximum κ score achieved for the data shown (blue).

C.5 Further investigation into the impact of training data size on ML classification results

In the data presented so far, analysis was conducted on data averaged over the repeat measurements, or with each of the ten repeat measurements used as a separate sample. This can be interpreted as an effective 10-fold increase in sample numbers. However, this increase did not have a significant impact (paired t-test $P > 0.05$) on any of the resulting ML model accuracies. An investigation was conducted to determine how changing the training data size, which can be achieved by changing both the number of patients used, and the number of repeat measurements used, affected the ML accuracy. This was conducted across the four tissue type classifications, and the four most successful tissue subtype classifications (fat and throbmus subtype groups). The effect of changing each variable was assessed, as well as the impact of changes to absolute training data size, which is a compound variable of both repeats and patients.

Changing training data size was achieved by running the ML models described in Section 5.2.2.3 over datasets with n patients in each class, where n is a value between 5 and 30 for tissue type analysis. The number of repeats used for each patient was also varied from 1 to 10. This amounts to 250 analysis combinations for each tissue type analysis (25 patient variations x 10 repeat variations). This equates to a large amount of computational time, so a reduced partition number of 20 was used. This will also reduce the probability of partitions being repeated for the smaller datasets with small patient numbers, but likely result in an increase in the accuracy error.

As in all previous methods, where multiple repeats were used for each patient, patient data was held within a single partition. The same hold-4-out cross validation method was used for this analysis as was used for the previous analysis. The analysis was then repeated with each repeat measurement used as a separate input, so that

10 repeat measurements were used for each patient sample. These were input as if they were individual samples. The same principle of ensuring that all repeat measurements relating to a single patient were used in the same validation group was applied when using the sample repeat measurements.

The effect of changing the number of repeats used, whilst keeping the number of patients used at the maximum, had very little effect on the accuracy of the ML models used. The bar chart in Figure C.6 shows how changing the number of repeats used impacted the wall:thrombus classification analysis for each ML method used. It can be seen that the κ accuracy of each ML model remains almost constant regardless of the number of repeats used. Similar results were achieved for the fat vs thrombus and fat vs wall analyses.

When the number of patients used is changed however, the ML classification performance is observed to change. The plots on the left of Figures C.5 and C.5 show examples of the trends observed in this analyses across different classifications, and shows an increase in accuracy as the n patients variable increases. The same trends in accuracy were observed across all the tissue type analyses used. The accuracy obtained however appears to be independent of the number of repeats used, shown by the coloured lines, which show very similar values at each point. The scatter plots on the right of these figures further show this trend. These figures show accuracy plotted against absolute training data size, which is the product of the number of repeats used and the number of patients used. Accuracy is seen to increase within each n repeats set (shown by the change in colour), to a common maximum once the maximum number of patients has been used.

This investigation shows that it is the number of patients included in the data, and not the repeats used, that impact the results. This is in agreement with a common ML observation that a small, good-quality dataset, that spans a large breadth of the data area, can be sufficient for ML models. Datasets that are larger, but do not cover the data space as well, will not generalise as well, and so model

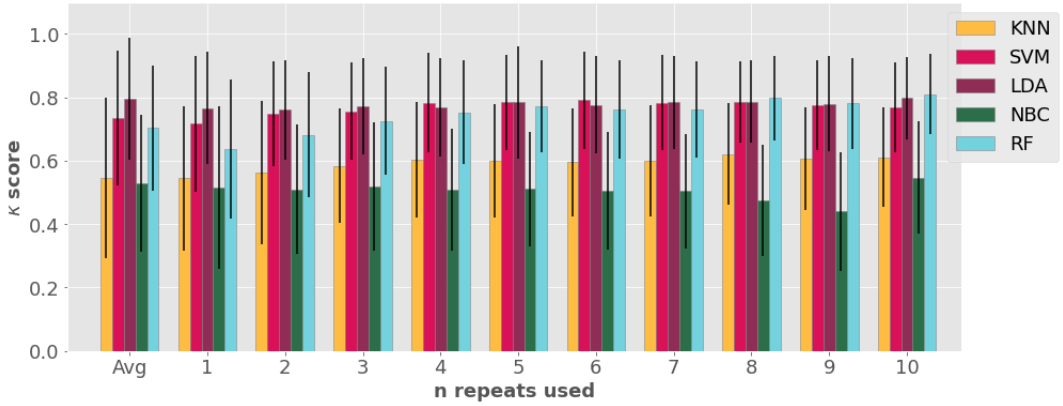

Figure C.6 The κ scores achieved for changes in n repeats with all patients used for wall and thrombus. Error bars show the standard deviation of κ score over the partitions used. The first group of bars, labelled 'Avg', represents the data obtained when all 10 repeats are averaged over.

accuracy against the test data will be poor. This is in agreement with what was observed previously when the repeat samples were used.

This analysis also shows that the datasets used for this test are sufficient to show the potential maximum accuracy for this study in the tissue type analysis, as the accuracy for all three classifications showed a characteristic logarithmic relationship between sample size and accuracy that is commonly observed for classical ML methods, known as the 'learning curve' (demonstrated in [379–382]). This suggests that data complexity is more influential than data size, and adding more samples to this analysis would not likely improve the accuracy [383]. There is the potential for a large time saving during measurements if a single repeat measurement is sufficient for ML accuracy, rather than having to take repeat measurements to obtain a representative sample. This shows that, for this particular analysis, the intra-repeat error is much less significant than differences observed in the spectra that are confounded with the variable of interest (tissue label in this case).

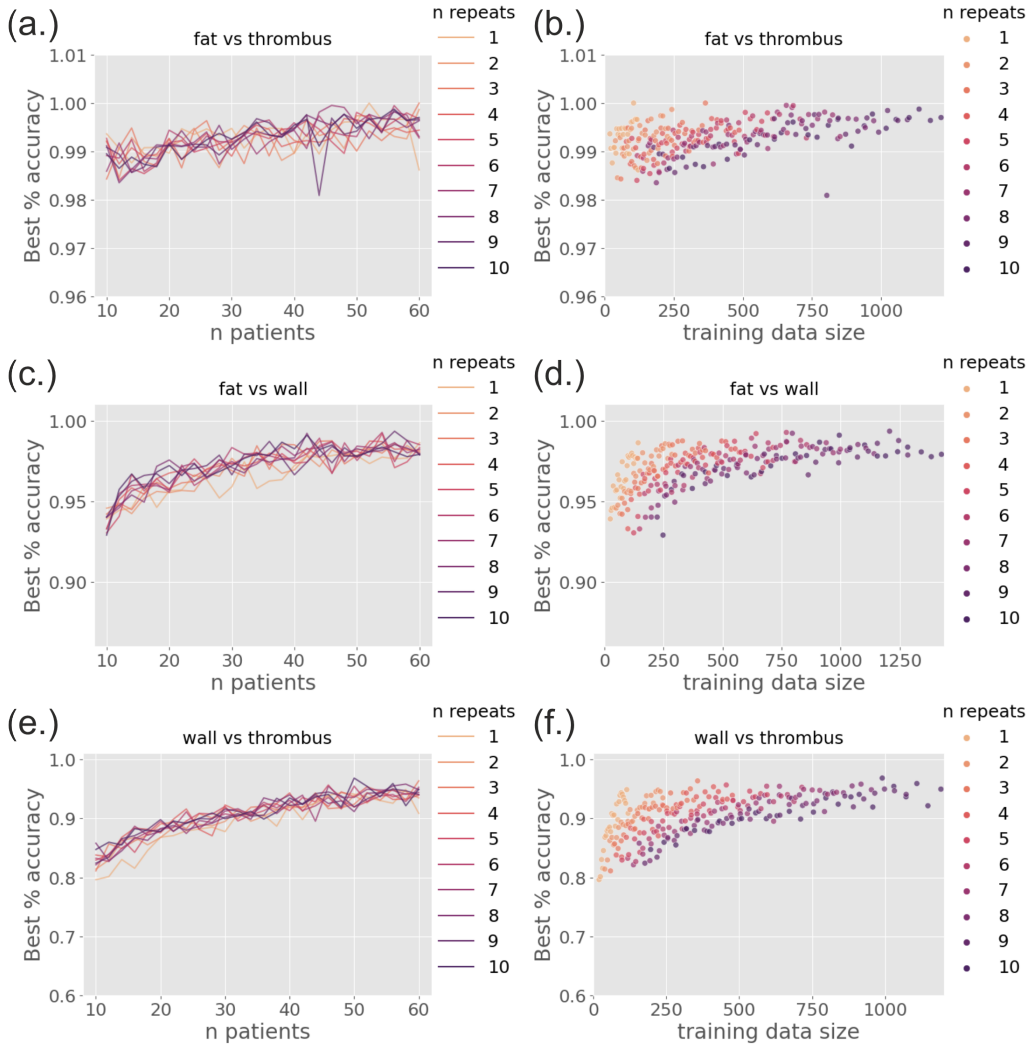

Figure C.7 The % accuracy obtained by the best ML model for the tissue type analysis conducted whilst varying the number of repeats used and the number of patients used. The number of repeat measurements used for each patients is represented by the changing colour of the plotted points. Plots on the right (a, c, e) show the change in accuracy as the number of patients used changes. Plots on the left (b, d, f) show how the accuracy changes as a function of absolute training data size. The % accuracy shown is the best accuracy obtained across the five ML models used for each n repeats and m patients combination.

C.5.1 The impact of the number of test patients in the hold-n-out validation method

As mentioned previously, a hold-n-out cross-validation method was used in the tissue subtype analysis, where $n = 4$ (2n per class) for each analysis. The impact of changing the number of test patients used was assessed using the tissue subtype analysis for the different fat and thrombus subtype classification. These classifications were used as they were moderately successful, but with higher error and loss than the tissue type analysis. Increasing the n-test patients will likely reduce the error in κ score achieved for small datasets as the cross-validation method converges with the K-fold cross-validation method, but the κ score is expected to decrease with increasing n.

Figure C.9 confirms this result, as it can be observed in each analysis that the error in accuracy decreases with increasing n patients, combined with a slight decrease in overall accuracy. The decrease in accuracy is a function of the training data size decreasing, but the effect of this appears to be low. The analysis discussed above used $n = 2$ training samples per class, which could be increased to achieve reduced error in accuracy in future analysis.

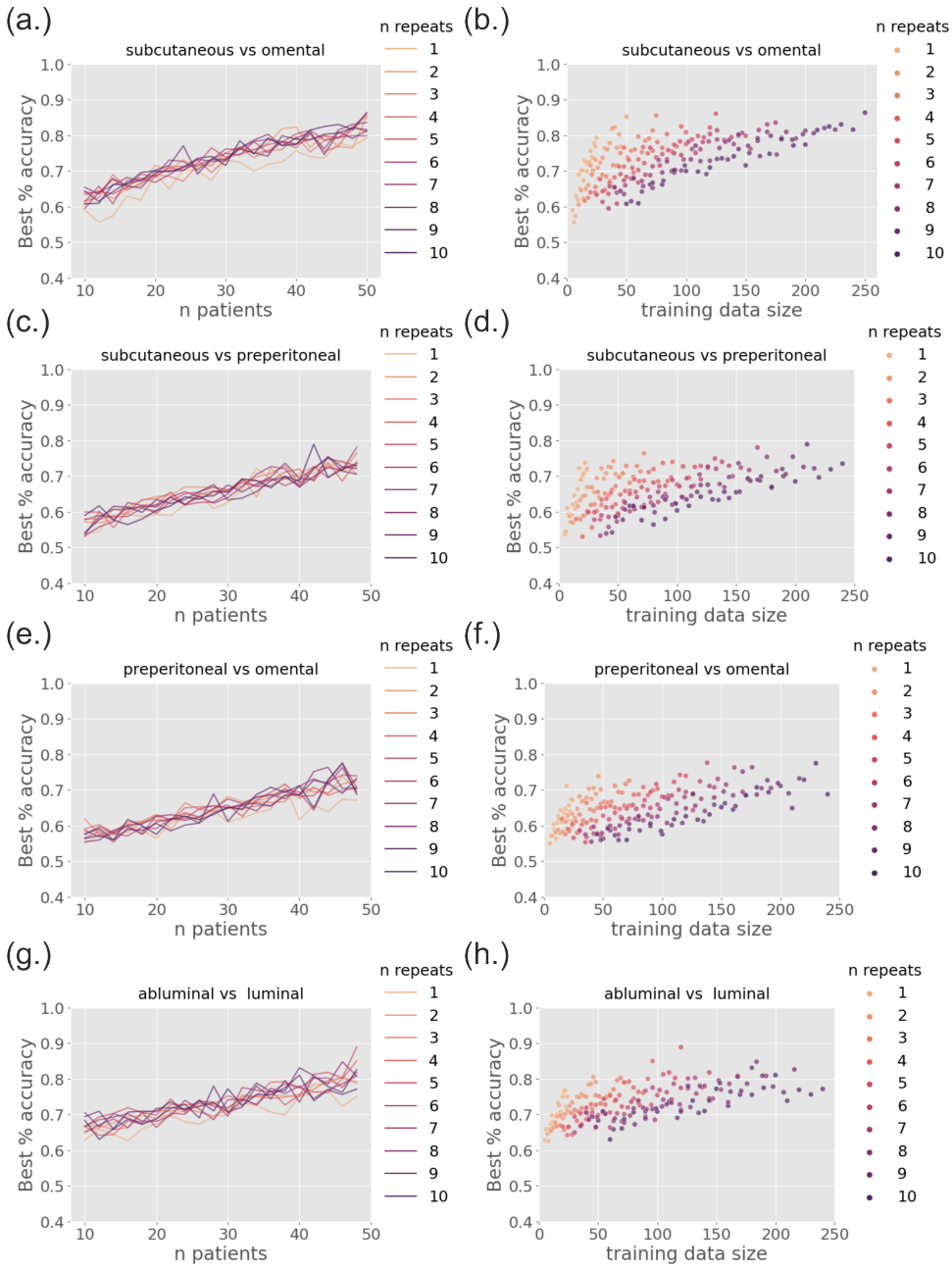

Figure C.8 The % accuracy obtained by the best ML model for the tissue subtype analysis conducted whilst varying the number of repeats used and the number of patients used. The number of repeat measurements used for each patients is represented by the changing colour of the plotted points. Plots on the right (a, c, e) show the change in accuracy as the number of patients used changes. Plots on the left (b, d, f) show how the accuracy changes as a function of absolute training data size. The % accuracy shown is the best accuracy obtained across the five ML models used for each n repeats and m patients combination.

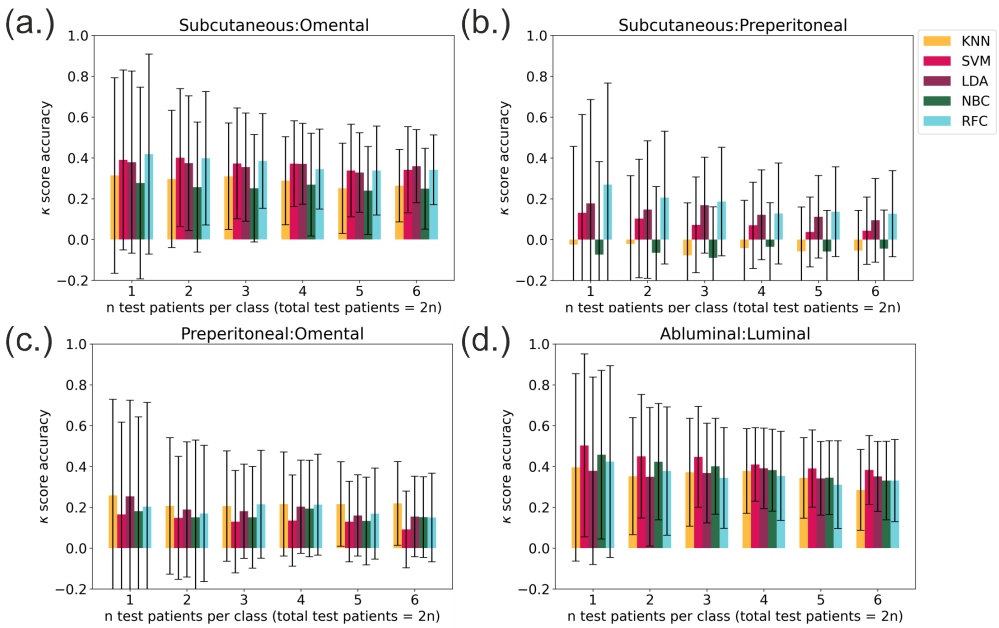

Figure C.9 The effect of changing the hold-n-out cross validation number on the κ score obtained for four different subtype classifications: subcutaneous vs omental fat (a), subcutaneous vs preperitoneal fat (b), preperitoneal vs omental fat (c), and abluminal vs luminal thrombus (d), using the Chi^2 FR method.

C.6 Classification of AAA patients by clinical variables

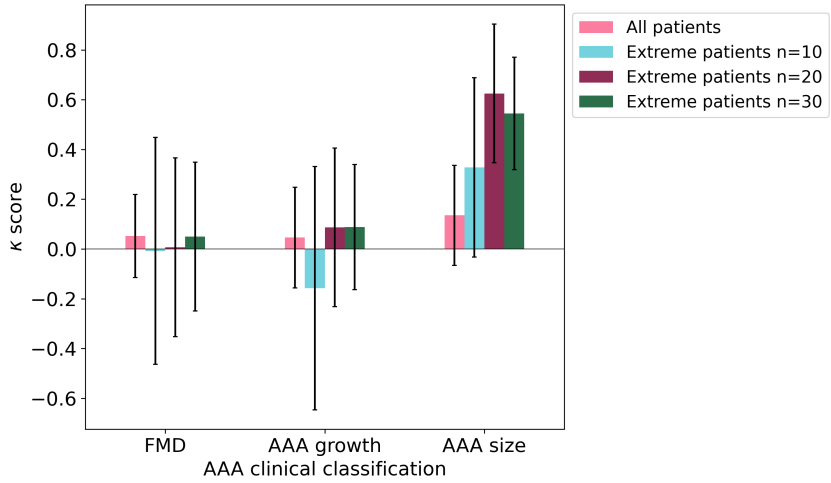

Figure C.10 The highest ML accuracy κ scores across the five ML algorithms used for the classification of patients by measured AAA clinical variables FMD, AAA growth and AAA size. Error bars show the standard deviation in κ achieved across the 100 partitions.

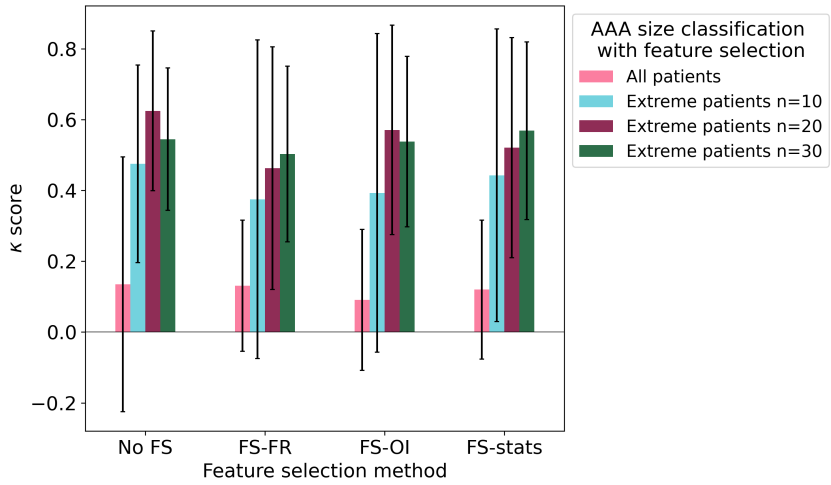

Figure C.11 The highest ML accuracy κ scores across the five ML algorithms used for the classification of patients by measured AAA size using four different FR algorithms: stats, OI, and χ^2 FR. Error bars show the standard deviation in κ achieved across the 100 partitions.

Table C.9 The ML results for the classification of AAA patients by clinical variables encoded into binary classifications.

Clinical variable	**KNN κ**	**SVM κ**	**LDA κ**	**NBC κ**	**RFC κ**
FMD - All patients	0.052±0.181	0.051±0.167	0.025±0.192	0.051±0.160	0.013±0.221
FMD N=10 Extreme	- 0.008±0.456	- 0.173±0.388	- 0.116±0.298	- 0.089±0.293	- 0.150±0.457
FMD N=20 Extreme	- 0.008±0.289	- 0.042±0.121	- 0.051±0.189	- 0.056±0.217	0.007±0.359
FMD N=30 Extreme	0.050±0.299	0.006±0.192	- 0.052±0.170	- 0.030±0.213	0.018±0.273
AAA growth - All patients	0.046±0.202	- 0.083±0.230	- 0.005±0.206	- 0.051±0.189	- 0.028±0.222
Growth N=10 Extreme	- 0.158±0.489	- 0.255±0.288	- 0.340±0.271	-0.246± 0.221	- 0.372±0.442
Growth N=20 Extreme	- 0.008±0.321	0.015±0.287	- 0.078±0.202	0.087±0.319	- 0.065±0.311
Growth N=30 Extreme	0.088±0.252	0.036±0.238	- 0.042±0.159	- 0.026±0.186	- 0.108±0.263
AAA size - All patients	0.143±0.202	0.035±0.161	0.075±0.188	0.135±0.187	0.025±0.208
Size N=10 Extreme	0.135±0.414	0.350±0.405	0.355±0.416	0.335±0.370	0.235±0.479
Size N=20 Extreme	0.425±0.344	0.600±0.313	0.562±0.297	0.451±0.345	0.356±0.321
Size N=30 Extreme	0.458±0.271	0.549±0.238	0.582±0.224	0.460±0.276	0.438±0.255

C.7 Identification of overlapping significant (P<0.05) peaks

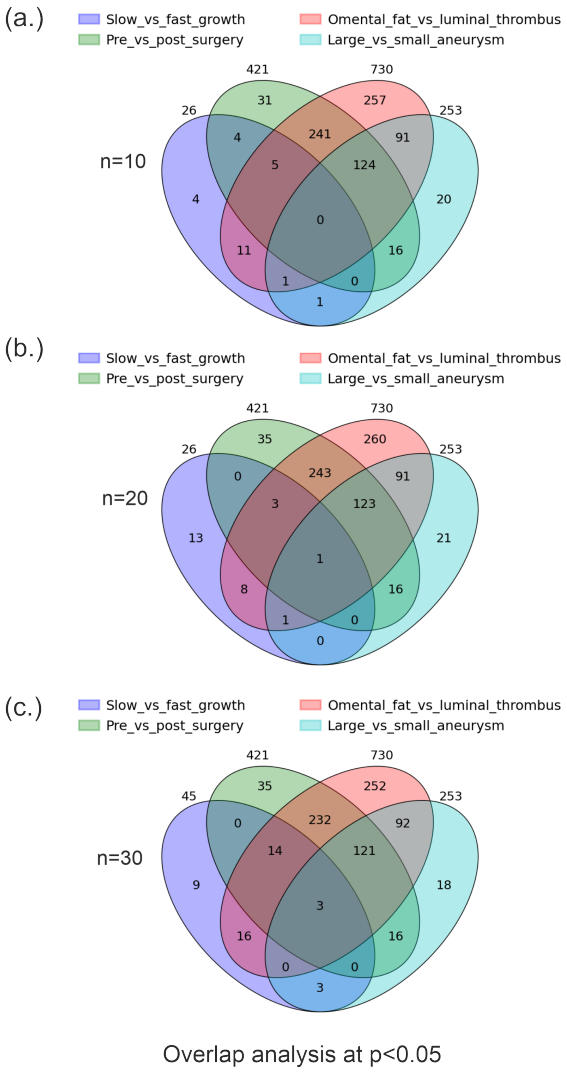

Figure C.12 The significant peaks at the lowest confidence interval used (p<0.05) found in across four different patient group comparisons: slow vs fast growth (blue), omental fat vs luminal thrombus (red), pre vs post surgery (green), and large vs small aneurysm (cyan), where the slow vs fast growth analysis has been conducted on extreme grouped patients at $n = 10$ per class (a), $n = 20$ per class (b), and $n = 30$ per class (c).

D

Appendix: Investigating different ASAP-MS instruments for clinical sample analysis

D.1 Scan speed analysis across Radian and CMS-S for tissue types

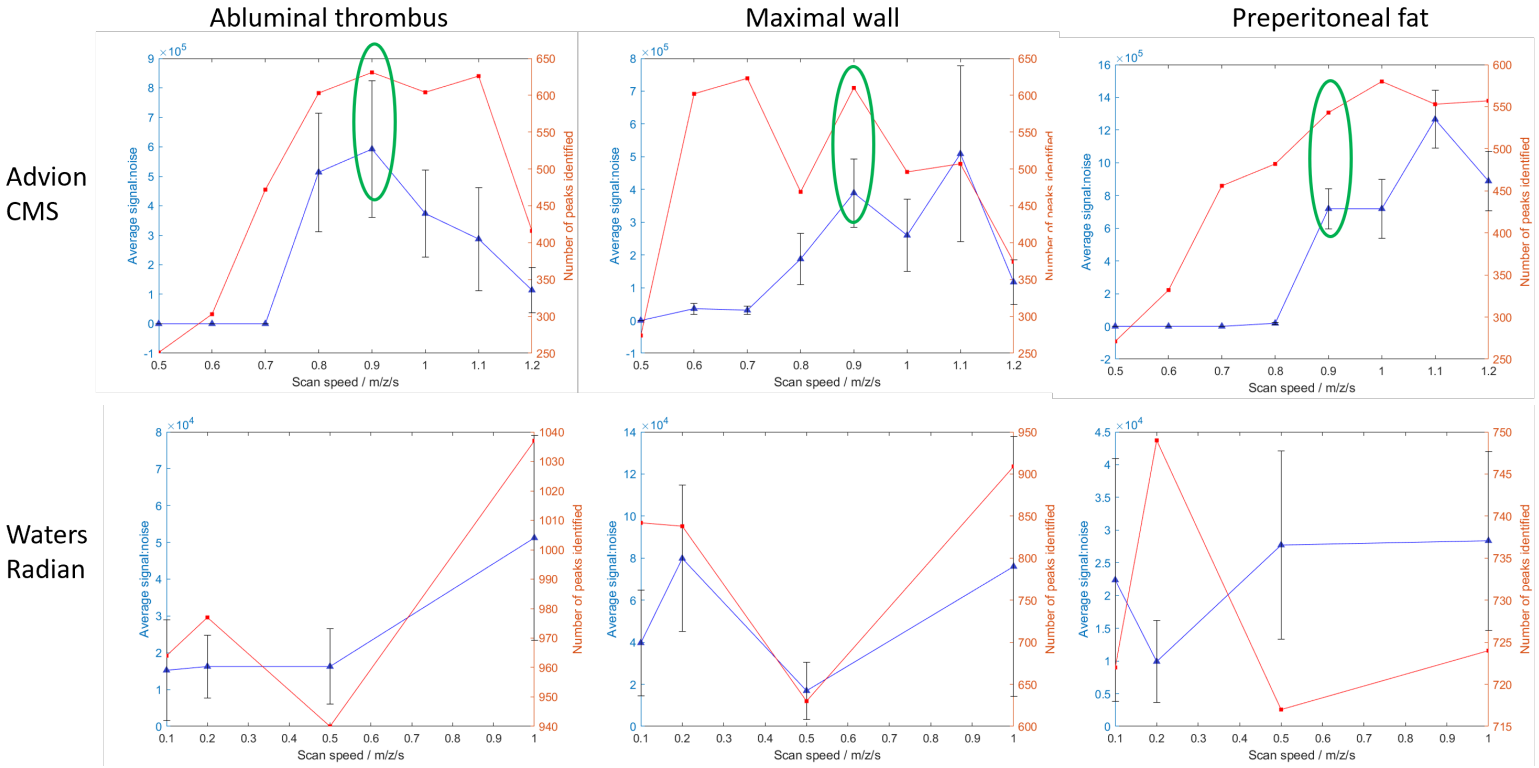

Figure D.1 Analysis of scan speed for multiple tissue samples for patient 77 across the Radian and Advion CMS-S instruments. The data is highlighted in green for scan speed 0.9 s, which was the chosen optimal scan speed.

D.2 Radian plasma measurements with changing cone voltage

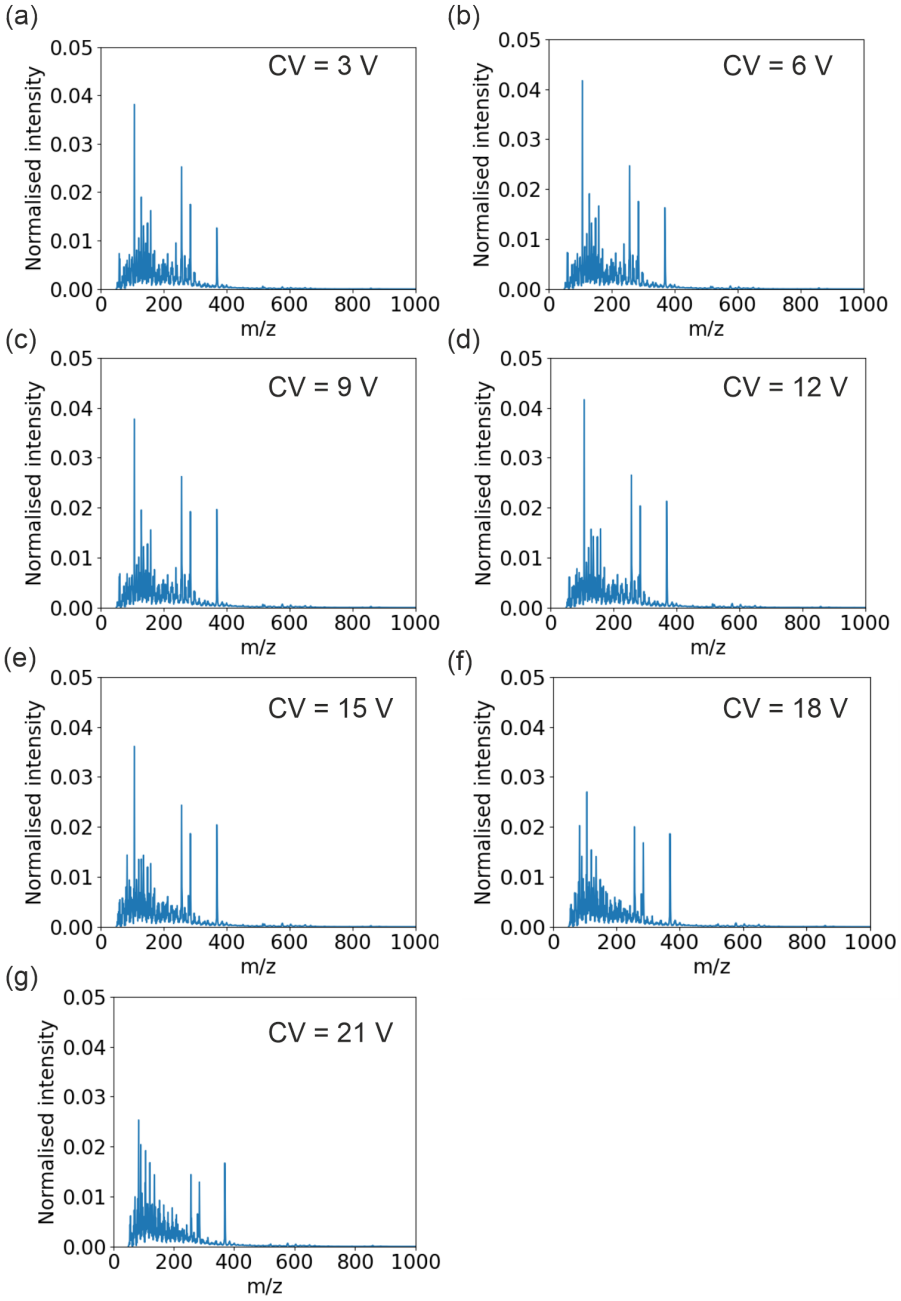

Figure D.2 The averaged plasma mass spectrum recorded from 3 – 21 V cone voltage on the Waters Radian instrument. The spectra are normalised to unit area under the curve.

D.3 Average artery wall measurements on the CMS-S and Radian instrument

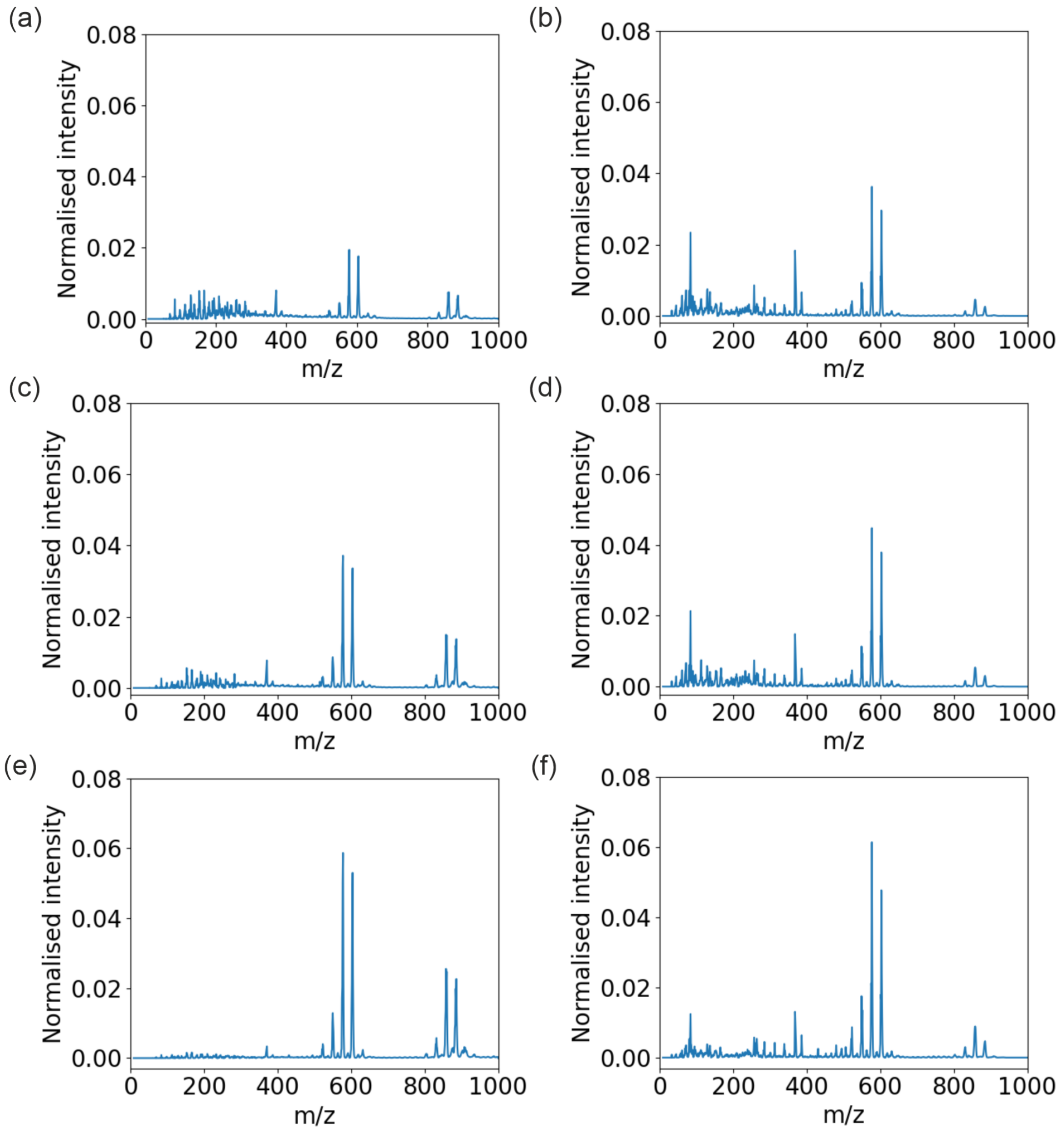

Figure D.3 Mass spectra for Patient 77: artery wall recorded on the Waters Radian (left, a,c,e) and the Advion CMS-S (right, b,d,f) representing the different artery wall tissue subtypes of proximal (top), maximal (centre) and distal (lower) wall. Peaks at m/z=42 and 369 have been removed.

D.4 Average fat measurements on the CMS-S and Radian instrument

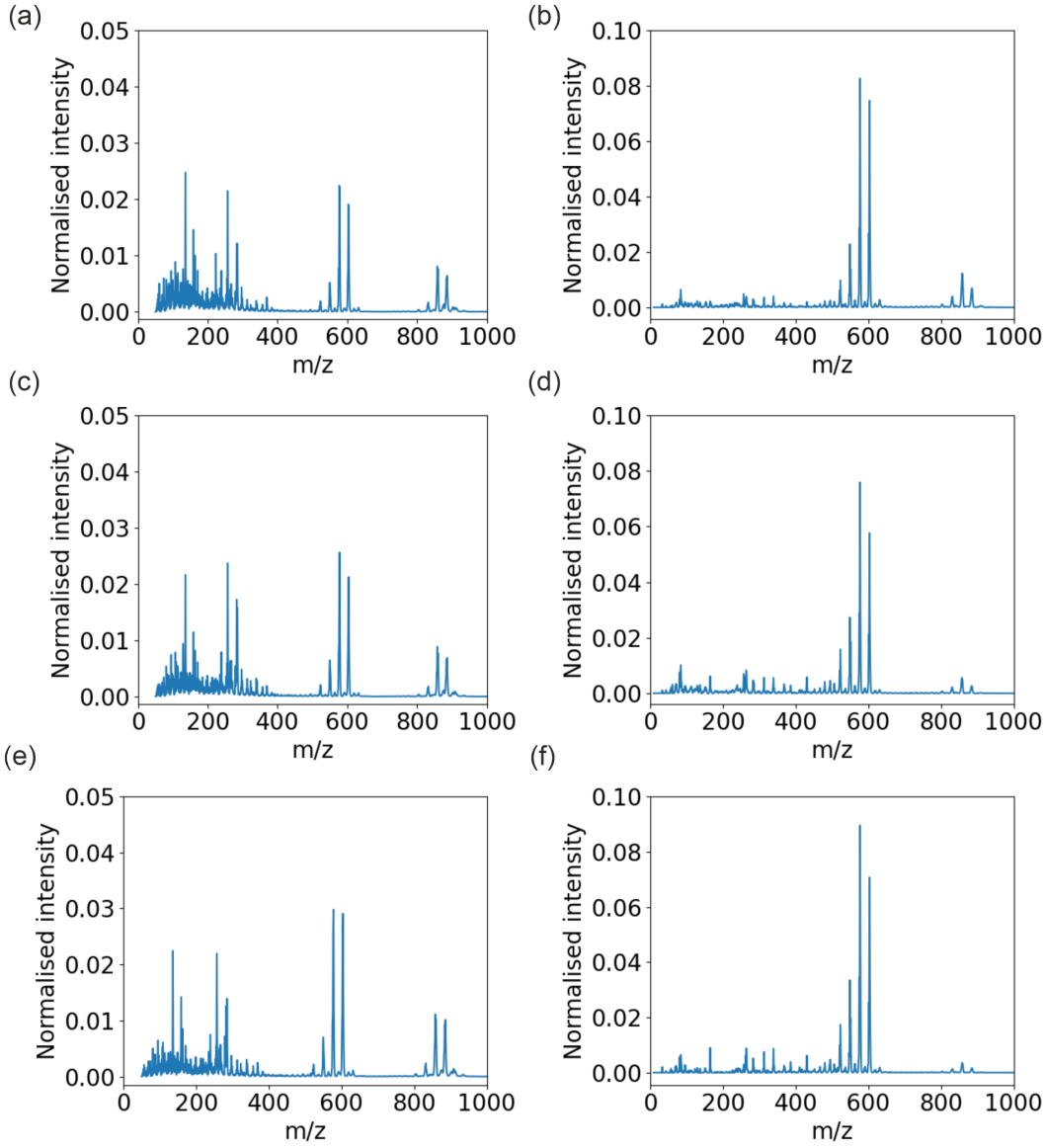

Figure D.4 Mass spectra for Patient 77: fat tissue recorded on the Waters Radian (left, a,c,e) and the Advion CMS-S (right, b,d,f) representing the different fat tissue subtypes of omental (top), subcutaneous (centre) and preperitoneal (lower) fat. Peaks at m/z=42 and 369 have been removed.

D.5 Average thrombus measurements on the CMS-S and Radian instrument

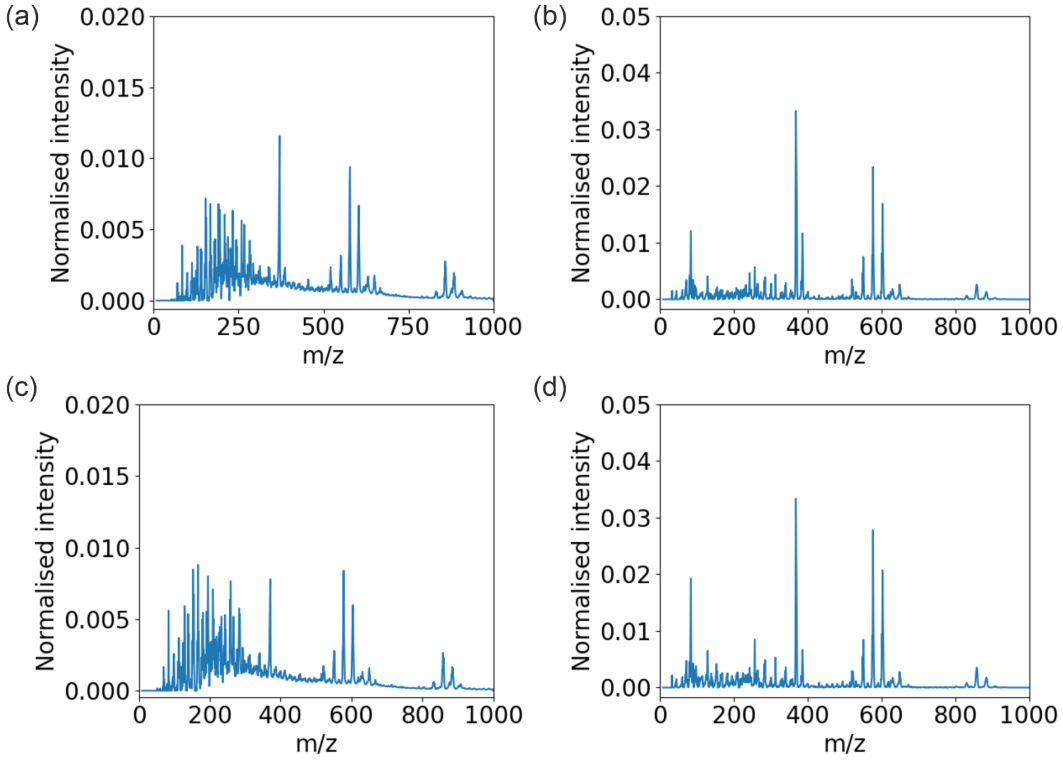

Figure D.5 Mass spectra for Patient 77: thrombus recorded on the Waters Radian (left, a,c,e) and the Advion CMS-S (right, b,d,f) representing the different thrombus tissue subtypes of abluminal (top), and luminal (lower) thrombus. Peaks at m/z=42 and 369 have been removed.

D.6 Inter-instrument comparison of tissue sample metrics

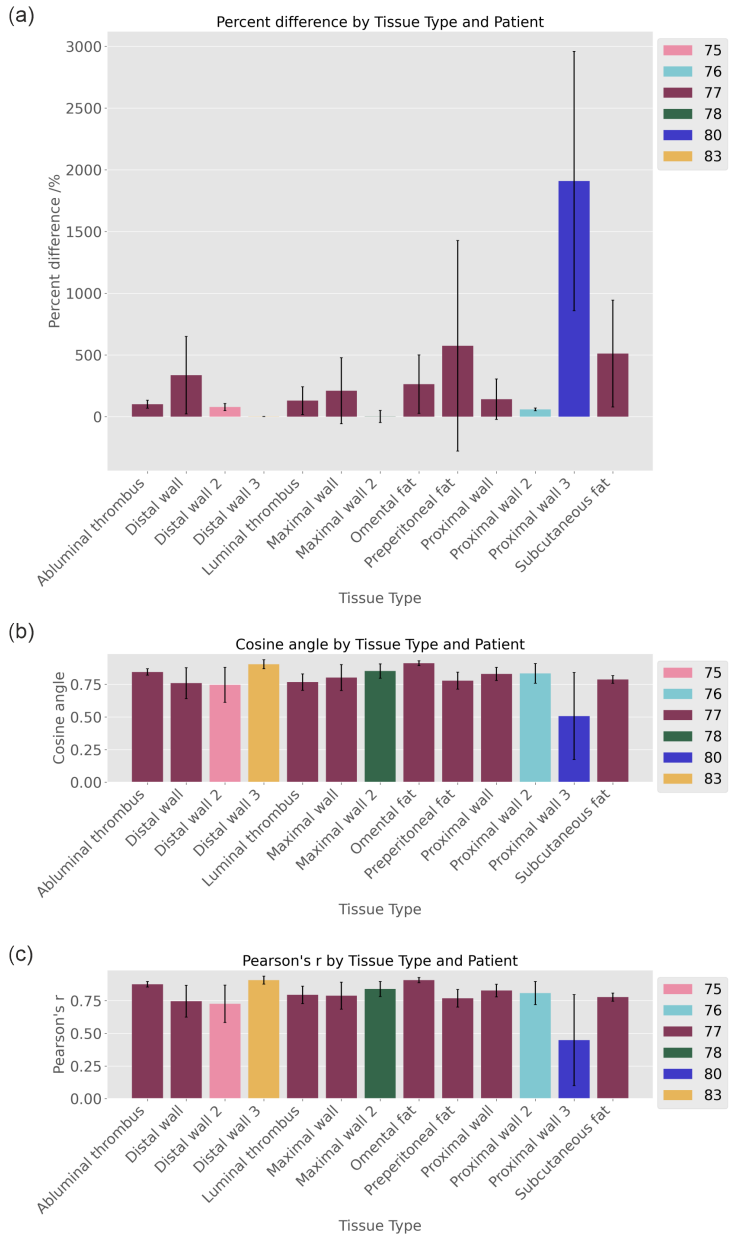

Figure D.6 Bar charts showing the similarity of spectra for tissue samples recorded on the Radian and CMS-S instruments. Values shown are the average over all repeat comparisons, with the standard deviation shown as error bars. The bar colours represent the six different patient IDs for each sample type. Metrics calculated are percentage difference (a), Pearson's r (b), and Cosine angle (c).

D.7 Inter and intra instrument reproducibility: correlation and cosine angle

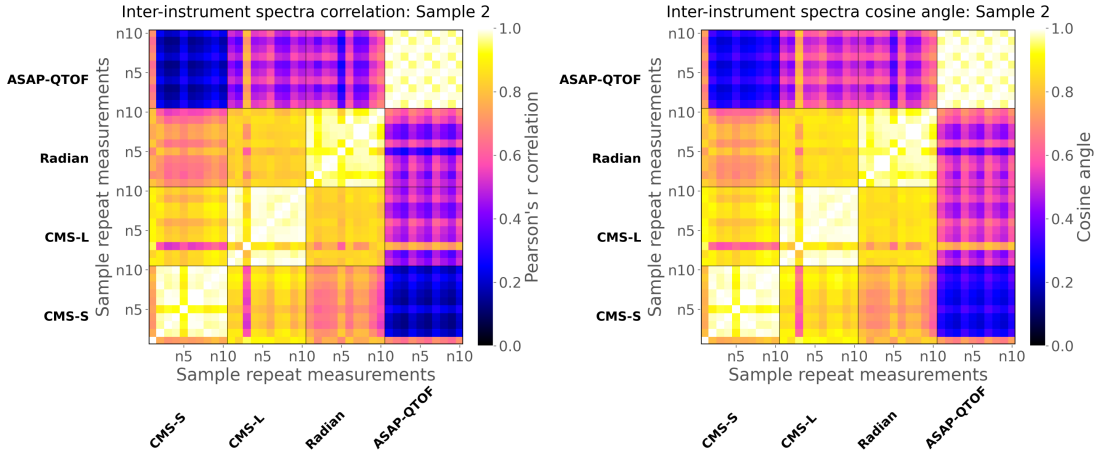

Figure D.7 Pearson's correlation (a) and Cosine angle (b) inter-instrument analysis for plasma sample 2.

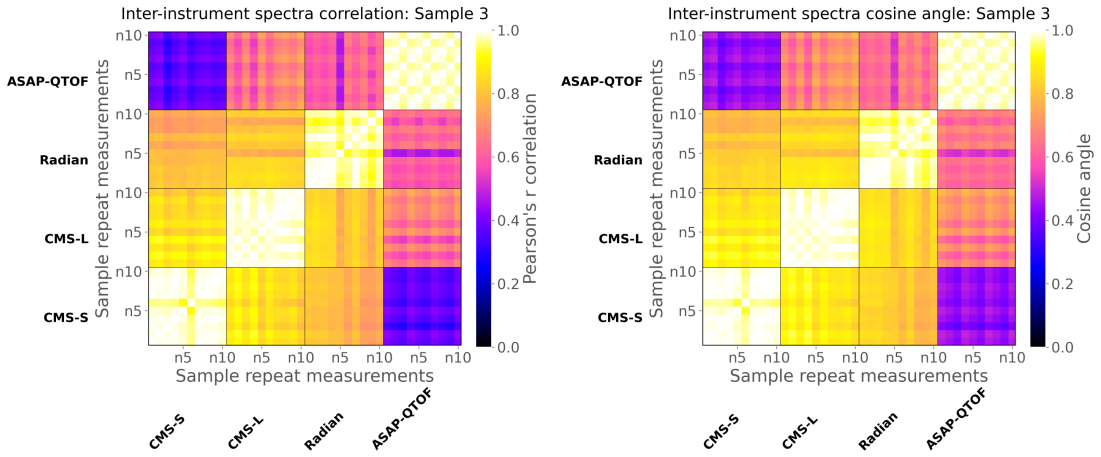

Figure D.8 Pearson's correlation (a) and Cosine angle (b) inter-instrument analysis for plasma sample 3.

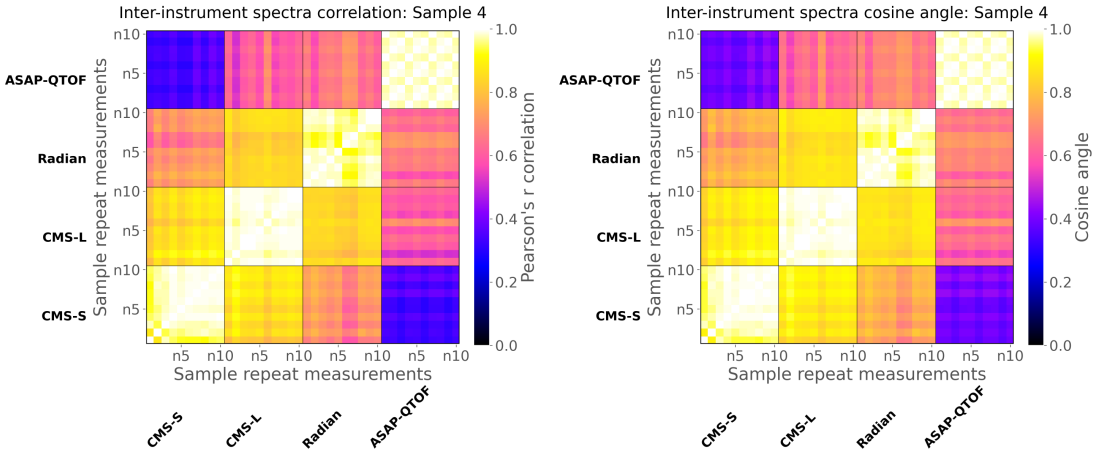

Figure D.9 Pearson's correlation (left) and Cosine angle (right) inter-instrument analysis for plasma sample 4.

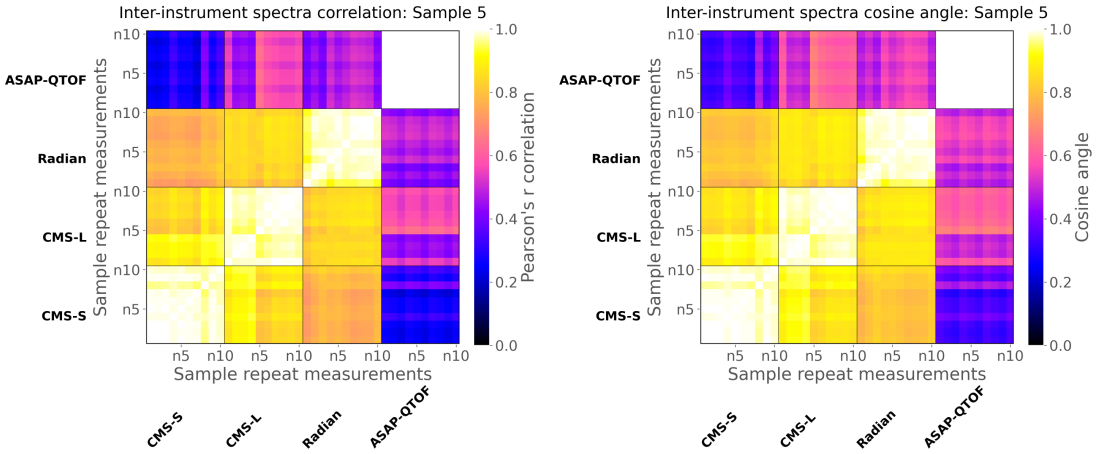

Figure D.10 Pearson's correlation (left) and Cosine angle (right) inter-instrument analysis for plasma sample 5.

D.8 Method for the generation of instrument functions

The steps for generating an instrument function between instruments A and B were as follows:

1. Collect the normalised repeat measurements of plasma samples taken on instruments A and B.

2. Split the data into a training and test set by holding back a set of repeat measurements for the same patient from each instrument.

3. Divide each repeat measurement for a single sample taken on instrument A in the training data by each repeat measurement for the same sample taken on instrument B. Repeat this step over the remaining samples in the training data. Fix all 'inf' and 'NaN' values caused by zero divisions to zero.

4. Average the resulting ratio spectrum over the repeats to generate a single ratio spectrum for instruments A and B.

5. Multiply each repeat measurement in the test set for instrument B by the ratio spectrum to generate the converted dataset.

6. Apply a 3 dimension PCA to the original training data across instruments A and B.

7. Apply the PCA transformation from the training data to the test data and the converted data.

D.9 Instrument function smoothing optimisation

Two common smoothing functions, the Gaussian filter, and a Savitzky-Golay filter, were applied to the raw instrument functions to assess the impact this would have on the conversion of the test data. The use of smoothing functions significantly decreased the accuracy of the data conversion, even with very low smoothing parameters. The sigma parameters for the gaussian smoothing function, and the window length and polyorder parameters for the Savitzky-Golay function, were optimised using a minimisation function by minimising for the distance between the centroid of the target dataset and the centroid of the converted data. The resulting optimised hyperparameter values are shown in Table D.1. Increasing the smoothing, through increasing the sigma value for a Gaussian function or the window for the Savitzky-Golay function, resulted in a decrease in the accuracy of the data conversion. An example of this is shown for the CMS-L to CMS-S conversion with the gaussian smoothing function in Figure D.11. There is a clear divergence of the target and converted data as the sigma value is increased. The optimised smoothing functions were frequently minimal in the extent of smoothing (sigma < 2 and window length < 7). Visualisation of these optimised smoothed instrument functions showed that they were also mass dependent with a non-smooth appearance, shown for all conversions in Figure D.12.

Table D.1 Inter-centroid differences for converted data analysis of ASAP-MS instruments

Target	**Converting**	**Sigma**	**Difference between target SD and target centroid to conversion centroid distance**
CMS-S	CMS-L	1.03	0.26
CMS-S	QTOF	0.7	0.001
CMS-S	Radian	1.05	0.19
CMS-L	QTOF	10	0.11
CMS-L	Radian	10	0.14
Radian	QTOF	10	0.14

Target	**Converting**	**Window length**	**Polyorder**	**Difference between target SD and target centroid to conversion centroid distance**
CMS-S	CMS-L	5	2	0.36
CMS-S	QTOF	5	2	0.002
CMS-S	Radian	5	2	0.026
CMS-L	QTOF	5	2	0.013
CMS-L	Radian	5	2	0.009
Radian	QTOF	5	2	0.003

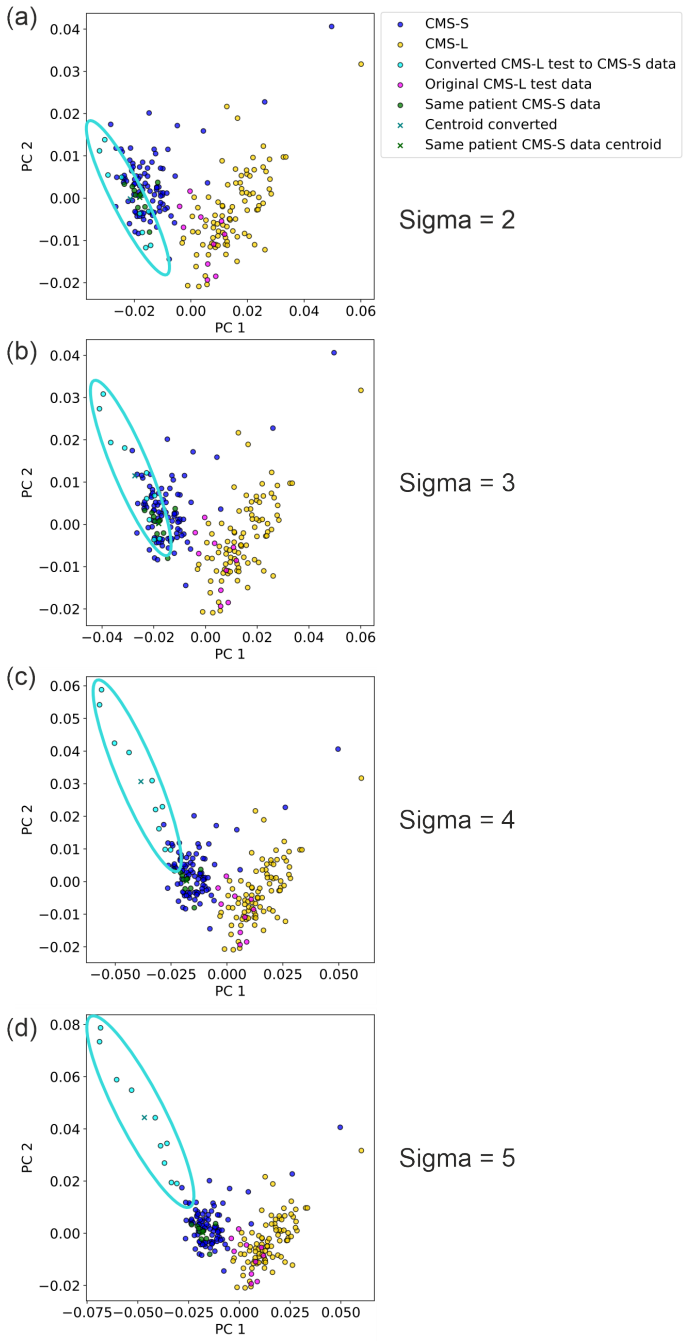

Figure D.11 The effect of increased Gaussian smoothing on the conversion of CMS-L data to CMS-S data. The plots show instrument spectra transformed using PCA, showing the 2D projections of PCs to determine clustering behaviour for the inter-conversion of plasma data. Training data recorded on instrument A is shown in blue, training data recorded on instrument B is shown in yellow, test data recorded on instrument A is shown in green, test data recorded on instrument A is shown in magenta, and the test data recorded on instrument B converted to instrument A is shown in cyan. The converted data is highlighted with a cyan ellipse.

D.10 Instrument function centroid distances

Table D.2 The metrics calculated for the interconversion of plasma data recorded across the four ASAP instruments.

instrument *A* **(Target)**	**Instrument** *B* **(To convert)**	**A-***B* **instrument distance**	**SD instrument** *A*	**Instrument function smoothing**	**Converted** *B* **to A data centroid distance**
CMS-S	CMS-L	0.030	0.008	Raw Gaussian filter Savitsky Golay	0.008 0.009 0.010
CMS-S	QTOF	0.105	0.019	Raw Gaussian filter Savitsky Golay	0.026 0.129 0.032
CMS-S	Radian	0.043	0.006	Raw Gaussian filter Savitsky Golay	0.008 0.265 0.364
CMS-L	QTOF	0.087	0.019	Raw Gaussian filter Savitsky Golay	0.020 0.162 0.029
CMS-L	Radian	0.035	0.006	Raw Gaussian filter Savitsky Golay	0.006 0.193 0.263
Radian	QTOF	0.088	0.019	Raw Gaussian filter Savitsky Golay	0.017 0.158 0.022

D.11 Instrument function PCA plots

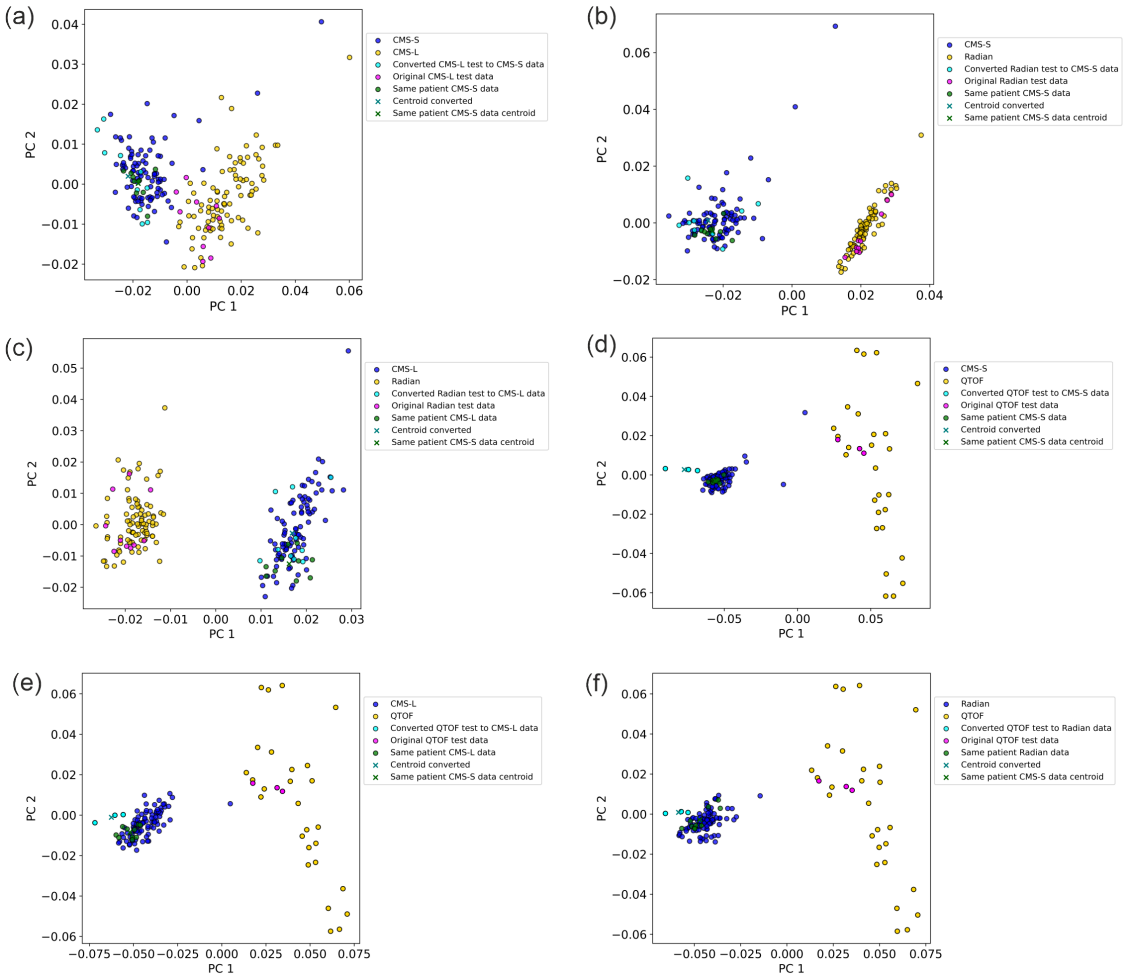

Figure D.12 Plots of instrument spectra transformed using PCA, plotted as PC1 and PC2 showing the interconversion of plasma data from one instrument (B) to another instrument (A). Plots show CMS-L to CMS-S converted data in (a), Radian to CMS-S converted data in (b), Radian to CMS-L converted data in (c), QTOF to CMS-S converted data in (d), QTOF to CMS-L converted data in (e), and QTOF to Radian converted data in (f). Training data recorded on instrument A is shown in blue, training data recorded on instrument B is shown in yellow, test data recorded on instrument A is shown in green, test data recorded on instrument A is shown in magenta, and the test data recorded on instrument B converted to instrument A is shown in cyan.

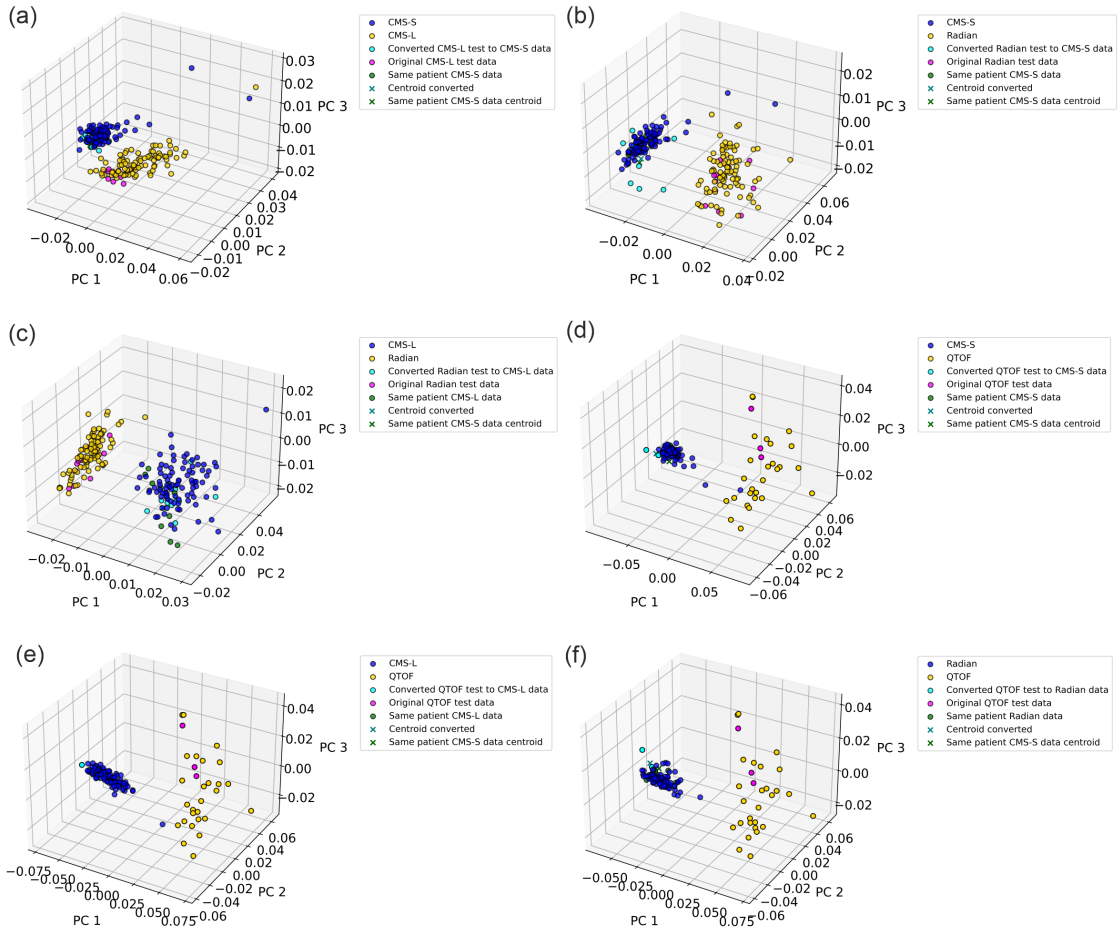

Figure D.13 Plots of instrument spectra transformed using PCA, plotted as PC1, PC2 and PC3 showing the interconversion of plasma data from one instrument (B) to another instrument (A). Plots show CMS-L to CMS-S converted data in (a), Radian to CMS-S converted data in (b), Radian to CMS-L converted data in (c), QTOF to CMS-S converted data in (d), QTOF to CMS-L converted data in (e), and QTOF to Radian converted data in (f). Training data recorded on instrument A is shown in blue, training data recorded on instrument B is shown in yellow, test data recorded on instrument A is shown in green, test data recorded on instrument A is shown in magenta, and the test data recorded on instrument B converted to instrument A is shown in cyan.

D.12 Instrument function with smoothing results plots

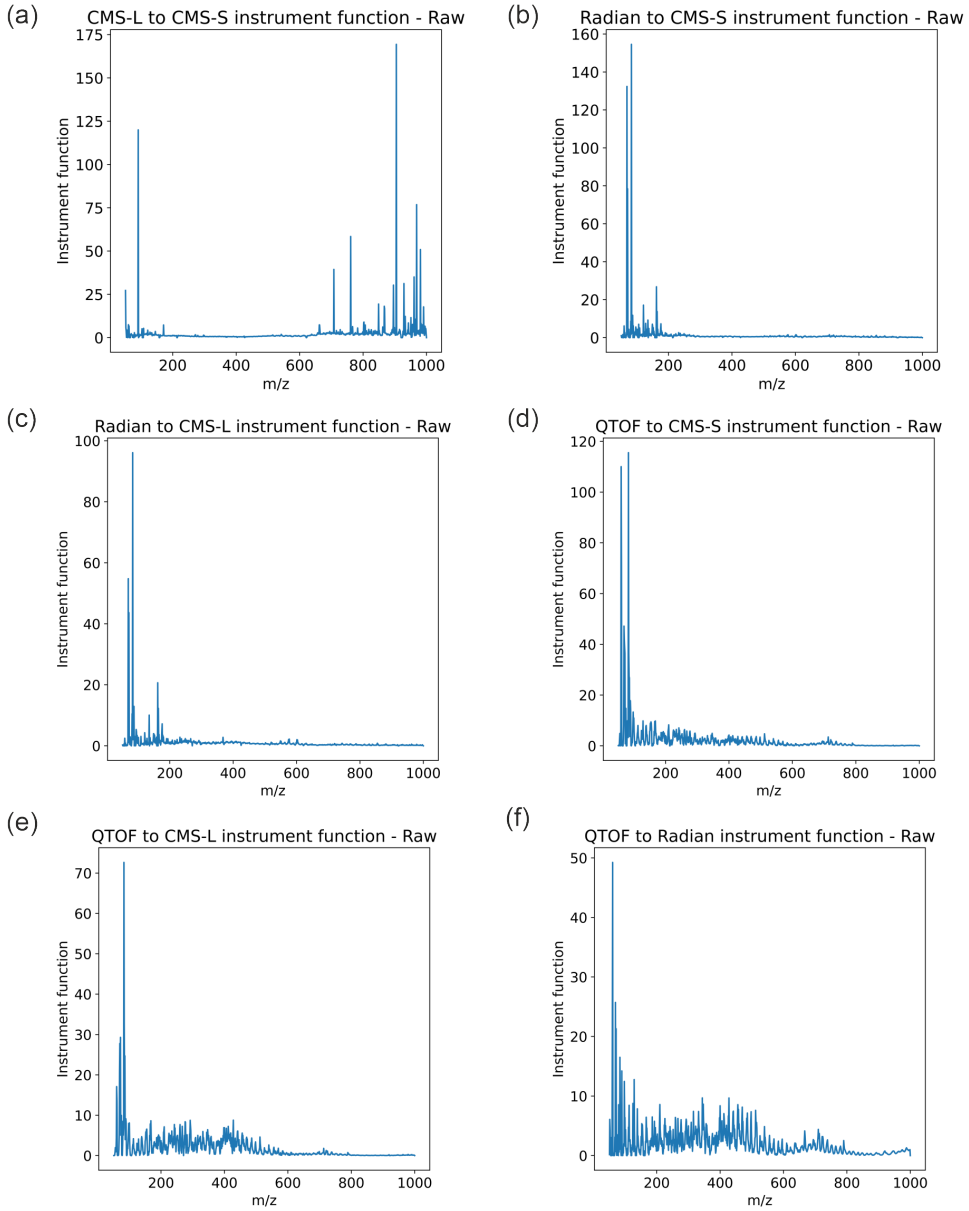

Figure D.14 Plots show the raw instrument functions plotted as mass spectra. The CMS-L to CMS-S instrument function in is shown in (a), Radian to CMS-S instrument function in (b), Radian to CMS-L instrument function in (c), QTOF to CMS-S instrument function in (d), QTOF to CMS-L instrument function in (e), and QTOF to Radian instrument function in (f). Note the significant peak dependence of the functions.

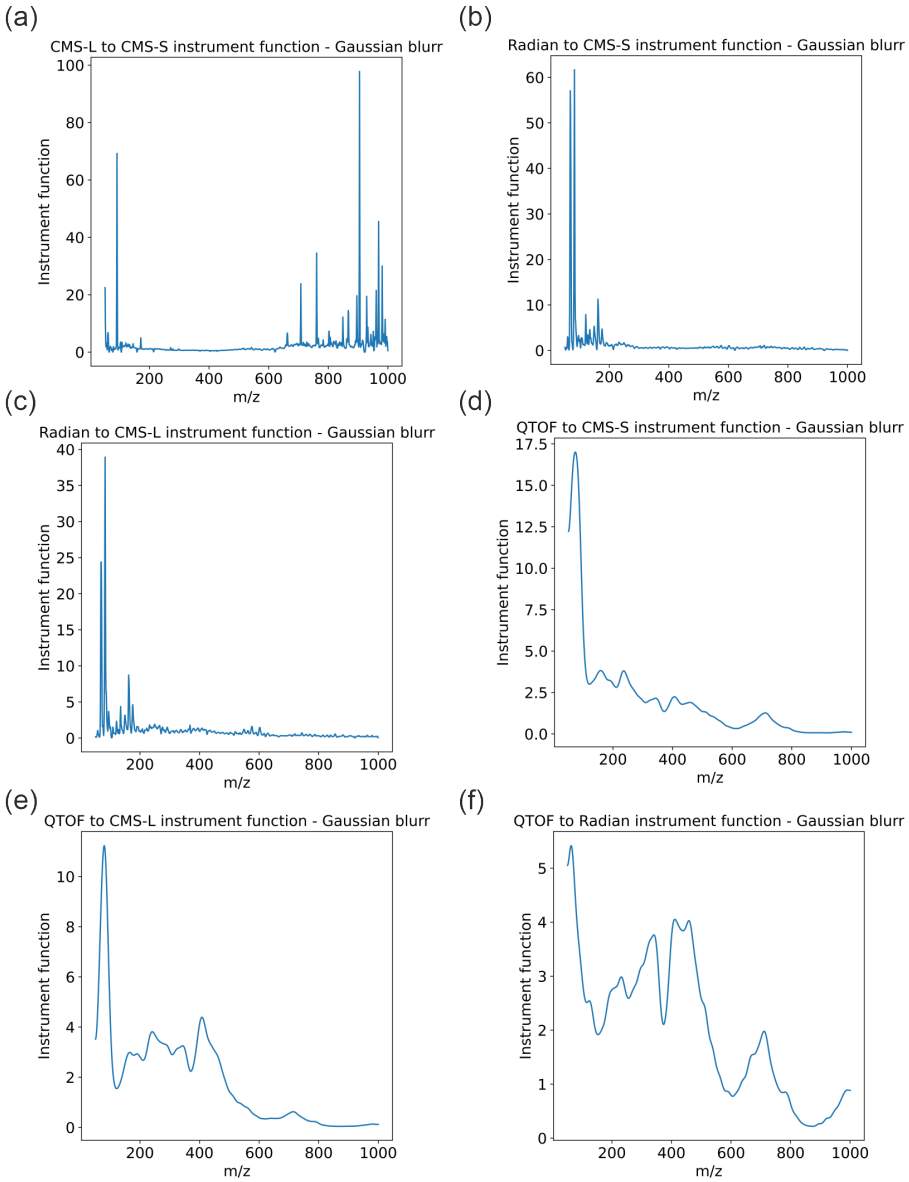

Figure D.15 Gaussian blurr applied to the instrument functions using the smoothing parameters defined in Table D.1. The CMS-L to CMS-S instrument function in is shown in (a), Radian to CMS-S instrument function in (b), Radian to CMS-L instrument function in (c), QTOF to CMS-S instrument function in (d), QTOF to CMS-L instrument function in (e), and QTOF to Radian instrument function in (f).

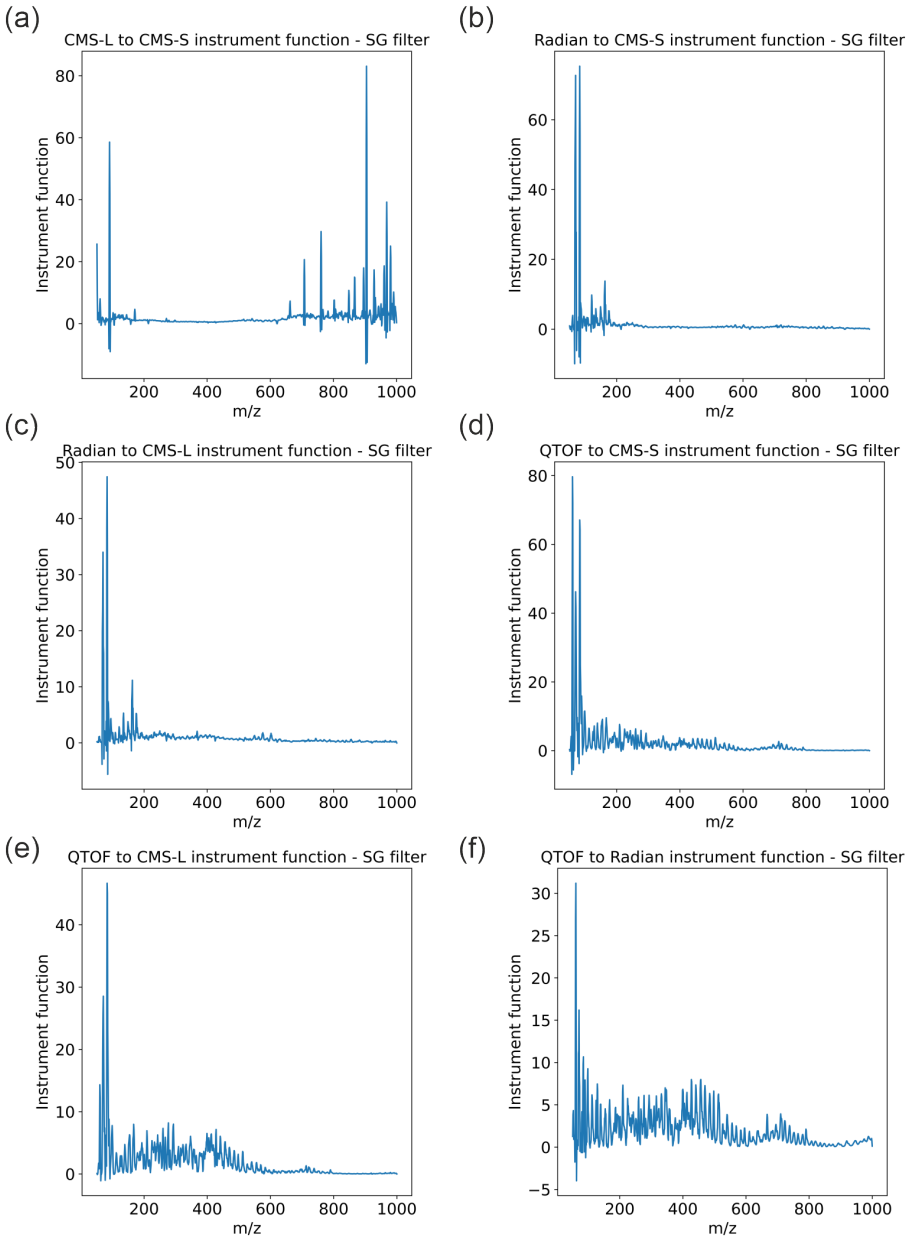

Figure D.16 Savitzky-Golay filter applied to the instrument functions using the smoothing parameters defined in Table D.1. The CMS-L to CMS-S instrument function in is shown in (a), Radian to CMS-S instrument function in (b), Radian to CMS-L instrument function in (c), QTOF to CMS-S instrument function in (d), QTOF to CMS-L instrument function in (e), and QTOF to Radian instrument function in (f).

D.13 Neural network ASAP-MS resolution mapping

D.13.1 Introduction: Using neural networks to map between low-resolution aASAP-MS data and high-resolution ASAP-QTOF-MS data

As was shown in the previous section, the ASAP-QTOF-MS data for clinical samples was shown to be highly complex, but due to the resolving power of the instrument being very high, the molecular formula of ion m/z peaks can be tentatively identified. In Chapter 6, we showed that it is possible to convert a spectrum recorded on one instrument into a format that matches the structure that would be present if it were recorded on another instrument. To do this from a low-resolution measurement to a high-resolution measurement would be potentially very valuable. It would allow for measurements to be taken on a cheaper, faster instrument, and then translated to an equivalent high-resolution spectrum. Due to the higher resolution and sensitivity, the number of features, or peaks, in a ASAP-QTOF-MS spectrum is high, much higher than in the single quadrupole ASAP-MS datasets. This makes mapping between the low and high-resolution data challenging, and cannot be done using a simple instrument function approach as discussed previously. One method of achieving this is to use neural networks, in a similar method that might be used to increase the resolution of an image. This uses an expanding network, in which the number of output neurons is greater than the number of input neurons, and upsampling or expansion layers used to learn increases in data resolution within the net. To determine if this would be feasible, the potentially 'easier' direction of converting the high-resolution data to low-resolution data was attempted using a more typically structured neural network in which the hidden layers decrease the number of neurons from the input to the output. A challenge of using neural networks is that they typically require very large datasets (thousands to millions

of samples) to accurately learn this type of mapping. Collecting a dataset on this scale is unfeasible in this context. However, as we have characterised the error and variation in plasma spectra on the instruments in question, artificial data could be created to build a large training dataset based on a smaller set of samples.

D.13.2 Methods

Plasma samples were selected at random from the AAA study discussed previously, consisting of 100 S1 and S2 patient plasma samples. Samples were thawed from -80°C to room temperature and mixed by vortex mixer prior to sampling. The low resolution ASAP-MS data was collected using an Advion CMS-S instrument, using the optimised plasma measurement protocol described previously in Section 2.4.4. The high resolution ASAP-MS data was collected using a Waters Xevo ASAP-QTOF-MS using the method described in Section 6.2.4. On both instruments, only a single repeat measurement was made for each plasma sample due to time constraints, and in the interest of generating artificial repeat data.

All data processing and neural network analysis was written and conducted in Python 3.7. Neural network analysis was run using Tensorflow version 10.2 on a Windows native system fitted with an NVIDIA A6000 GPU.

D.13.2.1 Data processing for high-resolution data

Preparation of the ASAP-TOF-MS data for analysis required additional steps compared to the CMS-S data due to the large size of each data file, and format in which the instrument data is reported. A workflow was generated to process the TOF data, which is discussed below, and summarised in the flow chart in Figure D.17.

As described previously, the CMS-S instrument data for a single acquisition are reported as an x, y, t data table, in which x is the m/z range in increments of 0.05, y is intensity, and t is time in scan time increments (seconds). This is straightforward to interpret and plot. The ASAP-QTOF-MS instrument has a different data export

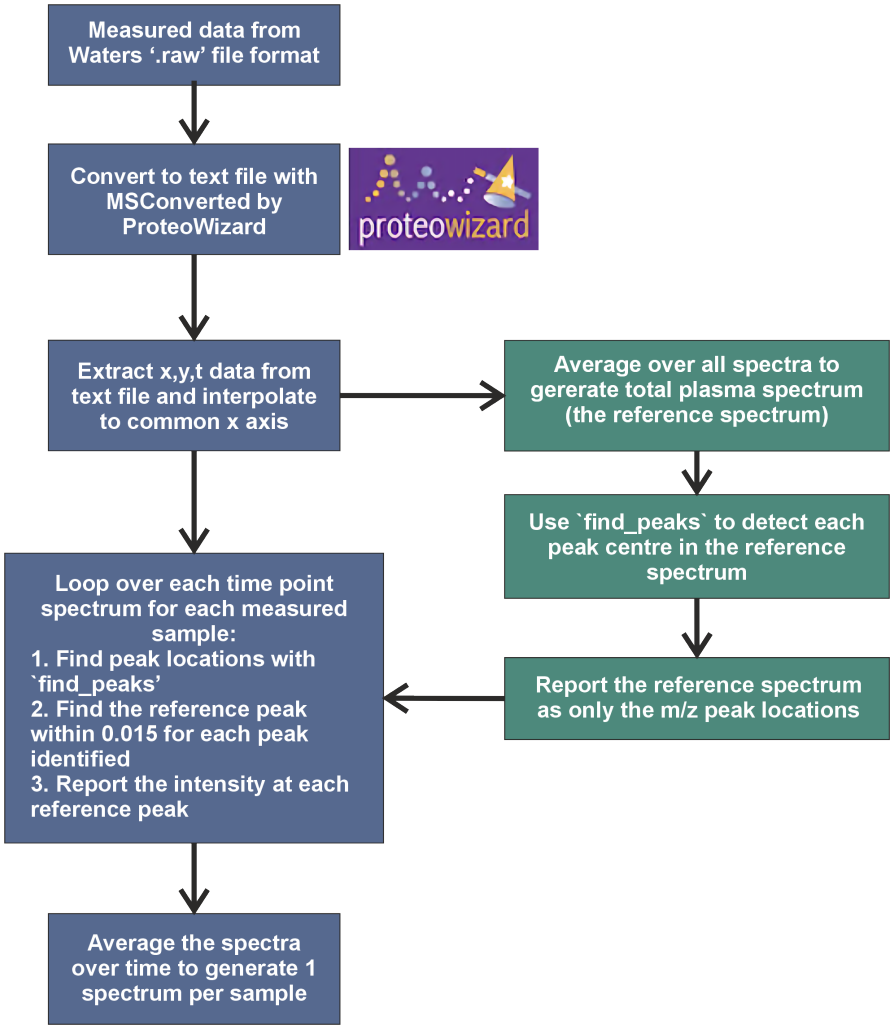

Figure D.17 The data processing protocol for the ASAP-QTOF-MS data from the raw data output of the Waters Xevo G2 XS instrument into a single spectrum per sample that can be used in further analysis.

format, the '.raw' format, which is used across many Waters platforms. This must be converted prior to analysis in non-specialised software packages. The open-source MS data support software, MSConvert by ProteoWizard (ProteoWizard version 3.0.24138) was used to convert each high resolution file from '.raw' to text [384, 385]. The high-resolution data is reported as x, y, t data only for intensity detected, rather than for every m/z increment possible. Peak drift (change in the m/z location of a peak maximum) occurs across both time and samples within the data. For the analysis to be run on this dataset, the peak positions must be corrected so that they occur on a common axis. A software package was written in Python to adjust the

data for this error. The functions within this package took the following steps.

1. Find the time point data in the text file and read the m/z and intensity data for that time point

2. Interpolate the data for that time point to fit to an m/z axis that includes every m/z increment from 50 to 1000 with an increment of 1×10^{-5}. This is the resolution of the instrument as reported in the manual. Intensity data added to the matrix have the value of 0.

3. Stack this data in the new sample file so that the file has the format x, y, t where x is the m/z range from 50 to 1000 with increments of 1×10^{-5}, y is intensity, and t is the time point in seconds.

4. Move onto the next time point, repeat steps 2-4.

The peak drift, or x-error, is corrected by aligning each spectrum to an 'ideal' reference spectrum. This was generated by taking an average of all the spectra recorded across the samples and time points, which shows the normal distribution of error in peak position, and allows the true peak maximum position to be found. An example of this is shown in Figure D.18. This was done using the peak detection function `'find_peaks'` to detect the true m/z peak location at the top of each Gaussian, with a minimum peak height of 250 intensity units, and minimum inter-peak distance of 0.01 m/z units between peaks. The error at the peak width half height for this instrument was found to be 0.015 from this analysis, shown by the FWHM measurement in D.18. This found 5954 peaks for an average plasma sample. Each sample spectrum was then fitted or aligned to this reference spectrum. The `'find_peaks'` function was used for each sample spectrum to identify the peak locations in the spectrum, with a minimum height of 100 and minimum inter-peak distance of 0.01 m/z.

The minimum height was reduced from the reference generator to account for peak height variation. For each peak in the reference spectrum, if a peak was

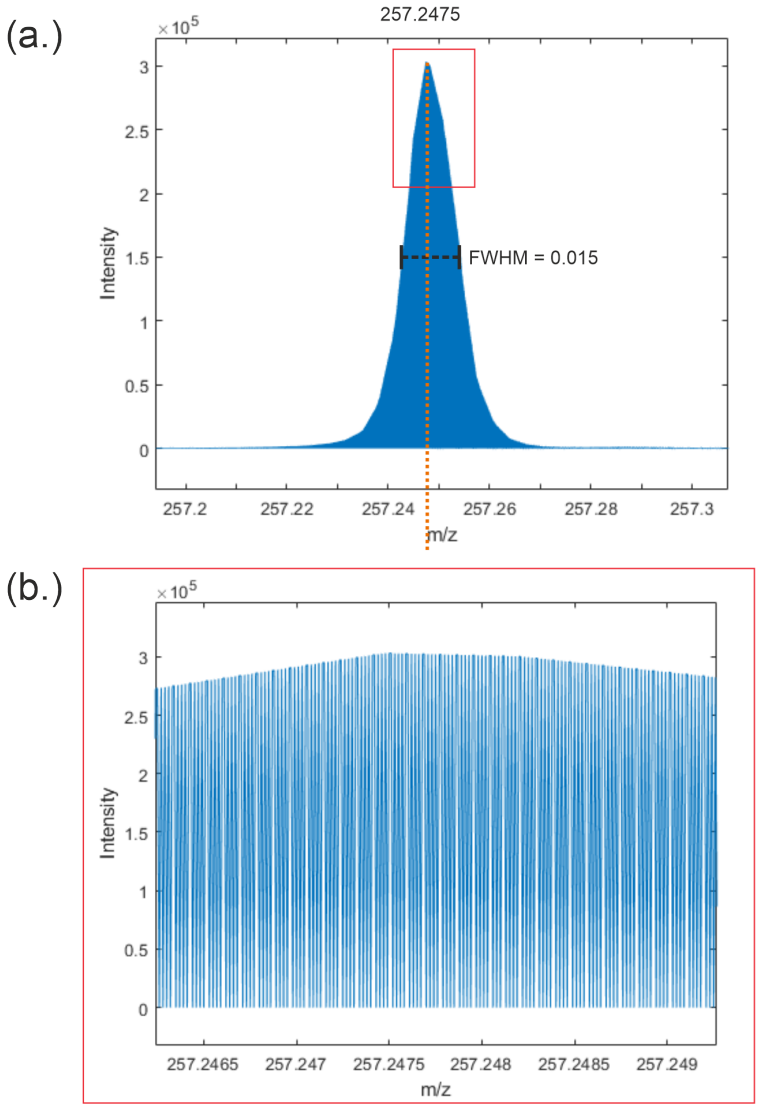

Figure D.18 The TOF reference spectrum generator, showing a Gaussian peak example. The peak in (a) shows the full-width half maximum (FWHM) of a typical peak with a peak maximum shown at 257.2475 *m/z*. The spectrum in (b) shows a zoomed-in view of the red box in (a), showing the spectral resolution signal at each *m/z* instance, showing how the peak width is spread over the mass range.

identified in the measured spectrum at less than 0.015 m/z away from the reference, the intensity at that peak was reported at the reference location. This resulted in the conversion of every spectrum recorded on the TOF instrument to a common x axis equal to the reference spectrum, reducing the size of the variable space from approximately 95×10^6 to 5954.

D.13.2.2 Data processing for low-resolution data

The low-resolution data, recorded on the CMS-S instrument, was processed as described previously for the OxAMI and OxAAA projects (peak selection, background subtraction, bin to unit mass). However, the same reference spectrum generation and fitting was then conducted to ensure that both datasets had been treated in the same manner. The `'find_peaks'` function was used for each sample spectrum to identify the peak locations in the spectrum, with a minimum height of 1000 and minimum inter-peak distance of 0.6. This reduced the size of the m/z variable space from 990 to 253. This also ensured that the mapping from input to output data would be peak to peak, preventing the neural network from also having to understand the distinction between peaks and background in this instance.

D.13.3 Generation of artificial data

Artificial data, or synthetic data, is data that is artificially generated using statistics, simulations, modelling, or ML models, rather than being collected in experimental methods. These models typically use combinations of known physical properties, predicted behaviours, experimental data, simulated data, and errors, as a starting point to generate a much larger set of data that replicates the patterns and features of the experimental data. This approach is frequently used to generate large training datasets for building ML models, as collecting such amounts of experimental data would be very time consuming and expensive, but large training datasets are commonly needed to generate generalised ML models. In its simplest form, artificial data can be generated by adding random noise to an experimental dataset. However, for more complex data, artificial intelligence (AI) models have been developed that are able to learn patterns in the original data and generate new data that contain those features. The most common example of this is the generative adversarial network (GAN), which uses two neural networks, a generator and a discriminator, to generate new data instances that resemble the training data [386, 387]. The

generator creates synthetic data based on the training data, while the discriminator tries to distinguish between the real and generated data. Both networks improve in performance with iterations, leading to a competitive training process in which artificial data becomes more similar to the experimental data [388].

During the analysis of the OxAMI data, the results of which are discussed in Section 4.1, a number of attempts were made to generate artificial data. Some of the methods employed included using the peak errors identified in the ASAP-MS data to generate statistical distributions of data for particular patient classes, Gaussian noise modelling, and using ML-based techniques such as GANs [387–389]. However, when artificial data was used in ML algorithms to attempt STEMI patient classification, the models would fail to achieve any accuracy above the lower threshold of $\kappa = 0.2$. It was determined that there was enough samples of experimental data compared to the dimensionality of the ASAP-MS data and distribution of errors to use these techniques, and so statistical and distribution-based methods were used in all subsequent tests.

Artificial data was generated for both the low and high-resolution data for the purpose of neural network training and validation using a statistical error-based method. The m/z-dependent error analysis of plasma spectra recorded on both the high and low-resolution instruments was conducted. The resulting mean peak-specific error is shown in Figure D.19. As observed in previous analysis, the error in the QTOF data was much lower, at an average of 10.42%, than that recorded on the CMS-S instrument at an average of 33.27%, and appears to be more consistent and stable across the m/z range. The mean intensity error and standard deviation at each peak was used to generate artificial datasets for each instrument dataset.

Three methods were used to generate artificial data from this error analysis: a low error dataset with Gaussian error added to each peak with a standard deviation of 5% of the peak intensity, a high error dataset with Gaussian error added to each peak with a standard deviation of 10% of the peak intensity, and a peak-specific

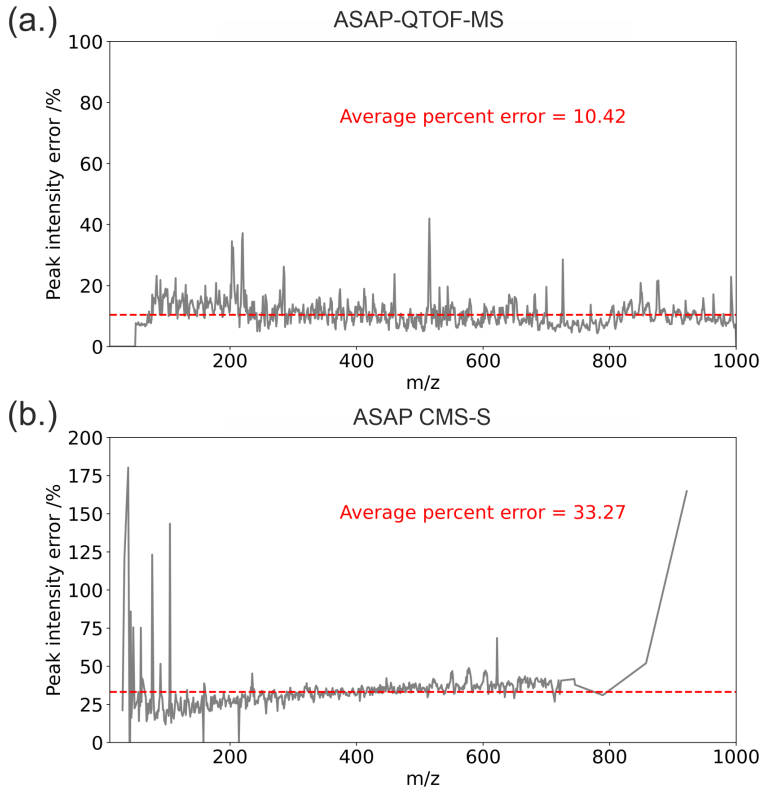

Figure D.19 The percentage error in normalised peak intensity calculated for plasma spectra as measured using the ASAP-QTOF-MS (a), and Advion CMS-S instrument (b). The spectra have been matched to the reference spectrum, cleaned to only the reference peaks, and normalised. The high *m/z* in the CMS-S spectrum is due to the lack of high-intensity peaks at these values.

error dataset with Gaussian error added to each peak with a standard deviation equal to that found for that peak in the error analysis of the peak intensity. For each patient sample, this was repeated 100 times, generating a dataset of 10,000 ASAP-MS spectra for each of the low and high-resolution datasets respectively, and for each of the three error levels. The artificial data differed from the measured data by error in the y (intensity) dimension only.

D.13.4 Neural network conversion of high-resolution data to low-resolution data

The combined artificial and experimental high-resolution data was partitioned into training and testing sets using an 80:20 ratio, ensuring that the artificial data built

based on the same samples were held in the same partition to prevent sample-based data leakage. An NN was built to take the high-resolution training data as the input, and the corresponding low-resolution experimental data for each as the output. [a] The artificial low-resolution data was not used here as these were not ground truths.

The architecture of an NN, which represents the structure of the net and includes features such as the number of neurons, hidden layers, and epochs, is very important for the success of a model, and needs to be tailored and optimised to fit the problem. Optimising an NN architecture is not a trivial process. For datasets with few dimensions, fairly simple nets can be used as starting points, and then changed by the user in a trail-and-error, or manual tuning, process to optimise the accuracy. When the problems are larger, however, more sophisticated methods may prove more useful, particularly when the number of input and output variables are different (for example, when changing the resolution of data). A number of methods exist to build and optimise NN architectures, including using genetic algorithms (also called Neuroevolution) in protocols such as NEAT (NeuroEvolution of Augmenting Topologies [390]), random and grid search algorithms [391], bayesian optimisation (BO) [391, 392], and reinforcement learning (a reward-based learning process) [393]. A number of methods were explored, with the BO method proving to be the most efficient in terms of computational costs. In this instance, the NN was optimised using BO in a method similar to that described in [394].

The architecture hyperparameters that were optimised are defined in D.4. The options input for optimisers, activation functions, dropout levels, and initial points are shown in the optimisation table in Table D.14. The error in the NN was determined as the RMSE between the predicted low-resolution data and the real low-resolution data.

[a] For example, 100 artificial high-resolution spectra were generated from the *patient A* spectra and are included in the training data, the label for each of those instances would be the single low-resolution measured experiential spectrum for *patient A*.

D.14 Bayesian optimisation for neural network architecture design - model setup

Table D.3 Neural network architecture optimisation hyperparameters that were used within the Bayesian optimisation method.

NN structural feature	**Input options**
Optimizer	'SGD', 'Adam', 'RMSprop', 'Adadelta', 'Adagrad', 'Adamax', 'Nadam', 'Ftrl','SGD'
Activation function	'relu', 'sigmoid', 'softplus', 'softsign', 'tanh', 'selu', 'elu', 'exponential', keras.LeakyReLU
Initial points	100
n iterations	1000
Dropout level	0.2

The BO and NN analysis was applied to the conversion of high-resolution data to low-resolution data (a downscaling transformation) at each of the error levels described previously, as well as from low-resolution data to high-resolution data (a super-resolution transformation). Once the BO process has completed, the output-optimised hyperparameters are used to build the NN, which is then trained on the training data. The trained model can then be applied to the test data, and the accuracy of the predicted data can be assessed. This analysis was conducted using Python 3.11.7 using the libraries `tensorflow` version 10.2, `keras` version 3.7.0, and `sklearn` version 1.5.2 on a Windows native system fitted with an NVIDIA A6000 GPU.

Table D.4 Neural network architecture optimisation hyperparameters that were used within the Bayesian optimisation method. Range values are input to the Bayesian optimisation as (lower limit ; upper limit) using the range of values defined.

NN structural feature	**Initial range in Bayesian optimisation**	
	Lower limit	**Higher limit**
n **neurons**	10, 50	1000, 6000
Learning rate	1×10^{-5}, 1×10^{-4}	1×10^{-3}, 0.1
Batch size	32, 64	256
Epochs	20	100
Hidden layers set 1	0	5, 10, 20, 50
Dropout layer	0	1
Hidden layers set 2	0	5, 10, 20, 50

D.15 NN resolution analysis: Results and Discussion

During the course of this analysis, no NN was generated in which the accuracy of the NN on the test data was greater than 0%. This was the case for both the downscaling and super-resolution directions, and for all of the different error levels used to generate artificial data. Table D.5 shows a summary of the different methods that were used for each analysis. At each analysis level, the BO was set up using the range of hyperparameter options discussed previously, creating an extensive search area. Unfortunately, it appears that this method is not appropriate for the conversion of these data types. It is likely that much more experimental data across a much wider range of samples would be required to generate a model that was capable of performing this transformation accurately.

The super-resolution problem is unlikely to succeed in this case, as this type of model is not the optimal way of doing this conversion. One way of describing this problem is like trying to get a model that classifies images of cats to suddenly generate a picture of a cat [395], as multiple high-resolution samples can correspond to the same low-resolution input [396]. Using GANs is particularly well suited to this problem, but as was shown in previous analysis, we do not currently have enough data to use these effectively.

Table D.5 Neural network conversion of data resolution results

Resolution change	Artificial data used	Error level %	Accuracy %
Downscaling	No	-	0
Downscaling	Yes	5	0
Downscaling	Yes	10	0
Downscaling	Yes	Peak specific	0
Super-resolution	No	-	0
Super-resolution	Yes	5	0
Super-resolution	Yes	10	0
Super-resolution	Yes	Peak specific	0

It was hoped that finding a method of converting between these instrument types would enable us to more accurately and confidently identify molecular peak identities in our CMS-S data using a high-resolution measurement. Unfortunately, by ASAP-TOF-MS there is still a high degree of fragmentation and many differences in the instrument setup and conditions that result in differences in the mass spectra. Tentative chemical formula assignments can be made for peaks, but without correlation to the lower resolution spectra, we cannot be certain that the same molecules are contributing to the same peaks in each case.

E

Appendix: Investigation of normalisation methods applied to ASAP-MS data

E.1 Normalisation results across different ML classification methods

Four other supervised ML classification models were used in the same way that the LDA model was used. The results for the no normalisation, AUC normalisation, and best combined model, are shown here for a KNN model (Figure E.1), SVM model (Figure E.2), NBC model (Figure E.3), and RFC model (Figure E.4). The results across different ML models were very similar, except the SVM model which performed very poorly when no normalisation was applied. The best scores obtained regardless of ML model are shown in Figure E.5. The methods used to generate these accuracies are reported in Tables E.1-E.4. The model that gave the best ML accuracy across all five models is shown in Table E.5.

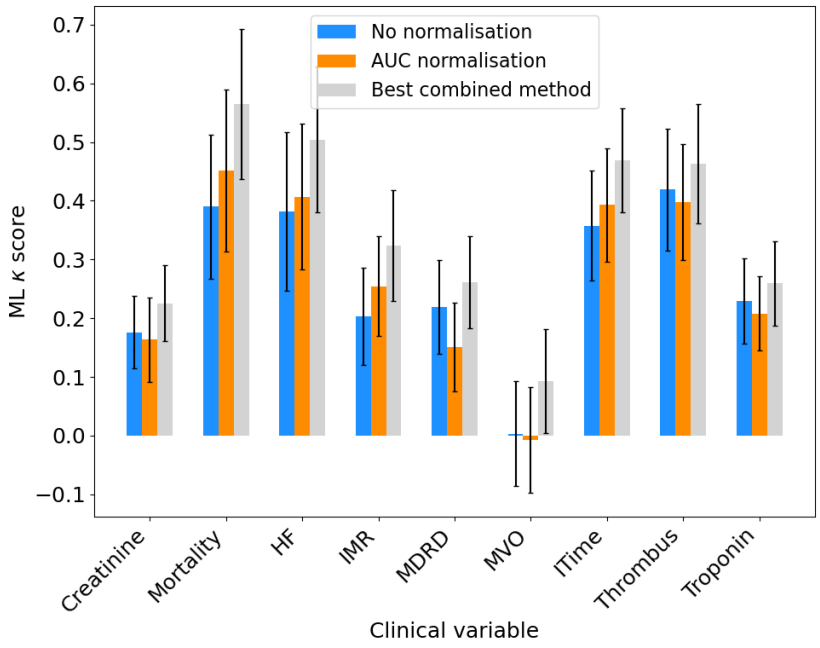

Figure E.1 Comparison of normalisation, scaling and transform methods for clinical variables evaluated by KNN classification accuracy κ score. Classification accuracy is shown with no normalisation, scaling and transformation (blue), with AUC normalisation with no other methods (orange), and the best normalisation, scaling and transformation combined method (grey). Errors show the standard deviation in κ over partitions.

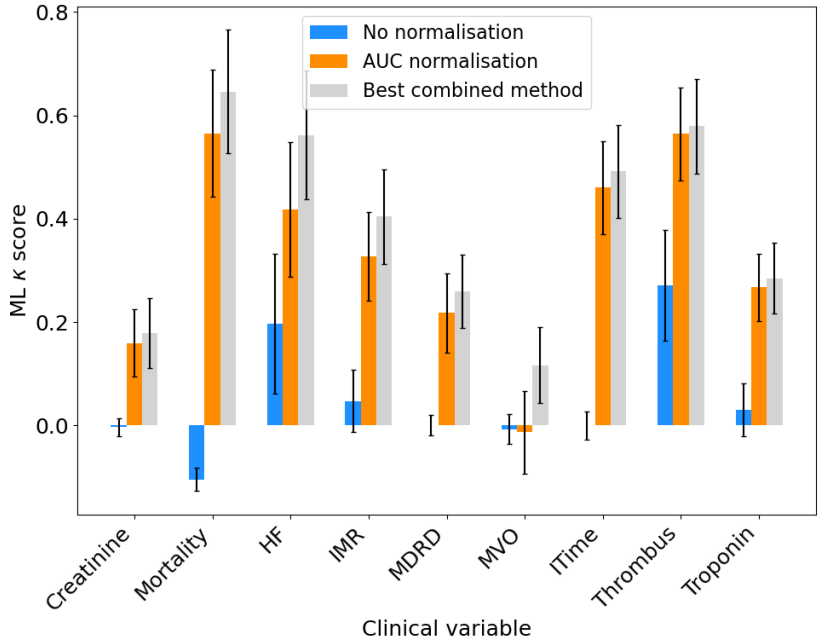

Figure E.2 Comparison of normalisation, scaling and transform methods for clinical variables evaluated by SVM classification accuracy κ score. Classification accuracy is shown with no normalisation, scaling and transformation (blue), with AUC normalisation with no other methods (orange), and the best normalisation, scaling and transformation combined method (grey). Errors show the standard deviation in κ over partitions.

Table E.1 Best normalisation, scaling and transform methods for OXAMI variables assessed by KNN ML classification accuracy

	KNN		
	Normalisation	Scaling	Transformation
Creatinine	Vector	None	Cube-root
Death	Vector	Mean centre	None
HF	AUC	Median centre	Log
IMR	Vector	None	Cube
MDRD	Vector	None	Cube
MVO	None	Mean centre	Log
Pain	None	None	Cube-root
Tscore	Quantile	None	Exp
Troponin	Quantile	None	Square

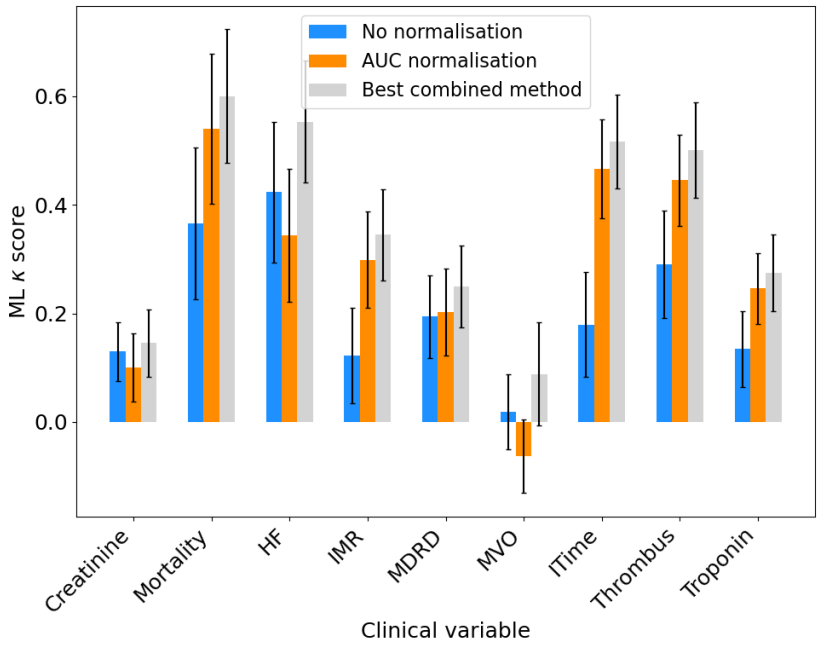

Figure E.3 Comparison of normalisation, scaling and transform methods for clinical variables evaluated by NBC classification accuracy κ score. Classification accuracy is shown with no normalisation, scaling and transformation (blue), with AUC normalisation with no other methods (orange), and the best normalisation, scaling and transformation combined method (grey).Errors show the standard deviation in κ over partitions.

Table E.2 Best normalisation, scaling and transform methods for OXAMI variables assessed by SVM ML classification accuracy

	SVM		
	Normalisation	Scaling	Transformation
Creatinine	mean	Mean centre	Exp
Death	None	None	Square-root
HF	AUC	Median centre	Log
IMR	Sum	Mean centre	None
MDRD	Vector	None	Square
MVO	Quantile	Range	Square
Pain	Sum	None	Log
Tscore	Sum	None	None
Troponin	Sum	None	Cube

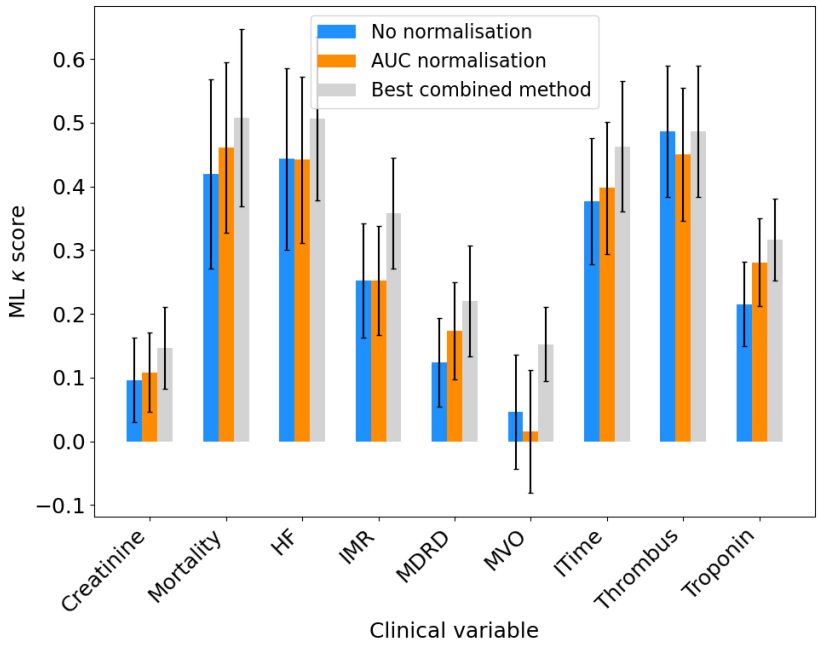

Figure E.4 Comparison of normalisation, scaling and transform methods for clinical variables evaluated by RFC classification accuracy κ score. Classification accuracy is shown with no normalisation, scaling and transformation (blue), with AUC normalisation with no other methods (orange), and the best normalisation, scaling and transformation combined method (grey). Errors show the standard deviation in κ over partitions.

Table E.3 Best normalisation, scaling and transform methods for OXAMI variables assessed by NBC ML classification accuracy

	NBC		
	Normalisation	Scaling	Transformation
Creatinine	None	None	Log
Death	Quantile	None	Square-root
HF	mean	None	Log
IMR	Vector	None	Log
MDRD	Vector	None	Cube
MVO	AUC	Median centre	Cube
Pain	AUC	Median centre	none
Tscore	Quantile	Median centre	Cube-root
Troponin	Sum	None	None

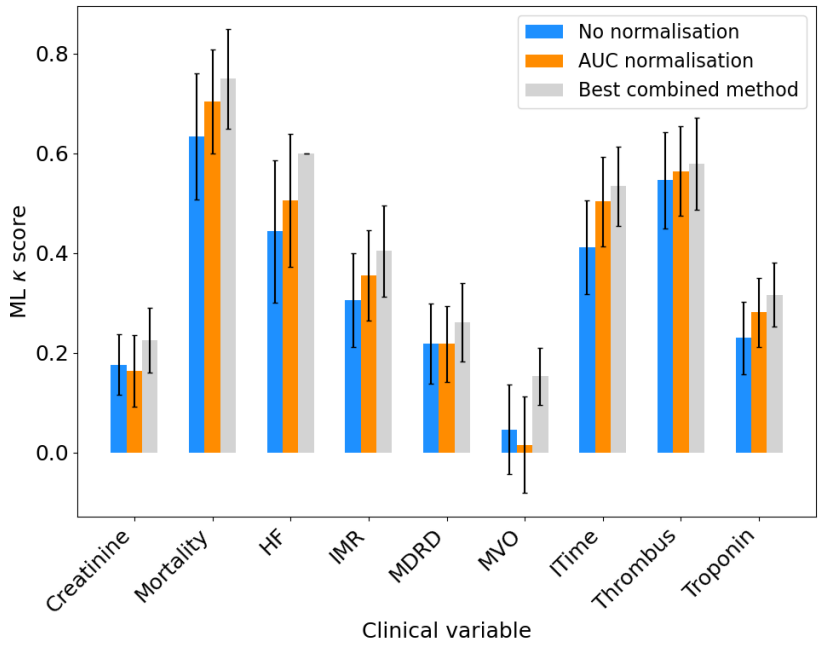

Figure E.5 Comparison of normalisation, scaling and transform methods for clinical variables evaluated by the best ML model across the five models used, shown as the classification accuracy κ score. Classification accuracy is shown with no normalisation, scaling and transformation (blue), with AUC normalisation with no other methods (orange), and the best normalisation, scaling and transformation combined method (grey). Errors show the standard deviation in κ over partitions.

Table E.4 Best normalisation, scaling and transform methods for OXAMI variables assessed by RFC ML classification accuracy

	RFC		
	Normalisation	Scaling	Transformation
Creatinine	Vector	Median centre	Cube
Death	AUC	Median centre	None
HF	Mean	pareto	Square-root
IMR	Vector	Median centre	Exp
MDRD	AUC	Mean centre	None
MVO	Min-vector	Auto	None
Pain	Sum	Median centre	Log
Tscore	None	None	None
Troponin	Vector	Median centre	Cube

Table E.5 Best normalisation, scaling and transform methods for OXAMI variables assessed by the ML classification accuracy, reporting the method that gave the highest accuracy across the five ML models.

	Best overall		
	Normalisation	Scaling	Transformation
Creatinine	Vector	None	Cube-root
Death	Vector	Mean centre	None
HF	Min vector	Mean centre	Log
IMR	Sum	Mean centre	None
MDRD	Vector	None	Cube
MVO	Min vector	Auto	None
Pain	Vector	Mean centre	None
Tscore	Sum	None	None
Troponin	Vector	Median centre	Cube