ENGINEERING SCIENCE

DPHIL THESIS

Intent Sensing for Assistive Technology

Joseph Russell

Worcester College

University of Oxford

Supervised by Professor Jeroen Bergmann

Submitted: August 7, 2023

This thesis is submitted to the Department of Engineering Science, University of Oxford,

in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Composition of the doctoral committee:

Prof. Perla Maiolino and Prof. Ravi Vaidyanathan

Keywords:

Intent Sensing, Assistive Technology, Wearable Technology, Sensor Networks

For my Dad, Iain, who never quite understood exactly what I did but never stopped supporting me.

Summary

This thesis aims to develop systems for intent sensing – the measurement and prediction of what it is that a user wants to happen. Being able to sense intent could be hugely beneficial for control of assistive devices, and could make a great impact on the wider medical device industry.

Initially, a literature review is performed to determine the current state-of-the-art for intent sensing, and identifies that a holistic intent sensing system that properly captures all aspects of intent has not yet been developed. This is therefore followed by the development of such a novel intent sensing system.

To achieve this, algorithms are developed to combine multiple sensors together into a modular Probabilistic Sensor Network. The performance of such a network is modelled mathematically, with these models tested and verified on real data. The intent sensing system then developed from these models is tested for sensing modalities such as Electromyography (EMG), motion data from Inertial Measurement Units (IMUs), and audio. The benefits of constructing a modular system in this way are demonstrated, showcasing improvement in accuracy with a fixed amount of training data, and in robustness to sensor unavailability – a common problem in prosthetics, where sensor lift-off from the skin is a frequent issue.

Initially, the algorithm is developed to classify intent after activity completion, and this is then developed to allow it to run in real-time. Different classification methods are proposed and tested including K-nearest-neighbours (KNN), before deep learning is selected as an effective classifier for this task. In order to apply deep learning without requiring a prohibitively large training data set, a time-segmentation method is developed to limit the complexity of the model and make better use of the available data. Finally, the techniques developed in the thesis are combined into a single continuous, multi-modal intent sensing system that is modular in both sensor composition and in time.

At every stage of this process, the algorithms are tested against real data, initially from non-disabled volunteer participants and in the later chapters on data from patients with Parkinson's disease (a group who may benefit greatly from an intent sensing system). The final system is found to achieve an accuracy of 97.4% almost immediately after activity inception, increasing to 99.9918% over the course of the activity. This high accuracy can be seen both in the patient group and the control group, demonstrating that intent sensing is indeed viable with currently available technology, and should be further developed into future control systems for assistive devices to improve quality of life for both disabled and non-disabled users alike.

Contents

1

Introduction

1.1 What is Intent?

Within the context of this thesis, intent is understood as "what it is that a person wants to make happen". This operates on many levels, both short and long-term. For example, a user might be sitting on the sofa, watching television. Their intent at this time might be to continue sitting and watching television. However, at some point, they might decide they are hungry, at which point their long-term intent might be to get a snack. At this point, the user might not yet have moved in any way, but while the activity has not yet begun, their intent has changed.

To achieve this goal, the user's short-term intent must first be to stand up. They might then start walking, and the intent of this motion is to get to the kitchen, after which they might plan to open the fridge. They may have planned what snack they would like, or might plan to decide based on what they find in the fridge. Retrieving the snack will

involve a number of specific arm movements and motor control tasks, with precise short-term intents. The user might then return to their sofa and continue watching television.

The continuous prediction and monitoring of all these different intent possibilities, throughout a user's activities of daily living, are what is known as "intent sensing". Intent has been discussed in existing literature [1,2], but the definitions and methodologies employed vary considerably. A clear definition of intent is required, along with an understanding of the extent to which it has already been explored, and this will be investigated in Chapter 2.

1.2 Why Is Sensing Intent Useful?

Understanding human intent has the potential to revolutionise the way we interact with technology. Rather than relying on user input, smart devices could learn to anticipate our needs, saving time and increasing user productivity. One of the biggest potential impacts, however, is in healthcare. Assistive technology is being developed to help support those with limited mobility, such as the elderly [3], amputee patients [4], and people living with illnesses such as Parkinson's disease [5]. Being able to predict user intent could make a huge difference to ensuring the safety of users of assistive technology, and improving quality of life.

Referring back to the example in Section 1.1 – if the user is elderly, then the act of standing is likely to lead to an increased fall risk [6]. Determining that the user is about to try to stand could be used to activate monitoring devices, to alert a carer, or to activate an

assistive smart brace or exoskeletal device. Lights could then automatically switch on, safely guiding the user to the kitchen. The television could automatically switch off to conserve power and save the user having to do this themselves. When the user is near enough, the fridge door could even open automatically, and close itself once the user has moved far enough away. These automated responses would support the user's need for independent living, allowing them to maintain their quality-of-life without requiring the intervention of others.

Intent sensing could also have particular benefits for amputees using smart prosthetic devices. Many of the most sophisticated prosthetic upper limbs [7], are limited in their functionality by the necessity for users to manually select which grip they would like to use. Usually, selection may be done with an app [7], a button interface [8], or gesture control [9], but in all these variations, it is indisputably clear this is nowhere near as intuitive as natural control of a healthy biological limb. It already has been suggested that humans are sensitive to small changes in the environment, which is reflected in our motor behaviour [10]. Previous work showed how object identification can potentially enhance grasping and create more natural interaction between user and environment [11]. With intent sensing, the usual control limitations of prosthetic devices could be overcome.

In the case of the above example, an intent-driven lower-limb device could automatically recognise the user's intent to stand and switch into the necessary mode, or an upper limb prosthetic could configure to the correct grip needed to open the fridge without any input from the user.

This kind of intuitive control could majorly improve the device's functionality, helping bring the user's physical ability more in line with that of a non-disabled person and reducing instances of prosthetic device abandonment [12].

1.3 How Can Intent be Sensed?

There are many different ways intent can be sensed. In fact, since intent is an intuitively challenging thing to predict, combining multiple sensing modalities could be key to predicting it accurately. In Section 1.1's example, wearable electromyography sensors could detect anticipatory activity in the muscles and determine that the user is about to move. Motion sensors could be used to categorise the user's movements to determine what activity they are doing. Gaze-tracking sensors could determine what objects the user might be about to interact with, and even contextual factors such as time of day and when the user last ate could be used to anticipate activity before the user even makes a decision. Further discussion into possible intent sensing modalities can be found in Chapter 2.

For this thesis, two hard restrictions are placed on possible sensing modalities. The first of these is that they must be passive; the objective is to measure what a user wants to do without them having to take additional action to deliberately indicate it. Methods such as voice control, gesture control, and interacting with an app interface may all be useful for many applications, but are not passive, so will not be considered for this project.

The second restriction is that the sensors must be non-invasive. Existing literature has shown promising developments in brain-computer-interfaces through implants, interacting with the nervous system and the brain itself to determine what a user is thinking of doing, but this kind of technology carries an inherent risk to health, ethical implications and often a huge cost [13]. The techniques developed in this thesis are intended to be as safe and accessible as possible, and therefore only non-invasive sensing modalities will be used.

1.4 Objective

The overarching objective of this thesis will be to develop systems for intent sensing. This will take the technology from its starting point in the literature, where relatively little research has taken place, to a stage where it can be applied to a group including both disabled and non-disabled individuals to accurately classify their intent (although using this to control an output is considered beyond the scope of this project). To achieve this, a clear scientific definition of intent will need to be detailed, an algorithmic framework will need to be mathematically modelled and built, the most suitable algorithms will need to be identified, and the system will need to be tested using data from real participants.

2

A Systematic Literature Review of Intent Sensing for Control of Medical Devices

Published by Joseph Russell and Jeroen Bergmann in *IEEE Transactions on Medical Robotics and Bionics*, vol. 4, no. 1, pp. 118-129, Feb. 2022, doi: 10.1109/TMRB.2021.3135704.

Chapter Abstract

The usefulness of medical devices, which require user input, is often limited by the control schemes that operate them. The recognition of user intent could enable far more intuitive control schemes that respond automatically to what the user wants the device to do. This chapter provides a definition for intent, and then aims to systematically review current methods for sensing intent. It compares the accuracy of different methods and discusses how they might be combined. A systematic literature search was performed using IEEE Xplore, PubMed and Web of Science databases. 2311 papers were

considered, reduced to 155 after review. All selected papers were assessed for quality using a checklist. The results identified and compared 15 sensing modalities used for intent sensing in a range of situations and applications that broadly fell into 12 distinct categories, with highly varying levels of accuracy. Several papers reached accuracy levels that could be suitable for everyday clinical application, but most work done on intent sensing to date has focused on activity transition classification, with fewer papers addressing task goal interference or predicting future actions. Further work can focus on the implementation of these kind of methods into a combined, context-aware intent sensing control system.

2.1 Introduction

Interactions between humans and technology are fundamental to daily twenty-first century life. Better methods of interacting with devices can enable more intuitive control, increase productivity and improve the overall user experience. Human computer interaction is a growing field of study, with possible implications in almost every workspace and living environment and many opportunities for improvement.

One particular area of great opportunity for research and development is that of intent sensing. Being able to infer what the user actually wants to happen, without them having to actively indicate it, could reduce or even eliminate the need for direct user input devices like buttons, hand-held controllers, etc. and has the potential to truly change the relationship between humans and technology.

This could be of particular interest for the medical technology industry. Clear examples of use cases consist of smart glucose control for diabetics, where research is underway to predict patients' activity in advance [14], as well as for the elderly and prosthetic device users, as discussed in Chapter 1.

The aim of this chapter is to systematically review the literature regarding intent sensing in humans and to summarise the current state-of-the-art in terms of accuracy of these methods. This includes signals from not only wearable devices, but also portable devices such as smart phones and static sensors, which could be found in smart environments such as smart homes (a developing field with increasing potential in the

healthcare industry [15]). The information generated by this technology could be used to assist in intent identification through the concept of context-awareness.

According to [2], intent sensing "includes recognising the current activity, inferring the task goal, and predicting future actions." In this review, the statement is refined slightly in that it refers only to positive activity leading to some change in the user's state. This could include low-level action – e.g. reaching for an object, grabbing it – or a high-level action, e.g. cooking, cleaning, driving, etc. It does not refer to the passive observation of states, such as using gait analysis to classify the current mode of ambulation (e.g. walking or running) or using a pressure sensor in a bed to detect that the subject is lying down. It does, however, include transitions between states – for instance, starting running, beginning to climb stairs, getting out of bed, or slowing down. This chapter therefore does not concern itself with simply monitoring and quantifying human activity/behaviour – but rather with the use of related sensors to infer when a human is transitioning, and if possible, predict what they are going to do next. Additionally, as stipulated in Chapter 1, intent sensing is not considered to include additional manual intervention from the user. Only "natural" actions from the person are considered – if the user is required to indicate their intent in some way, e.g. by pressing a button or performing a specific gesture, this will not be considered true intent sensing.

Intent sensing is therefore defined as predicting future actions, passively recognising an activity state transition and inferring the subsequent task goal. Furthermore, to compare and contrast the performance and potential of each method of intent sensing, papers reviewed in this study must in some way quantify the accuracy of their system.

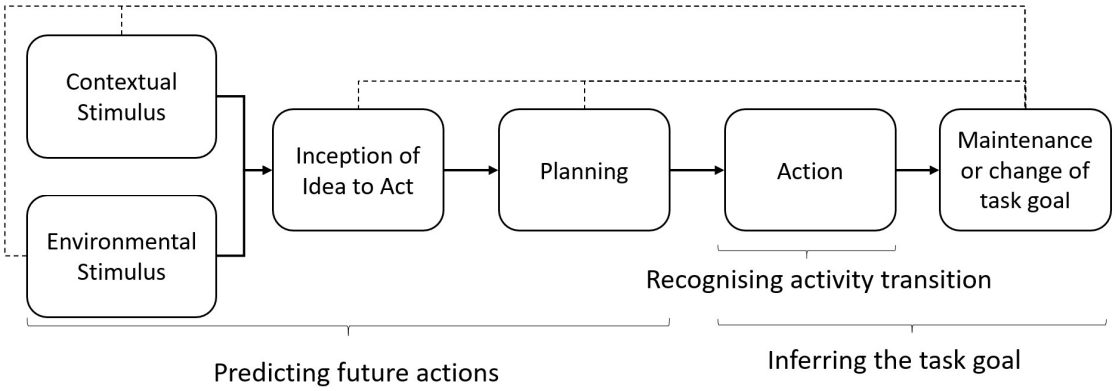

Figure 2.1. *Flowchart detailing the proposed theoretical flow of task completion. The figure highlights the three aspects of intent sensing defined in the scope of this chapter. Contextual stimulus refers to factors such as the user's routine, emotional state, previous experience, desires and instincts which may lead to action. Environmental stimulus encompasses external factors such as proximity to stairs, indicating a user might want to climb them. The maintenance or change of task goal involves continuous re-assessment of the previous steps which may or may not lead to a change in the current action. Feedback loops between the stages (dotted lines) are not considered for the intent definition at this stage. While all elements of this flowchart are important for an overall understanding of intent, the studies performed in this thesis predominantly focus on the region after the inception of the idea to act, and do not attempt to incorporate contextual and environmental stimuli.*

This literature review will therefore summarise the existing methods of human intent sensing, compare their accuracy, and subsequently assess their potential for the design of an intent sensing operating system.

The components for a new definition of intent (see Figure 2.1) are further described below to provide additional clarity beyond the originally proposed definition.

2.1.1 Predicting Future Actions

The first key aspect of intent sensing identified in the definition is prediction – determining what a user is about to do before it happens. Less work has been done on this than activity transition recognition, and performance results on this aspect are highly mixed. The prediction of future actions is essential, as it allows the system to make the necessary changes on time. The transition and goal detection can be determined as it happens, but the prediction needs to be completed beforehand. This is the main difference of this component compared to the following two.

Some studies have integrated prediction into a smart-home activity transition recognition system by tracking user movement and building a profile of the user's typical routines. This allows some degree of high-level prediction, e.g. "the user normally goes to the bathroom before bed." [16]

On a lower level, wearable sensors have been used to predict upper-limb reaching actions through the use of electroencephalography (EEG), but only over time periods of up to a second [17]. Locomotion events can also be predicted through wearable sensors, but again, the timeframe for this prediction is rather short [18].

Increasing the time span would allow for the system to be more prepared regarding the intent of the user, but this is likely to result in an increased uncertainty in terms of accuracy.

2.1.2 Recognising the Activity Transition

Another important part of intent sensing is the concept of activity transition recognition – identifying actions that a user is currently doing or has recently done, specifically when they are changing from one activity state to another. Much of the work in this field has focused exclusively on activity transition recognition for a whole range of applications [19].

Some activity transition recognition techniques leverage the concept of a "smart home," using networks of sensors placed strategically around the user's home to detect changes in their current Activity of Daily Living (ADL) [20]. Often this is done by tracking the user's location, as they move from bedroom to kitchen to living room – however, this gives only a high-level classification of the activity transition without observing specific movements [21].

Individualised sensors – e.g. pressure sensors in the bed, smart detection of kettle turning on, etc. allow more precise classification of activities, but are obviously fundamentally limited, in that each sensor is specifically designed for sensing one particular task [22].

Portable devices have also had their data leveraged to perform activity transition recognition. Smart phones and watches contain sensors such as accelerometers and microphones which can be used to passively gather data to classify user activities throughout the day. These are primarily used to identify periods of exercise, but it is feasible that

data from mobile devices could contribute towards an activity transition recognition sensor network [23].

In intent sensing, any activity recognition should feed into the recognition of an activity transition. The user will always be in a certain state when sensing is starting, and the system will be preparing itself to deal with the next state when a transition is detected. This is a fundamental step of intent detection.

2.1.3 Inferring the Subsequent Task Goal

The final component of intent recognition as defined above is inference of the task goal, which includes identification of the motor goal – the neurally planned outcome of a current physical activity, which may be (automatically) corrected in the event of a perturbation [24]. However, the concept of task goal as considered in this definition extends beyond the motor goal, towards the higher-level goal of a current activity transition, e.g. reaching out in order to pick up a cup, approaching the stairs in order to climb them, or even moving the hands in order to turn a steering wheel and therefore park a car. In a sense, it is a short-term prediction at the end of the current activity. This is distinctly different from predicting future actions. Inferring the task goal refers exclusively to identifying the objective of a task which is currently underway, whereas prediction of future actions relates to anticipating tasks which have not yet begun.

Several studies have taken place into the inference of task goals using eye tracking, to identify objectives in ADLs [25] or to predict the toy a child would like to play with [26]. Work has also been done to understand the goals of drivers when manoeuvring on the

road, with warnings to indicate if the inferred manoeuvre could be unsafe [27]. This part of the definition is concerned with identifying what the goal of the activity will be once the transition phase has been entered. The transition phase allows the system to be aware of a change, whilst the activity goal will indicate what the system needs to do directly after transition.

2.1.4 Example

To more clearly explain the three aspects of intent sensing identified in this definition, an example is provided. A camera can be used to detect a glass of water on a table, which can represent an environmental stimulus for a person. The person can reach for the glass because they would like to drink from it. However, the time of day and length of time since their last drink (contextual stimuli) can be used as well to predict that they may reach for the glass. This can happen before the idea has even occurred to the person. While the confidence of this prediction may initially be low, it is likely to increase over time as the action becomes more imminent. Cues taken from a person's body language and perhaps even brain activity can indicate the person is starting to act on their planning. All these aspects can be considered for intent sensing.

Then, when the person begins to reach, the activity transition from a rest state to a reaching motion may be recognised, for example by an electromyography (EMG) monitoring system on the arm.

The task goal of the arm movement is to reach the glass, which may be inferred through eye-tracking, or by estimating the trajectory of the reach. One might even be able to detect a goal change – for instance, the glass is pushed away instead of picked-up.

2.1.5 Limitations

The proposed definition of intent works well in the context of controlling medical devices, but could be further optimised for broader use beyond the scope of the current chapter. The literature offers conflicting use of the word "intent", and ambiguities may arise in certain applications as to whether or not they should be considered "true" intent sensing. Examples of this include wheelchair navigation through doorways [28] or the prediction of web links that a computer user may intend to click on [29], which by this chapter's definition would be considered intent sensing, but might not be currently referenced as such.

2.1.6 Objective

This review will focus on studies related to intent (as previously described) in some way. As such, it should be noted that the review will not include wider contextual tools which may have potential applications for intent detection, such as the recognition of obstacles or the learning of daily routines. Literature will only be included if attempts were made within those studies to apply measurements and observations directly related to intent sensing. In addition, this review accepts articles that utilise sensor information in their predictions of intent. Purely theoretical papers, or papers which exclusively use self-reported intent, will not be included.

The overall objective of the review will be to assemble all existing literature on intent sensing as previously defined, assess its quality and categorise it according to which sensors it uses. The literature will also be clustered according to the components of the intent sensing definition it addresses, and the general distribution of types of intent measured will be determined. The results identified will then be used to assess each sensor type's suitability for intent sensing and highlight areas in the literature where more research is needed. Mann-Kendall tests (a well-established statistical test for trend analysis [30]) will also be employed to quantify trends over time in numbers of papers related to each sensor type and intent sensing aspect to indicate areas of particular growth.

2.2 Methods

2.2.1 Search Strategy

This chapter will perform a systematic review of existing methods of intent sensing, following the guidelines and principles of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [31]. The search strategy consisted of the keywords: (Accuracy OR Sensitivity OR Specificity) AND (((Predict OR Prediction) AND (Activity OR Behaviour)) OR Intent) AND (Sensors OR Sensor network OR BSN OR Wearable Sensors). This was used to search three databases: IEEE Xplore, Pubmed and Web of Science, returning results that included these search terms in their keywords, title or abstract. The search covered all published papers up to and including January 2020.

2.2.2 Study Selection

All results were first screened by reviewing the abstract and eliminating those which did not satisfy conditions for relevance according to the definition of intent sensing discussed in Section 2.1. As this definition is quite nuanced, a simple selection of keywords alone was not sufficient to fully determine each paper's relevance. A second reader also reviewed the results, and in instances where the readers' opinions on relevance conflicted, a discussion took place until both readers were in agreement. If no agreement could be reached, a third reviewer was required to adjudicate.

The inclusion criteria were:

1. Must be written in English.
2. Must be peer reviewed.
3. Must refer (either implicitly or explicitly) to at least one of the three aspects of intent sensing identified in this chapter's definition – predicting future actions, activity transition recognition, and goal inference.
4. Must not refer only to activity recognition without identification of transitions between states. However, the inclusion of activity recognition alongside valid components of intent sensing is acceptable.
5. Must contain some measure of accuracy or reliability, i.e. not simply a description of a theoretical method.

Each study had the following data recorded: Title, First Author, Year, Number of Participants, Participant Details, Aim, Outcomes, Sensors Used and Intent Aspects Identified.

2.2.3 Quality Assessment

The quality of the papers was then reviewed using an eight-item checklist. The criteria used were based on those in [32], which were themselves based on the Joanna Briggs Institute critical appraisal checklist for cross-sectional research [33], "adapted to be applicable for Machine Learning applications". The final modified checklist is shown in Table 2.1, in the Results section. For some papers, not all eight criteria were relevant, so they were discounted whenever they were not applicable. For each article, the number of relevant checklist criteria satisfied was counted, and divided by the total number of relevant criteria to give a percentage score. This percentage score was considered to represent the paper's quality, and the papers were then ranked in order of these scores.

The quality assessment was then confirmed by a second reader, and again, any disagreements were discussed until both readers agreed or a third adjudicator was used.

2.2.4 Categorisation

The papers were then categorised according to the sensors used, the intent components identified, and the type of action intent measured, with the numbers of articles falling into each category counted. The accuracies reported by papers in each of the categories

were recorded. It should, however, be noted that the accuracy metrics used varied from paper to paper and are not always directly comparable.

For example, [34] used EMG and kinematic sensors to classify between seven different activity transitions with 95% accuracy, whereas [35] used EEG to distinguish between a reach motion and rest with 72% accuracy. These two studies differ in complexity, and so the accuracy comparison should be used as a general guide only.

2.2.5 Trend Analysis

The numbers of papers published each year involving i) each aspect of intent sensing and ii) each sensor type were recorded, and the general trend analysed using a Mann-Kendall test, assessing paper count per year between 2005 and 2019. The resulting values were tabulated for comparison.

2.3 Results

A total of 2808 papers were screened across the three databases, with 497 duplicates identified and removed. This left 2311 papers for abstract-reviewing, which subsequently reduced the number to 155. None of the papers were excluded based on their main body contents, indicating the keywords chosen were effective. Figure 2.2 shows a flowchart of the reviewing process. Qualitative synthesis was then performed on these remaining papers.

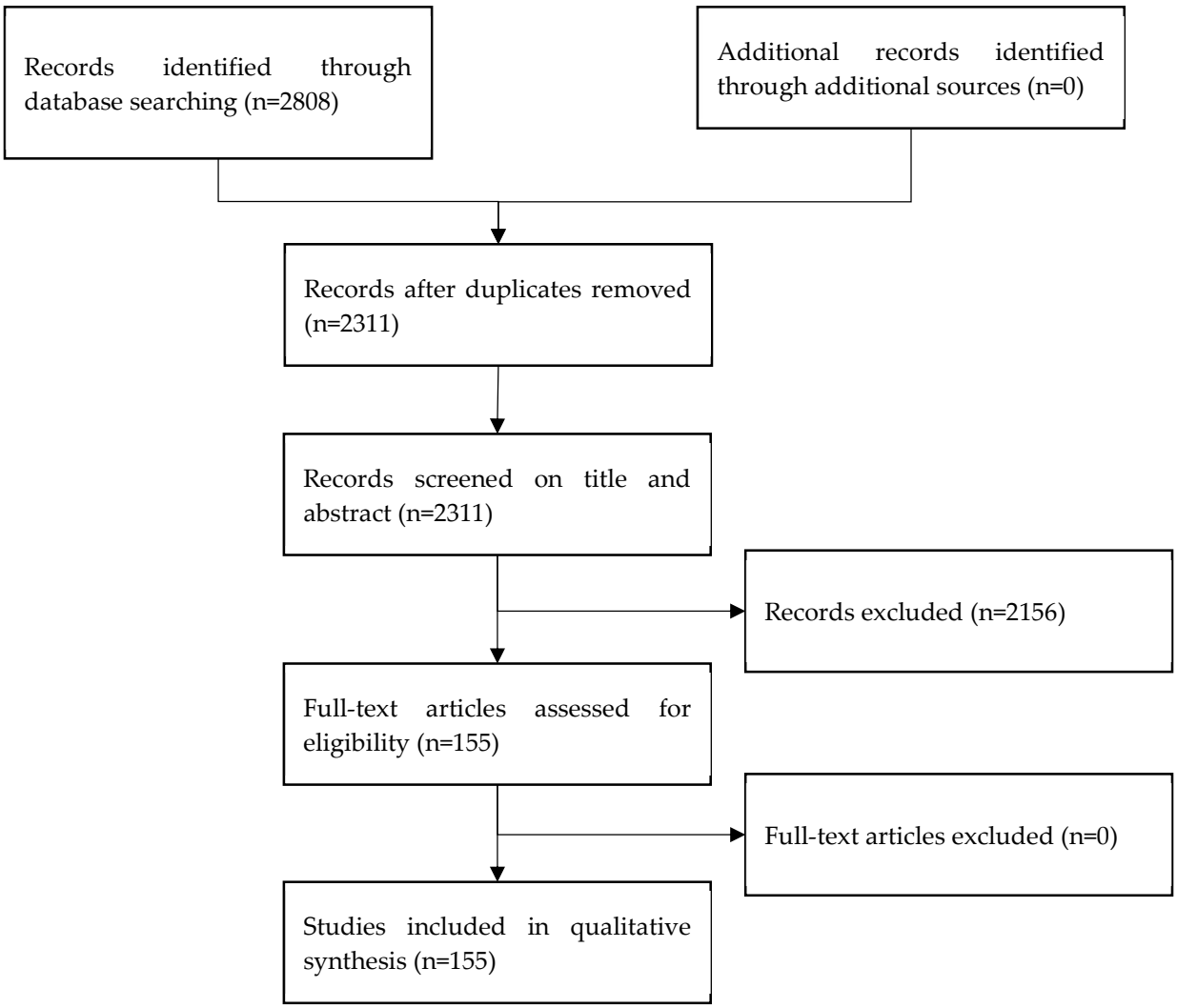

Figure 2.2. *Flowchart of the inclusion/exclusion decision-making process used to determine which papers should be used in the study.*

2.3.1 Quality Assessment

The quality ratings of the papers varied, with a median of 81%, and an interquartile range of 63%-100%. 44 papers had a perfect 100% score, and no papers scored 0% - the lowest quality score recorded was 31%. All papers reviewed had at least some description of aim/objective, though this varied in clarity, satisfying or partially satisfying item 1. Items 2 and 5 were not applicable to papers which did not gather data (as they used

a pre-existing dataset), so these were only evaluated for 104 of the papers. The percentages of papers satisfying each part of the intent definition are shown in Table 2.1.

Item	Description	Valid for (# of papers)	Satisfied by (% of papers)
1	A clear objective and description of inclusion criteria of the study population (i.e. selection of subject group, or time series data) or statement of aim prior to data collection?	155	90%
2	A detailed description of the study population (i.e. how are the subjects recruited?)	103	51%
3	A clear description of the data source and how data was collected (i.e. clearly describe method of measurement.)	155	97%
4	A valid and reproducible data collection and measurement method (i.e. was the data measured in a reliable way?)	155	63%
5	Attainment of ethical approval. Were ethical issues (e.g. subject confidentiality) considered?	100	37%
6	Were findings and implications discussed in detail?	155	74%
7	Were the outcomes measured in a valid and reliable way?	155	97%
8	Was an appropriate cross-validation and evaluation method used?	155	90%

Table 2.1. *Quality assessment checklist. The number (#) of papers for which a certain item was valid is given, as well as the percentage of valid papers that satisfied the criteria.*

2.3.2 Categorisation

Dividing the papers into categories based on the earlier definition revealed that 88% discussed the activity transition recognition aspect of intent, 6% discussed goal inference, and 17% discussed prediction of future actions. There was little difference in quality across the three categories, with the median quality remaining at ~75% for all three.

Separating the papers reviewed by the sensing method revealed that 15 different sensing methods were used to detect intent. These were: computer vision, Electroencephalography (EEG), Electromyography (EMG), Mechanomygraphy (MMG), Force Myography (FMG), angular rotation measurement, motion tracking, gaze tracking, accelerometers, smart phone sensors, smart watch sensors, smart home sensors, force/pressure measurements, self-reporting and Electrocardiography (ECG). The distribution of these is displayed in Figure 2.3, and a comparison of the accuracies in the studies for each sensor type is shown in Figure 2.4.

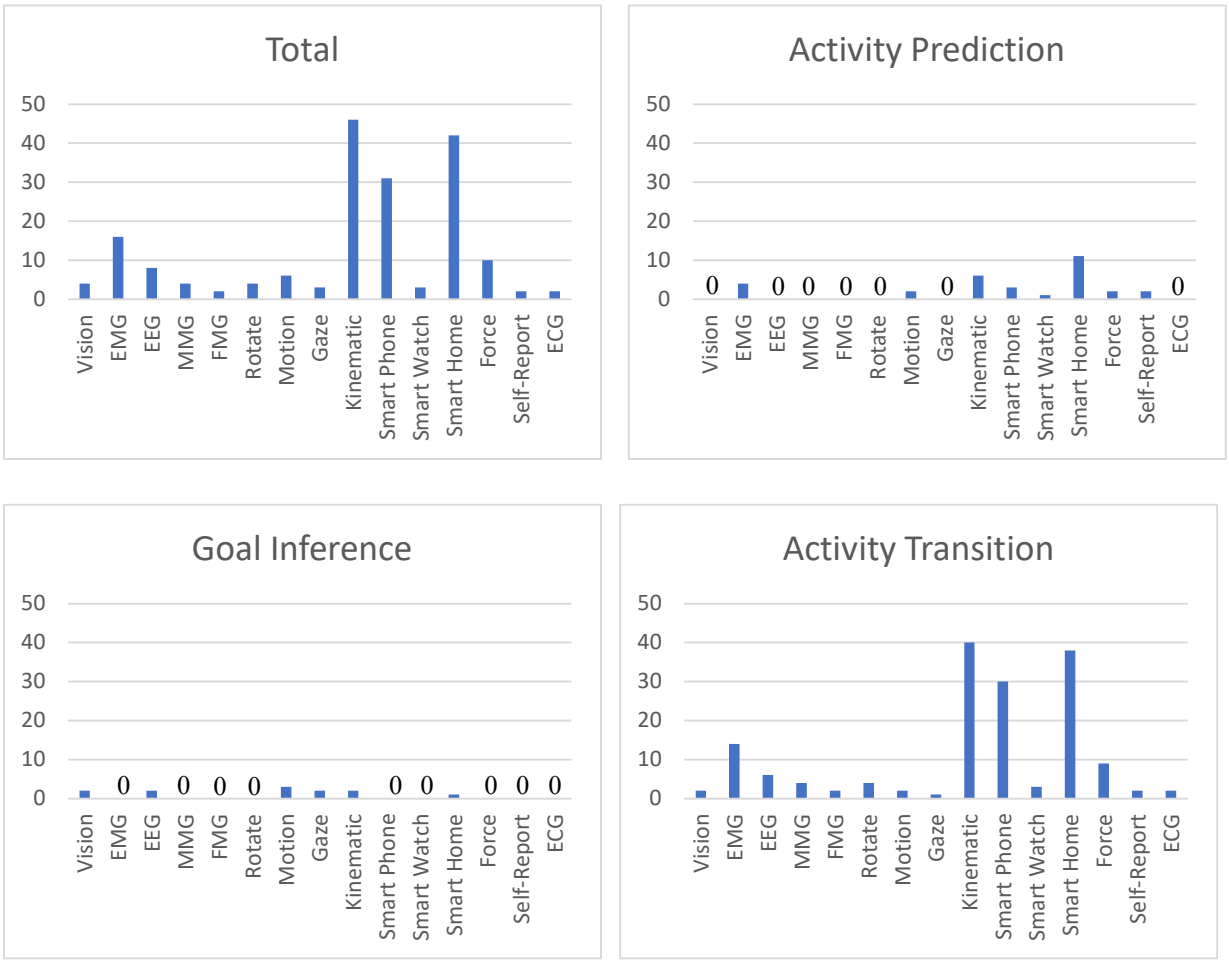

Figure 2.3. *Breakdown of the number of total papers (vertical axis) according to sensor type (horizontal axis), as well as the number satisfying each component of the intent definition. Multiple sensors can be*

used in the same paper. Vision relates to computer vision; Electromyography (EMG); Electroencephalography (EEG); Mechanomygraphy (MMG); Force Myography (FMG); angular rotation measurement (Rotate); motion tracking (Motion); gaze tracking (Gaze); kinematic sensors (Kinematic); Smart phone sensors (Phone); smart watch sensors (Watch); smart home sensors (Home); force/pressure measurements (Force); Electrocardiography (ECG) and self-reporting (Self). 0s are used to highlight sensors which were not used in any papers for a particular intent sensing aspect.

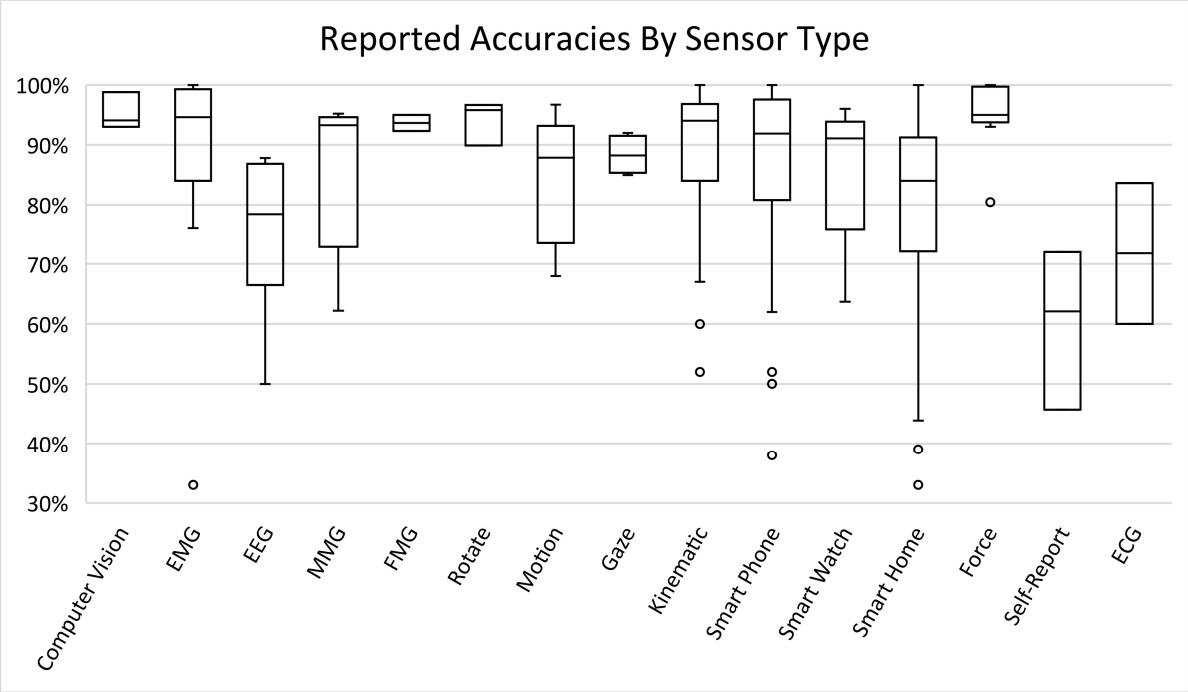

Figure 2.4. *A box-and-whisker plot to show the range of accuracies reported in the results across the 155 papers analysed, with accuracy (%) on the vertical axis and sensor type on the horizontal. The boxes show the upper and lower quartiles, the middle line shows the median, the whiskers show the range, and the points indicate outliers. A point is considered an outlier if it is larger than 1.5 times the IQR of the third quartile or 1.5 times smaller than the IQR of the first quartile. Care should be taken when interpreting this figure, as conditions and methods vary between papers and are not "like-for-like".*

In total 24 (15%) of the papers reviewed used more than one type of sensor, and only one paper used more than two (three sensor types were used). It should be noted that this occurrence of multiple sensors in some papers will mean the values displayed in Figure 2.3 add up to greater than the total number of papers.

Dividing the papers into categories to show the type of action intent used revealed twelve broad groups: locomotion, sitting/standing, general ADLs at home and outside the home, reach/grasp tasks, hand gestures, estimating the user's trajectory, predicting posture intent, load-carrying tasks, and tasks related to exercise, driving and computer use. The distribution of these is shown in Figure 2.5.

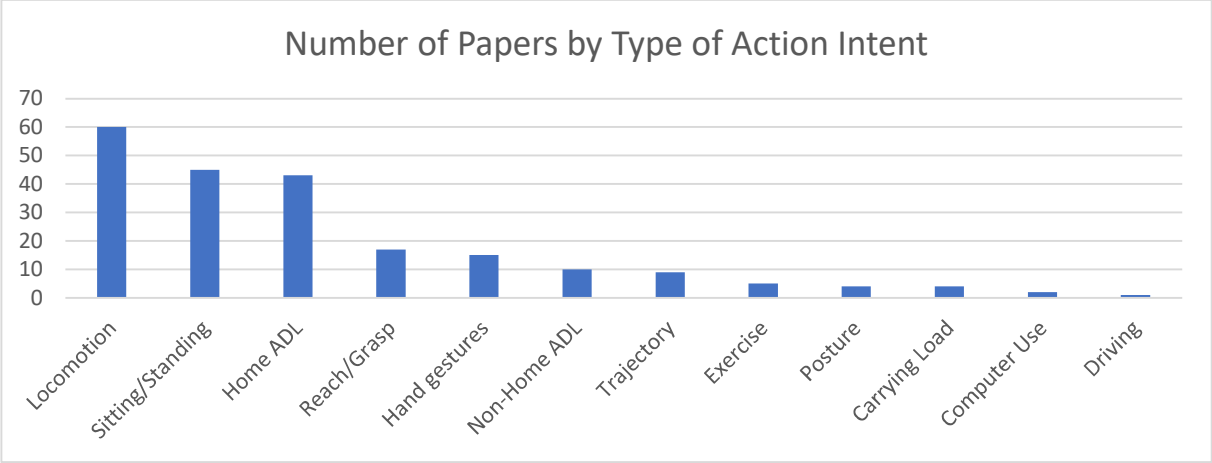

Figure 2.5. *The distribution of papers, grouped into categories by type of action intent studied. The number of papers is displayed on the vertical axis. Note that some papers included multiple types of action intent, so the sum of the columns is greater than the total number of papers.*

Across the three aspects of intent sensing discussed, the accuracies reported by those papers concerned with activity transition recognition ranged from 33-100%, for goal inference, 73-98.8%, and for prediction, 40-100%.

2.3.3 Trend Analysis

The cumulative number of papers published per year involving each intent sensing aspect is shown in Figure 2.6.

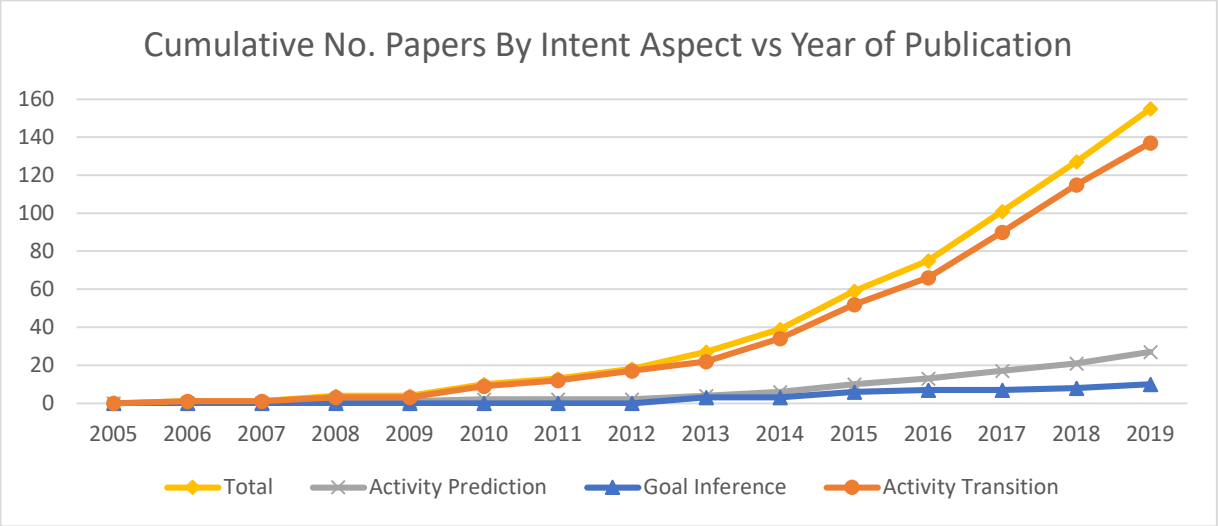

Figure 2.6. *The cumulative number of papers utilising each aspect of intent sensing vs the year of publication. The vertical axis shows the cumulative number of papers and years are displayed on the horizontal axis.*

The results from the Mann-Kendall tests performed on the number of papers published per year according to the intent sensing aspect and sensor type categories are shown in Table 2.2.

Category	N (no. papers)	z-value	p-value
Total	155	4.33557	0.00001**
Intent Aspect			
Prediction	27	3.72137	0.00020**
Goal	10	2.20593	0.02739*
Transition	137	4.08640	0.00004**
Sensor Type			
Environment Vision	4	2.36520	0.01802*
EMG	16	2.95832	0.00309**
EEG	7	1.59631	0.11042
MMG	4	1.43346	0.15173
FMG	2	0.78840	0.43046
Rotate	4	0.96357	0.33526
Motion	6	2.58378	0.00977**
Gaze	3	0.42894	0.66797
Kinematic	45	4.02611	0.00006**
Smart Phone	30	3.62825	0.00029**
Smart Watch	3	2.03810	0.04154*
Smart Home	42	2.98249	0.00286**
Force	10	1.74850	0.08038
Self-Report	2	0.78840	0.43046
ECG	2	1.50513	0.13229

Table 2.2. *Results from Mann-Kendall tests performed on the number papers reviewed by each year of publication, categorised by intent aspect and sensor type, to indicate trends in interest over time. Note that some papers included multiple sensor types, so N total is less than the sum of the sensor types. * indicates p-values <0.05. ** p-values <0.01.*

2.4 Discussion

2.4.1 Screening

The results of the literature review revealed much about the current state of the field of intent sensing. Many of the papers discarded during screening refer to intent, but do

not align with the definition given in this review. Instead, they refer to the use of a control interface – e.g. the use of particular pre-programmed gestures – in order to operate a device.

This does not align with the "passive" requirement of the intent sensing system stipulated in this thesis, and is not considered by this review to be true intent sensing, as it requires the user to perform a particular (additional) action in order to inform the sensing system of their intentions.

2.4.2 Aspects of Intent Sensing

Referring back to the three aspects of intent sensing discussed in the introduction: activity transition classification, goal inference and activity prediction, the existing literature varied in number of publications across the three fields. It should be noted that while a higher minimum accuracy was reported for prediction and goal inference than was reported for activity transition recognition, this is more likely to be simply due to the papers having different assessment criteria. It is also intuitively easier to make more accurate observations about current events than to predict future events.

No studies showed evidence of combining the prediction and goal inference aspects, and by extension none attempted to combine all three aspects of the intent detection definition into one unified system, so a system which does integrate all three would be a strong contribution to the literature.

2.4.3 Activity Transition Classification

The vast majority of the work that has been done to date is on activity transition classification, identifying changes in what the person is currently doing at a particular time. 88% of the papers reviewed focused on this aspect of intent sensing. While identifying a subject's current transition between activities is an important stage of intent identification, it only captures information up to the present moment, and does not consider what the user will do next or why they are doing it, so this process must be combined with other methods to more completely identify user intent. The recognition of gait transitions seems a particularly well-developed field with many high-accuracy methods available, and multiple studies report accuracies in the region of >99% [36,37].

2.4.4 Prediction

The prediction aspect of intent sensing is the second-most explored of the three, with 17% of papers referring to this. Results were mixed in this area, and the size of the time window selected seemed to have a strong impact on the performance of the system – predictions that are only a second in advance of a particular event understandably tended to have a higher accuracy than those predictions performed five seconds in advance. This was demonstrated in particular by [38], with prediction accuracy dropping from >80% to >60% when the time window was doubled. Alongside these studies, a number of new models, including Hidden Markov machine learning [39] and graph-based [40] models to learn typical patterns of user behaviour, have been developed for use in predictive classification, which have shown promising results.

2.4.5 Goal Inference

The goal-inference aspect of intent sensing was explored very little, with only 6% of papers referring to it, indicating there is room for more research and exploration. In a similar way to the prediction component, it should be noted that accuracy of the goal inference increases as the action being studied progresses in time – in general, it is harder to predict an action's goal moments after its beginning than it is when it is nearly complete. This is demonstrated in [41], a study inferring the goal of a driving manoeuvre to evaluate it for safety, where accuracy of the inference increases from 80% at 0.5 seconds after activity inception to 90% at 1 second after. It is also shown that a lower classification accuracy is found when more possible goals are available [42,43]. Notably, while some studies have attempted goal inference in non-medical settings, no system has yet been developed to incorporate task goal inference into an intent recognition system for a medical device, indicating there is significant scope for development in this field.

2.4.6 Comparison Over Time

Figure 2.6 shows the cumulative number of papers related to each aspect of intent sensing plotted against the year of publication. Since the first intent sensing paper published in 2005, papers in the activity transition category have been by far the most numerous, followed by activity prediction and goal inference. This order has been maintained, with the gap widening over time. The total number of new intent sensing papers every year has generally increased, with the cumulative graph getting steeper and the Mann-

Kendall test showing a positive trend, indicating a growth in interest in the field of intent sensing in more recent years. Trend analysis suggests the fields of all three aspects of intent are growing year on year, but goal inference is growing more slowly. As this is an essential component of a complete intent sensing system, accelerating research in this relatively undeveloped field would be of benefit to the development of intent sensing as a whole – more so than activity transition classification, which already appears to be developing rapidly.

2.4.7 Sensing Methods

2.4.7.1 Computer Vision

Computer vision is the automated interpretation of data from visual sensors, e.g. cameras. Here, it is used to refer to the observation of the environment around the user – the use of visual sensors to track the motion of the user has been categorised as Optical Motion Tracking. Environmental computer vision has been utilised in the areas of classification and goal inference [44]. Work on it has been relatively sparse, with only four papers reviewed incorporating it. It has been shown to have potential in goal inference, identifying whether a user's goal was to climb stairs with a ~99% accuracy [45] and planned interaction with objects yielding a ~93% accuracy [43], but further studies are needed to apply it to more general classification of ADLs, rather than to individual, specific situations. Depth sensing was also able to perform activity transition classification between locomotion modes such as walking, running and stair ascent with an

accuracy of ~94% (quality score 75%) [46], but computer vision is rarely used for this purpose.

2.4.7.2 Electromyography (EMG)

EMG is the process of recording and evaluating the electrical signals produced by muscles during activity, driven by the motor neurons [47]. EMG's use for intent is growing rapidly according to trend analysis, and is used in sixteen papers in total, dividing into twelve papers to detect activity transitions, and two papers with some predictive element (and two papers using both). However, no papers have used it for goal inference. Results for activity transition classification in the papers reviewed showed consistently high-accuracy (>90%), and while results in the four papers considering prediction were mixed, of particular note was the use in a high quality paper (score 88%) of EMG to predict sit-to-stand transitions ~130ms before the subject leaves their seat with an accuracy of ~99.5% [48], supporting EMG's potential in predictive control.

2.4.7.3 Electroencephalography (EEG)

EEG is the monitoring of electrical signals from the scalp, which are analysed to extract information about the underlying brainwave signals [42]. EEG was used in seven papers, primarily to perform activity transition classification, where both results and quality of papers were mixed. It was not used in prediction, but in two papers goal inference was attempted, with accuracy varying between 40% and 80% in one – a brain computer interface [42]. In the second goal-focused paper, the anticipated force of an in-progress grasp was predicted, but the best median correlation coefficient between prediction and

reality was only 0.51 [49]. Across the seven papers, the highest accuracy reported was ~86% [50] – high, but not high enough for reliable control of a medical device. In general, the papers show positive results, suggesting EEG has already gained traction as a possible intent sensing method. Nonetheless, the variance and limitations in the reported accuracy in the papers reviewed do not support this as a practical intent sensing method for medical devices at the moment.

2.4.7.4 Mechanomyography (MMG)

MMG is the measure of physical vibrations within the muscles during activity [51]. MMG was used in four papers to recognise activity transitions, with accuracy results ranging between ~60 and ~95%. No papers used it for prediction or goal inference. This low number of papers and exclusive focus on only one aspect suggests limitations in the existing work that has been done with MMG for intent sensing and that other sensing types may be more readily applicable to identify intent, based on the current available evidence.

2.4.7.5 Force Myography (FMG)

FMG is the use of force sensors to capture the expansion and contraction of a surface muscle [52]. FMG was used in only two papers, again to recognise activity transitions, but not to predict or to infer goals. One was able to determine postures during drinking with an accuracy of 92% [53], and the other classified positions of a robotic glove with an accuracy of 95% [54]. These are promising results for activity transition classification, but again the low number of papers and exclusive focus on activity transition suggests

more studies are needed to determine the extent of FMG's applicability to a full intent sensing system.

2.4.7.6 Angular Rotation

The angular rotation of joints was used as a sensing input for four of the papers, all aiming to perform recognition of transitions between activities. Results were positive, with one good quality paper (score 75%) classifying motion patterns with an average accuracy of ~96% [55], but again the work was limited entirely to activity transition and no evidence is presented in the papers to suggest potential application in goal inference or prediction.

2.4.7.7 Optical Motion Tracking

Optical motion tracking – the use of visual sensors to detect and analyse human movement, often through the tracking of markers [56] – featured in six papers, covering all three aspects of intent sensing and achieving >93% accuracy in all areas. This successful, high-accuracy coverage of all aspects of intent indicates motion tracking has strong potential for intent sensing, and more research should be undertaken on this topic - especially, exploring its utility in more complex and realistic environments, as the required cameras are unlikely to be available outside of controlled spaces.

2.4.7.8 Gaze

Gaze tracking monitors the user's eye movement to determine their region of focus and intention [42]. Human gaze analysis was employed in three papers, once to identify

activity transition and twice to infer the goal of the current activity. Trend analysis does not support significant growth in the field. It was used successfully in identifying current computer tasks to successfully speed up preparation of materials, but accuracy results for this are unclear and the quality score is low at 43% [29]. The remaining two papers performed goal inference with accuracy >80%, high but not at an acceptable level for reliable control. This suggests there is evidence for gaze as a possible input for goal inference, but no strong evidence for use in prediction or activity transition.

2.4.7.9 Kinematic Sensors

Kinematic sensors detect and measure motion, and include accelerometers and Inertial Measurement Units (IMUs). This was the most common type of sensing input and is growing the most rapidly according to trend analysis, appearing in 45 papers across all three aspects of intent sensing, with 37 papers using them for activity transition classification, two for goal inference and three for prediction. A total of three papers used them for both activity transition classification and prediction. Naturally, across so many papers, accuracy results covered a wide range depending on the scenario investigated. Nonetheless, many reported a performance of >90%. The versatility of kinematic sensors (across all three aspects) and the breadth of the already existing research on them makes them a strong candidate for use in an intent sensing system.

2.4.7.10 Smart Phone

Modern smart phones include a variety of sensors, including accelerometers, location tracking and microphones, and are the second-most rapidly growing field for intent

according to trend analysis. Smart phones were employed in 30 studies, with 27 of them focusing solely on activity transition classification and three incorporating both activity transition classification and prediction. Results varied, with some papers reporting accuracies >95% and some as low as 60% or even ~40% depending on the situation. While this suggests that they may be useful for activity transition sensing with only a small amount of evidence towards application in prediction, the ubiquity of the smart phone means it is highly likely to be available to users at any given time. Therefore it may provide additional (if limited) sensing information to an intent sensing system at little-to-no additional cost.

2.4.7.11 Smart Watch

Much like smart phones, smart watches include a variety of sensors, but are only worn on the wrist. They were used in only three of the papers reviewed, with two of these for activity transition classification and one with both classification and prediction (though the accuracy results for the prediction are unclear). The placement on the wrist provides superior identification of arm movements, with a paper classifying physical activity transitions with accuracy ~96% [57]. However, more research is needed before clear statements can be made about the potential of smart watches in general for intent sensing.

2.4.7.12 Smart Home

Smart homes are a developing technology involving the placement of sensors around the user's living space, with applications in activity transition recognition, energy

monitoring and healthcare [58]. Trend analysis suggests smart homes are a rapidly growing field for intent, and a total of 42 papers used smart homes/environments across all three aspects of intent sensing (though only one paper focused on goal inference). Once again, the large number of papers resulted in a wider range of accuracies for activity transition classification, but many of these were in the >90% accuracy range. For prediction, accuracies ranged between 60% and 90%, and for the goal inference paper, the aim was identified by mining location trajectories as the user moved around the environment, with accuracy between 84% and 96% [59] (though the quality score for this paper was only 43%). The breadth of research regarding smart homes and their ability to cover all three aspects of intent sensing makes them good candidates for use in a unified intent system.

2.4.7.13 Force/Pressure

Ten papers used strategically placed force and/or pressure sensors, some positioned in shoes and some included in tactile arm braces. All of these focused on activity transition recognition, though two also incorporated prediction – although the accuracy of the prediction component in these cases is unclear. Accuracy of classification using these sensors is consistently high, achieving >90% in all but one of the papers, which achieved only 80% (quality score 71%) [60]. The use of force sensors shows high potential in activity transition classification – particularly if they can be built into a wearable device – but there is not clear evidence in support of them for the other intent sensing aspects.

2.4.7.14 Self-Reporting

Self-reporting of activities, in which the user manually records what activities they perform throughout the day, was used in two studies to build up a data set of the user's daily routine and provide context for current activity identification, as well as predictions of the future. While results were highly accurate (>95%), self-reporting of activity is not a viable option for automatically identifying user intent without the need for manual intervention. Self-reporting was only included for articles that already reported on a sensing technology, thus only a limited set of the self-reporting literature is contained in these results.

2.4.7.15 Electrocardiogram (ECG)

ECG is used to monitor the rhythm, rate and electrical activity of the heart [61]. In two studies, ECG was used to help identify physical activity, determining whether a person is currently exercising or not, which is not considered by this chapter to be intent sensing. While this alone would not satisfy conditions for inclusion in this review, in both studies, ECGs were also used in combination with accelerometers which were being applied in an intent sensing context. As a result, however, there is little information on the intent sensing performance of ECG sensors by themselves. However, the anticipatory response in heart-rate shortly prior to physical activity is a well-documented phenomenon [62] and this could be a promising sensor modality to research further in future dedicated intent sensing studies.

2.4.7.16 Multi-Modal Sensing

Many studies used a single sensing method to identify user intent under controlled conditions - however, combinations of two or more independent sensing methods should be able to provide higher accuracy (this will be discussed further in Chapter 3). Relatively few studies attempted to do this, with only 15% of studies using more than one sensing method and only one with three methods, again indicating there is room in the research for exploring multi-modal intent sensing, and a multi-modal network featuring several of these methods would be unique in the literature.

2.4.8 Comparison

The reported accuracies of the fifteen sensing methods were compared in Figure 2.4. Care should be taken interpreting this figure, as the methods varied greatly from paper to paper and accuracy metrics used are not "like-for-like". However, the graph does indicate the general trend of the results in the literature. Excluding outliers, the greatest range was observed in papers using smart home sensors, but this was also one of the most widely-used sensing methods, occurring in forty two papers. The smallest range was found for FMG, but as this was only used in two papers, this is not necessarily unexpected.

By comparing median, upper and lower quartiles, papers utilising self-reporting and ECG showed the lowest accuracies, whereas computer vision, EMG, rotation and force sensing showed the highest.

Table 2.2 also indicates the trend in no. papers published using each sensing modality. It suggests the most rapidly growing sensing fields for intent are EMG, optical motion tracking, kinematic, smart home and smart phone sensing.

2.4.9 Types of Action Intent

The types of action intent measured in the papers reviewed fell broadly into twelve categories, shown in Figure 2.5. By far the most common intent measured was locomotion – transitions between walking, running, ascent and descent. Sit/stand transitions were often included with locomotion, but also separately in many cases. The next most popular category was ADLs at home, which refers to the classification of a wide range of complex task transitions, such as cooking, cleaning, watching television, etc. This featured commonly in smart home systems. Upper-limb intent sensing was less common but often used with reach/grasp tasks and gesture recognition. Other categories were rarer, including ADLs outside the home (e.g. commuting, or at work), prediction of subject movement trajectories, posture intent, and specific activities related to exercise, carrying loads, using computers and driving.

These distinct categories demonstrate the breadth of actions that contribute towards intent. The relative focus on the top three most popular categories suggests there is scope for more research into more general intent sensing systems covering a wider range of possible actions in detail.

2.5 Conclusion

In conclusion, it has been shown that much preliminary work has been done on certain components of intent sensing, such as the classification of transitions between activity states. However, comparisons between research outcomes remain difficult due to a lack of a well-established definition. A common framework for the definition of intent sensing can further support this growing field. The definition provided in the current chapter can be built upon by other researchers.

Fifteen different sensing methods were used in the papers identified, for a range of contexts and applications that broadly fell into twelve categories. Of these sensing methods, only motion tracking, accelerometers and smart-home sensors were shown to be studied across all three aspects of intent sensing. Currently, these fields show the greatest potential for intent sensing covering the full definition of intent. The accuracies reported were mixed and depended on the research aim of each paper reviewed, so care had to be taken to avoid drawing misleading conclusions by contrasting incomparable accuracy assessments. More standardisation in testing (perhaps involving pre-determined sets of actions) would allow for more insightful comparisons. Nonetheless, many of the papers reported high levels of accuracy, which gives confidence in intent sensing as a promising, practically applicable field, and observations were made that prediction accuracy increased firstly as the length of anticipatory time decreased, and secondly as the number of possible options decreased.

This review has focused on a very specific scope, covering only papers with direct relevance to intent sensing as previously defined. Further literature review studies addressing related fields and with wider, more general scope can be useful to place intent sensing within the broader concept of human behaviour.

To further the field of intent sensing, it is suggested that future work could attempt to combine data from multiple sensing systems or scenarios to further increase the ecological validity of these systems. Particular attention should be paid to the goal-inference and prediction aspects of intent sensing, as these have been explored relatively little despite forming an important part of the concept of intent sensing.

This would allow a medical device to really work together with the patient across a range of activities and offer a new way of bringing value to the user. True device symbiosis creates unique challenges for the research community, but it will also offer a more holistic development pathway for medical devices. It could enhance device responsiveness and ease of use, whilst enabling more precise, advanced control without the need for additional control interventions. These improvements could reduce device abandonment and increase quality of life for users, with a potential for positive long-term impact across the (medical) device industry and beyond. To this end, the following chapters of this thesis will combine data from multiple sensing modalities suggested in this study, and consider a holistic concept of intent covering all three aspects, rather than just activity transition recognition.

The proposed definition introduced in this chapter is the first endeavour to provide a more robust framework for the development of intent sensing technology. The definition does not aim to be suitable outside of the biomedical engineering field, but it could be further specified to increase the utility in other domains. Nonetheless, this proposed definition will hopefully spur further discussion in terms of creating a robust framework for intent sensing assessment, which could accelerate progress towards true intent sensing becoming widely used in the not-too-distant future.

Statement of Authorship for joint/multi-authored papers for PGR thesis

Title of Paper	A Systematic Literature Review of Intent Sensing for Control of Medical Devices
Publication Status	■ Published □ Accepted for Publication □ Submitted for Publication □ Unpublished and unsubmitted work Written in a manuscript style
Publication Details	J. Russell and J. Bergmann, "A Systematic Literature Review of Intent Sensing for Control of Medical Devices," in IEEE Transactions on Medical Robotics and Bionics, vol. 4, no. 1, pp. 118-129, Feb. 2022, doi: 10.1109/TMRB.2021.3135704.

Student Confirmation

Student Name	Joseph Russell		
Contribution to the Paper	Performed all investigation, analysis and writing. J. Bergmann provided supervisory guidance, editing of the manuscript, and acted as a second reader in the literature review.		
Signature		Date	06/08/2023

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Prof Jeroen Bergmann		
Supervisor comments		
Signature 	Date	06/08/2023

3

Probabilistic Sensor Design for Healthcare Technology

Published and presented by Joseph Russell and Jeroen Bergmann at *2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, New York, NY, USA, 2019, pp. 0585-0590, doi: 10.1109/UEMCON47517.2019.8993086.

Chapter Abstract

Many different sensing modalities were identified in Chapter 2 that can be used to identify intent. An effective system should use a combination of a number of sensors working together. The problem of intent detection can be modelled as a Probabilistic Sensor Network (PSN). PSNs could also have wider applications in healthcare technology – for example, failure to accurately detect a medical event, such as electromyographic (EMG) activity or the onset of a heart attack, could have serious consequences for the user. The

application of sensor networks to optimize event detection is therefore a key area of biomedical engineering research. The optimum number of sensors for event detection can be difficult and time-consuming to determine experimentally. A novel probabilistic model could be used to estimate the appropriate number of sensors required and is described in this chapter. A simple statistically independent model (naïve approach) is introduced, which may be useful but relies on a set of assumptions that do not reflect most real-world applications. From there, a more practical, revised model that can be more easily applied to real-life systems was developed. Both models were subsequently tested against experimental data from a real EMG system. While the naïve approach was practically unachievable, the model was found to be mathematically sound. Predictions from the models were compared and the revised model was found to be more accurate than the naïve model. The revised model was verified against a majority voting system and it successfully predicted sensor network performance with only a small margin of error. This work presents a probabilistic sensor network approach to virtually exploring possible networks and informing optimum design of sensor networks used in healthcare and industry, and the principles developed will form a vital part of this thesis' intent sensing system.

3.1 Introduction

Chapter 2's literature review identified many possible sensing modalities to measure intent. Each had its own advantages and disadvantages, and each provided distinct types of data. A holistic intent sensing system covering all three of the intent sensing aspects identified in Chapter 2 will need to combine multiple sensing modalities in order to be effective. This chapter will explore the theory behind one method of achieving this - the probabilistic sensor network (PSN).

Research into body area networks has become increasingly important in recent years, with rising medical costs and an ageing world population motivating the development of cheaper, more efficient technologies to assist in the care of patients with a wide variety of conditions [63]. One such area is artificial limbs, where control interface technology such as electromyography (EMG) depends on the use of accurate, reliable sensor information from an inherently noisy, unreliable source [64]. EMG is the practice of monitoring the electrical activity in a set of muscles and is widely used in medical applications such as prosthetics. It is particularly applicable as potential input to an intent sensing system for prosthetic devices.

As well as forming a foundation for an intent sensing system, the use of sensor networks to combine results from multiple sensors into a single, more reliable measurement could be of great benefit in EMG and other wearable technology applications [65]. Establishing a framework to model a sensor network's performance, combine its sensor information, and predict the optimum number of sensors would be particularly useful in

sensor network design, enabling virtual prototyping [66] and reducing the need for extensive experimentation. This is a relatively novel field, and there is therefore little prior literature available. The report will explore a "naïve" model that does not take into account false positives and negatives and will propose a "revised" model that does. An error between the experimentally determined values and those of the models can be calculated to capture model performance.

The aim of this chapter is to determine the root mean square error (RMSE), absolute and mean error between the measured and predicted probabilities across different models. The event detected will consist of the activation of the lower-arm muscles (active when closing the hand), an event often used in EMG-based trans-radial prosthetic control. The procedures developed will be kept as broad as possible to allow for generalisation to other event-detecting scenarios.

3.1.1 Probabilistic Sensor Networks

A sensor network is a system of multiple sensors cooperating to sense physical phenomena [67]. Before such a system can be designed, a model can be produced to mathematically analyse the probabilities of detecting the event. The simplest model would be a naïve model that assumes statistical independence [68].

Figure 3.1 shows a diagram of the (naïve) model, demonstrating how the detection event is broken down into different stages. Each stage is assigned its own probability, indicating the chance that a signal will successfully pass through the stage without being lost. A signal must successfully pass through all four stages in order to register as

detected. Figure 3.1 also shows examples of how the network can be modified – by sensing from multiple environments, adding extra sensors, using multiple conditioning methods, or combining decisions from multiple processing methods.

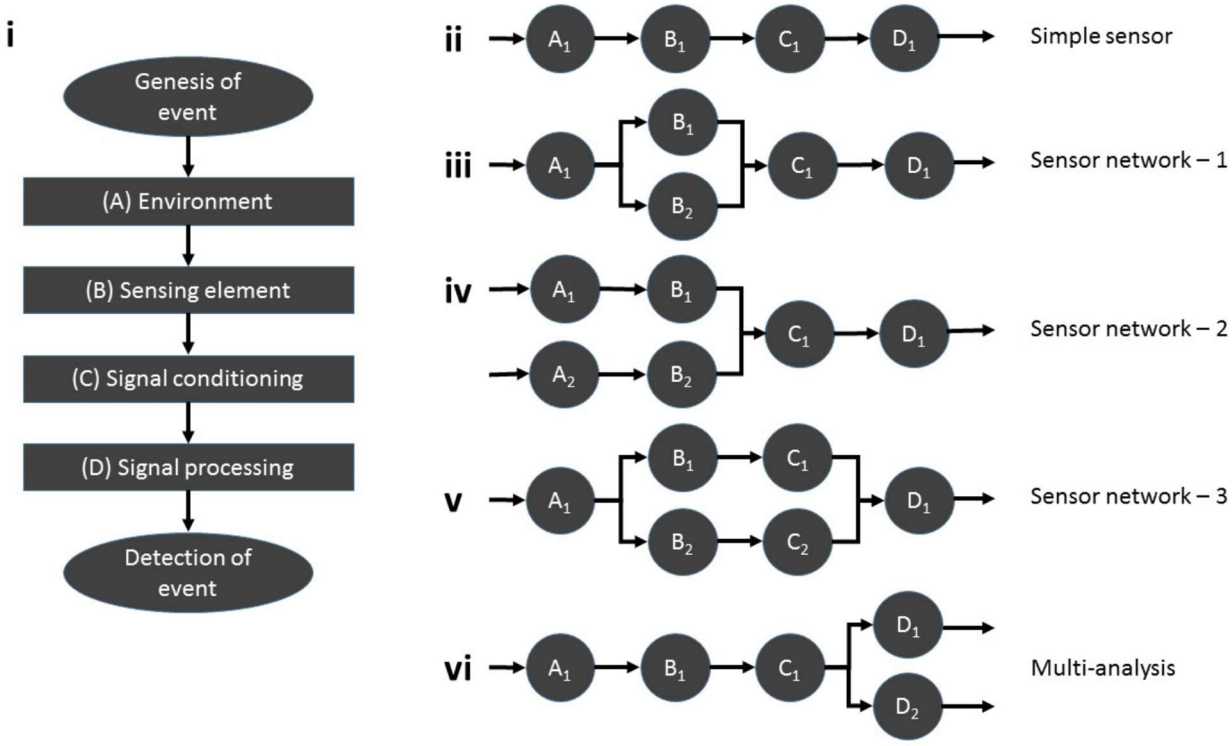

Figure 3.1. *Model for event detection (taken from* [68]*). (i) shows the steps that lead up to the detection of a given event. (ii) shows the simplest design. (iii) displays a network design that consists of two sensors. (iv) is a network that has sensors that sense different sensing environments. (v) shows a network containing two different sensors, which require different signal conditioning. (vi) shows a sensing configuration that is processed in two different ways.*

Assuming independence between stages, the probability of event detection (P) for a design with a single sensor node (Figure 3.1.ii), where $P(A1)$, $P(B1)$, $P(C1)$ and $P(D1)$ are the probabilities associated with each stage, is given by:

$$P(A1 \cap B1 \cap C1 \cap D1) = P(A1) \cdot P(B1) \cdot P(C1) \cdot P(D1) \tag{3.1}$$

The probability will change depending on how the design affects the different stages.

3.1.1.1 Stage (A) – Environment

The first stage of the sensor network detection process is the environment stage. This is the probability that, when an event occurs, a measurable signal identifiable as the event is produced in the region being sensed. The choice of sensing environment is very important in any sensor network.

In EMG, each muscle being sensed might be considered to be its own environment. The placement of the sensors in the environment will influence the probability of correctly detecting an event, such as muscle activation [69]. Furthermore, if these muscle sites are close to one another, there may be the possibility of cross-talk [70].

3.1.1.2 Stage (B) – Sensing Element

The second stage of the process is the sensing element. This is the probability that the sensor used is capable of picking up the signal from its environment. The naïve model assigns a single probability to the proportion of correctly identified events (sensitivity), but does not consider the possibility of false positives (specificity). These are pliable in relation to each other, as it is impossible to determine specificity merely from information about the sensitivity [71].

Typically, in the context of EMG, if the sensor is chosen well, this sensitivity probability is very high (~1). Choosing a sensor that has insufficient precision or accuracy will produce a lower probability.

3.1.1.3 Stage (C) – Signal Conditioning

The third stage of the process is signal conditioning, which can consist of filtering or amplification [72]. It can be described as the probability of an event passing through any conditioning, such as noise filtering, which takes place.

However, this stage should also encompass the probability that not enough filtering has been done. For instance, a system might detect a noisy signal perfectly from the environment in stage A with a perfect sensor in stage B, but if no filtering takes place, stage C might still output an imperfect detection probability. Adding an appropriate filter (e.g. notch filter for single-frequency noise) could completely remove the noise, increasing the Stage C probability to almost 1. As such, a more precise definition of the signal conditioning probability would be the probability that a signal containing the event information, once passed through any conditioning steps, is still measurable and not corrupted by noise.

3.1.1.4 Stage (D) – Signal Processing

This final stage, signal processing, may be defined as the manipulation of random signals to extract information about the underlying mechanisms that generate or transform them [73]. It covers any operations performed using the signal after conditioning, be that comparison to a threshold value, pattern recognition, etc. It is the probability that the software correctly detects that an event has occurred, using the signal that has passed through the other stages. For a sensor network, any fusion of signals can take place in this stage.

3.1.2 Mathematical Modelling

3.1.2.1 Accounting for Dependency

The naïve model assumes the sensor data readings to be independent. This assumption does not hold in most real-world scenarios. For two sensors acting together, with individual probabilities of $P(B1)$ and $P(B2)$, the total probability of the sensor stage, $P(B)$, can be given by:

$$P(B) = P(B1 \cup B2) = P(B1) + P(B2) - P(B1 \cap B2) \tag{3.2}$$

This may be rewritten [74] as:

$$P(B1 \cup B2) = P(B1) + P(B2) - P(B1|B2) \cdot P(B2) \tag{3.3}$$

The sensors may be dependent on one another, due to, for example, crosstalk. Commonly, this will be a positive dependency, i.e. when one sensor is high, the other is also likely to be higher. As such, it can be expected that:

$$P(B1|B2) \geq P(B1) \tag{3.4}$$

Therefore, the combined sensor probability predicted by the naïve model will be an overestimate and provides an upper bound for the true value – the magnitude of the difference will depend on the degree of dependency.

If the sensors are totally dependent on one another – i.e. B1 is only 1 if B2 is also 1 - adding sensors will not increase the combined sensor network probability.

3.1.2.2 Accounting for the Confusion Matrix

A confusion matrix is a mathematical construct used for evaluating a classification model [75]. An example of a two-class confusion matrix is shown in Figure 3.2.

		Actual Values	
		Positive	Negative
Predicted Values	Positive	True Positive	False Positive
	Negative	False Negative	Negative

Figure 3.2. *An example of a binary, two-class confusion matrix.*

The naïve model considers only the top-left quadrant of the confusion matrix , assuming that if any of the sensors in the network detects an event, then it is a true positive.

In an EMG system, however, false positives are just as much of a consideration as true positives. The noisy nature of EMG signals means that, in a threshold-based classifier, in order to achieve required sensitivity levels, the specificity must consequentially suffer. The choice of activation threshold is a trade-off between these two factors, which are equally important.

3.1.2.3 Combining Sensor Readings

One established method of combining sensor readings is majority voting, which simply counts how many sensors support each possible outcome and selects the result

supported by the most sensors. This technique is readily applicable to a range of scenarios [76], including EMG, and so a majority voting system will now be discussed in detail.

The majority voting system used for this application will have two possible outcomes: either the event is occurring, or it is not. Therefore, at least half of the sensors must agree on the outcome. If equal numbers of sensors detect and do not detect an event, an arbitration system will be employed.

For the purpose of simplification, it will be assumed that the sensors used have the same specifications, and that the sensitivity and specificity of the sensors are the same.

Majority voting is most effective when the first assumption holds true, and it is then possible to establish a mathematical rule predicting the network's performance. This model will therefore incorporate this assumption, and a more effective technique for situations where this assumption is not true will be considered at a later stage.

The probability $P(S)$ that an individual sensor correctly detected whether or not the event has occurred, is assumed to simply be equal to both sensitivity and specificity.

The probability $P(B)$, of the combined sensor network correctly detecting whether an event has occurred, varies depending on the number of sensors (n). For a single sensor, it is obvious that the probability of the sensor network choosing a value that is correct is equal to the probability of the individual sensor correctly detecting a signal, i.e.

$$P(B) = P(S) \tag{3.5}$$

For two sensors or more an arbitration system is used.

3.1.2.4 A Revised Model

A consequence of the assumption that all sensors are the same is that in the case of a majority voting system where equal numbers of sensors predict two different events, there is no information on which of the two predictions is correct. The selection must therefore be made arbitrarily, and so one of the sensor predictions must be discarded to break the deadlock. This means that for a number of sensors n, where n is even, the sensor network has the same probability of being correct as a system of $n-1$ sensors.

Following these principles for increasing numbers of sensors, the resultant probability of a sensor network being correct with 3 sensors is:

$$P(B) = P(S)^3 + 3P(S)^2\big(1 - P(S)\big) \tag{3.6}$$

For 5 sensors:

$$P(B) = P(S)^5 + 5P(S)^4\big(1 - P(S)\big) + 10P(S)^3\big(1 - P(S)\big)^2 \tag{3.7}$$

and 7 sensors yields:

$$P(B) = P(S)^7 + 7P(S)^6\big(1 - P(S)\big) + 21P(S)^5\big(1 - P(S)\big)^2 + 35P(S)^4\big(1 - P(S)\big)^3 \tag{3.8}$$

These equations demonstrate a clear pattern, which can be summarised into a final equation as follows:

$$P(B) = \begin{cases} \sum_{k=0}^{\frac{n-1}{2}} {}_nC_k \left(P(S)\right)^{n-k}\big(1 - P(S)\big)^k & if\ n\ is\ odd \\ \sum_{k=0}^{\frac{n}{2}-1} {}_{n-1}C_k \left(P(S)\right)^{n-1-k}\big(1 - P(S)\big)^k & if\ n\ is\ even \end{cases} \tag{3.9}$$

where n is the number of sensors, k is a dummy variable for summation and ${}_nC_k$ is the binomial coefficient. In a real-life situation, where sensors can be wrong, this model gives a much more accurate estimate of the sensor network's performance than the naïve model.

3.2 Methods

3.2.1 EMG Hardware and Software

The EMG sensor kit used was the BITalino (BITalino Freestyle, PLUX Wireless Biosignals, Lisbon, Portugal), a low-cost body-sensing board. It can sense up to five EMG inputs at a time – enough to test the sensor network system in action. Each input is connected to two sensing electrode pads placed on the muscle and a ground reference pad, placed on the collarbone. It runs at a sample rate of 1000Hz [77]. Data was recorded using OpenSignals (r)evolution (Public Build 2018-12-04, PLUX Wireless Biosignals, Lisbon, Portugal). All subsequent processing was done in MATLAB (R2015b, Mathworks Inc, Natick, MA, USA).

3.2.2 Designing an Experiment

To test the model, a method was needed for producing EMG signals under controlled, repeatable conditions, alongside an absolute ground truth indicating whether or not the event was occurring.

The action to be detected was the periodic activation of muscles that close the hand. The rest of the arm was allowed to rest on a desk, being kept entirely still. Surface electrodes

were placed on five different major muscle sites – the extensor digitorum, the flexor carpi radialis, the triceps, the biceps and the lower deltoids, following standard placement recommendations provided by the Surface Electromyography for the Non-Invasive Assessment of Muscles project (SENIAM) [78]. Maximum activity would be expected in the flexor and extensor muscles, with the magnitude (and therefore accuracy) expected to decrease with distance from the hand.

To create a ground truth, a simple online metronome [79] was used. Set to produce an audible beep at a rate of 0.5Hz, this allowed the test subject to open and close their hand at regular two second intervals. Every time the metronome "ticked" when their hand was open, the subject closed their hand such that the springs of the finger exerciser were fully compressed. This position was then held until the metronome "ticked" again, when they fully relaxed their hand. This was repeated for three minutes, or until the subject was no longer able to fully compress the springs.

The data obtained from this experiment was used to identify the probabilities of each stage of the naïve model and compare the accuracy of the combined system with the model's predictions. This was done by constraining the experiment's conditions to set the probability of each stage to ~1, and then relaxing the constraints one by one and measuring the resultant probabilities, which were assumed to be approximately equal to the product of the four stages.

3.2.3 EMG Sensor Processing

Data from the BITalino was imported into MATLAB, and artificial noise was added using a pseudo-random sequence, with tuneable magnitude. The data was processed using standard EMG methods [80]. An optimal voltage threshold was selected and when the smoothed signal exceeded this threshold, it was assigned the value of "1", and otherwise, "0".

3.2.4 Sensor Fusion

There were up to five data channels originating in five sensors, each individually filtered and processed (using general techniques) into 1s and 0s. The data was combined using three methods that are described below.

3.2.4.1 Naïve Model Fusion

The naïve model assumes that if any of the sensors detects an event, then the event has happened, with no possibility of error. To recreate this situation in the EMG example, each of the sensors in the network was compared to the ground truth. If any of their measurements was found to be correct, then the correct value was used as the network output. If not (i.e. they all agreed on an incorrect value), then that value was used instead.

This technique formed an upper bound on the accuracy of the sensors – the best theoretical result the network of sensors could possibly have achieved, as only one sensor needed to be correct for the system to be correct.

3.2.4.2 Majority Voting

As previously discussed, majority voting is a straightforward, well-established method of combining multiple readings for any discrete "multiple-choice" system.

A disadvantage of majority voting is that it does not account for differences in accuracy between measurements. A badly placed sensor which is only accurate 60% of the time makes the same contribution as an excellent sensor with a 99.9% accuracy. This reduces sensor network performance when sensor sensitivities and specificities are not the same.

3.2.4.3 Bayesian Confidence Fusion

An alternative method of combining data from multiple sources is to apply Bayesian techniques to estimate an overall confidence, given the data, whether an event has happened. This can be done using Bayes' rule [81] where $P(E)$ is the probability of the event occurring and $P(V)$ is the probability of getting a particular data set of sensor values:

$$P(E|V) = \frac{P(V|E) \cdot P(E)}{P(V)} \tag{3.10}$$

The data set probability can be obtained using the total probability rule [82], and therefore the confidence can be written as follows, where $P(E')$ represents the probability of the event not occurring:

$$P(E|V) = \frac{P(V|E) \cdot P(E)}{P(V|E) \cdot P(E) + P(V|E') \cdot P(E')} \tag{3.11}$$

$P(V|E)$ is obtained from the sensor sensitivities. It is the probability of all of them getting the results they did at the same time. If they are independent, this is the product of their probabilities. $P(E)$ is a prior for the probability of the event occurring, and knowledge of this is therefore required. One method of obtaining this is to capture the event's frequency – over a time period containing multiple instances of the event, the fraction of time that the event is occurring.

For the metronome-controlled EMG example, the user opened and closed their hand at a rate of 0.5Hz, with a mark space ratio of 1. Over a single 4-second period, both the time that the hand was closed ($T1$) and the time that the hand was open ($T0$) were equal to 2 seconds. Therefore, the probability of the event occurring (i.e. the hand being closed), $P(E)$, at any given time is given as follows:

$$P(E) = \frac{T1}{T1 + T0} = \frac{2}{2 + 2} = 0.5 \tag{3.12}$$

$P(V|E')$, similar to $P(V|E)$, is obtained from the sensor specificities, and $P(E')$ is simply given by:

$$P(E') = 1 - P(E) \tag{3.13}$$

All the information needed to perform this calculation can be obtained with the use of calibration data – a section of experimental data acquired before the main experiment where the user performs the same task, where the ground truth is known – and therefore this technique can be used.

The output of the system, V_{out}, is governed by:

$$V_{out} = \begin{cases} 1 & if\ P(E|V) \geq 0.5 \\ 0 & otherwise \end{cases} \quad (3.14)$$

This system has an advantage over the majority voting in that the accuracy of each sensor is linked to the amount that it contributes toward the sensor network output. Sensors which provide no information, with an accuracy of 50%, are ignored entirely. Sensors with a 100% accuracy are always used over all other sensors. The system uses established statistical mathematics to calculate the probability the event took place, so always outputs the most likely eventuality, given the data from and relative accuracies of the sensors. If applied correctly, using accurate probabilities, it is the theoretically optimal method for combining independent sensors, but calibration data is required.

3.2.5 Analysis

To assess and compare the models, it is important to have a method of gauging their performance. The metrics used were sensitivity (true positive rate) and specificity (true negative rate).

To measure these metrics, a ground truth was required. As previously described, the experiment devised for this chapter required the user to open and close their hand in time with a 0.5Hz metronome. This is a simple, easily repeatable action with a very low possibility of significant error. While the opening and closing of the hand was not expected to perfectly align with the metronome due to factors such as inconsistent reaction time, it was still expected to do so to a very near margin.

Additionally, using a larger time window (i.e. the maximum of 2 seconds) would majorly reduce the impact of these effects, allowing the ground truth to be assumed as a square wave of period 4 seconds. This ground truth was required to be aligned with the experiment's start, but once aligned, it was possible to measure sensitivity and specificity.

If $N1$ is the total number of ground truth 1s and $N1^*$ the number of correctly measured 1s, $N0$ the total number of ground truth 0s and $N0^*$ the number of correctly measured 0s, then sensitivity (SS) and specificity (SP) are defined by:

$$SS = \frac{N0^*}{N0} \tag{3.15}$$

$$SP = \frac{N1^*}{N1} \tag{3.16}$$

The mean of the sensitivity and specificity was obtained and plotted for sensor networks of between $n = 2$ and $n = 5$ sensors. For each value of n, the individual sensor probabilities were varied (by altering artificial noise intensity) to produce 80 points, showing how sensor network probability varies with individual sensor probability.

The measured probabilities were compared to the predictions of the models. The mean, absolute mean, maximum, minimum and root mean squared errors ($RMSE$) for each model were calculated. For each sample i out of N total samples, $\hat{y}_i$ represents the prediction and y_i, the measurement [83].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}} \tag{3.17}$$

The experiment was repeated 10 times, with results represented graphically with scatter plots, to which third-order polynomials were fitted to show the trend of the results.

3.3 Results

3.3.1 Naïve Model

The naïve model prediction was plotted against the measured outcomes (Figure 3.3). Table 3.1 shows the errors between the model outcomes and measured results.

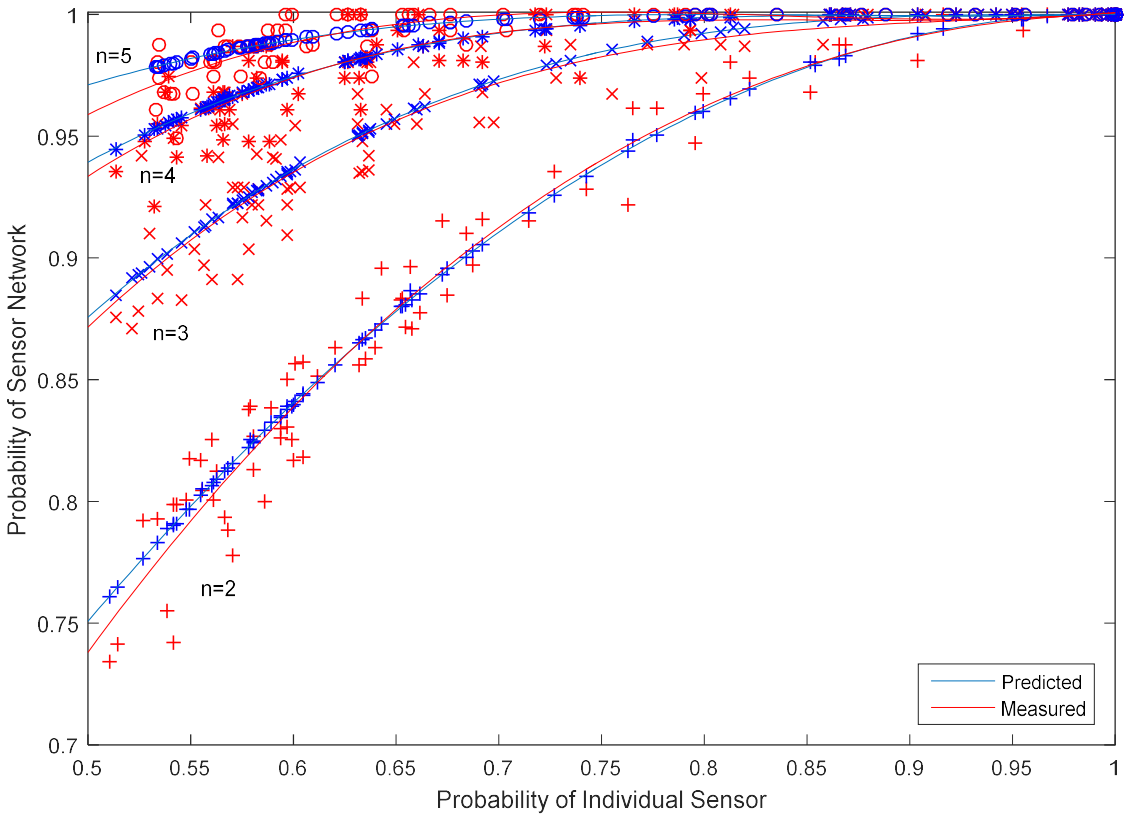

Figure 3.3. *Comparison of the sensor network probabilities measured vs those predicted by the naïve model for n=2 to n=5 sensors, as the mean probability of the individual sensors varies between 0.5 and 1 (controlled by adding artificial noise). Blue points/lines represent the model's prediction. Red points/lines represent the actual measured results. Third-order polynomials have been fitted to demonstrate the trends.*

n	RMSE	Absolute Error			Mean Error
		Mean	Max	Min	
2	0.0142	0.0106	0.0486	0	-0.0015
3	0.0137	0.0100	0.0484	0	-0.0015
4	0.0100	0.0067	0.0315	0	-8.9748e-04
5	0.0069	0.0041	0.0314	0	-0.0013

Table 3.1. *Error between the naïve model prediction and the measured results for a naïve model fusion network of n sensors.*

3.3.2 Revised Model

The revised model prediction was plotted against the measured outcomes (Figure 3.4). Table 3.2 shows the errors between the model outcomes and measured results.

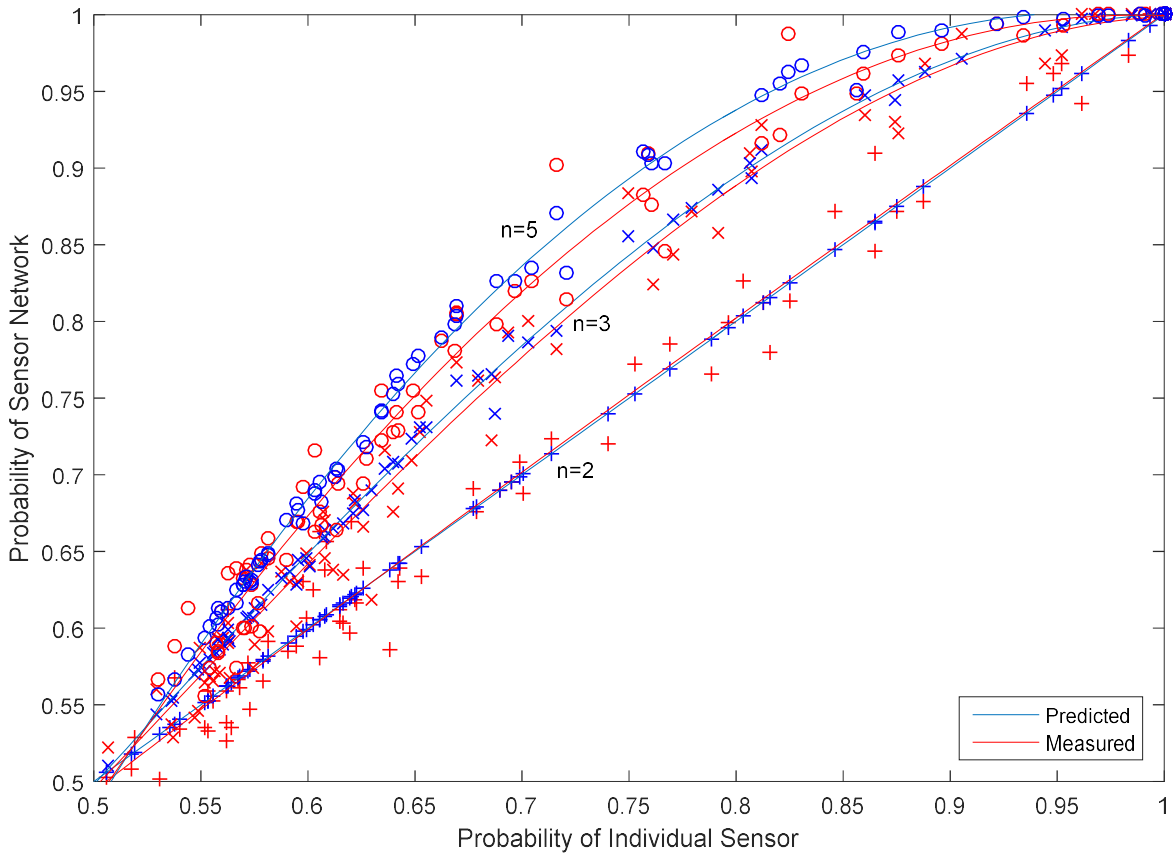

Figure 3.4. *Comparison of the sensor network probability measured vs those predicted by the revised model for n=2, n=3 and n=5 sensors, as the mean probability of the individual sensors varies between 0.5*

and 1 (controlled by artificial noise). Blue points/lines represent the model's prediction. Red points/lines represent the actual measured results. Third-order polynomials have been fitted to demonstrate the trends of the plot. The n=4 results has been omitted, as they overlapped with n=3.

n	RMSE	Absolute Error			Mean Error
		Mean	Max	Min	
2	0.0207	0.0150	0.0575	0	-3.4121e-04
3	0.0188	0.0141	0.0707	0	-0.0053
4	0.0246	0.0174	0.0761	0	-0.0110
5	0.0207	0.0157	0.0575	0	-0.0093

Table 3.2. *Error between the revised model prediction and the measured results for a majority voting network of n sensors.*

3.3.3 Confidence Fusion Method

Figure 3.5 compares the confidence fusion method with the majority voting method.

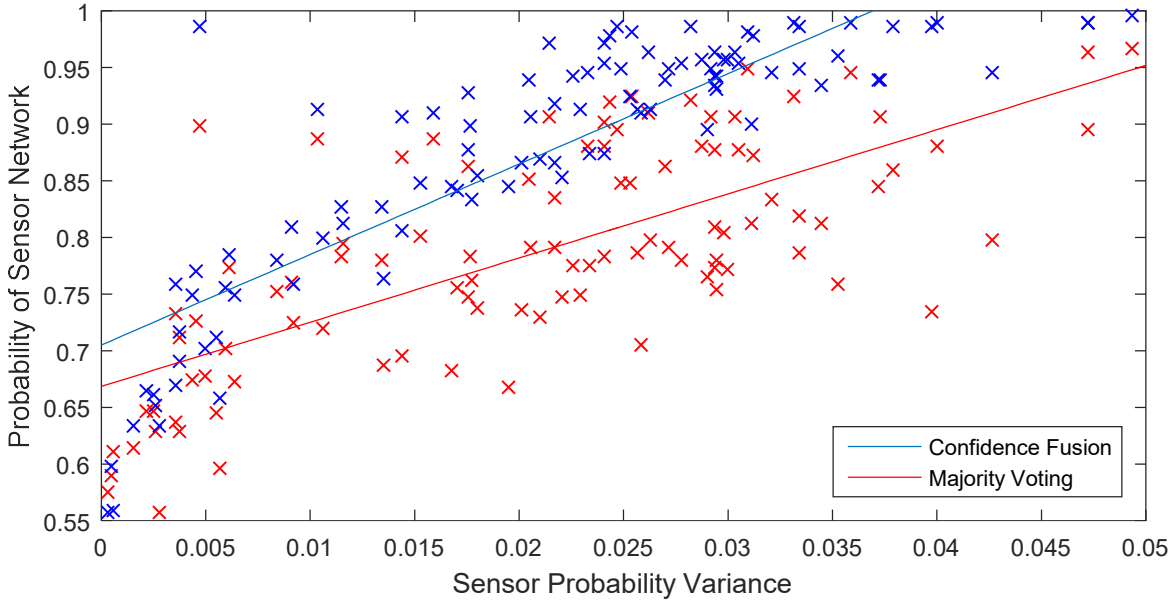

Figure 3.5*. Comparison of the new confidence fusion method with the majority voting method for varying selections of sensor probabilities (n=5), plotted according to their variance. Linear best-fit plots have been overlaid in order to show the general trend.*

3.3.4 Comparing the Models

The naïve and revised models were compared to data from majority voting and confidence fusion, averaged over 80 variations of sensor noise intensities (Figure 3.6).

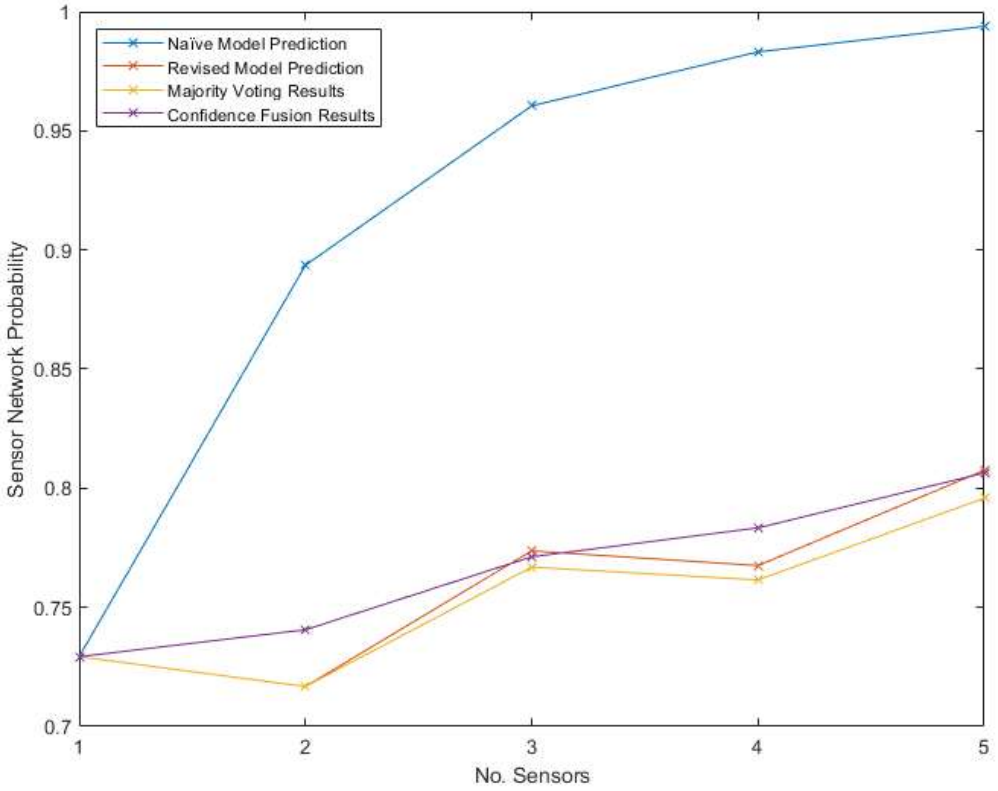

***Figure 3.6**. Comparison between naïve and revised model predictions, along with the measured results of the majority voting and confidence fusion methods, plotted against the number of sensors.*

Model	RMSE	Absolute Error			Mean Error
		Mean	Max	Min	
Naïve	0.2119	0.1580	0.4075	0	0.1580
Revised	0.0146	0.0082	0.0527	0	0.0049

***Table 3.3**. Error between the two model predictions and the measured results for a majority voting network of 5 sensors.*

Model	RMSE	Absolute Error			Mean Error
		Mean	Max	Min	
Naïve	0.1982	0.1459	0.3986	0	0.1459
Revised	0.0226	0.0150	0.0860	0	-0.0072

Table 3.4. *Error between the two model predictions and the measured results for a confidence fusion network of 5 sensors.*

3.4 Discussion

The naïve model shows a good fit with an RMSE of <1.5%. The model predicts the detection based on EMG data and the measured results are comparable to that of other studies [84]. It should be noted that there was a small negative mean error for all choices of n, i.e. the model prediction was a slight overestimate. This kind of error is to be expected due to co-dependency between sensors, as the naïve model assumes that they are independent.

When predicting the performance of the two practically achievable combination methods (majority voting and confidence fusion) the revised model produced far more accurate predictions than the naïve model. As shown in Table 3.3, when analysing EMG data with 80 different combinations of sensor noise in a majority voting system, the revised model gave an RMSE of ~1.5%, whereas that of the naïve model was ~21%. The revised model was slightly less accurate for the confidence fusion method (see Table 3.4), with an RMSE of ~2.3%, but still far closer to the measured value than the naïve model, which had an RMSE of ~20%.

As expected, the difference between the majority voting and confidence fusion methods was smallest when all sensor probabilities were the same i.e. when the two methods were identical. As the variance of the sensors increased, the benefit gained from confidence fusion increased, verifying that the confidence fusion method is the preferred fusion method. However, it requires accurate calibration data, which may impede its practical applicability.

The study has shown experimentally an application of the naïve model [68]. Descriptive statistics have been generated to compare the model's predictions with real EMG data. These results have shown that, when data can be processed as required, the naïve model is a good estimate. However, implementing it practically is often impossible, as it can require knowledge of the ground truth. The model is, however, useful as an estimate of an upper bound on the best possible performance of a system. A revised model has been developed, and while it is a slight overestimate for majority voting and a slight underestimate for confidence fusion (the theoretically optimum method), it is a far more effective tool for estimating sensor network performance than the naïve model and is therefore the best available general model to predict the best network configuration for event detection.

3.4.1 Recommendations for Future Work

This study made extensive use of thresholding in the signal processing stage for reasons of simplicity and ease of mathematical analysis, but more complex alternative methods could be implemented. Future research could focus on this and determine how the

choice of processing techniques impacts the optimum number of sensors and environments in a system. The subsequent chapters of this thesis utilise far more complex classification methods than simple thresholding, for which this framework still applies.

Additionally, the experiments performed used only one subject. While it is expected that similar results would be obtained from repeats using other subjects, EMG signals vary from person to person [85], so a study involving multiple participants from a range of demographics would provide additional verification for the models. Chapters 4 and 5 of this thesis go on to use EMG analysis with much larger subject groups.

Once the models have been further verified, they could be applied in industry with virtual prototyping, to provide estimates for the number of sensors required without experimentation, and to determine potential bottlenecks in a sensor network. Building upon the foundations established in this chapter, it is hoped that future work will push the boundaries of what is achievable through the use of PSNs, accelerating product design with and using the new models described to bring wearable sensor networks into widespread use, supporting the healthcare industry and thus the wellbeing of the public as a whole.

Statement of Authorship for joint/multi-authored papers for PGR thesis

Title of Paper	Probabilistic Sensor Design for Healthcare Technology
Publication Status	■ Published □ Accepted for Publication □ Submitted for Publication □ Unpublished and unsubmitted work Written in a manuscript style
Publication Details	J. Russell and J. Bergmann, "Probabilistic sensor design for healthcare technology," 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 2019, pp. 0585-0590, doi: 10.1109/UEMCON47517.2019.8993086.

Student Confirmation

Student Name	Joseph Russell		
Contribution to the Paper	Performed all investigation, data collection, software development, analysis and writing. J. Bergmann provided supervisory guidance and editing of the manuscript.		
Signature		Date	06/08/2023

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Prof Jeroen Bergmann		
Supervisor comments		
Signature 	Date	06/08/2023

4

Towards Dynamic Multi-Modal Intent Sensing Using Probabilistic Sensor Networks

Published by Joseph Russell, Jeroen Bergmann and Vikranth H. Nagaraja in *Sensors* 2022, *22*, 2603, doi: 10.3390/s22072603

Chapter Abstract

Intent sensing has many potential technological applications. In particular, assistive medical devices such as prosthetic limbs, could benefit from intent-based control systems, allowing for faster and more intuitive control. The accuracy of intent sensing could be improved by using multiple sensors sensing multiple environments. As users will typically pass through different sensing environments throughout the day, the system should be dynamic, with sensors dropping in and out as required. An intent sensing algorithm that allows for this cannot rely on training from only a particular combination of sensors. It should allow any (dynamic) combination of sensors to be used.

Therefore, the objective of this study is to develop and test a dynamic intent sensing system under changing conditions. A method has been proposed that treats each sensor individually and combines them using Bayesian sensor fusion. This approach was tested on laboratory data obtained from subjects wearing Inertial Measurement Units and surface electromyography electrodes. The proposed algorithm was then used to classify functional reach activities and compare the performance to an established classifier (k-nearest-neighbours) in cases of simulated sensor dropouts. Results showed that the Bayesian sensor fusion algorithm was less affected as more sensors dropped out, supporting this intent sensing approach as viable in dynamic real-world scenarios.

4.1 Introduction

4.1.1 Intent Sensing

Chapters 1 and 2 established that intent sensing has great potential for the development of technology involving human–system interaction. Devices controlled through intent sensing would not need a direct input from the user. Instead, using a Probabilistic Sensor Network (PSN) as described in Chapter 3, the user's needs could be passively measured or even anticipated, and the device of interest could be activated automatically.

This could be of particular benefit in medical devices, where the accurate understanding of the user's intent could help improve their quality of life. The potential applications are wide-ranging, including assistive robotics for stroke patients [86] and detecting physical activity for personalised drug delivery in those with diabetes [14]. One application of particular interest is prosthetics.

As of 2017, there are estimated to be 57.7 million people worldwide living with limb amputation due to traumatic causes [87], many of whom use or are in need of a prosthetic device. One of the principle causes of device abandonment is reported to be difficulty of use [88], and as such, an intent sensing system that could improve the intuitiveness and user experience of a prosthetic device has the potential to be extremely valuable.

As established in Chapter 2, intent sensing is composed of three aspects: (i) the recognition and identification of activity transitions, (ii) the inference of task goals, and (iii) the prediction of future activities. See Figure 4.1 for an illustration of this.

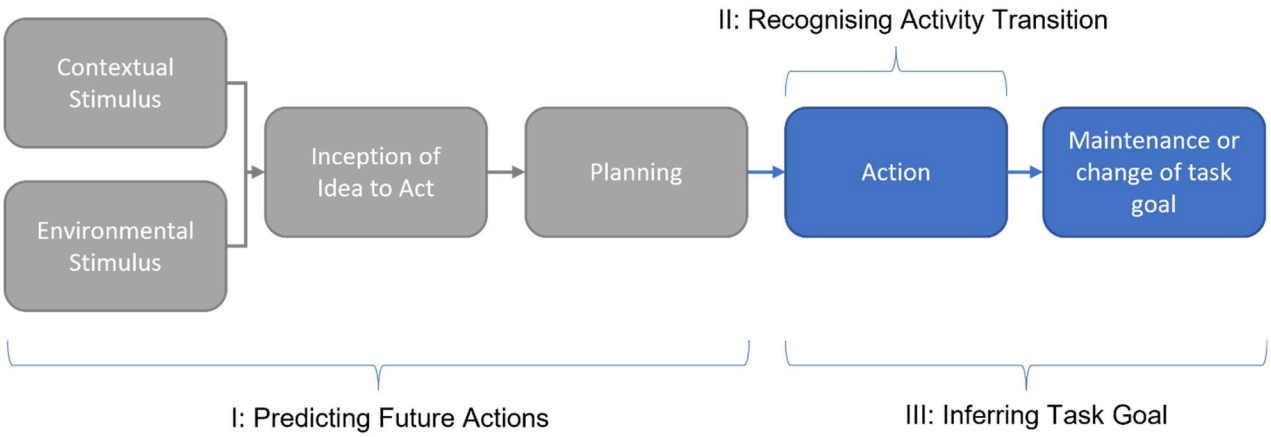

Figure 4.1. *Framework representing the three aspects of intent sensing. Adapted from Chapter 2. Grey regions indicate stages not utilised in this study. Environmental Stimulus indicates change in the user's surroundings (such as sound or the arrival of another person) that may trigger a response. Contextual Stimulus includes wider factors such as the time of day, typical routine, previous actions etc. The Inception of the Idea to Act represents the user's conscious decision to take action and predominantly concerns the brain and nervous system. The Planning phase includes any preparation that may take place before the activity begins (such as a visual inspection of the path ahead when about to start walking or the pre-tensing of muscles before attempting a timed grasp). The Action phase covers the real-time execution of the activity, including any changes observable while the activity is being performed. The Maintenance or Change of Task Goal phase looks ahead to the objective of the activity, considering why it is being performed and whether this objective alters over the course of the activity.*

As shown in Chapter 2, current intent sensing literature contains a well-established body of work related to activity transition recognition, with less research on task goals

and predictions. This chapter will predominantly focus on transition recognition, followed by an inference of the task goal. While attempts will be made to perform classification as early on as possible in an activity, outright prediction before an activity begins is not within the scope of this study.

4.1.2 Sensor Networks for Intent

The accuracy of input detection is particularly essential in assistive medical devices such as prosthetics, where errors could lead to both frustration and injury. Provided effective sensor fusion algorithms are used, accuracy can be improved by combining information from multiple sensors. To produce the highest possible accuracy, the intent sensing system should take advantage of all sensors available at any time.

A range of embedded sensors might be available in a typical prosthetic device. These can often include surface electromyography (sEMG) sensors integrated into the socket, measuring electrical activity in the residuum of the user's superficial muscles [89]. Kinematic sensors are also often available, providing measurements of the device's orientation and acceleration through Inertial Measurement Units (IMUs) and combining together information from accelerometers, gyroscopes and magnetometers [90].

Multiple sensors could be added to a prosthetic to increase accuracy. However, there are potential disadvantages to adding more and more sensors, such as increases in cost, weight, power consumption and difficulty donning/doffing the device. To further improve the accuracy of the system without introducing these issues, information from external sources could also be included.

Smart phones and smart watches, for instance, include a whole range of sensors, such as audio and GPS sensors and are commonly used throughout the day. When these devices are in use, they provide additional sensory information largely independent from the sensors that are embedded in the medical device [91].

Smart home technology is also becoming increasingly prevalent [15], including smart doorbells, speakers, home assistants and more. In-home monitoring of patients using (both wearable and contactless) sensors for medical purposes is also an increasingly common practice. Smart environments are also being developed in workplaces, on public transport and within private vehicles, each providing a range of sensory information that could be used to provide predictions of intent.

These are examples of how device users pass through varying sensor-rich environments, with sensors becoming available or unavailable throughout the day. To take advantage of these uncertain information sources, an intent sensing system should be dynamic, following a "drop-in/drop-out" structure, incorporating information from sensors only when they are available and weighting their contribution according to their accuracy, which would have to be pre-learned on a sensor-by-sensor basis rather than as a complete, fixed network. This approach would allow for any set of sensors to be combined for intent sensing.

This would not only enhance the performance of the system by utilising all available resources whenever possible, but would also improve system robustness in case of the failure of some of the sensors. A system that is still able to run (albeit with reduced

accuracy) with only a subset of its sensors will have a huge advantage over systems that require all their sensors to work in unison to function.

This study therefore proposes to investigate the use of a new, modular approach to networked intent sensing that is not trained on any particular combination of sensors, but instead is able to freely add in or remove sensors to produce robust intent predictions.

4.1.3 Probabilistic Sensor Networks

One approach to creating the proposed robust, drop-in/drop-out system is to model the process as a PSN [68]. As discussed in Chapter 3, this method breaks the process of detection down into four stages, each with their own independent probability of the signal correctly passing through them: Environment, Sensing, Conditioning and Processing (See Figure 4.2). Assuming independence between the stages, the total probability of correctly detecting an event is the product of these four probabilities.

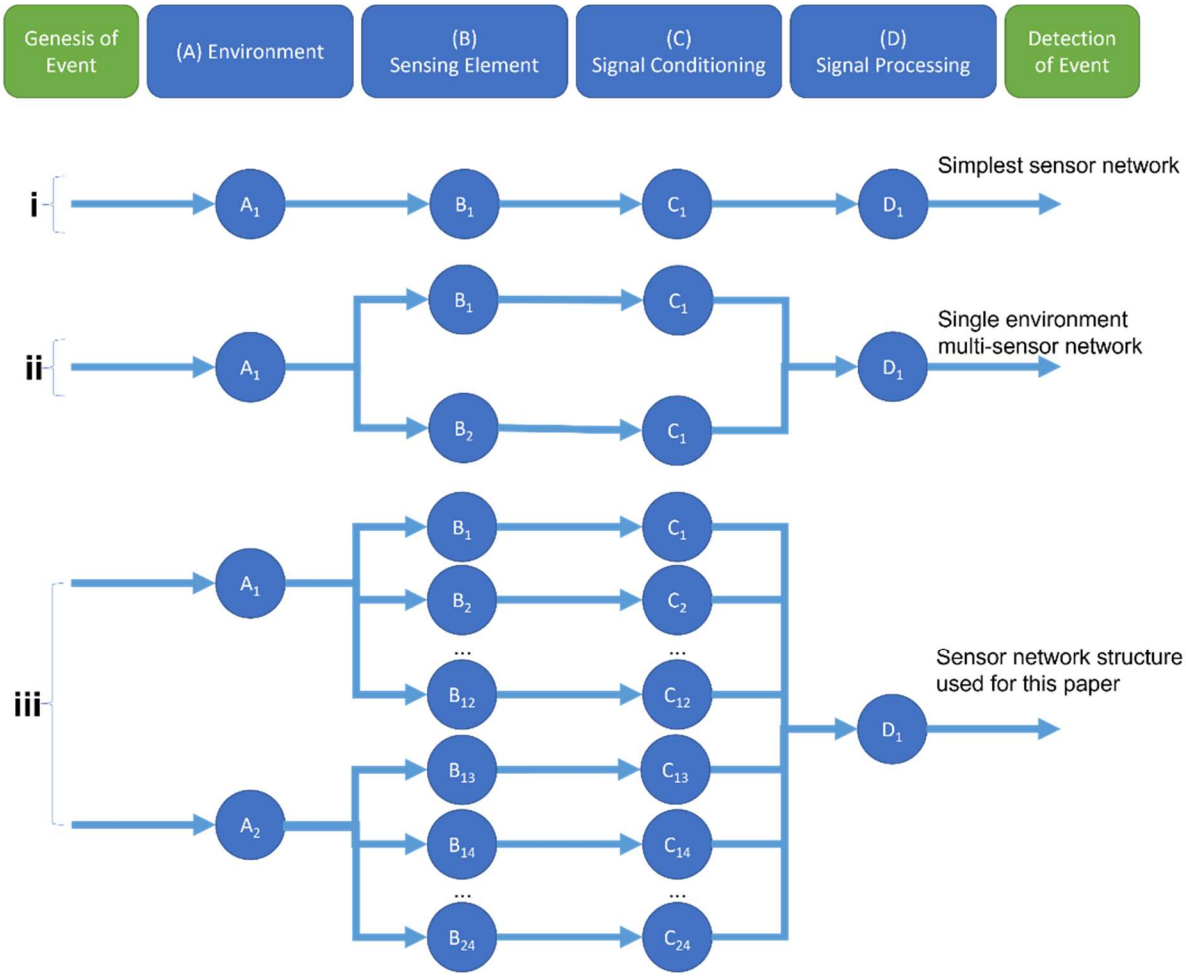

Figure 4.2. *Model of a Probabilistic Sensor Network for event detection (adapted from* [68]*). The four steps that lead from the genesis of a given event up to its detection are shown at the top of the figure. (**i**) shows the simplest design, with only one node at each stage. (**ii**) shows a more complex network, featuring two sensors sharing the same sensing environment, each with their own conditioning step. (**iii**) displays the network used in this study, with two sensing environments (EMG and IMU sensing), each with 12 sensors, conditioned individually and combined together in the processing step.*

Since the probability associated with each stage is naturally less than or equal to 1, the total probability of correctly detecting an event (in this case, a particular user intent) will never be greater than the highest probability of the four stages. Therefore, each stage acts as a potential "bottleneck" for the system. For instance, if the probability of

an intent being detectable from the electrical activity in the muscle is only 0.8, then the Environment probability is 0.8, and so it does not matter how accurate the Sensing, Conditioning or Processing stages become—the total probability will never exceed 0.8.

Each stage can contain multiple sensing nodes. Adding a second sEMG sensor to another site on the muscle and combining the two sensors together with a sensor fusion algorithm increases the probability of the Sensing stage. However, both sensors are operating in the same Environment, and so the total probability will still be limited to 0.8. Adding sEMG sensors to a second Environment, such as a different muscle, helps circumvent this "bottleneck" and allows for the total probability of the system to increase.

In order to keep the network dynamic without being limited to any particular configuration of sensors, each sensor node will be considered on its own and only combined together at the final Processing stage.

In the proposed intent sensing system, sensors that are added may have very high accuracies for identifying some intents and low accuracies for others—such as sEMG sensors placed on muscles in the lower left leg when used to detect walking versus using the same sensors to detect a reaching motion of the arm. A simple majority voting system would not account for this, so it would be unsuitable.

Instead, the more versatile Bayesian approach proposed in Chapter 3, Section 3.2.4.3 can be taken. Where $P(E)$ is the probability of the event occurring and $P(V)$ is the probability of getting a particular set of sensor values, Bayes' rule [81] gives the probability that an event has happened given a set of sensor values, $P(E|V)$, as:

$$P(E|V) = \frac{P(V|E) \cdot P(E)}{P(V)} \tag{4.1}$$

The dataset probability can be obtained using the total probability rule [82], and so the confidence can be written as follows, where $P(E')$ represents the probability of the event not occurring and $P(V|E')$ represents the probability of getting a set of sensor values given that the event has not occurred:

$$P(E|V) = \frac{P(V|E) \cdot P(E)}{P(V|E) \cdot P(E) + P(V|E') \cdot P(E')} \tag{4.2}$$

$P(V|E)$ is obtained from the sensor sensitivities. $P(E)$ is a prior for the probability of the event occurring and could be obtained from contextual information, such as user routine and time of day.

Calculating a confidence value in this way for each possible intent and choosing the intent option with the highest confidence is a more effective method of combining sensor outputs. Effectively, it is giving an optimal "weighting" to each sensor's contribution according to its individual accuracy. Combining sensor information according to this method means that, provided the accuracy of each sensor is precisely known, adding sensors can only monotonically improve the overall accuracy of the system, even if only by a very small amount as in the case of sensors that are close to random in their predictions.

To apply such a method, accuracy estimations for each sensor will need to be obtained for each possible intent option, i.e., each sensor will require a known confusion matrix. This should list the probability of each intent being true given the sensor's prediction of

a particular intent. The entries for this confusion matrix can be populated through calibration. This can be completed individually for each user, for all users or for some combination of the two, starting with a general estimation and adjusting it to become more personalised over time.

In the case of sensors dropping in and out over time, all that is required in this Bayesian method is for their confusion matrices to be added to or removed from the equation. No retraining of the other sensors is needed to compensate for the change, and therefore this method is interesting as a potential solution to the need for a dynamic network.

Other, more advanced methods of combining sensor measurements together do exist. Many of these employ machine learning techniques, such as decision trees [92], random forest classifiers [93] and support vector machines [94]. These techniques do not treat each sensor as an individual "black box" and instead consider, for example, the relationships between sensors. Exploiting this extra dimensionality has the potential to provide additional information, suggesting it is a more suitable technique. However, it also means that they must be trained on specific combinations of sensors and that adding and removing dimensions from the trained classifiers "on the fly," as sensors drop in and drop out of the network, is not within the capability of established machine learning techniques (if not trained for it).

It is therefore proposed that while combined machine learning techniques are theoretically able to perform better than the Bayesian fusion technique, they will rapidly drop

in performance when sensors are removed, and repeated experiments simulating sensors dropping out should show this effect.

4.1.4 Objective

The objective of this study is to develop a Modular Method (MM) suitable for the dynamic environments described previously and to compare it to a Non-Modular Method (NMM), utilising a combined machine learning algorithm representative of current commonly used methods. The experiments performed should measure classification accuracy early in the activity cycle, well before any activity example is completed, in order to demonstrate the goal inference aspect of intent sensing. The change in accuracy with increasing time allowed between the activity's inception and classification should be determined.

Then, the two algorithms should be compared in their accuracy versus a varying number of sensors when they were allowed to train on the exact combination of sensors they are being tested on. The hypothesis is that the new MM will not perform as well as the NMM.

Subsequently, a scenario of random sensor dropout will be introduced, where the sensor combinations required are not trained on in advance. The hypothesis is that the MM will perform better overall than the NMM, with the difference in accuracy between the two techniques increasing as more sensors drop out.

4.2 Methods

4.2.1 Data Collection

Data used for this study were originally gathered under laboratory conditions as part of a prior study [95]. All data were anonymised, and prior informed consent was obtained from each individual. The study was approved by the institutional ethics committee (Reference Numbers: 16/SC/0051 and 14/LO/1975).

Five adult non-disabled participants wore full-body MVN-Awinda [96] IMUs (Xsens Technologies B.V., Enschede, The Netherlands) during task execution (sampling frequency: 60 Hz). Passive retro-reflective markers were placed on their body, which were tracked with a 16-camera Vicon (Vicon, Oxford, UK) motion tracking system. This Optical Motion Capture (OMC) data (sampling frequency: 100 Hz) were used for the verification of activity labels.

To minimise issues with the magnetometer component of the IMUs, it was ensured that the only sources of local magnetic interference were the data collection devices themselves and the laptop used to receive the recorded data, and the layout of the room was kept exactly the same throughout the study.

A wireless 10-channel Zerowire (Aurion Srl, Milan, Italy) EMG system was used to collect sEMG data at a sampling frequency of 1000 Hz for the five selected superficial muscle groups: Pectoralis major (Clavicle), Biceps brachii, Triceps (Long head), Deltoid (Medial) and Brachioradialis.

Subjects were asked to perform three trials each of eleven different reach/grasp activities (Reach to grasp: (i) forward, (ii) left, (iii) right and (iv) up; Reach: (v) forward, (vi) left, (vii) right and (viii) up; (ix) hand to mouth, (x) hand to the top of head and (xi) hand to contralateral shoulder), which were grouped into three categories: Reach tasks, Reach-to-Grasp tasks and Gross Motor Skill tasks.

These actions were performed using a specially built test rig, shown in Figure 4.3. A total of 165 datasets were collected. Each participant also performed a Maximal Voluntary Isometric Contraction (MVIC) test with each measured muscle group to allow for normalisation of the corresponding experimental sEMG data.

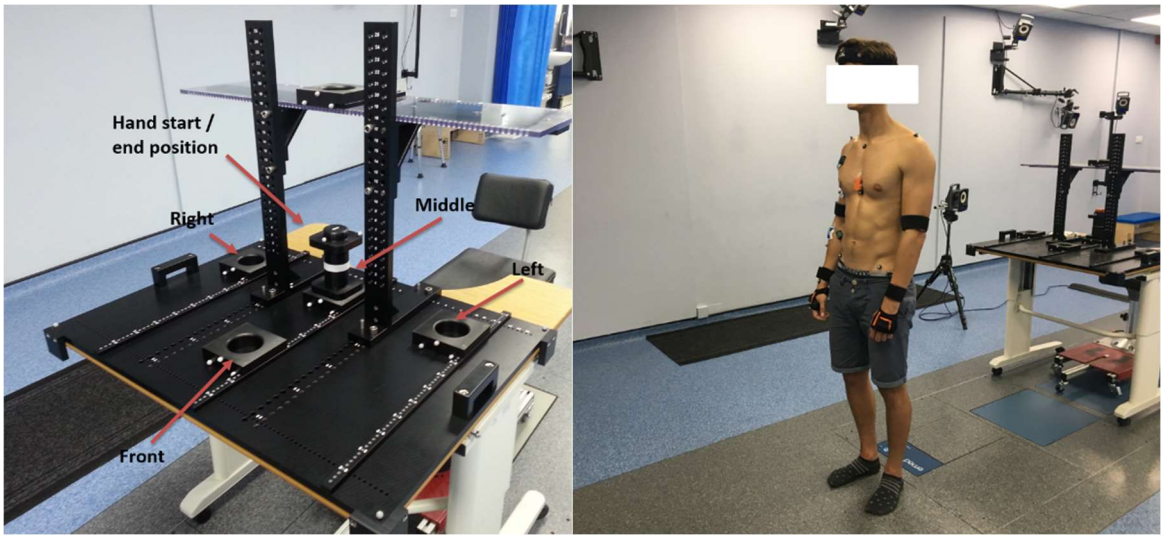

Figure 4.3. *Photographs of the experimental setup used for the acquisition of data used in this study. The configuration pictured is for the reach-grasp activity.*

To ensure the experiment was representative of intent sensing rather than pure activity classification, only the first 1000 ms of each dataset were used.

4.2.2 Processing

The affiliated Xsens MVN Analyze software [97] was used to process the IMC data and export it as MVN open XML (.MVNX) files to be used as inputs for the algorithms.

The OMC data corresponding to marker trajectories was processed and exported in C3D or Coordinate 3D (C3D.ORG; https://www.c3d.org/index.html; accessed on 27 March 2022) format using the Vicon Nexus 2.5 software [98]. Further processing took place in MATLAB R2020b (Mathworks, Natick, MA, USA). The sEMG data was synchronised with the IMU data from the Xsens IMC system using the recommended protocol [99].

The sEMG data was filtered using a 10–500 Hz band-pass fourth order Butterworth filter and normalised according to the maximum signal measured in the MVIC tests [100].

4.2.3 Feature Extraction

A breakdown of the features extracted from the IMU and sEMG sensor signals is shown in Table 4.1.

Measured Value	No. Features
Orientation	4
Accelerometer	3
Magnetometer	3
sEMG	11
Total	**21**

Table 4.1. *Breakdown of the number of features used from each sensing input available.*

The three IMU channels included were orientation, accelerometer and magnetometer signals. The raw values for these were used directly in the training step. These three distinct data channels allowed each IMU sensor to be more closely approximated as

probabilistically independent by the Bayesian model. The orientation measurement contained some co-dependency on the accelerometer and magnetometer readings.

For the sEMG signals, a more complex feature extraction method was required. The process detailed in [101] was followed to ensure standard methods were applied. The data for each participant were divided into 200 ms segments and shifted by 50 ms increments (such that consecutive segments overlapped by 150 ms). Within these segments, the following features were extracted: Integrated EMG, Mean Absolute Value, Mean Absolute Value Slope, Variance of EMG, Root Mean Square, Waveform Length, Autoregressive Coefficients (to the fourth order), Frequency Median and Frequency Mean.

This full set of features was carried forward for analysis. An investigation of feature reduction, not used in the final algorithm, may be found in Appendix A.

4.2.4 Data Separation (MM/Bayesian Fusion Only)

To allow as large a training set as possible, leave-one-out cross-validation was used. As such, for each repetition of the analysis, one sample was held back for testing (the Testing Set), leaving 164 samples for training.

In order to learn the probabilities associated with each sensor and thereby populate each sensor's confusion matrix (as is required for the Bayesian sensor fusion method) without introducing any element of bias, the remaining samples were divided again. Half of the 164 samples were pseudo-randomly selected and used to train an MM classifier for each sensor (the Classifier Training Set), and the remaining half were used to test the

MM classifiers and measure their accuracy for each activity (the Probability Learning Set). To complete this, the number of successful classifications was divided by the number of samples for each activity to estimate the probability, which was recorded in the confusion matrix. The sums of the diagonal entries of the confusion matrices were then used to approximate each sensor's overall accuracy.

The choice of which samples were placed in the Classifier Training Set and which in the Probability Learning Set could have an impact on the performance of the classifier. To produce results representative of all subjects and activities, a selection algorithm was used to pseudo-randomly place an approximately equal number of data samples from each participant and activity type in the Classifier Training and Probability Learning Sets. Where multiple examples were available (each participant provided three samples of each activity), the set they were placed in was randomly selected.

An optimisation step also took place here, repeating the previous steps five times, with a different pseudo-randomly selected split between the sets each time. The classifier chosen to take forward was whichever resulted in the highest mean accuracy across sensors, as measured in the Probability Learning Set.

For the comparison, a Combined K-Nearest Neighbours (KNN) technique was used in the NMM; this data-separation step was not required, as all data in the training set from all sensors were used to train a single classifier, with no estimation of the classifier's accuracy.

4.2.5 Learning Classifiers

For both the MM (Bayesian Fusion algorithm) and the NMM (Combined KNN algorithm), KNN classifiers were used. The distinct difference between the two was that, for the MM, one KNN classifier was trained for each sensor and then combined, whereas the NMM trained a single classifier using all the sensors as inputs (see Figure 4.4 for a graphical representation).

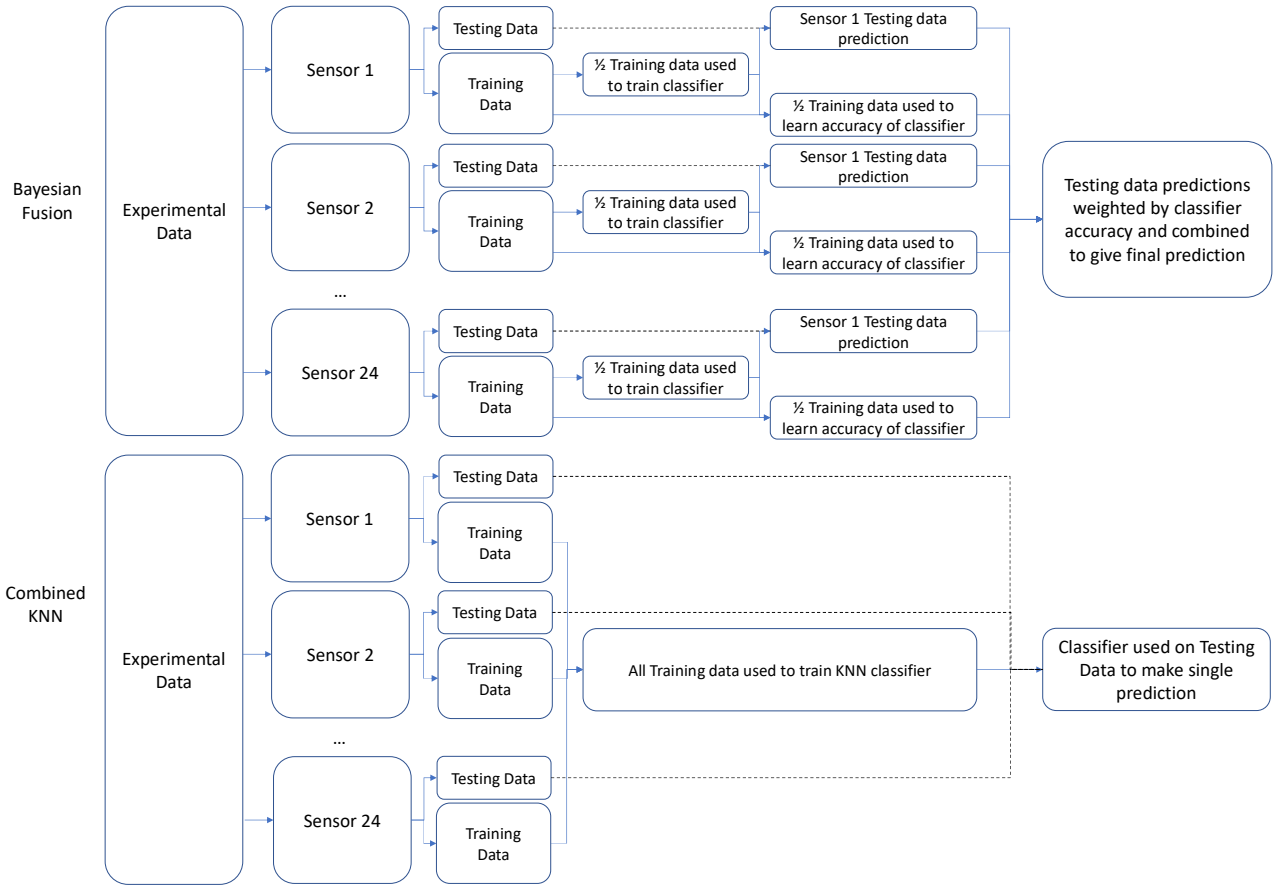

Figure 4.4. *Overall pipeline of training and testing for the two algorithms. The Bayesian Fusion (MM) algorithm divides the training dataset into Classifier Training and Probability Learning subsets. The Classifier Training Set is used to train the classifier for each sensor, and then the Probability Learning set is used to populate the confusion matrix for each classifier. The confusion matrix then provides weightings for each sensor's contribution to the overall network output, which is used to predict the class of the testing set. The Combined KNN (NMM) does not subdivide the training set, instead using all sensor*

data to train a single classifier, which then predicts the class of the testing set. Dashed lines indicate testing data.

The features described in Section 4.2.3 were used as inputs to train the KNN classifier [102], with hyperparameter optimisation selecting a K value (no. neighbours) of 1 and a Gaussian distance metric with an exponent of 0.5. The same number of features was used for both algorithms. For the Combined KNN classifier, all features were used for one algorithm, whereas for the Bayesian Fusion algorithm, the features were evenly distributed across the individual sensors.

4.2.6 Sensor Fusion (Bayesian Fusion Only)

In the Bayesian Fusion algorithm, each sensor produced its own independent classification of the activity. Equation 4.2 was used with the confusion matrices populated in the Probability Learning Set in order to calculate a probability of each activity being the true activity given the sensor values. Whichever activity had the highest probability was selected as the output of the combined system.

For the comparison with the Combined KNN algorithm, all the sensor inputs were fed into the KNN algorithm previously used exclusively within each EMG sensor. This is a well-established supervised learning technique [103] used in many EMG-driven intent sensing studies and should be representative of the general performance of machine learning techniques.

4.2.7 Testing

The performance of the two algorithms was compared by testing their classification of the data sample in the testing set. The total number of correct classifications across the 165-fold leave-one-out cross-validation method was divided by the number of trials (165) to result in an accuracy measure for each algorithm.

4.2.8 Time Variation

The first goal to be investigated was the effect on the accuracy of intent classification when varying amounts of time were allowed to pass after the activity's inception before intent classification was performed. To measure this, the experiment was repeated with all 24 sensors active, making the prediction using only the first X milliseconds of each sample, with X increasing from 200 in 50 ms increments up to the full 1000 ms allowed. The resulting accuracies were plotted against the time allowed to show the trend. The trend for both methods was then quantified using a Spearman's rank correlation coefficient, where $R(P_i)$ and $R(T_i)$ are the ranks of each (i-th) sample in accuracy and time, respectively, and n is the number of samples:

$$r_s = 1 - \frac{6\sum\left(R(P_i) - R(T_i)\right)^2}{n(n^2 - 1)} \tag{4.3}$$

This gives a result between −1 and 1, where 1 is a perfectly monotonically increasing pattern, −1 is a perfectly monotonically decreasing pattern and 0 indicates no monotonic relationship. This is an appropriate measure, as it will indicate to what extent the

hypothesis is true, namely that accuracy will increase when more time is allowed to pass [104].

4.2.9 Variant No. Sensors

The second area to be investigated was the effect on intent classification accuracy when the number of sensors was varied, where in each instance, the algorithms were trained only on the sensors that were active. It should be noted that this is not testing robustness to sensor dropout (this is described in Section 4.2.10) but instead showing the effect of increasing the number of sensors.

The number of sensors used as inputs, R, was varied from 1 to 24. For each number, 50 randomly selected combinations of R sensors were tested and the mean accuracy was recorded, along with the 95% confidence intervals. These were plotted and compared graphically.

Referring back to the PSN model discussed in Section 4.1.3, as both algorithms use the same sensors in this experiment, in the same sensing environments and with the same signal conditioning, this suggests the difference in probability comes entirely from the Stage D, Signal Processing step.

If the probability associated with the MM (Bayesian Fusion) method is $P(D1)$ and the probability associated with the NMM (Combined KNN) method is $P(D2)$, and the probabilities of the Environment, Sensing and Conditioning stages are $P(A)$, $P(B)$ and $P(C)$ for both algorithms, respectively, the ratio of the Processing probabilities for the two algorithms is given by:

$$\frac{P(MM)}{P(NMM)} = \frac{P(A) \cdot P(B) \cdot P(C) \cdot P(D1)}{P(A) \cdot P(B) \cdot P(C) \cdot P(D2)} = \frac{P(D1)}{P(D2)} \tag{4.4}$$

This ratio therefore quantifies the relative benefit of using the MM over the NMM for intent detection, and so this was estimated using the data points plotted and then used to compare the algorithms. This ratio was not expected to be consistent for all data points, and so the range of values was given.

4.2.10 Simulated Dropout

The third (and most important) element to be investigated was the effect on the intent classification accuracy of the algorithms only trained with all sensors active when a number of sensors begin "dropping out". To test this, the experiment was repeated again, this time by randomly selecting N sensors to be set to a constant 0. For the Bayesian algorithm, these sensors' predictions were not used. On the other hand, the K-nearest-neighbours algorithm trained on all the sensors cannot have an input removed, so they continued to use the 0 value.

The number of dropped sensors, N, was increased from 0 to 23, with the predictions made only 1000 ms into each activity—well before their completion, making this analysis a "goal inference" task. The accuracies measured were averaged over 50 random combinations of N sensors dropping out. The mean accuracies over twenty repetitions of the MM (Bayesian Fusion) and NMM (Combined KNN) were plotted against N, along with the upper and lower bounds of the 95% confidence interval of each.

As in Section 4.2.9, the ratio of the accuracy of the MM to the NMM was used to quantify the relative benefit. This will vary as N increases, and so the range was given.

4.3 Results

The accuracies of the two classification methods versus an increasing amount of time allowed after activity inception with all sensors used are shown in Figure 4.5. Both showed trends of increasing accuracy over time, with a Spearman's rank correlation coefficient of 0.9 for the MM (Bayesian Fusion) and 0.6 for the NMM (Combined KNN).

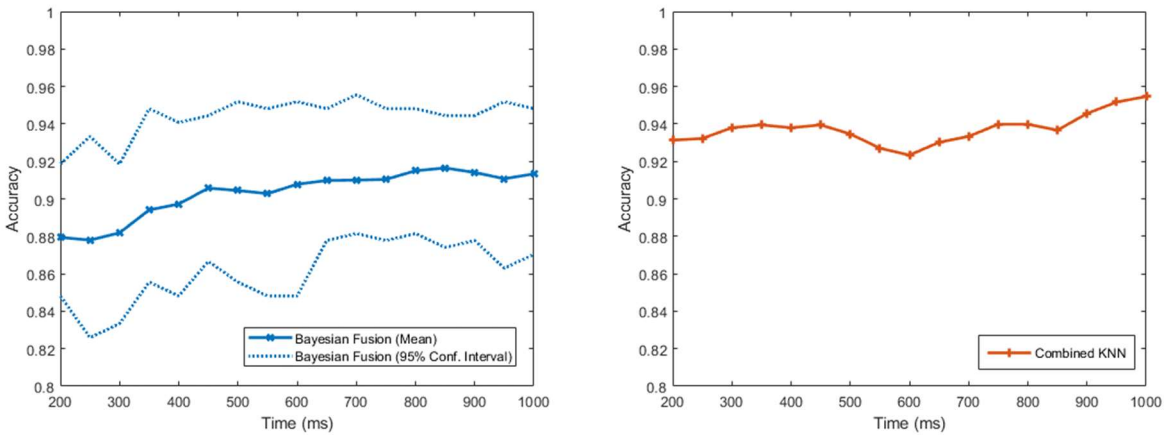

Figure 4.5. *Comparisons of the accuracy of the intent-classification system against the time allowed to pass after the activity's inception before classification was performed, up to the 1-second limit and with all 24 sensors active (no dropout). These graphs are shown separately to clearly illustrate the presence of a general trend for each algorithm, but comparisons between the two in this context should be avoided (see Section 4.4.1 for discussion on this). The Combined KNN method does not have confidence intervals, as all the sensors are included and there is no subdivision of the training data, so its performance is entirely reproducible.*

The accuracies of the MM and the NMM trained and tested on a varying number of sensors from 1 to 24 with no dropout are shown in Figure 4.6. The NMM had consistently higher accuracy than the MM in this case, with no overlap between the 95% confidence intervals until the number of sensors reached 21, at which point the upper confidence interval of the MM exceeded the lower confidence interval of the NMM. The accuracy ratio of the MM to the NMM method ranged between 0.8 and 0.9, with a mean of 0.85.

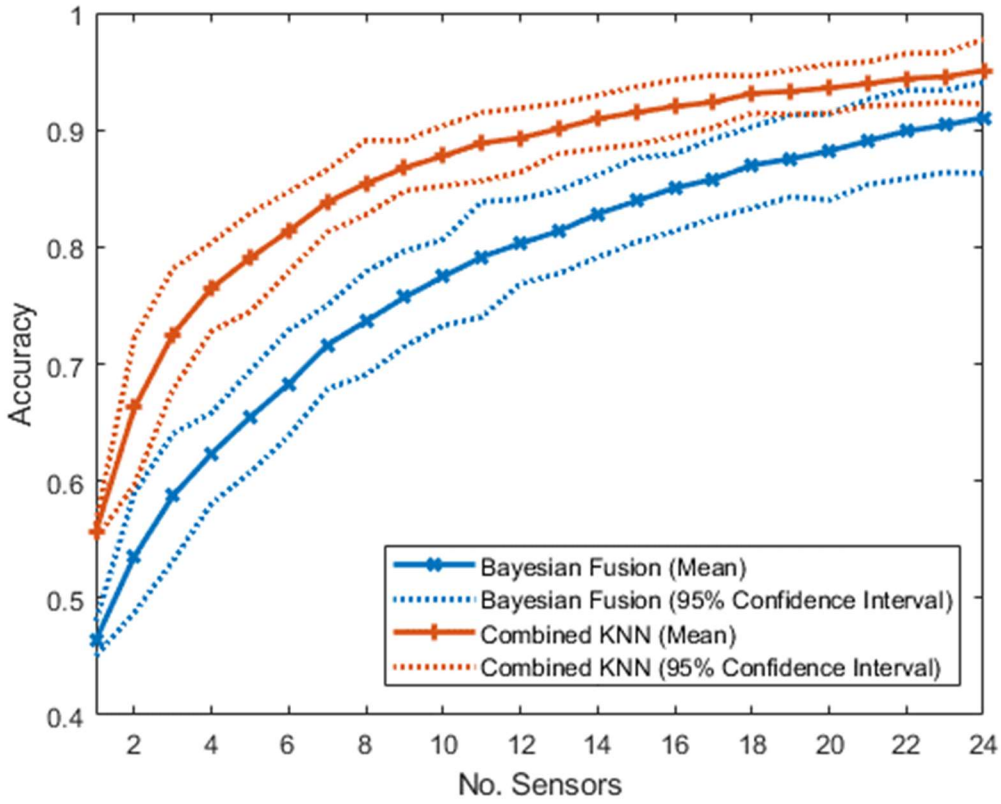

Figure 4.6. *Comparison of the accuracy of the intent-classification system using the Bayesian Fusion method (treating each sensor separately and then combining them) and the combined method (putting all sensor information into a single KNN classifier) as the number of sensors increases. No sensors dropped out—instead, the number of sensors was varied from 1 to 24, and the algorithms were trained on the number of sensors active in each case.*

The accuracies of the MM and the NMM all trained on the complete set of sensors when the number of sensors dropping out varies from 0 to 23 are shown in Figure 4.7. Initially, with no dropped sensors, the NMM resulted in a higher mean accuracy than the MM. As more sensors dropped out, this difference decreased, until with 10 sensors dropping out, the MM mean accuracy exceeded that of the NMM. From this point on, the accuracy advantage of the MM over the NMM continued to increase, resulting in an accuracy ratio of MM to NMM ranging from 0.9 at 0 dropout to 1.4. Initially, there was a large amount of crossover between the 95% confidence intervals of the two methods, but after 17 sensors had dropped out, the lower bound of the MM became greater than the upper bound of the NMM, and remained so for all greater numbers of dropped sensors.

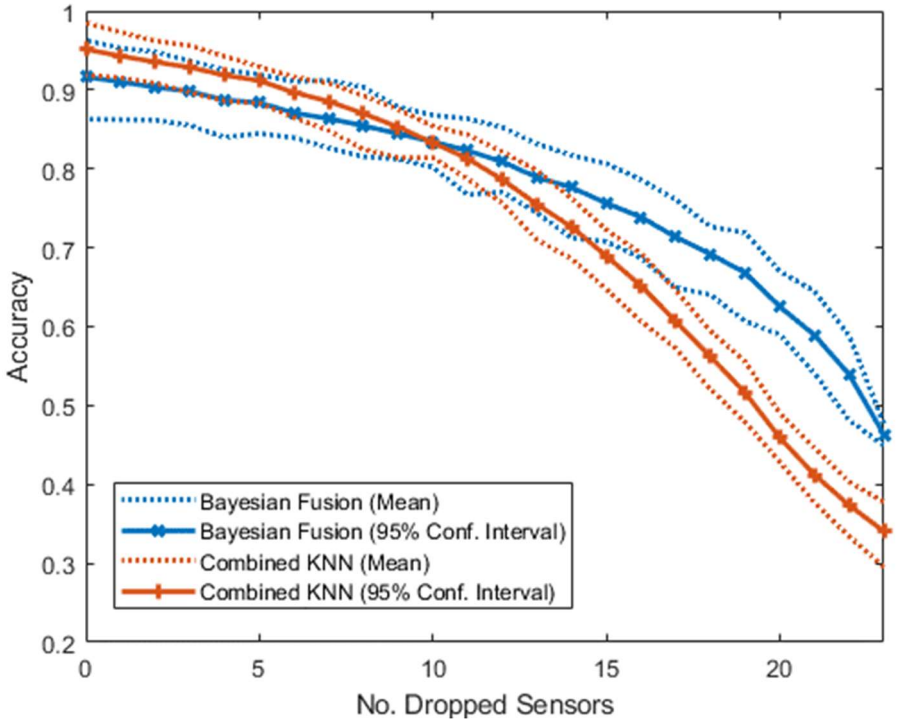

Figure 4.7. *Graph showing the accuracy of the intent-classification system with increasing number of sensors dropping out. The Bayesian Fusion method (treating each sensor separately and then combining them) and the combined method (putting all sensor information into a single KNN classifier) are shown.*

4.4 Discussion

4.4.1 Time-Dependent Classification

The results obtained from the experiment demonstrated a link between the time allowed to collect data and the accuracy of the classification, with the accuracy increasing approximately monotonically over time (see Figure 4.5). This link was shown more strongly for the MM than for the NMM, with a higher Spearman's correlation coefficient for the former than the latter. There is no reason in the algorithms that the time allowed should have a different effect on the MM than the NMM. It is possible that the difference was due to the experiment being performed with all 24 sensors active—in this case, the NMM was expected to perform better than the MM, and was therefore closer to the maximum accuracy achievable by the sensors. As a result, there was less capacity for an increase as more time was allowed. Regardless of the difference, the experiment indicates that intent classification is more difficult earlier on in the activity cycle.

This was an expected result, as the activities all started in approximately the same position, with the differences between them increasing as they progressed. Similar findings were shown in [38], though in a somewhat different context, and when viewed in combination with this study, it is suggested that the monotonic increase in classification accuracy begins even before activity inception (where it is purely predictive) and continues throughout the activity cycle.

4.4.2 Variant Number of Sensors

A clear relationship was also seen between the total intent classification accuracy and the number of sensors used as inputs in the algorithms (see Figure 4.6). It should again be noted that, in this test, the algorithms were trained on each possible combination of sensors, which is not possible in real-world applications where the sensors that may or may not be available are not known in advance.

For both algorithms, the classification accuracy increased with the number of sensors. However, as there is no "sensor drop out" in this scenario, the NMM (Combined KNN algorithm) showed consistently better performance than the MM (Bayesian Fusion algorithm).

Again, this was an expected result that aligns with previous studies [105], as the Combined KNN algorithm was able to exploit relationships between the features from different sensors, whereas the Bayesian Fusion algorithm was limited to only exploiting relationships between features within each modular sensor. The Combined KNN algorithm therefore had access to more information than the Bayesian Fusion.

The ratio between the accuracy of the MM and NMM ranged between 0.8 and 0.9, suggesting that while the modular approach did show lower performance than the non-modular approach with no sensor dropout, this difference was relatively small, and much of the accuracy was still retained.

Both algorithms followed a similar pattern of initially rapid increases in accuracy as the number of sensors increases, with a decreasing slope as the number of sensors becomes

large, approaching an asymptote. This followed the general pattern expected by a PSN with increasing sensor numbers (see Chapter 3).

4.4.3 Simulated Dropout

The case of simulated dropout is the one most directly pertaining to the proposed real-world application, and is the main focus of this study. While the Bayesian Fusion algorithm was not designed to outperform standard combined methods in situations where the specific combination of available sensors is known and trained on, this pre-training will not be possible in dynamic real-world scenarios.

The situation of sensors dropping out reflects the fact that, in a dynamic sensing environment, sensors which were initially available will no longer become available as the user moves away from them. It also applies to the issue of maintenance, where sensors on a user's device may fail over time with use, often requiring regular follow-up appointments [106], which may be costly and unfeasible in developing countries [107].

By treating the individual sensors as modular, the Bayesian Fusion algorithm allowed any combination of sensors to still function together as a system, rather than relying on all of them. In this scenario, therefore, both algorithms were trained on all available sensors, and then increasing numbers of sensors were randomly set to 0. The hypothesis was that the modular Bayesian Fusion approach would be much more robust to this dropout than the Combined KNN approach, which was dependent on all sensors working together and should therefore drop in accuracy more rapidly.

This was supported by the results, which showed that while the NMM was superior with no dropout (as expected from previous studies [108]), as the number of dropped sensors increased, the MM overtook its accuracy and became increasingly superior. Given that a real-world intent sensing system might involve hundreds of different sensors dropping in and out throughout the user's activities of daily living, these results indicate the MM as a more appropriate choice than non-modular alternatives.

Furthermore, an NMM would have to be pre-trained on every possible combination of sensors, which becomes prohibitively complex and computationally expensive with hundreds of sensors in play. Conversely, an MM would only need to train a single classifier for each sensor.

4.4.4 Limitations of the Study

While the algorithms used in this study were designed to still be valid for real-time application, the tests described were performed on data "after the fact." A practical application would require real-time testing, factoring in elements such as processing speed.

This study involved only five participants, each performing three trials. While this was sufficient to demonstrate the general ideas discussed, small datasets with high dimensionality can lead to bias in performance estimates and inaccuracies in classifiers [109]. As such, a larger dataset would allow the study of more effective classifiers, with more accurate confusion matrices and more precise final accuracy measurements.

This study classified intent between three different classes, but in daily life, intent options are far more diverse.

This study used only IMU and EMG sensors as inputs, in order to demonstrate the combination of different sensing environments for networked intent sensing. Many other sensor types would also be valid.

In terms of sensor numbers, this study was limited to 24 sensors. While this is far more than may be found on a typical prosthetic device, compared to a network composed of wearable, smart-phone and smart home sensors across multiple environments including home, travel, work, etc., it is relatively small. Given the pattern established by this study, it is likely that the difference between the algorithms will only increase with larger numbers of sensors dropping out.

This study only concerned itself with the scenario of sensors dropping out, and did not consider the possibility of sensors dropping in. It would have been impossible to include sensors in the NMM that were not originally trained on, as this would result in more input dimensions than the trained classifier allows. The MM, however, would have had no issue with this, so long as confusion matrix entries for the new sensors were provided.

4.4.5 Suggested Future Work

While the data set used for this work was a good starting point, applying the methods established here to a much larger data set (such as [110]) would also be of interest,

allowing the study of more accurate classifiers and more accurate estimations of the confusion matrix entries.

Secondly, the analysis performed here took place offline, with all data recorded in advance. The next stage of the algorithm's development would be to adapt it to run in real-time, perhaps in a scenario similar to [111], which would be much more representative of practical use in a prosthetic device. A real-time scenario for intent sensing is investigated in Chapter 5.

More sensors, particularly from multiple sensing environments, would further improve the algorithm and allow more precise classification into finer, more detailed classes. A future study could combine the sensors used in this study (perhaps built into a wearable device) with smart home sensors and built-in sensors on devices such as smart phones and smart watches to further expand the network and provide a better simulation of the proposed real-world application.

The MM also has the benefit of reducing the dimensionality of the classifiers trained compared to the NMM. The NMM, using the same number of training samples, must train a classifier to distinguish between features from every sensor at once. The MM splits this problem up into R individual classifiers (where R is the number of sensors), each with the number of dimensions reduced by a factor of 1/R. It is possible that this could reduce the amount of training data required to train the needed classifiers—a future study could investigate, verify and quantify this potential advantage. Chapter 7 of this thesis showcases this benefit particularly strongly.

Finally, the maintenance applications of the modular algorithm could be further explored by applying the algorithm to the number and types of sensors found on actual prosthetic devices for current industry-standard classifications, to investigate advantages in continued prosthetic viability as sensors cease to function.

4.5 Conclusion

Firstly, it has been shown that intent classification is easier the later on in the activity cycle it is attempted, and that a high classification accuracy (~96%) can be reached using only the first 1000 ms of data after activity inception for simple tasks.

Adding more sensors has been shown to produce a strong improvement in accuracy regardless of which intent sensing algorithm was used, well beyond the relatively small number of sensors typically used in prosthetic devices. Networking larger numbers of sensors together is supported by this study as a potential method for improving device input detection accuracy, which could not only lead to better performance in devices but also allow the option of more precise, complex input actions that previously have not been detectable with a high-enough accuracy through existing methods.

The proposed modular approach to sensor fusion supports a dynamic intent sensing network more effectively than the comparison combined approach, with higher accuracy under conditions of major sensor dropout and the possibility for “drop-ins”, which are not viable for a non-modular system.

This approach is not only useful in a dynamic context where users move from one sensing environment to the other, but also in maintenance scenarios, to allow devices to

continue to function with a reduced set of sensors where repair is costly or impossible. It could even be possible to use a combined algorithm where all sensors are available, and then switch to a modular algorithm when sensor dropout is detected, maximising accuracy with a "best of both worlds" approach.

In summary, this is an early exploration into the requirements and viability of a dynamic intent sensing system, and it is hoped that subsequent research will push this technology further, towards a future where intent sensing approaches ubiquity in medical devices and beyond.

Statement of Authorship for joint/multi-authored papers for PGR thesis

Title of Paper	Towards Dynamic Multi-Modal Intent Sensing Using Probabilistic Sensor Networks
Publication Status	■ Published □ Accepted for Publication □ Submitted for Publication □ Unpublished and unsubmitted work Written in a manuscript style
Publication Details	J. Russell, J. H. M. Bergmann, and V. H. Nagaraja, "Towards Dynamic Multi-Modal Intent Sensing Using Probabilistic Sensor Networks," Sensors, vol. 22, no. 7, p. 2603, Mar. 2022, doi: 10.3390/s22072603.

Student Confirmation

Student Name	Joseph Russell		
Contribution to the Paper	Performed all investigation, software development, analysis and writing. V. Nagaraja provided access to the data set. J. Bergmann provided supervisory guidance and editing of the manuscript.		
Signature		Date	06/08/2023

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Prof Jeroen Bergmann		
Supervisor comments		
Signature 	Date	06/08/2023

5

Real-Time Intent Sensing for Assistive Devices with Implications for Minimising Maintenance

Published by Joseph Russell, Jeroen Bergmann and Vikranth H. Nagaraja in *Prosthesis* 2023, 5, 453-466, doi: 10.3390/prosthesis5020031

Chapter Abstract

It has been established in the previous chapters that in order to maximise accuracy of intent sensing, particularly in the context of controlling a prosthetic device, as many sensors as possible should be utilised. The availability of sensors may change over time due to changing surroundings or activities, sensors failing, and electrode contact being lost. The sensor network should be dynamic and modular in nature, continuing to function even when some sensors are unavailable. The management of sensor unavailability may help to reduce the need for device maintenance, particularly in developing nations with limited availability of these services. An algorithm is proposed to classify intent

using networked sensors in real-time. Data are gathered using human participants wearing four surface electromyography sensors and performing a pseudo-random sequence of grasps. The relationship between time offset and prediction accuracy is investigated, with the algorithm predicting future intent actions up to half a second in advance. Sensor dropout is simulated by randomly setting sensor readings to zero. The new algorithm is compared to existing algorithms and shown to be more accurate in situations of sensor dropout, with the difference increasing as more sensors become unavailable. This suggests that when reductions in sensing capabilities are likely to occur over time, the modular method is more appropriate for control.

5.1 Introduction

5.1.1 Background – Establishing the Problems

As discussed in the previous chapters, intent sensing is particularly applicable for assistive technologies where it could allow more responsive, intuitive control, reducing cognitive loading and enhancing the user's overall experience. Prosthetic technology provides an ideal use case for this research, as accurate control of the artificial limb is essential for the user's safety and quality of life.

Modern active upper limb prosthetic devices are typically controlled with electromyography (EMG) sensors, detecting electrical activity present during muscle contractions. However, the experience of using this technology is not intuitive; the muscles must be deliberately contracted in order to activate the device in a manner that must be learned, and the number of degrees of freedom is limited [112].

Recent advancements have suggested neural interfaces as a potential solution, involving integration between a prosthetic device and the nervous system to produce more natural control. However, this kind of invasive technique carries an unavoidable risk of infection and tissue damage [113], which limits its potential for widespread uptake. As stated in Chapter 4, if a system could be devised to increase prosthetic functionality and enable natural, intuitive control using non-invasive sensing, then this could be of great benefit.

Prosthetic maintenance is also an issue. Medical Center Orthotics & Prosthetics (MCOP), a leading US prosthetic out-patient clinic, recommends that prosthetic patients return to their local clinic for a maintenance check-up at least twice per year [106]. While this may be possible for some, patients with limited mobility or those who live far from the clinic and do not have access to the necessary funds or transport facilities may find this to be challenging. This is particularly problematic in developing nations, where clinics may be sparsely available and patients may have to travel several hours to reach them [107]. If the need for prosthetic maintenance could be reduced, this could have positive impact on these patients' livelihoods.

Additionally, a common problem with myoelectric prosthetics is electrode lift-off, where movement and a change in limb volume, both throughout the day and over longer periods of time, can result in the quality of electrode contact being reduced to the point where no clear signal can be extracted [114]. A system that is robust to electrode lift-off could improve the prosthetic usability and further reduce need for visits to clinics.

This study attempts to address these problems, proposing an algorithm to non-invasively measure users' intent in real-time in a way that is robust to sensor availability, failure, or electrode lift-off, enabling intuitive control and reducing the need for device maintenance.

5.1.2 Intent as a Solution

As stated in Chapters 2 and 4, intent can be defined as consisting of three aspects: (i) Prediction of Future Activities (ii) Activity Transition Recognition and (iii) Inference of Task Goal [2]. Prediction of Future Actions involves looking ahead to an activity that has not yet begun to take place (e.g. detecting a staircase ahead, which a user may soon be about to climb). Activity Transition Recognition involves identifying the user's current activity and detecting when it changes (e.g. starting walking). Goal Inference is concerned with identifying the objective of the current activity (e.g. the user is walking towards the staircase). An overview of these aspects is shown in Figure 5.1.

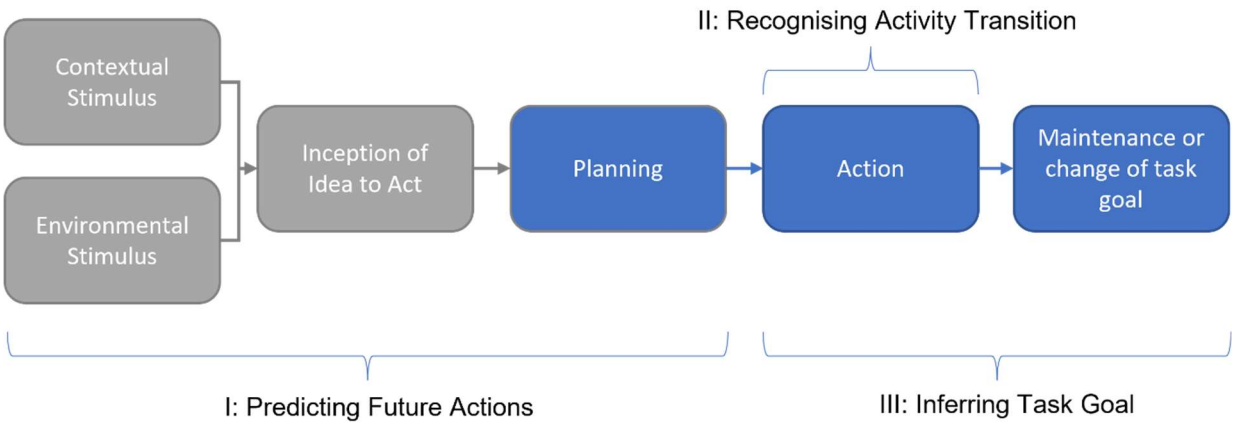

Figure 5.1. *Overview of the phases of intent sensing, adapted from Chapter 2. This study concerns itself with the Planning, Action and Maintenance or change of task goal phases (highlighted in blue).*

The effect of sensor unavailability (through failure, lift-off, etc.) on intent prediction accuracy remains unclear and also directly impacts on the maintenance question. A technology that no longer reaches a certain level of performance needs to be serviced; so, keeping the performance level as high as possible even when components fail will result

in a better experience for the user. This study explores how a new modular algorithm compares to more established methods of intent prediction when sensors drop out.

This study investigates all three aspects of intent sensing, with elements of activity transition recognition, inference of a task's goal, and activity prediction. Analysis was performed in real-time, with the intent estimations continuously updating over the course of the activity. Participants were given advance warning of the actions they were required to perform to allow the possibility of an anticipatory response to occur [115], which was used to predict the action itself. The activity transition was then detected, and any change in task goal was monitored, resulting in a holistic intent sensing system. At the moment, these systems are still rare in the literature (see Chapter 2), and different algorithmic approaches can be used to predict the action. The different options for this are discussed and compared.

5.1.3 Intent Sensing Methods: Modular and Non-Modular

Current multi-sensor control systems for prosthetics generally tend to favour a combined, non-modular method (NMM), where the system is trained on a specific combination of sensors that are available, and therefore, suffers from a serious reduction in accuracy when the sensors become unavailable. The choice of algorithm used can affect the extent to which this occurs: a K-Nearest Neighbours (KNN) classifier, for instance, is generally more robust to sensor dropout than a Decision Tree classifier is. Chapter 4 showed that a modular method (MM) might be the most optimal choice for dealing with sensor dropout, as each sensor is trained individually and makes its own prediction,

which is then combined with the others at the end of the process using Naïve Bayes techniques. Removing a sensor naturally reduces the accuracy of the system, but the reduction is dependent on the contribution of that specific sensor and does not have a wider impact on the classification algorithm. As such, the decrease in accuracy from sensor dropout is expected to be smaller than it is in a combined, NMM system.

However, without dropout, combined NMMs, which use all sensor inputs to train a single classifier, should perform with a higher accuracy than the equivalent modular approach does, as a combined method can exploit the relationships between sensors.

Once sensors start to drop out, the equivalent MM should decrease in accuracy much less rapidly than that of the NMM does, until it becomes the more accurate method, with this difference subsequently increasing as more sensors drop out.

This was initially investigated in Chapter 4, which tested an MM against an NMM retrospectively using laboratory gathered motion data. This study aims to apply an enhanced version of these techniques in a more realistic, real-time scenario.

5.1.4 Intent in Real-Time

Unlike in Chapter 4, the analysis performed in this study took place in a simulated real-time scenario. This means that all techniques used were carefully chosen to ensure that they would work in real-time, and the simulated real-time was controlled to ensure the algorithm was never given access to "future" data before the appropriate time.

Using simulated real-time enabled the data collection to be performed remotely. Running the study fully "online" would have been more ideal, as it would have enabled the

quantification of the processing speed and would have allowed for the possibility of "user-in-the-loop" behaviour being captured. However, both these factors are beyond the scope of this study, which aims to develop algorithms for an intent sensing input and does not extend to testing hardware implementation.

Rather than retrospectively classifying a set of data as one activity or another, the algorithms used in this study classify data continuously; at each time step, they update a set of intent predictions. Each update will include an estimation of the current action intent, predictions for actions in the immediate future, and improvements on the estimations from the immediate past.

It is understood that it is more difficult to predict what the intent will be in the future than it is to estimate what is happening at the present moment and that this in turn is more difficult than estimating what the intent was in the past. Chapter 4 suggested this relationship between the accuracy of intent estimation and time offset should be approximately monotonic, with the accuracy increasing the later on in the activity cycle intent is estimated.

This study will investigate this relationship further, exploring how accuracy increases from the predictive to retrospective time offsets.

5.1.5 Experimental Concept

The proposed real-time Modular Method (MM) was compared to a more standard Non-Modular Method (NMM) in a trial involving human participants. To clearly showcase the potential performance of the proposed algorithms, the experimental scenario was

kept as simple as possible, while still satisfying the conditions for sensing intent and remaining representative of a practical prosthetic device. Binary classification was used, selecting between two possible intent states – the right hand being open or the right hand being closed, which the participant switched between according to a pattern shown to them on a screen. A fine time resolution was used, classifying the open/closed state for each 50ms window, allowing precise measurements of the effect of time on classification accuracy.

The sensing input used was surface electromyography (sEMG), a standard sensor type included in typical prosthetic devices [116]. Up to four of these sensors were used – which is more than a prosthetic device would usually include - to be representative of the best-case scenario as a starting point.

As this study aims to investigate applications of the proposed algorithms to reducing the need for prosthetic maintenance, sensor dropout was simulated to represent a sensor becoming unavailable (from causes such as sensor failure or electrode lift-off). To achieve this, the inputs of sensors were replaced with recorded pure sensor noise when they failed (unlike in Chapter 4, where dropout was simulated by setting the sensor output to 0). This means that the signal contained no information, so could not be used for classification. Established methods exist for detecting sensor signal loss through a variety of causes [117]. The simulated dropout/unavailability was implemented from the start of recording; as for the vast majority of the time, a prosthetic device is in a steady state, with a certain number of sensors remaining active. The event of a sensor

actually transitioning from working to not working is very rapid, with minimal contribution to the sensor's overall accuracy; so, the transition to failing was not considered.

5.2 Methods

5.2.1 Data Collection

The data used here were specifically gathered for this study from seven non-disabled human participants (2 females and 5 males; age 23–60; weight 58–91 kg). The study was performed in accordance with the Helsinki Declaration, and ethical approval was obtained from the Medical Sciences Interdivisional Research Ethics Committee (IDREC) of the University of Oxford (Reference Number: R68585/RE001). All data were anonymised.

The methodology employed here builds on the method first employed in Chapter 3. Each participant wore four sEMG sensors (voltage differential measurement, gain 1009, range ±1.65mV, CMRR 80dB, input impedance 10GOhm) connected to a low-cost BITalino (r)evolution body-sensing toolkit, with a sample rate of 1000Hz [118]. Each sensor module consisted of two gelled self-adhesive disposable Ag/AgCl electrodes (diameter 24mm, thickness 1mm, and coated in conductive hydrogel) to be placed on the muscle, and an additional electrode was placed on the collar bone as a ground reference. The sensors were placed on four major muscle sites on the right arm, following standard recommendations from SENIAM – biceps brachii, and triceps brachii (long head and lateral head) [78], along with the extensor carpi radialis. These are shown in Figure 5.2.

Each participant held a finger exerciser [119], with their forearm flat on a table and their palm facing upwards.

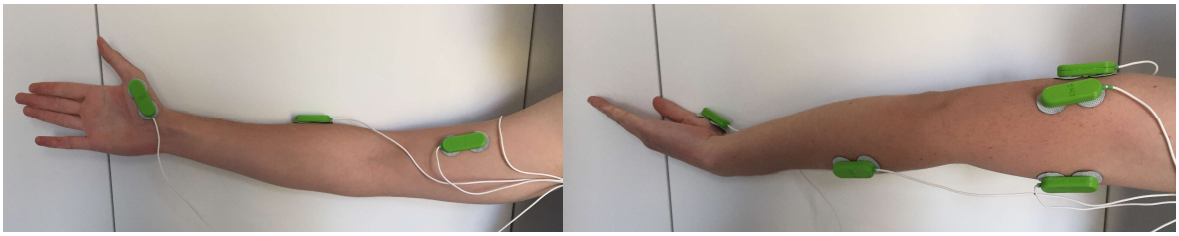

Figure 5.2. *Photographs of the positions of EMG sensors (green boxes) on an anonymised participant. The left image shows the frontal plane of the arm. The right image shows the arm internally rotated to provide a view of the sensors placed on the triceps. In order to be representative of a prosthetic use case, the sensor placed on the hand was not used.*

A pseudo-random sequence of 1s and 0s was generated and presented on a screen in front of the participant as a line graph. This sequence was constrained so that exactly 50% of the sequence was 1 and 50% was 0, with the first 3 entries always being 0. The left edge of the graph was labelled as the "present", with the future actions the participant would be required to take being shown on the right. Every second, the "future" plot would advance by one time step towards the "present". The participant was asked to close their hand when the "present" line intersected with 1 and to open their hand when the line was at 0. An audio cue was also used, with a continuous high-pitched tone sounding when the participant was required to close their hand, and a low-pitched tone when they were required to open it.

The sequence used was saved, along with timestamps, to be used as ground truth when the accuracy of the intent classification was assessed.

In order to standardise the strength of muscle activation, participants were instructed to close their hand with enough force to fully compress the finger exerciser and no more and to maintain this for the duration of the "1".

Each trial lasted three minutes, and each of the seven participants was asked to perform three trials, resulting in twenty-one sets of data. To prevent muscle fatigue, the trials were separated by five minutes of rest time.

Additionally, a thirty-minute recording was made to represent sensor lift-off, in which the sensor on the biceps brachii was rotated 180 degrees, so that the electrodes pointed away from the surface of the skin. This meant that the recording was pure noise, with no underlying EMG information. Trials were repeated during this period with the hands opening and closing as in the ordinary trials, so that motion artifacts could also be included. Sections of this recording were used when sensor drop-out was simulated.

The BITalino was connected through Bluetooth to a personal computer running Open-Signals (r)evolution (v2.1.1, Plux Wireless Biosignals, Lisbon, Portugal), which was used to record EMG data.

5.2.2 Signal Processing

All signal processing was performed in MATLAB (R2022a, Mathworks Inc, Natick, MA, USA). Raw EMG signals were initially filtered with a 10-500 Hz band-pass third-order

Butterworth filter and normalised by the maximum filtered signal recorded during training for each sensor on each participant. [100]

5.2.3 Feature Extraction

To interpret the EMG data, protocols established in [101] were followed to represent the industry standard for EMG feature extraction, selecting key features identified in that study. Data were segmented into windows of 200 ms, each shifted by 50 ms, so that consecutive segments overlapped by 150 ms. From each segment, the features extracted were: Integrated EMG, Mean Absolute Value, Mean Absolute Value Slope, Variance in EMG, Root Mean Square, Waveform Length, Autoregressive Coefficients (to the fourth order), Frequency Median, and Frequency Mean.

As this study was performed in real-time, an intent prediction was made at the end of each segment, i.e., every 50 ms. For each prediction, the features from the previous ten segments (650 ms) were used, resulting in a total of 110 features for each sensor. These were used to ensure that the full pattern of the changing EMG signal could be recognised and did not introduce input lag, as each segment was labelled according to its delay. This means that, once trained, the system weighted the segments' contributions appropriately and was predicted to increase accuracy approximately monotonically over time as an activity transition occurred. Further evidence of this is included in Appendix B.

5.2.4 Real-Time MM Intent Algorithm

The MM algorithm used in this study was adapted from that in Chapter 4. Sensors were treated individually, with no consideration for the relationships between sensors. Training data were used to train an individual KNN classifier for each sensor and to populate a confusion matrix to quantify its accuracy in classifying each possible intent outcome. During testing, each sensor made its own individual prediction of the intent, and these predictions were combined by effectively weighting them according to their confusion matrix entries. This was achieved using Bayes' rule [81]:

$$P(E|V) = \frac{P(V|E) \cdot P(E)}{P(V|E) \cdot P(E) + P(V|E') \cdot P(E')} \tag{5.1}$$

$P(E)$ is the prior probability of a particular intent being true, and $P(E')$ is the prior probability of that intent not being true; in this study, both of these are always set to 0.5. $P(E|V)$ is the probability of that intent being true given the set of sensor values currently being measured. $P(V|E)$ is the probability of measuring the current sensor values given that the intent being considered is true. Assuming that there is probabilistic independence between the individual sensors, this can be approximated as the product of the probabilities of each individual sensor. $P(V|E')$ is the probability of measuring the current sensor values given that the intent being considered is not true.

This equation was used to determine the probability of each possible intent being true, and then the algorithm simply selected the intent with the maximum likelihood given

the data. When a sensor dropped out, its contribution was not included in the equation, and no re-training was required.

5.2.5 Comparison NMM Algorithm

The NMM algorithm used for comparison did not require the learning of a confusion matrix, and it simply required all training data to be fed into a KNN classifier. When sensors dropped out, their entries were replaced with pure sensor noise, but their dimensions cannot be removed from a combined algorithm; so, the accuracy was expected to rapidly decrease.

5.2.6 Data Separation

Leave-one-out cross validation was used, which separated data into a training set of twenty samples and a testing set of one sample. Each sample contained three minutes of data. For the NMM, all data in the training set were used along with the recorded ground truths to train a single KNN classifier, which was then used to classify data in the testing sample.

For the MM, as was the case in Chapter 4, a separate classifier had to be trained for each sensor, and each classifier also required a confusion matrix, indicating its sensitivity and specificity. To ensure these were accurate and non-biased results, the same data could not be used both to train the classifier and populate the confusion matrix. The training set was, therefore, once again subdivided into a classifier training set and a probability learning set.

The classifier training set was used to train a KNN classifier for each sensor. This was then tested on the probability learning set and compared against the ground truth. The results of each classification were tallied and divided by the number of samples to produce the entries for the confusion matrix for each classifier.

The confusion matrix entries then gave the weighting for the contribution of each sensor to the Naïve Bayes sensor fusion algorithm, combining together the individual predictions from each sensor to produce a final prediction for the MM algorithm.

The two algorithms were then used to predict the intent for each segment in the testing set, with the number of successful classifications divided by the number of segments to give the accuracy. This is the metric that has been plotted in the graphs in the Results Section.

5.2.7 Time Offset

In order to measure the effect of changing time offset on classification accuracy, the recorded ground truth was shifted in increments (T) of 100 ms from −500 to +500. This had the effect of training the system to use each set of ten segments to predict the user's intent, T segments in the future for negative values and in the past for positive values.

The classification accuracy of samples in the testing set was measured for each value of T using both algorithms, with the results plotted. A Spearman's rank correlation coefficient was calculated (again using Equation 4.3) in order to quantify how well the trend may be described as monotonic. This value was used to determine to what extent the following hypothesis is true: that accuracy will increase with greater values of T [104].

5.2.8 Sensor Dropout

A varying number of sensors, n, in the testing set were then replaced with noise from the pure noise recording to simulate sensor dropout. Both algorithms were trained on the full set of four sensors, but the MM was able to simply not include contributions from dropped sensors in its classification system. The NMM, on the other hand, cannot have dimensions removed; so instead, continued to use these entries, which had been replaced with pure noise.

The number of dropped sensors, n, varied from 0 to 3, with classification accuracy recorded at all values of T. This was repeated for every possible combination of dropped sensors, and the mean accuracy was taken and plotted for both the MM and NMM algorithms for all T values. First, the lines of accuracy vs time offset for values of n were overlaid on two graphs for the MM and NMM, respectively. No confidence intervals were plotted to provide clarity. Subsequently, individual plots were created, comparing the accuracy vs time offset for the NMM and MM, with one plot for each value of n. For these graphs, the 95% confidence interval for the MM is shown, illustrating the variation in accuracy resulting from the choice of subdivision between the classifier training and accuracy learning sets. No interval is shown for the NMM, as there was no subdivision within its training set, and so, the results are entirely reproducible. The performances for both algorithms were then compared.

Following this, the time offset, T, was fixed at 0 ms to represent a device predicting the current intent at any given moment. The number of dropped sensors, n, varied from 0

to 3 with all possible combinations for both algorithms. Again, the 95% confidence interval was also plotted for the MM, with no confidence interval plotted for the NMM due to the entirely reproducible results.

5.3 Results

The results of varying the time offset using the MM and NMM are shown in Figure 5.3. The maximum accuracies found were 0.91 for NMM and 0.86 for MM (both at zero dropouts), at which point, the gradient had become small, suggesting that these are approximately the maximum accuracies achievable in each algorithm with a positive time offset.

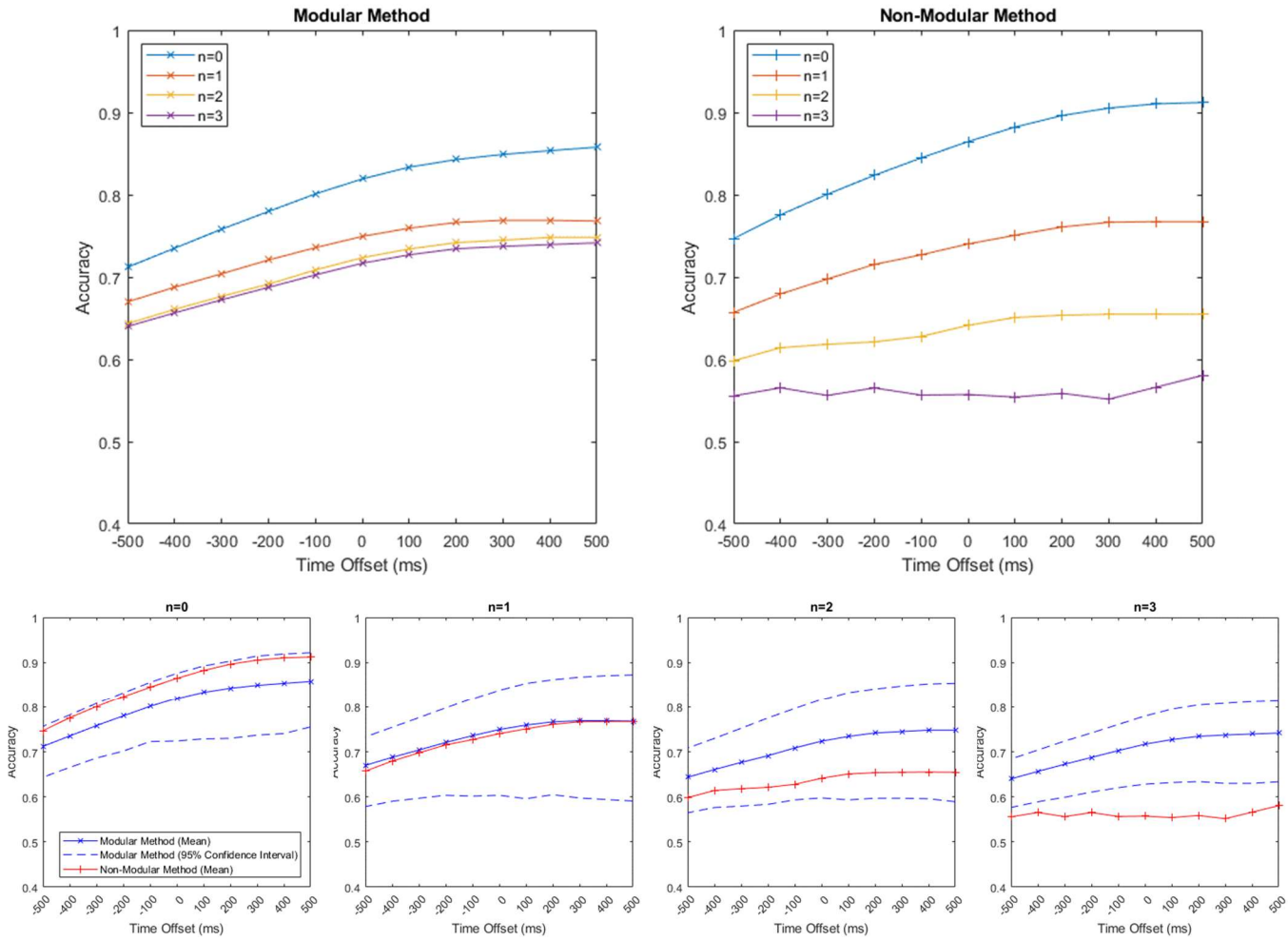

Figure 5.3. *Graphs to show accuracy against the time offset, which is defined as the difference between the time the prediction is made and the time that the intent is being predicted for, which was measured in milliseconds. Results for the Modular Method (MM) are shown on the top left, and results for the Non-Modular Method (NMM) are shown on the top right. Lines from n = 0 to n = 3 are shown, where n is the number of dropped sensors. On the bottom, overlaid plots for both the MM (blue) and NMM (red) are shown, with separated values from n = 0 to n = 3. The 95% confidence interval for the MM is included here (the NMM does not have a confidence interval, as its performance is entirely reproducible). Negative time offsets represent predictions about what the intent will be in the future. Positive time offsets represent historical estimations of past intents. An offset of 0 is a prediction of the current intent at the time of prediction.*

Spearman's rank correlation coefficient was calculated for each line in order to quantify the extent to which the relationship between accuracy and time offset was monotonic. This is shown in Table 5.1.

MM	
N	r_s
0	1
1	0.96
2	1
3	1

NMM	
N	r_s
0	1
1	0.99
2	0.97
3	0.27

Table 5.1. *Spearman's rank correlation coefficients (r_s) calculated for accuracy vs time offset from n=0 to n=3 using the Modular Method (MM) and the Non-Modular Method (NMM). An r_s of 1 is a completely monotonic increase, 0 is no monotonic increase or decrease, and -1 represents a perfect monotonic decrease.*

The resulting accuracy of the MM and the NM, with a time offset fixed at 0 ms, with varying numbers of dropped sensors ranging from $n = 0$ to $n = 3$, is shown in Figure 5.4. Both algorithms were trained on the full set of four sensors. Both plots showed a Spearman's rank correlation coefficient (r_s) of −1, which indicates a completely monotonic decrease. The mean gradient of the NMM was −0.10/sensor, and for the MM, it was −0.03/sensor.

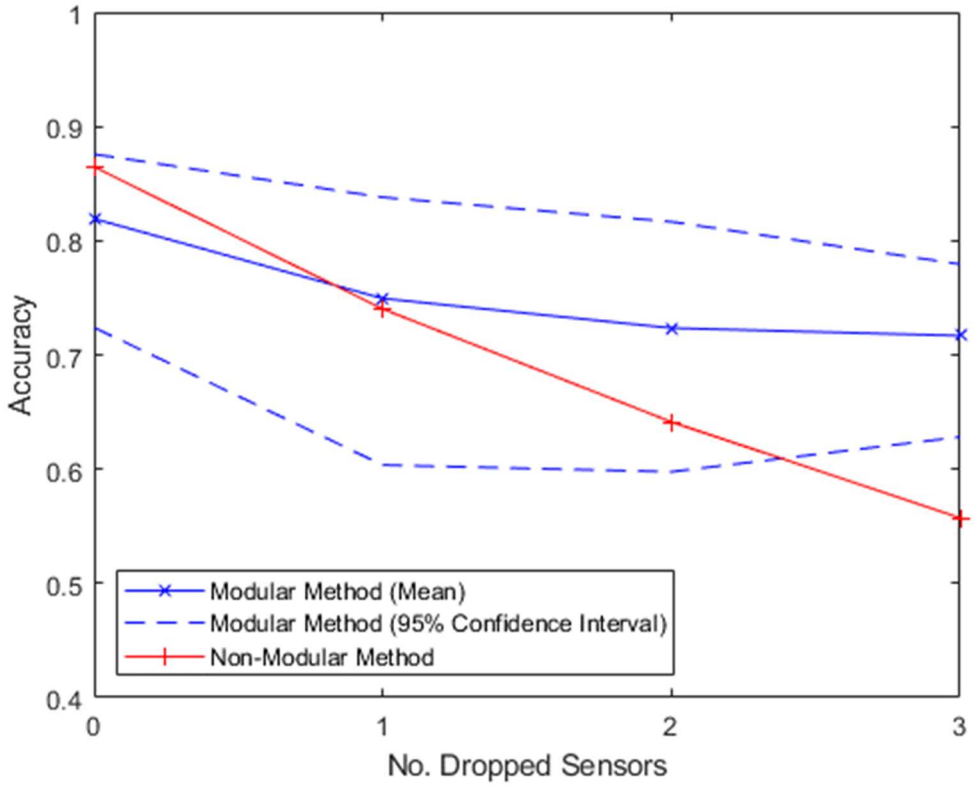

Figure 5.4. *Graph to show accuracy vs the number of dropped sensors with a fixed time offset of 0ms. All algorithms were trained on all four sensors.*

5.4 Discussion

5.4.1 Time Offset

Varying the time offset resulted in a clear monotonic relationship between accuracy and time offset. This demonstrated that the classification of intent is easier when it is conducted retrospectively, and it becomes more difficult to accurately perform it in the earlier phases of the activity cycle. Predicting intent before the activity has begun is particularly difficult, and the MM and NMM could only initially achieve accuracies of 71% and 75%, respectively, compared to the maximum accuracies of 86% and 91%,

respectively, 500 ms after onset. This difference may be somewhat explained by the nature of the sensing apparatus. EMG sensors operate by detecting the electrical activity in muscles, which can only be used to predict future activity through the anticipatory response: the slight pre-tensing of muscles just before an activity the subject knows they are about to perform [115]. This is not a reliable response, and it only occurs in a short window of time before the activity begins, resulting in low accuracy far in advance of the activity.

This might be improved in a future system by combining sEMG sensors with another sensing modality that is more effective in the predictive region, such as electroencephalography (EEG), which could provide additional information in the planning phase by detecting activity in the brain. Contextual factors such as the time of day could also be employed in a real-world scenario, or environmental stimuli may be detected, which may trigger the idea to act. It has been shown that even small changes in the environment can lead to clear differences in motor control [10]. The nature of the experiment performed in this study did not consider these, but future studies performed in situ in a participant's home, rather than under control conditions in a lab, may be able to incorporate factors of this nature.

The monotonic relationship between accuracy and time offset has implications for practical intent sensing systems. At a basic level, it highlights a design trade-off between accuracy and responsiveness in any intent-controlled device. Utilising a predictive classifier could totally eliminate the lag between a user attempting an activity and an assistive device engaging, which is used to aid them. However, this results in a lower

accuracy of intent prediction than that which might otherwise be achieved by waiting slightly longer.

Different applications could benefit from different choices of accuracy versus delay. Systems that require a rapid response, but have low consequences for incorrect classification, might benefit from a "lower" time offset, thereby accepting the lower accuracy level. Systems requiring a higher accuracy level might be better suited to a "higher" time offset, thereby accepting the increased lag to achieve a desired outcome.

This is something which could be dynamically shifted during device operation. If the accuracy has been reduced due to sensor degradation, failure, or electrode lift-off, increasing the time offset could help alleviate the reduction. However, the resulting increase in lag would have to be carefully considered, particularly in a human-in-the-loop control system, where the permissible lag may be strictly limited. In prosthetic devices in particular, other studies found that an increase in delay reduced the control effectiveness, with a delay of 100–125 ms having been shown to be optimal [120].

A holistic intent sensing system covering all three aspects of intent could utilise multiple classifiers with different time offsets in parallel. This would enable the device to initially make a prediction as to what the intent might be in the future, allowing it to perform any necessary preparations ahead of time. As the activity onset occurs, this intent prediction could be updated with a new estimation of the current intent, which the monotonic relationship indicates would be more accurate than the prior prediction would be.

As the action then continues, the intent could again be updated using an "after the fact" classifier, thereby recording the intent with an even higher accuracy.

A continuously updating intent detection system of this nature could achieve high levels of responsiveness and accuracy, anticipating and preparing for upcoming actions, and then responding accurately to them once they occur. The "after the fact" classifier could even be used to gradually train and improve the predictive classifier, adjusting and tuning it to the specific user. This would be an interesting area to explore in a future study.

5.4.2 Sensor Dropout

The hypothesis was also shown to be true that while the NMM initially had a higher mean accuracy in the condition of no sensor dropout, the MM began to display a better mean performance as more than one sensor began to drop out. This occurred for all values of time offset, as observed in Figures 5.3 and 5.4.

With $n = 1$ sensor dropping out, the mean accuracy of the MM was similar to that of the NMM. At all the time offsets, the MM performed slightly better than the NMM did, with this difference decreasing as the time offset increased. However, the NMM's accuracy remained entirely within the 95% confidence interval for all time offsets at $n = 1$, indicating that the difference is small and the methods are approximately comparable.

For $n = 2$, the mean accuracy of the NMM is lower than that of the MM (while it is still within the 95% confidence interval). For $n = 3$, the NMM accuracy is well below even

the lower bound of the 95% confidence interval of the MM, showing a strong difference in accuracy between the methods.

In Figure 5.4, the two methods are compared at a time offset of 0 ms, which is representative of the common method of predicting current intent using all prior information. This clearly showcases the accuracy of the NMM decreasing much more quickly than that of the MM as the sensors drop out, with the NMM decreasing at a rate of −0.10/sensor compared to the MM's rate of −0.03/sensor.

These findings indicate that standard NMM techniques, such as those commonly used in existing commercially available systems, deteriorate rapidly when sensors start to become unavailable, justifying the need for more frequent maintenance check-ups.

Conversely, the MM technique experiences much smaller accuracy losses when sensors start to become unavailable. Even with three sensors dropped out, 72% classification accuracy was retained at 0 ms time offset, while the NMM's accuracy was essentially random, supporting the MM as a viable technique for extending the prosthetic's functional lifetime beyond the current standards.

In regard to the trade-off between accuracy and lag described in the previous section, if an increase in time offset is used to retain accuracy in the event of sensor dropout, a smaller increase is required in the MM than in the NMM. For instance, with $n = 2$ sensors dropping out, if a 65% accuracy is required, the MM can achieve this with a time offset of −400 ms, whereas the NMM can only reach this accuracy level with a time offset

of 100 ms. The human reaction time is typically in the region of 150–350 ms [121]; so, this difference is comparatively large.

In practice, it is possible to combine these two techniques into an Adaptive Algorithm, which would utilise the NMM when all sensors are available and switch to the MM when a pre-determined number of sensors have dropped out. This could be employed in an assistive device so that when the device is new and fully functional, the NMM is utilised, and then, as sensors start to drop out, when the predicted accuracy of the MM exceeds that of the NMM, the Adaptive Algorithm switches to use the MM. This "best of both worlds" approach could extend the lifespan of a prosthetic device without compromising the system accuracy when all sensors are active, and it may prove to be very effective for practical prosthetic control.

5.4.3 Limitations of the Study and Recommendations for Future Work

This study took place under carefully controlled conditions, with all participants performing exactly the same, simple actions: opening and closing their hand. This was useful to clearly illustrate the performance differences between the MM and the NMM algorithms. However, it is not representative of the real-life, day-to-day use of an assistive medical device such as an active prosthesis. A future study might use data gathered over a number of days in participants' homes to be more representative of the algorithms' practical performances.

The classifiers used in this study labelled data in only two classes: open or closed. In reality, there are many more possible intents a person might have during their daily activities, and so, the classification problem is much more complex. Dealing with this will require more sensors and more training data. Future work could explore the classification of a larger number of possible intents using a larger data set. Chapters 6 and 7 investigate this to some extent, using a larger number of participants to classify three different, more complex possible intents.

In order to be representative of the sensors available on a prosthetic device, only EMG sensors were used in this study. However, a wider range of sensors, including Inertial Measurement Units and Smart Home sensors, might be available to users in practice, especially as sensing technologies improve in the coming years. Future work on a holistic intent sensing system should utilise a wider variety of sensors and sensing environments to improve the performance.

The prediction of future actions in this study only extended back to 500ms before activity onset. While the wider understanding of intent as introduced in Chapter 2 began with environmental and contextual factors and continued through the inception of the idea to act, the short time window used here only allowed for some detection of the planning phase, attempting to pick up the anticipatory response in the muscle activity immediately prior to the moment of activity onset. To incorporate environmental and contextual factors, future studies would have to collect data in a more unconstrained, "natural" setting to understand their normal daily routines. Detection of the inception of the idea to act would require a wider variety of sensing inputs such as EEG, as EMG

alone is insufficient to detect this. Incorporating these factors into future studies would make the necessary data collection much more challenging, but would provide a more cohesive measure of intent.

All processing in this study was performed on a PC. If it were to be performed in real-time on a real medical device, other factors, such as processor speed, battery consumption, and weight, should be considered. Studies exploring the application of these techniques to real medical devices should take these design limitations into account.

Finally, while the K-Nearest Neighbours approach taken for classification in this study is appropriate for the size and feature richness of the data set, more advanced techniques, such as Convolutional Neural Networks (CNNs), exist, which could use deep learning techniques to automatically extract features and take advantage of emerging properties. The MM algorithm could utilise these CNN classifiers within its sensor modules and still combine them together using a maximum likelihood method, such as the one performed in this chapter. Future studies should explore the performance improvement resulting from the use of state-of-the-art techniques, which would more precisely determine the maximum accuracy achievable in an intent sensing device. Chapters 6 and 7 of this thesis will go on to apply deep-learning neural networks in this way.

5.5 Conclusion

This study has contributed a novel method for combining sensors in a modular fashion to produce a continually updating real-time prediction of intent. This method has been shown to be robust to sensor dropout, a key requirement of the proposed ideal holistic

intent sensing system. Newly gathered data were used to verify the theory that accuracy of intent prediction increases approximately monotonically with time after activity inception.

The results obtained support the MM over the NMM in all measured performance metrics under the condition of sensor dropout. In the context of maintenance, this shows potential for increases in the lifespan of prosthetic devices, maintaining their useability even when one or more of their sensors has become unavailable.

This could have major benefits for those with limited access to clinics, particularly in developing nations, and for those that fund their medical devices privately, for whom extending the prosthetic's lifespan could result in large financial savings.

There are also environmental implications for a device having an increased lifespan. This partly applies to prosthetic devices, but the intent sensing principles discussed in this chapter could also be applied to a wide variety of applications, from healthcare to gaming. Increasing the lifespan of all these devices could reduce wastage and landfill use.

Future studies should look to apply these real-time intent sensing principles in more practical environments, where different sensor modes are available. A particular opportunity for accuracy improvements lies in networking with smart home technology, for which the MM algorithm can be used to allow sensors to "drop in/out" as and when they become available.

The ideal intent sensing system should take advantage of every available sensor at any given time, as with enough training data, the algorithm used should weigh the sensors optimally, so that adding any sensor, however small its information contribution is, should only ever improve the performance of the system.

Future studies should also look to test these algorithms in a useability study with a "user-in-the-loop" scenario to attempt to quantify the benefits of intent sensing as a control input.

It is believed that, in time, as these techniques are developed, intent sensing systems of this nature will grow to become commonplace in modern technology, with benefits for both the users and designers, as well as the environment.

Statement of Authorship for joint/multi-authored papers for PGR thesis

Title of Paper	Real-Time Intent Sensing for Assistive Devices with Implications for Minimising Maintenance
Publication Status	■ Published □ Accepted for Publication □ Submitted for Publication □ Unpublished and unsubmitted work Written in a manuscript style
Publication Details	J. Russell and J. H. M. Bergmann, "Real-Time Intent Sensing for Assistive Devices with Implications for Minimising Maintenance," Prosthesis, vol. 5, no. 2, pp. 453–466, May 2023, doi: 10.3390/prosthesis5020031.

Student Confirmation

Student Name	Joseph Russell		
Contribution to the Paper	Performed all investigation, data collection, software development, analysis and writing. J. Bergmann provided supervisory guidance and editing of the manuscript.		
Signature		Date	06/08/2023

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Prof Jeroen Bergmann			
Supervisor comments			
Signature		Date	06/08/2023

6

Developing a Deep-Learning Approach for Intent Sensing Using Inclusive Design

Submitted by Joseph Russell, Jemma Inches, Camille Carroll and Jeroen Bergmann to *Nature Scientific Reports*. Currently undergoing peer review.

Dataset published by Joseph Russell, Jemma Inches, Camille Carroll and Jeroen Bergmann on *Dryad* 2023, doi: 10.5061/dryad.fbg79cp1d

Chapter Abstract

Intent sensing has potential applications throughout human-machine interaction, bringing possible benefits to both disabled and non-disabled users. Developing intent sensing systems that are designed for inclusivity requires a new set of algorithms, for which deep learning classifiers are an obvious progression from the K-Nearest Neighbours classifiers used in Chapters 4 and 5. Such a system should continuously update in real-time, constantly adjusting whilst improving its predictions as new data becomes

available. For a system to continuously update over long periods of time, large amounts of training data are normally needed. A new, time-series voting method is proposed that combines intent measurements taken over time to generate a single intent prediction, with accuracy increasing monotonically as time passes. This is tested on real data from patients with Parkinson's disease, as well as healthy controls, who wear a simple wrist-based inertial sensor. The results showed an increase in intent classification accuracy in comparison with other deep-learning methods, reaching ~97% overall (96% in the patient group and 99% in the control group). Time-segmented methods also showed the desired monotonic increase in accuracy over time, supporting time-segmented deep learning as an effective method for intent classification.

6.1 Introduction

6.1.1 Intent

As discussed in the previous chapters, intent sensing has potential applications in a wide range of fields across human-machine interactions, from prosthetics to virtual reality gaming. Here it is being considered for the development of assistive technology within the home.

The previously identified three aspects that make up intent all constitute important parts that need to be measured by an intent sensing system. A true, holistic intent sensing system should cover all three – it should make some attempt to predict a future action before it starts, recognise the transition, and continue improving in accuracy as it estimates the task goal. This will reduce the input lag of any device that is controlled with this method. It is also an important requirement for a human-in-the-loop system and allows for the preparation of upcoming activities in advance. No such a system has yet been reported within the scientific literature.

This study aims to include all three aspects, beginning with a prediction immediately after inception of the idea to act and refining this prediction through activity onset and monitoring the task goal. Intent sensing in general, while a growing field, is still a relatively new idea within the sensing community (see Chapter 2). The development of a holistic intent sensing system represents a novel contribution to the field.

6.1.2 Inclusive Design

Assistive technology driven by intent sensing could be useful to many people, including users with a range of (dis)abilities. The concept of inclusive design refers to technology developed to be applicable to many groups, rather than one specific target group. [122] As well as the wider social benefits of not excluding any particular user group, inclusively designed technology can have significant savings in development costs by requiring only one product design to serve many users, in addition to having an inherently larger customer base and potential uptake. [123]

In the previous chapters, only data from non-disabled participants was used. This chapter uses a training set consisting of both disabled and non-disabled participants. By doing so, this study will develop an intent sensing system that works with high accuracy for both. This will highlight the wide applicability of intent sensing, and showcase its potential for use in inclusively-designed products in the future.

6.1.3 Parkinson's Disease

An appropriate example of a group for which inclusive design would be of benefit is for those with Parkinson's disease. This is a debilitating neurological condition affecting millions of people, particularly the elderly, across the world, with over 90,000 new cases diagnosed every year in the US alone [124]. It is characterised by a steadily increasing loss of motor control [125], with few treatment options available [126]. Assistive technology is being developed to help extend patients' mobility and independence, with

research taking place into tremor-suppressing wearables [127] and exoskeleton technology [128] to achieve this.

Parkinson's disease does not initially affect the cognitive functions of patients in a major way [129]. There could be a loss of fine motor skills while still having a "clear idea of what [to] do" [130]. As such, the user's intent does not necessarily change – merely their ability to act on it physically. The development of intent sensing technology could therefore prove invaluable in allowing patients to express their intention, so that assistive technology can accurately help them perform their desired actions.

Designing this technology inclusively will allow it to be used by a large, general user base, while also improving quality of life for those with Parkinson's disease and other conditions.

6.1.4 Smart Watches

Smart watch devices are becoming increasingly available, with over 100 million Apple Watch users and over 111 million Fitbit users recorded by 2021 [131,132]. They typically perform many functions including activity monitoring and the acquisition of health data. Typical sensing modalities available include a microphone and an Inertial Measurement Unit (IMU), which provides kinematic data usually used to quantify the user's level of motion.

Smart watches usually have the capacity to download apps, which can be developed by third parties. Apps are already frequently used to support patient health in the home, with the Blatchford Linx [133] integrated prosthetic leg interacting with a smart phone

app, and the StackCare system [134] being used to monitor elderly users' movements and to notify family members in the event of an accident.

It would therefore be possible to develop an intent sensing application, which could interface with assistive healthcare devices to support patients with Parkinson's disease in their activities of daily living (ADLs).

6.1.5 Deep Learning

Deep learning is a tool used within the field of artificial intelligence that has grown rapidly in recent decades [135]. Its applications are vast, and while simpler techniques can sometimes be more effective, when supplied with enough training data, deep learning can out-perform most other methods in classification tasks [136].

One of the key benefits of deep learning is the automatic extraction of features from raw data, without requiring a developer to manually indicate these. [137] In fact, it is able to determine emerging features and correlations within data that a human might have great difficulty in detecting, and therefore has great potential in a task as challenging as predicting human intent before activity actually commences.

Deep learning has not previously been applied to intent sensing in existing literature. In Chapter 5, promising results were obtained in simulated real-time using a K-Nearest Neighbours classifier. The use of deep learning is hypothesised to produce higher accuracy results across a larger set of possible intent options, with a reduced number of sensor inputs.

However, a requirement of the intent sensing system is that it should make an initial prediction of intent, and continue improving this prediction over time. In a standard deep learning system, utilising time-series data over longer periods means more training data is required to learn an effective model. For a system to continuously update over a long period of time, extremely large amounts of training data are needed using conventional methods.

In order to overcome this, a new, time-series voting method is proposed, in which the data is segmented into time windows of a fixed length. Each window is used to make an individual deep-learning-based prediction, and these are combined using weighted voting algorithms, with more votes being included as time passes. This limits the complexity of the learned model, and is hypothesised to be effective with a much more feasibly-sized data set.

6.1.6 Objective

This study aims to develop a deep-learning-based algorithm for inclusive intent sensing, where accuracy of prediction increases over time.

To achieve this, this study will compare different algorithms, including a new algorithm that segments time-series data into fixed-size time windows, each of which provides a weighted vote towards a continuously updating overall prediction. An algorithm will be considered effective as an intent sensing system only if a) it predicts intent with high accuracy, b) the accuracy continues to increase as activity begins and continues and c) both prior objectives are achieved for both the control and the patient group.

6.2 Methods

6.2.1 Data Collection

Data used in this study was obtained from 34 volunteer participants, 15 of whom had been diagnosed with Parkinson's disease and 19 of whom had not. Demographic information on both the control and patient groups is shown in Table 6.1, and disease progression information for the patient group is shown in Table 6.2, including the original 1987 Unified Parkinson's Disease Rating Scale (UPDRS) [16] and the Hohen and Yahr Stage [17].

	Patients	Control	Total
Number of Participants	15	19	34
Age	67 ± 9	64 ± 10	65 ± 9
Sex	10 Male, 5 Female	11 Male, 8 Female	21 Male, 13 Female

Table 6.1. Number of participants, age (mean and standard deviation) and sex for the patient and control groups.

Disease Duration (Years)	5 ± 3
UPDRS	44 ± 19
Hohen and Yahr Stage	2 ± 0.5

Table 6.2. Disease progression information for the patient group, including duration in years since diagnosis, Unified Parkinson's Disease Rating Scale (UPDRS) and Hohen and Yahr Stage. All metrics include mean and standard deviation.

All volunteers signed an informed consent form and ethical approval for the study was obtained from the NRES Committee South West (REC reference 13/SW/0287).

The protocol consisted of initially asking participants to perform a calibration pose, standing up straight with their arms relaxed by their sides, followed by three different ADLs based on those utilised in the Motor Activity Log, as tested in previous studies [18,19] – opening and closing a door, buttoning and unbuttoning a cardigan, and making toast. Each activity was repeated three times without pausing within a trial. Research nurses supervised the activities throughout, providing verbal guidance where required.

During the activities, the participant was engaged in conversation by the supervising research nurses, but were asked not to talk about the activity they were performing. This engagement was aimed at making the motor behaviour more natural and to better represent ADLs in which cognitive loading is increased due to the application of multitasking.

Each subject wore a three-axis nine-channel IMU (MTx, Xsens Technologies B. V., Enschede, Netherlands) secured to their right wrist to represent a smart watch. This provided accelerometer, gyroscope and magnetometer data, along with a nine-entry rotation matrix (generated by Xsens software) defining the IMU's orientation, giving an 18-

row time series data stream with a sample rate of 50 Hz. The trials were also recorded using video and audio, which were used to further check compliance to the protocol and identify any issues with the sensors (e.g. sensor loss).

To prevent errors with the IMUs due to magnetometer interference, the layout of the experiment room was kept the same throughout the study, with the only sources of local magnetic interference being the data collection devices themselves and the laptop used to receive the recorded data.

6.2.2 Data Processing

All data processing was performed using MATLAB (R2022b, Mathworks Inc, Natick, MA, USA).

There was some variation in the directions participants faced while performing the trials. The study was intended to represent typical use in the home, and so it was important to ensure the starting direction the subject was facing was not used as a learned feature. Additionally, the orientation of the IMU on each patient's wrist was representative of real-world placements and therefore was not always consistent – pre-processing was implemented to correct and standardise the obtained data.

First, the average rotation matrix of the IMU during calibration was calculated for each subject. There is no single standard method for taking an average of a three-dimensional rotation matrix – the method used here involved converting the matrix to a quaternion, finding the mean of each of the quaternion's four entries throughout calibration, re-normalising the quaternion, and then converting it back to a rotation matrix. This

method is valid, so long as the variation in rotation is small – which is true for calibration data from a user standing still.

Each three-dimensional vector from the accelerometer, magnetometer and gyroscope throughout the trials, along with the 9-entry orientation rotation matrix, was then pre-multiplied by the inverse of the calibration rotation matrix, standardising the data by ensuring the Z-axis pointed vertically downwards, and compensating for differences in starting orientation of the IMU on the wrist.

Then, to remove any information about the initial direction the user was facing for each trial, a random rotation about the vertical axis was applied to all data points for each trial by way of pre-multiplying by another randomised rotation matrix. This ensured that the features learned were agnostic towards the user's starting direction, while retaining information about any change in rotation during the activity.

6.2.3 Time Segmentation

Each trial was then segmented into time windows of width 500ms, with each shifted by 250ms, such that a 250ms overlap existed between each consecutive window (this allowed patterns of length of up to 250ms to be captured regardless of temporal alignment). The selected 250ms is approximately equal to typical human reaction time [138], and as such was considered appropriate to capture a large portion of any intent response after inception of the idea to act and before activity actually begins.

Each time window was treated as a separate 18-row time-series data sample, to be fed into the deep learning classifier in training and in testing.

6.2.4 Data Separation

Data was separated into training and testing sets according to leave-one-out cross-validation. This was done by participant, with all time windows associated with all nine of each participant's trials being held back or included at once.

6.2.5 Deep Learning

As the data in each time window is time-series, and contains information based on the signal patterns, it is essential that the classifier used is also appropriate for time-series data. A well-established time-series deep learning method is the bidirectional Long Short-Term Memory (LSTM) Neural Network [139–141], which is provided in the MATLAB Deep Learning Toolkit. The 18 IMU features from the data were inputted into an LSTM network, with network architecture shown in Figure 6.1.

Figure 6.1. *Deep learning architecture used to classify ADLs in the study.*

The samples in the training set, along with the corresponding ground truths, were used to train a single LSTM network based on all activities and time windows in the set. 15 hidden units were used, over 100 epochs, with a learn rate of 0.001 and a mini-batch size of 512. Mini-batches were shuffled every epoch.

A customised loss function was used in the classification layer, defined as the mean of a) the overall classification error and b) the maximum error across all three classes. This meant that with each iteration, the LSTM network would learn to increase overall accuracy, but only if this did not result in a larger decrease in the ability of the network to classify any of the classes. This prevented the network from "sacrificing" accuracy in one class in order to more accurately classify the other two.

The data from the participants in the testing set was then fed into the trained LSTM network. For each trial, each time window was classified, in chronological order. This resulted in many, often differing predictions of intent for each trial – one for each time window.

6.2.6 Time-Series Voting

The individual predictions for each time window were then combined. This was performed using three different approaches, which were compared: 1) majority voting, 2) constant weighted voting and 3) time-variant weighted voting. In all three methods, the overall prediction at each time step was based on the individual predictions of all previous time steps. For the majority voting method, the overall prediction was simply the intent option with the most votes.

For the weighted voting methods, a similar technique was used to the Modular Method employed in Chapters 4 and 5, only in this case the "modules" were not separate sensors, but instead, readings from the same sensor at different times. In this way, the training set was subdivided in half into two smaller, randomised subsets. As both the classifier

and its confusion matrix were assumed to be equally important, these subsets were chosen to be the same size. The effect of varying the ratio of this subdivision was also investigated in Appendix C.

The first of these subsets was the classifier training set. This was used to train the LSTM classifier. The second was the probability learning set. This was used to produce a confusion matrix for the learned classifier, by predicting the intent of all trials in the probability learning set and recording the classifications and misclassifications of intended activities.

For the constant weighted voting method, only one confusion matrix was learned, by pooling and classifying all data from all time windows into one large set, before subdividing. For the time-variant weighted voting method, a separate confusion matrix was learned for each time step (i.e. one confusion matrix was learned to represent the first 500ms of each trial, another was learned for the second 500ms, etc.)

The predictions from each time step up to the given time were then combined according to Bayes' rule [81]:

$$P(E|V) = \frac{P(V|E) \cdot P(E)}{P(V|E) \cdot P(E) + P(V|E') \cdot P(E')} \tag{6.1}$$

$P(E)$ is the prior probability of a particular intent being true, and $P(E')$ is the prior probability of that intent not being true – in this study, the prior was assumed to be uniform, making $P(E) = 1/3$ and $P(E') = 2/3$. $P(E|V)$ is the probability of that intent being true given the set of sensor values currently being measured – this is the objective to be found. $P(V|E)$ is the probability of measuring the current sensor values

given that the intent being considered is true. Once again assuming probabilistic independence between the individual sensors, this can be approximated as the product of the probabilities of each individual sensor. $P(V|E')$ is the probability of measuring the current sensor values given that the intent being considered is not true.

This method used the probabilities from the confusion matrices to effectively weight each prediction according to its "confidence." For the constant weighted method, the same confusion matrix was used every time, whereas for the time-variant method, a unique confusion matrix was used for each time step.

This was performed for each time step across each trial, so that each overall prediction was based on one more vote than the previous prediction. This was hypothesised to result in accuracy of overall prediction increasing approximately monotonically over time as the activity continues.

As the majority voting method was not subdivided, this meant its classifier was trained on the full data set, rather than only half as in the two weighted methods. This was expected to give the method an accuracy advantage. To determine how much of the difference between the methods' accuracies was due to this effect, the majority voting method was repeated using only half its data set, to make it comparable to the weighted methods.

6.2.7 Comparison Methods

To investigate the effect of segmenting and re-combining time-series data in this way, several alternative approaches were tested for comparison.

The first method, Comparison A, used the same network as before, trained on the individual time windows. However, at each time step of the test set, only the most recent time window was used as an input into the classifier. All previous time windows were ignored.

Comparison B again used the same network, but this time the test set was not divided into time windows. Instead, at each time step, the entire data stream up to that time step was taken as one long input – much longer than the 500ms windows the network was trained on. This was fed into the classifier and used to make the prediction for each time step.

Comparison C did not use the same network as before. A network with the same architecture was trained on training sets that did not use time windowing. Instead, each trial's data was taken as one large, time-series data sample, such that each subject contributed only nine data samples. These were used to train the network, which was then used to classify the testing set which was also not time-windowed, in the same manner as Comparison B.

6.2.8 Tests

Initially, the four time-segmented methods (the two weighted methods, and the majority voting method using the full and half dataset) were compared in accuracy over the first thirty seconds of each activity, taking the mean accuracy across all participants in both groups. The comparison methods were also compared in the same way.

The time-segmented method with the highest average accuracy was compared to the comparison method with the highest average accuracy, with the central 95% confidence intervals across all repeats also included.

To investigate the effect of the size of the data set used to train the classifier, the majority voting method trained on the full training set was compared to the same method trained on only half the training data set, as is required for the weighted methods. The results were plotted, again including the central 95% confidence interval.

A Spearman's Rank correlation test was performed on the mean accuracy of all methods, again using Equation 4.3, in order to quantify to what extent their increase in accuracy over time can be considered monotonic.

Finally, the accuracy of the time-segmented methods was then calculated individually for the control group and for the patient group, showing the difference in behaviour for the algorithms across both groups (trained on a single data set containing both groups). This was used to quantify to what extent the developed system was suitable for both disabled and non-disabled individuals, and therefore identify its applicability for inclusively-designed products.

Additionally, the LSTM deep learning method was briefly compared to a Support Vector Machine classifier, with results shown in Appendix D.

6.3 Results

The different time-windowing methods were compared, with their mean accuracy plotted against time after activity inception in Figure 6.2.

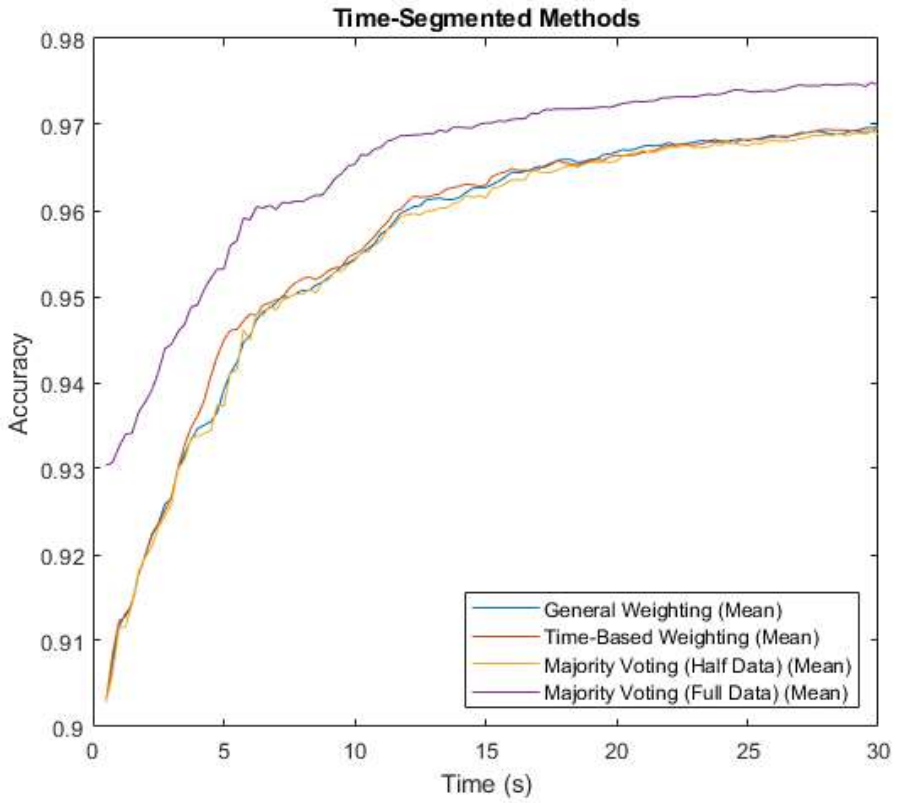

Figure 6.2. *The four time-windowed methods compared. Majority Voting with both half and full data is included as a comparison against the general and time-weighted methods, which are trained on only half the data (with the other half used to learn confusion matrices).*

The three comparison, non-time-windowed methods were compared and plotted in Figure 6.3.

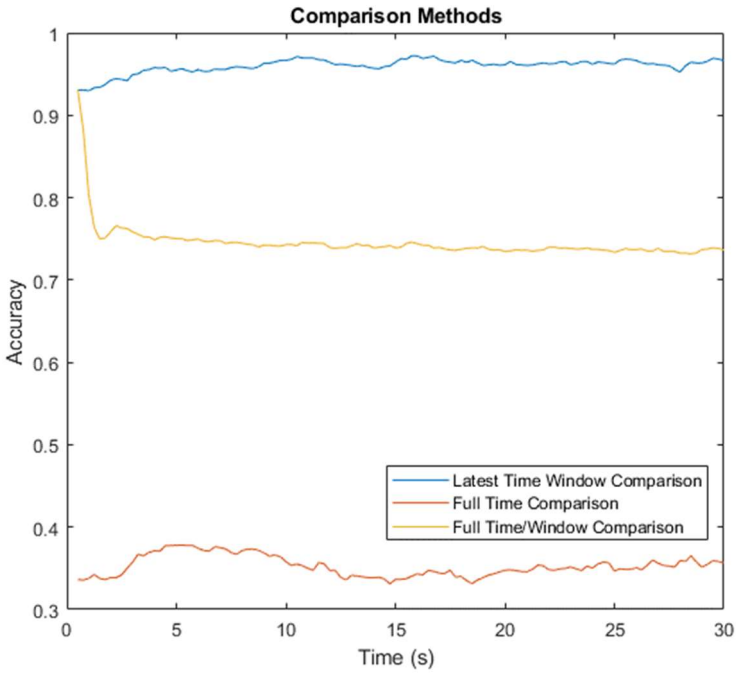

Figure 6.3. *Accuracy vs time for the three non-time-windowed comparison methods.*

The highest-accuracy time-windowing method was compared to the highest-accuracy non-time-windowed method in Figure 6.4, with the central 95% confidence intervals included for both.

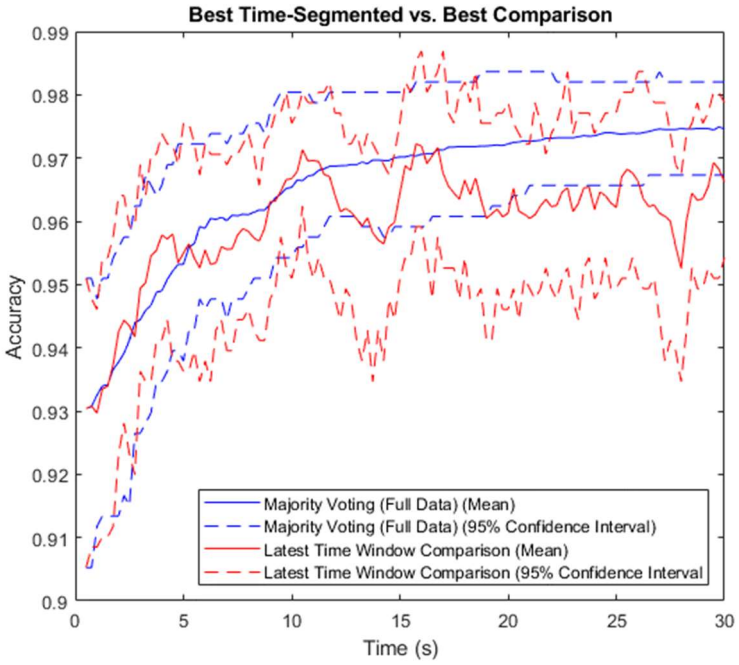

Figure 6.4. *The best time-segmented method, the Majority Voting with Full Data, vs the best comparison method, the Latest Time Window Comparison, both shown with their central 95% confidence intervals.*

The difference in accuracy between majority voting trained on the full data set vs majority voting trained on a half data set (as used in the weighted voting methods) is shown in Figure 6.5.

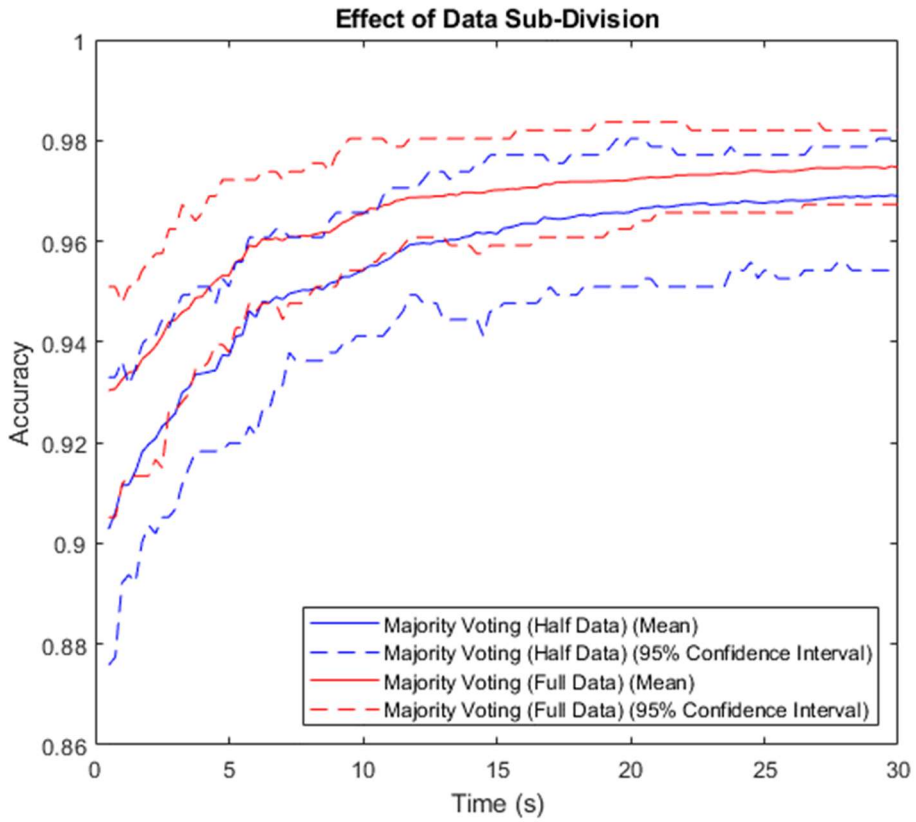

Figure 6.5. *The Majority Voting method compared when using the full data set for training, vs using only half the training set as is performed for the weighted methods.*

The Spearman's Rank correlation coefficient was calculated for each of the methods to describe the relationship between accuracy and time after activity inception, with results shown in Table 6.3.

Method	r_s
General Weighting	0.999
Time-based Weighting	0.999
Majority Voting (Half Data)	0.999
Majority Voting (Full Data)	0.999
Comparison A	0.555
Comparison B	-0.905
Comparison C	-0.095

Table 6.3. *Spearman's Rank correlation coefficients (r_s) calculated for accuracy vs time after activity inception for the methods tested. An r_s of 1 is a completely monotonic increase, 0 is no monotonic increase or decrease, and -1 represents a perfect monotonic decrease.*

The classification accuracy of each of the weighted methods was also investigated for both the patient and control groups, with results shown in Figure 6.6.

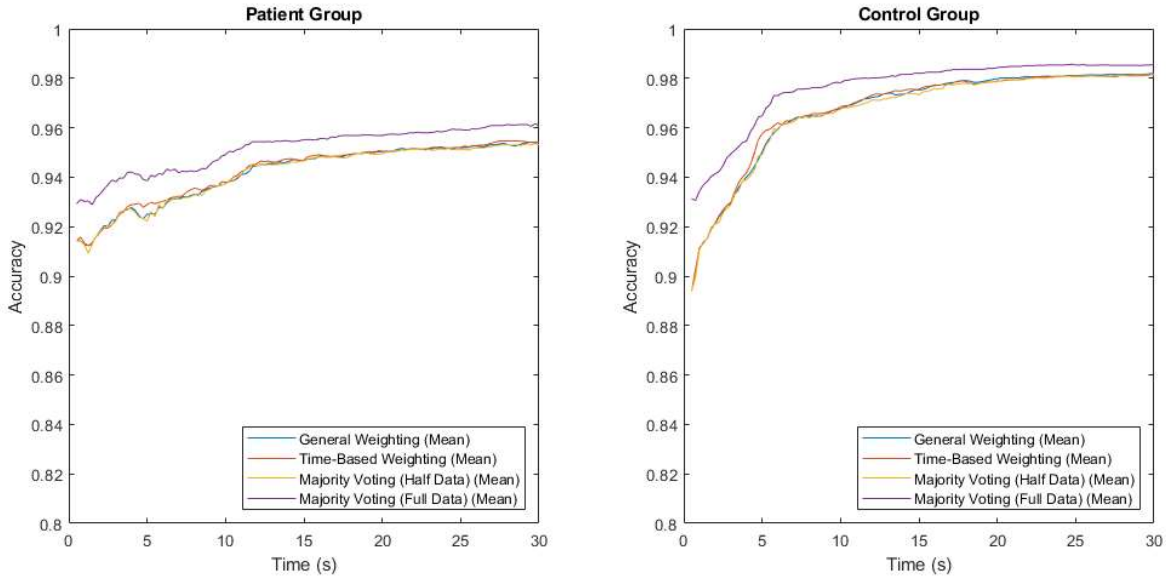

Figure 6.6. *Accuracy vs time for the time-windowed control methods, compared between the patient group and the control group.*

The Spearman's rank coefficients were also calculated for each group, with results shown in Table 6.2.

Method	r_s (Patients)	r_s (Control)
General Weighting	0.997	0.999
Time-based Weighting	0.996	0.994
Majority Voting (Half Data)	0.993	0.998
Majority Voting (Full Data)	0.995	0.981

Table 6.2. *Spearman's Rank correlation coefficients (r_s) calculated for accuracy vs time after activity inception for the time-segmented methods tested on the Patient and Control Groups. An r_s of 1 is a completely monotonic increase, 0 is no monotonic increase or decrease, and -1 represents a perfect monotonic decrease.*

6.4 Discussion

All four of the time-windowed algorithms tested produced the hypothesised monotonic increase in accuracy over time that was required for a true intent sensing system, with a strong positive Spearman's Rank correlation coefficient for each, compared to the comparison methods, which showed weaker or even negative correlations. This suggests that, using the proposed time-segmentation, deep learning can be applied to intent sensing to make an initial prediction ahead of activity inception, and to continue to refine this as the activity begins and progresses.

Of the four time-windowed methods, majority voting trained on the full training set showed a consistently higher accuracy than the other methods, especially in the first few seconds after activity inception. This difference appeared to be due to the fact that

it was trained on twice as much training data as the other methods. This is illustrated in Figure 6.5, showing the difference in using majority voting trained on the full training set, vs only half the data set (as is used for the two weighted methods). This is a limitation of the data set used in this study. For future technology developed using these techniques, a much larger data set can be collected. As the size of the data set increases, the relative benefit in retaining the full data set for training will decrease. For a very large data set, which may be collected from users themselves if intent sensing is implemented in a product with a wide uptake, this difference could effectively disappear.

For the methods trained on only half the data set, the difference is small, but it can be seen in Figure 6.2 that the mean performance of the half-data majority voting method is in general lower than that of the two weighted methods. The two weighted methods were approximately comparable, though the time-weighted method showed slightly higher performance earlier on, with the generally-weighted method providing high accuracy results later on in the activity.

By far the highest accuracy comparison method tested was the Latest Time Window method, which classified each time step of the trial based entirely on its most recent time window. Whilst this method did at times show a higher accuracy than the majority voting, it did not continuously increase over time, with a Spearman's Rank correlation coefficient of only 0.555. This does not satisfy the requirement of an intent sensing system to continue improving its prediction as the activity progresses.

Figure 6.6 shows that the time-segmented methods achieved high levels of accuracy in both the patient group and the control group. Accuracy was higher in the control group – this could be due to the motion of the control patients being more consistent and predictable, whereas the patient group contained greater variety, with different patients at different stages of disease development. Table 6.2 shows that a broadly monotonic increase in accuracy with time was observed with all time-segmented methods in both groups.

However, it is interesting to note that the full-data majority voting method displayed a slightly lower Spearman's Rank Correlation Coefficient in the control group than the patient group (0.981 vs 0.995). All the other time-segmented methods returned coefficients that were more similar between groups.

As a monotonic increase in accuracy over the duration of an activity is highly important to a holistic intent sensing system, this suggests that the majority voting method may be less suitable than the weighted methods for an inclusively-designed system. The weightings provided by the other methods may help compensate for differences between sensor sensitivity and specificity (and variations in this between time steps) to produce a more consistent intent prediction, making them more applicable to both patient groups and healthy groups. It is suggested, therefore, that if a large enough data set can be obtained to overcome the difference in training on half vs all the training data, the weighted methods may be more suitable for use in an inclusive design.

6.5 Conclusion

This study has shown that time-series segmentation provides superior deep learning performance for intent sensing than comparison methods, and in particular, provides a measure of accuracy that increases monotonically over time, which was seen in all time-segmented methods, satisfying the requirements of the study.

Of the methods tested, majority voting was shown to have the highest overall accuracy. However, deeper investigation suggested its increase in accuracy came from the fact that it was effectively trained on a data set twice the size of the weighted methods. In addition, the majority voting method's degree of monotonic increase over time was not as strong as the weighted methods. It is essential that the developed intent sensing system is generalisable to both disabled and non-disabled users, and so the suitability of the majority voting method may not be as strong as the weighted methods for an inclusive intent sensing system.

The weighted methods showed similar performance, with a weighting based on individual time-steps showing slightly higher accuracy earlier on in the activity. It is therefore suggested that, if a larger training data set could be obtained, this may be the most suitable method for a deep-learning based time-series intent system.

The methodology developed in this chapter will be utilised in future research, forming the core of a holistic intent sensing system that could be applied in many fields of human-machine interaction, improving quality of life for disabled and non-disabled individuals alike.

Statement of Authorship for joint/multi-authored papers for PGR thesis

Title of Paper	Developing a Deep-Learning Approach for Intent Sensing Using Inclusive Design
Publication Status	□ Published □ Accepted for Publication ■ Submitted for Publication □ Unpublished and unsubmitted work Written in a manuscript style
Publication Details	J. Russell, J. Inches, C. Carroll and J. H. M. Bergmann, "Developing a Deep-Learning Approach for Intent Sensing Using Inclusive Design," Nature Scientific Reports.

Student Confirmation

Student Name	Joseph Russell		
Contribution to the Paper	Performed all investigation, software development, analysis and writing. J. Inches and C. Carroll performed data collection. J. Bergmann and C. Carroll provided editing of the manuscript. J. Bergmann provided supervisory guidance.		
Signature		Date	06/08/2023

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Prof Jeroen Bergmann		
Supervisor comments		
Signature 	Date	06/08/2023

7

A Modular, Deep-Learning Based Holistic Intent Sensing System

Published by Joseph Russell, Jemma Inches, Camille Carroll and Jeroen Bergmann in *Frontiers in Neurology* 2023, 14, 1260445, doi: 10.3389/fneur.2023.1260445.

Dataset published by Joseph Russell, Jemma Inches, Camille Carroll and Jeroen Bergmann on *Dryad* 2023, doi: 10.5061/dryad.fbg79cp1d

Chapter Abstract

In the preceding chapters, algorithmic systems have been proposed, developed and tested for measuring user intent through a Probabilistic Sensor Network, allowing multiple sensors to be dynamically combined together in a modular fashion. In Chapter 6, a time-segmented deep-learning system was developed to accurately predict intent continuously. This study combines these principles, and so proposes, develops and tests a novel algorithm for multi-modal intent sensing, combining measurements from IMU

sensors with those from a microphone and interpreting the outputs using time-segmented deep learning. It is tested on an expanded version of the data set from Chapter 6, consisting of a mix of non-disabled control volunteers and participants with Parkinson's disease, and used to classify three activities of daily living as quickly and accurately as possible. Results showed intent could be determined with an accuracy of 97.4% within 0.5s of inception of the idea to act, which subsequently improved monotonically to a maximum of 99.9918% over the course of the activity. This evidence supports the conclusion that intent sensing is viable as a potential input for assistive medical devices.

7.1 Introduction

In Chapter 5, a real-time system for predicting intent using a modular sensor network was established, and in Chapter 6, a technique for predicting intent using deep learning over multiple segmented time steps was developed. This chapter combines these principles, developing a holistic, dynamic multi-modal intent sensing system as envisaged in Chapter 2.

To minimise risk and maximise compliance [142], intent prediction should be performed non-invasively (unless the patient already has an implanted device). Information that can be used for prediction can be obtained from a range of sensors. Measurements from wearable and non-wearable sensors can be individually classified using deep learning as in Chapters 5 and 6, before being combined as a Probabilistic Sensor Network to accurately determine user intent.

As stated in Chapter 4, to ensure robustness and independence between sensors, multiple sensing modalities should be used. Many possible sensing modalities have been explored for intent, including electromyography (EMG), electroencephalography (EEG) and gaze-tracking, as identified in Chapter 2. This study, however, will focus on motion data from Inertial Measurement Units (IMUs) and audio data from a microphone, as these modalities are representative of what might be found in typical consumer devices such as smart watches and smart phones, [143] and are included in currently available wearable Parkinson's disease-monitoring devices such as the Kinesia 360. [144]

Chapters 3 and 4 have shown that there are many benefits to constructing an intent sensing Probabilistic Sensor Network (PSN) using a modular method. They allow sensors to be freely added and removed from the network as they become available, without any retraining being required. This enables the possibility of a system where a user can move around a smart environment and take advantage of any wearable and non-wearable sensors they may encounter at any given time to always produce the most accurate prediction of intent.

Modular methods were also shown in Chapter 4 to be far more robust to sensor unavailability, due to causes such as failure or, in the case of wearable sensors such as Surface EMG, sensor lift-off.

The benefit of modularity that this study will focus on, however, is the ability to add sensors to a network without increasing the complexity of the learned models, and therefore without requiring an exponentially increasing amount of data to properly train them.

To elaborate – if each sensor provides 18 features, and there are six sensors, as in this study, then combining all the features from all the sensors to train a single classifier requires learning of a model in 108 dimensions. Attempting to do this with only a small amount of training data will lead to overfitting, as separating data in that many dimensions is very easy for a classifier to do "by chance", without learning any actual pattern that will reoccur for data that is not part of the training set.

This study, however, proposes to instead train one deep-learning classifier for each of the six sensors. With this approach, each classifier learned is only 18-dimensional, requiring much less training data to avoid overfitting. However, the same number of training data points are available as there were for the 108-dimensional classifier; it is simply the number of features that are reduced. As such, six much more effective classifiers are able to be learned, without discarding any of the features which may contain relevant information. The predictions from each of these classifiers can then be combined as part of a PSN.

A similar benefit is also gained by time-segmenting the data used for the deep learning classifier as shown in Chapter 6, reducing the complexity of the learned classifiers and increasing the number of available data points for training, and therefore increasing the overall accuracy of the classifiers. In this study, the system will be modular in both sensors and time.

The objective of this study is to utilise deep-learning-driven, time-segmented classification algorithms to develop a system to determine user intent through six sensors across two sensing environments, and to quantify its performance. The work also aims to show the potential of an intent sensing system that is agnostic to the kind of user (abled or disabled). The study will determine the accuracy of the intent prediction for both patients and controls at the early stages of the activity.

7.2 Methods

7.2.1 Data Collection

This study uses an expanded version of the Parkinson's disease-based data set seen in Chapter 6. The same 34 volunteers were used, under the same conditions, and recorded a calibration pose, followed by three activities of daily living (ADLs) – unlocking and opening a door, buttoning and unbuttoning a cardigan, and making toast. As before, each activity was repeated three times, without a break.

The participants each wore five Xsens IMU three-axis nine-channel IMUs (MTx, Xsens Technologies B. V., Enschede, Netherlands). These were secured to the participants lower and upper arms (both left and right), and to their head, as shown in Figure 7.1. In Chapter 6, only the data from the IMU sensor on the lower right arm was used, but in this study, data from all five sensors was included. A 44.1khz microphone on a nearby laptop (Lenovo Thinkpad X1, Dynamic Range 95 dB, Signal-to-Noise Ratio 19 dB) was also used to record audio throughout the activity.

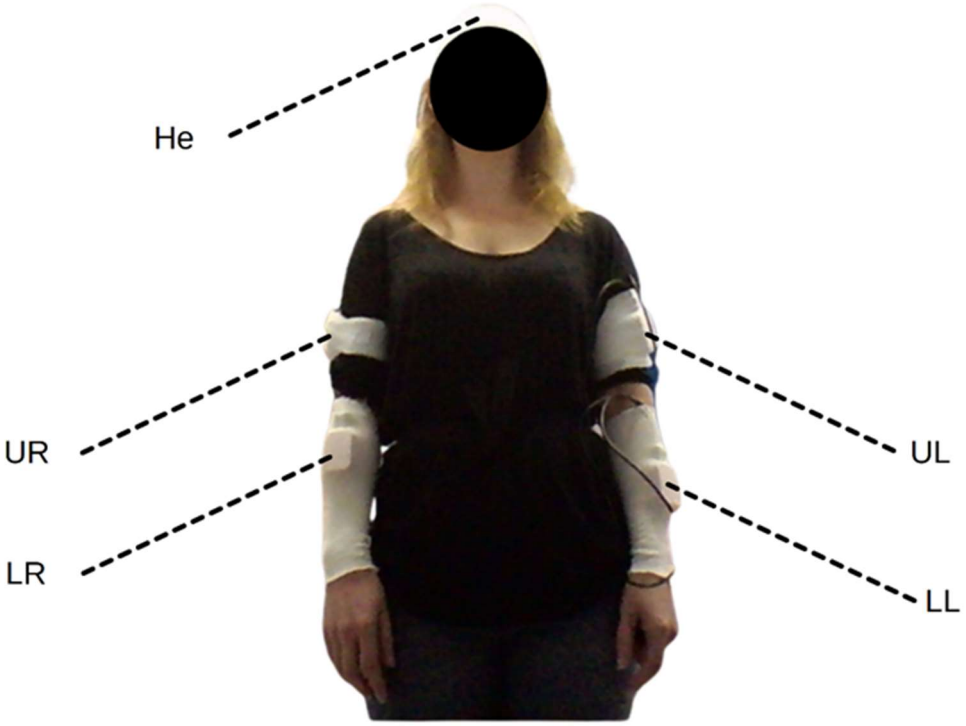

Figure 7.1. *An anonymised participant wearing all five IMU sensors – on the upper left (UL) and right (UR) arm, the lower left (LL) and right (LR) arm, and the head (He).*

Each IMU provided magnetometer, gyroscope and accelerometer data, along with a 3x3 rotation matrix provided by the XSens software. The microphone positioned in front of the participant provided a 44.1khz .wav audio signal.

7.2.2 Processing

As in Chapter 6, data from each IMU was rotated by the inverse of the mean rotation matrix collected during calibration, to correct for any misalignments in the axes of each sensor. The sensor data was then multiplied by a random rotation for each trial, to prevent the system from being able to use the participant's starting direction to determine their intent. All data analysis was done using MATLAB (R2022b, Mathworks Inc, Natick, MA, USA).

The audio data from the microphone was segmented into groups of 882 samples – 1 group of samples for every single sample from the IMU sensors. This was then processed using a Hamming Window [145] of length 882, followed by Matlab's AudioFeatureExtractor function to determine the first 18 Mel-Frequency Cepstral Coefficients (MFCCs), which are representations of the power spectrum of the sound [146], for each group. Standard numbers of MFCCs used in similar studies vary between 13 and 25 [147]. The number 18 was chosen here in order to match the number of features contributed from each IMU sensor, to ensure the system did not initially weight any one sensor more heavily than the others (weights were determined and refined during training).

No speech analysis processing was performed, so the system did not attempt to determine the words said during conversation with the supervising nurse, as speech would not be a reliable feature in a real-world scenario where the subject is on their own. Instead, only information about the general nature of the sound, such as power and frequencies, was used – it was anticipated that this would allow detection of events such as the sound of the key turning in the lock, or the buzz of the toaster operating, to more accurately determine the activity.

7.2.3 Deep Learning

Time-segmented deep learning was employed to classify the intent as quickly and as accurately as possible. As in Chapter 6, each trial was divided into time windows of width 500ms, each with an overlap of 250ms, enabling a maximum learned pattern length of 250ms (approximately equal to typical human reaction time [138]). Each time-

window was taken as a separate, 108-feature sample for training. Long-Short Term Memory (LSTM) neural networks [141] were trained for each individual sensor, and for all six sensors together. These were trained with the same parameters and architecture as in Chapter 6, only now over 50 epochs instead of 100, selected experimentally to reduce likelihood of overfitting in the Non-Modular Method. In total, 298.758 minutes of data were included in the dataset.

Leave-one-out cross-validation was used, such that all the trials for one subject at a time were withheld as a testing set, with the other thirty-three subjects used for training. This was repeated 34 times so that each subject was withheld once, with the results averaged across the set of repeats. To prepare for use in the weighted methods, elaborated on in Section 7.2.4, this training set was randomly subdivided into a Classifier Training Set and a Probability Learning Set, with half the subjects being included in the former and half in the latter. This was necessary in order to train the LSTM networks for each sensor and then assess their performance both for each sensor and at each time-step, with the results being used to weight the sensor contributions in testing.

A majority voting method was also used which assumed that all sensors and time steps had equal weight – this meant that the full training set could be used to learn the classifier, effectively doubling the size of the training set, at the cost of not being able to have weights specific for each of the sensors.

7.2.4 Modular Method

At every time step, each sensor made a prediction using the time-segmented deep-learning method. All the predictions from each sensor at all preceding time steps were then combined. The two weighted methods were based on Chapter 5's Modular Method, using Bayes' rule (see Equation 6.1), with their contributions effectively being weighted according to the confusion matrices obtained during training. This produced a probability for each of the three possible intents – the intent with the maximum probability was then selected.

A majority voting method was also tested, effectively giving all sensors and time-steps equal weight. While the loss of the weightings obviously inhibits the ability of the network to incorporate any sensor without risking accuracy loss, no subsets are required within the training set, as there are no confusion matrices to be learned – therefore, a majority voting system will be trained on twice as much data as the weighted methods.

In order to determine how much of the change in accuracy (when using a majority method) comes simply from the access to the larger training set, a majority voting method where half of the training set is discarded is also tested, in order to make it comparable to the weighted methods.

7.2.5 Non-Modular Method

To provide a comparison, a non-modular method was also used, where the features from all six sensors were included in one single LSTM network. This was also time-

segmented, as Chapter 6 established that time-segmentation was essential in a viable intent sensing system.

7.2.6 Comparison

To determine how well the system compares to the theoretically possible accuracies, the naïve model from Chapter 2 was used with the confusion matrices obtained during training in order to predict an upper bound for the accuracy of the resultant classifier during testing. The measured accuracy was compared to this upper bound to determine how close performance is to the theoretical maximum.

In this case, the equation for the naïve model is simply:

$$P(B) = P(S_1 \cup S_2 \cup \ldots \cup S_6) = 1 - P(S_1 \cap S_2 \cap \ldots \cap S_6) \tag{7.1}$$

7.2.7 Monotonic Test

As an approximately monotonic increase in accuracy over time is required for an effective intent sensing system, a Spearman's Rank test was again applied (see Equation 4.3), quantifying to what extent this requirement was fulfilled by each method.

7.2.8 Additional Tests

To further challenge the classifier, the three-class problem was artificially inflated to a six-class problem with results shown in Appendix E, and the effect of shortening the time window from 500ms to 200ms was determined and shown in Appendix F.

7.3 Results

Measuring sensor network accuracy using the four time segmentation methods for a system with modular sensors, and for a system with combined sensors, produced Figure 7.2. Both sets contain data from both the patient and control groups.

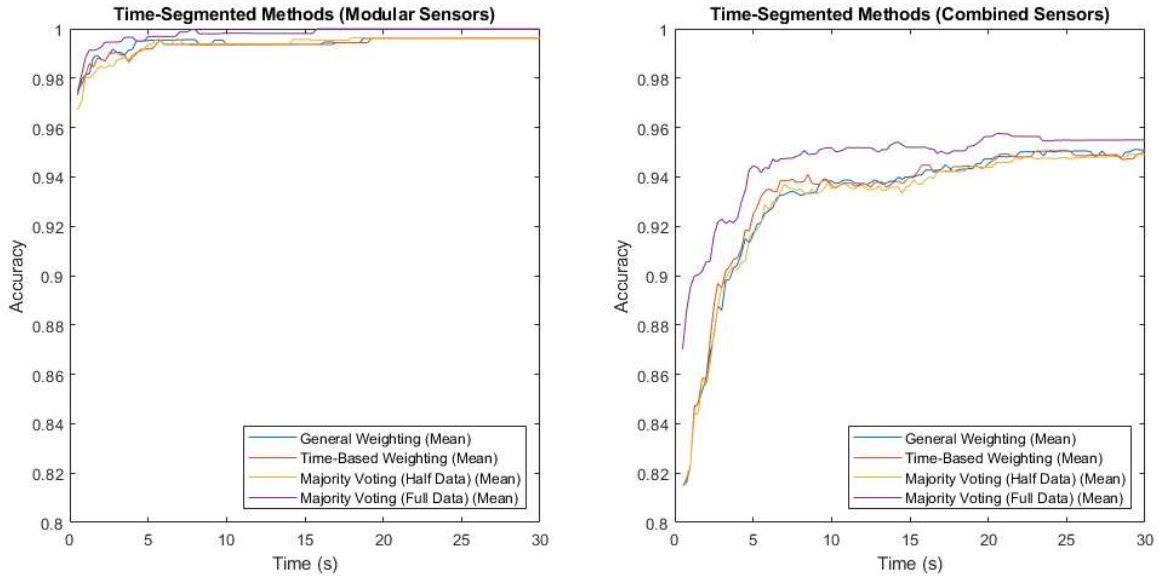

Figure 7.2. *Accuracy vs time for the modular sensor method and the combined sensor method, using the four different time-segmentation techniques originally proposed in Chapter 6, now applied across multiple sensing modalities.*

The method with the highest accuracy for both of these, at all time steps, was Majority Voting, where the classifiers were trained on the full data set, but no weightings were used, in both sensors and time. This reached a maximum of 0.999918 for the modular method, and 0.957516 for the combined method.

To investigate the extent of the benefit that Majority Voting gains by training on twice as much data, a comparison of the majority voting method trained on the full data set

vs trained on half the data set is shown in Figure 7.3. Both sets contain patient, as well as non-patient data.

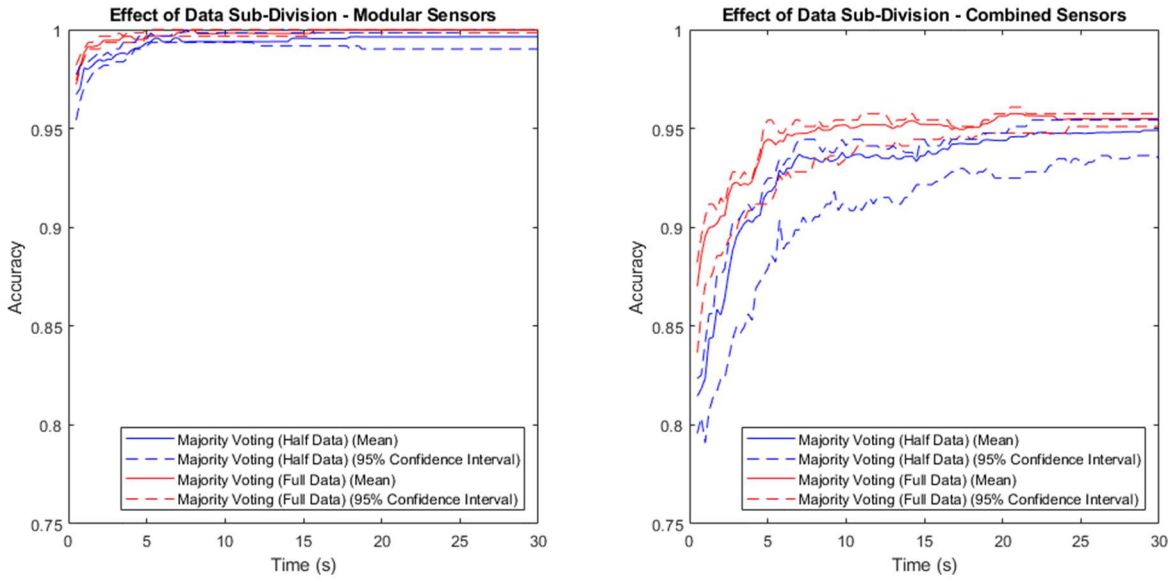

Figure 7.3. *Accuracy vs time for the modular sensor method and the combined sensor method, combined over time using the majority voting methods with the full and half data sets.*

Both methods showed a higher mean performance with the full data set than with the half data, with a larger difference in the combined method than the modular method.

As the Majority Voting method resulted in the highest accuracy for both methods, it was then used to compare the modular method (sensors) to the non-modular method (sensors), with results shown in Figure 7.4.

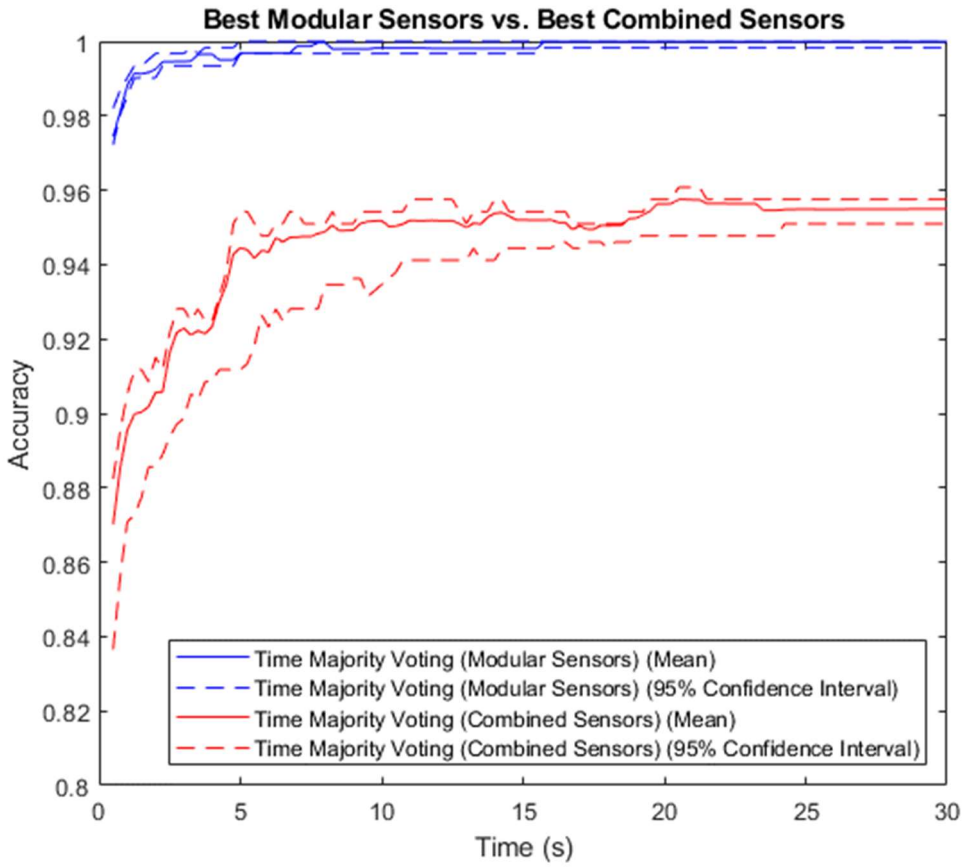

Figure 7.4. *Accuracy vs time for the modular and non-modular methods, with error shown.*

The modular method showed both higher accuracy and lower variance than the combined method. The Spearman's Rank Coefficient of the modular method was 0.89, and for the non-modular, combined method, 0.62.

Comparing the accuracy of the modular and non-modular methods across the patient and control groups produced the results shown in Figure 7.5.

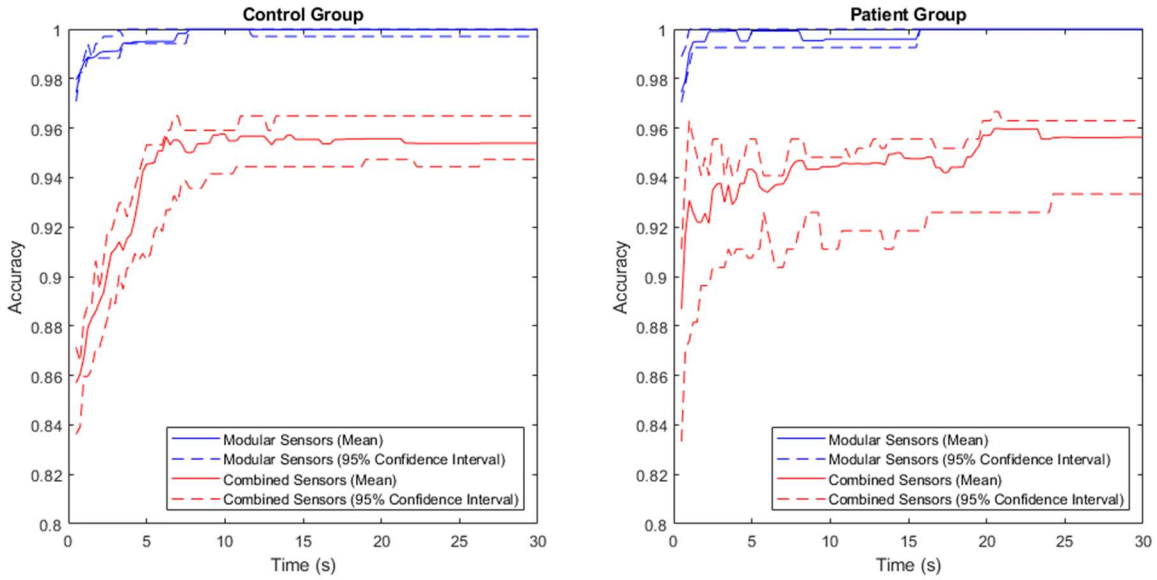

Figure 7.5. *A comparison of accuracy between the modular and combined methods for the patient and control groups, with the central 95% confidence interval plotted for both.*

The accuracy of the modular method was consistently high in both groups, rapidly approaching 1. The combined method had a larger variance, and lower maximum accuracy in both groups.

The performance of the majority voting (time), modular method (sensors) network was then compared to the performance of each individual sensor. This is shown in Figure 7.6.

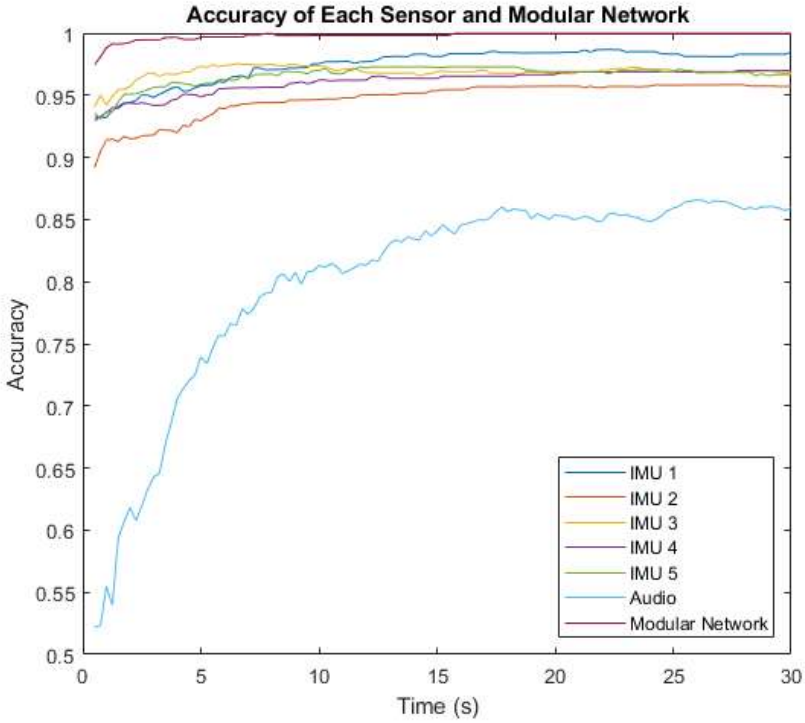

Figure 7.6. *Variation of intent classification accuracy with time after activity inception, shown for the six individual sensors, and for the complete modular intent sensing network.*

The highest accuracy sensor was IMU 3 in the first 9.5 seconds, overtaken by IMU 1 for the remaining time. The lowest accuracy sensor was the audio at all time steps.

The theoretical maximum accuracy predicted by the Naïve Model was initially 0.999996 at 0.5s, which is almost 1. The performance of the modular network method began close to this, at ~0.974. and reached a maximum of 0.999918 within the first 16 seconds.

7.4 Discussion

7.4.1 Analysis of Results

Using the majority voting time-segmented method, trained on the full data set, the holistic system was able to classify user intent to an accuracy of 97.4% within only 0.5s of

the inception of the idea to act. This is exceptionally high accuracy, out-performing the systems in the previous chapters and strongly supporting a modular-sensor, time-segmented deep learning approach for intent classification. While intent is a different goal to activity recognition, these accuracy levels are comparable to those from similar studies, but are achieved in a much faster time [148,149]. This accuracy increased approximately monotonically, with a Spearman's Rank Coefficient of 0.89, and 16 seconds after activity inception reached an accuracy of 99.9918%, meaning the system was able to correctly classify almost every trial for every subject by this point.

By comparison, the non-modular, combined method achieved a mean accuracy of only 87.0% in the first 0.5s, increasing to a maximum of 95.8% after 20.5 seconds – far lower than the mean accuracy of the modular method. In addition, the Spearman's Rank Coefficient for this was a lower value of 0.62, suggesting that not only was the accuracy of the non-modular method lower, but it also did not nearly as effectively satisfy the requirement of accuracy increasing monotonically over time.

Figure 7.6 showed the individual performance of the sensors compared to the overall performance of the modular network. The IMU sensors each showed higher accuracy than the audio. However, the inclusion of the audio modality had the major benefit of it being totally probabilistically independent from the IMU measurements. While the IMUs were located at different sites, they were all constrained by the probability of the sensing environment. A lower bound for this is $P(A1) = 0.987$, the highest accuracy recorded by any of the individual IMU sensors. The highest recorded accuracy for the audio sensor, and therefore the lower bound for the environment probability of audio

was $P(A2) = 0.866$. However, as they are both entirely different sensing environments, the overall network is constrained by neither of these limits, and thus outperforms all the individual sensors, and is able to approach 1.

The majority voting method trained on the full data set was once again shown to be the best-performing of the time-segmentation methods trialled. Figures 7.2 and 7.3 show that this contrast is due to the difference in size of training set, as artificially withholding half of the training set for the majority voting method, in order to make it comparable to the weighted methods, results in very similar recorded accuracies.

The exceptionally high performance of the modular method was observed in both the patient group and the control group, though a 100.00% (to 2 decimal places) classification accuracy was achieved 8 seconds later in the patient group than in the control group. This aligned with expectations, as large variations have been observed in the physical activity of patients with Parkinson's disease [151], suggesting that this would make classifying the patient group harder than classifying the control group.

Even with this difference, the accuracy of classification in the patient group was still very high, strongly supporting intent sensing as a viable method for interpreting user activity. This opens up possibilities for a number of possible clinical applications to support those with Parkinson's disease, such as assistive exoskeleton technology, which could predict users' intentions and provide motor support in achieving their task goals that they might not otherwise be able to complete themselves. Alternatively, intent sensing systems could predict activities which might be considered high risk, and rapidly

alert carers of the increased possibility of danger to the patient. Intent could also be used as an input for human-computer interfaces, providing more intuitive control to patients over devices which could allow them to communicate and maintain their quality of life as the disease progresses.

The high accuracy in both groups shows the modular intent sensing method as a case example for inclusive design, with ability to apply such a system for both disabled and non-disabled users. Responding to user diversity with appropriate performance across the full range of potential users will bring benefits, such as scalability of technology. An inclusive design approach also provides additional advantages related to desirability and user satisfaction even if their own physical and/or cognitive ability is changing [152].

7.4.2 Limitations of the Study

Caution should be taken in the interpretation of these results, as only 3 ADL classes were considered (unlocking and opening a door, buttoning and unbuttoning a cardigan, and making toast). The class prediction is likely to change as more activities are considered – a larger number of classes will lead to a reduced classification accuracy [30].

Additionally, the size of the data set is limited. A future study could be performed with hundreds of participants, increasing the training and testing accuracies and potentially reducing the advantage gained by the majority voting system by using the full data set.

Furthermore, while the wearable IMU system should be applicable in many real-world scenarios, there may be implementation issues with the microphone, the accuracy of which may vary dramatically when used outdoors, or in noisy environments. However,

previous studies have shown the proposed sensor fusion algorithm to be robust to sensor dropout [9,11], meaning that if large amounts of noise are identified, it should be possible to dynamically remove the microphone input from the system. This may also be of benefit if there is any issue with the IMU sensors, such as interfering vibrations from heavy machinery, or a technical fault.

While the algorithm tested is theoretically scalable to large numbers of sensors, the number of sensors included would be limited by the hardware used. Larger numbers of sensors would require greater processing power and lead to increased lag, cost and weight of any wearable device. This study did not investigate these implementation issues, but future studies should explore the trade-off between these factors.

While this study did perform all three aspects of intent sensing, its predictive ability extended only as far back as the inception of the idea to act. The nature of the experiment meant that contextual and environmental factors could not be taken into account, as the participants were instructed which activity to perform rather than deciding for themselves. Future studies could introduce elements of participant choice and the development of routines to extend the intent prediction even further back in time.

It should also be noted that the activities themselves, whilst performed without constraints, were using the exact same objects for all volunteers. It has been shown that small changes in objects could lead to different motor patterns [31]. Further work is needed to determine the accuracy of this approach in truly free living conditions. Nonetheless, the high accuracy found in this study is promising.

7.5 Conclusion

This study introduces a novel holistic multi-modal intent sensing system. A continuously-updating system was able to predict a user's intent almost immediately after activity inception, and to continue refining that prediction as time passed. This was done using a modular network of sensors, including two entirely unrelated sensing environments, that might realistically be available to a patient. The system was shown to be highly effective in both the patient and control group, demonstrating it as an effective example of inclusive design.

The results shown in the study highlight intent sensing as an achievable, highly accurate method of classifying what a user is trying to do, with potential applications throughout many fields of inclusive design within science and technology.

Statement of Authorship for joint/multi-authored papers for PGR thesis

Title of Paper	A Modular, Deep-Learning Based Holistic Intent Sensing System
Publication Status	■ Published □ Accepted for Publication □ Submitted for Publication □ Unpublished and unsubmitted work Written in a manuscript style
Publication Details	J. Russell, J. Inches, C. Carroll and J. H. M. Bergmann, "A Modular, Deep-Learning Based Holistic Intent Sensing System," Frontiers in Neurology, vol. 14, 1260445, 2023, doi: 10.3389/fneur.2023.1260445

Student Confirmation

Student Name	Joseph Russell		
Contribution to the Paper	Performed all investigation, software development, analysis and writing. J. Inches and C. Carroll performed data collection. J. Bergmann and C. Carroll provided editing of the manuscript. J. Bergmann provided supervisory guidance.		
Signature		Date	01/12/2023

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: Prof Jeroen Bergmann			
Supervisor comments			
Signature		Date	01/12/2023

8

Discussion

8.1 Overview

Prior to this DPhil, intent (as defined in this thesis) had not been significantly explored in previous literature, and to the best of my knowledge, no attempt had been made to develop a holistic intent sensing system. This thesis has developed novel algorithms to detect user intent and compared them to more basic alternative methods, showing benefits in a number of situations.

To achieve this, a model was developed for the use of Probabilistic Sensor Networks (PSNs), which was then applied to retroactively determine user intent after the fact in a study. This was further developed into a simple pseudo-real-time system, and then into a continuously-improving, holistic system demonstrated on data from volunteers with Parkinson's disease. This was then expanded into a multi-modal, modular sensing system, networking together six sensors across two sensing modalities to predict intent as

quickly and as accurately as possible. The results shown were extremely promising, and provide a strong case for a drop-in/drop-out, passive system using probabilistic methods to determine user intent.

8.2 Retrospective

The systems developed in this thesis have been novel, providing intent as a new way of looking at human-machine interaction. The literature review in Chapter 2 established that intent as defined in this thesis had not been widely investigated, and highlighted a gap in the literature for a system that continuously predicts intent from before activity inception and improves this prediction monotonically throughout the activity itself (see Figure 8.1 for a graphical representation). The most viable sensing modalities were identified and utilised in the subsequent chapters, though there is great potential in exploring the remaining modalities in future research. Over the chapters of this thesis, the necessary techniques for the development of the proposed system have been laid out, and a basic holistic intent sensing network established.

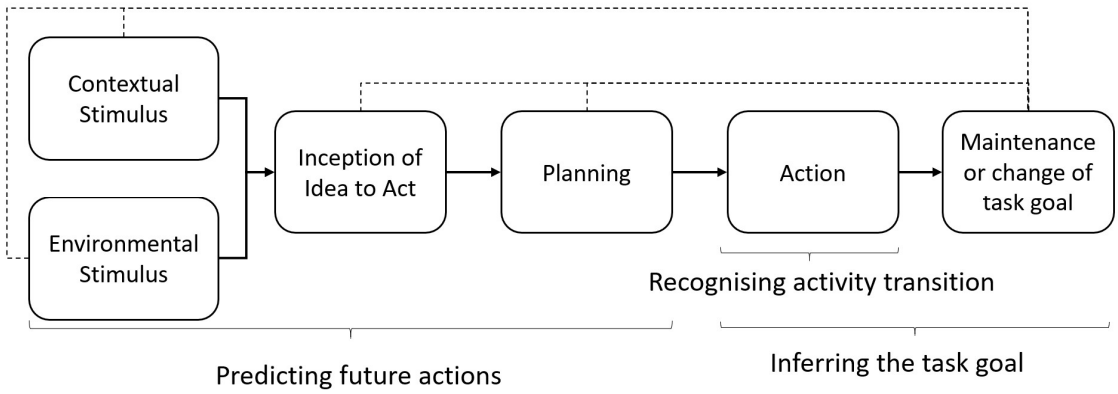

Figure 8.1. *Overview of a holistic intent sensing system, re-printed from Chapter 2 for reference.*

The work on PSNs in Chapter 3 was essential, as the concept of networking sensors is at the heart of this thesis. Many alternative, cutting-edge methods for human machine interfacing dismiss wearable sensors because of their relatively low accuracy, instead focusing on implanted sensors. For example, Össur, a leading prosthetics company, are driving research into implantable electromyography (EMG) electrodes for intuitive control of their lower-limb prostheses [153], and Elon Musk's Neuralink project is aiming to allow control of devices using thought alone by implanting directly into the brain [154]. These technologies carry with them inherent issues of risk and cost [13].

These problems can be avoided by instead combining multiple non-invasive sensors to probabilistically infer what the implanted device would otherwise directly sense. The PSN models provide a mathematical framework, essential for developing the methods for this, and were used in each following chapter.

In order to achieve the plurality of sensors required to sense intent to an effective accuracy, it was proposed that the system should be able to take advantage of any sensors that may be available to the user at any given time, and that this sensor availability will change throughout the day. The original intent sensing system developed in Chapter 4 therefore took real data involving human participants and classified a selection of reach/grasp activities using a modular sensor network, and explored the effect of sensors dropping out. The system designed was shown to be robust to sensor dropout in comparison to alternative methods, indicating it was more likely to be viable in a practical scenario. This was also predicted to be useful for assistive devices in developing

nations, where access to a clinic for device maintenance may be limited [107], allowing continued functionality with a reduced sensor set.

The analysis in Chapter 4, however, was performed entirely "after the fact", and so was not representative of a real situation where data is being continuously gathered and analysed as the user performs activities. The methods proposed in Chapter 5 were therefore developed to be viable in real-time, and while the analysis itself was performed "offline", it was done so in pseudo-real-time, with classification performed at each time step using only data from that and preceding time steps.

This experiment used only EMG sensors, and classified between two intent states under a highly controlled scenario. It was intentionally envisaged as a "minimum viable product" setup to perform intent classification in the most basic possible sense, and successfully established the possibility of a continuous intent sensing system.

In these studies, a relatively basic classification method, K-Nearest Neighbours (KNN) was used, [155] ensuring the results were predictable and explainable while the intent prediction process was established. In Chapter 6, the classification method was replaced with a far more state-of-the-art Long-Short Term Memory deep learning neural network, which has been shown to be effective in time-series applications such as this [139,141].

This was much more representative of the accuracy levels achievable in a real intent sensing product, and was used for the remainder of the project.

The primary contribution in Chapter 6, however, was the development of a time-segmenting intent sensing classifier. This took the modular sensor network concept used

in the early chapters and applied it instead to time, with each 0.5s time segment effectively treated as a separate sensor contribution. This was necessary as it allowed the deep learning classifier to be trained on a finite amount of data and used in continuous time, rather than requiring an ever-increasing amount of training data to classify increasingly long time periods. This was applied on a new data set with patients with Parkinson's disease (a mobility-limiting condition [130], the effects of which could be mitigated through intent-controlled assistive devices [5]), classifying them using a single sensor intended to represent a commonly available smart watch [91].

The result of this was an intent sensing classifier that began at the moment of activity inception and improved monotonically over time, as the activity itself began and then progressed. This satisfied the requirements of a holistic intent sensing system as established in the literature review.

The final study, Chapter 7, took the principles of modularity of sensors from Chapter 5 and modularity in time from Chapter 6 and combined them, networking five wearable sensors to produce a highly accurate, monotonically improving intent sensing system. This was a highly sensor-rich environment, intended to showcase the possible effectiveness in a best-case scenario. It is unlikely a real user would be willing to wear six sensors at once – however, the modularity of the PSN system means that the number of sensors can vary dynamically during use, as both wearable and non-wearable sensors drop in and out of range.

The results of this study were highly promising, though it should be noted that only three possible intent classes were tested. In reality, a much higher number of possible intents exist. Extending the system to a higher number of classes is trivial, however, the accuracy of the classifier will inevitably decrease as the number of possible classes increases, increasing the required length of time after activity inception before a specified accuracy level is reached.

Nonetheless, the fact that the system was able to achieve an accuracy of 97.4% within the first half second after inception of the idea to act, increasing to 99.9918% over the course of the activity, is extremely positive, and a satisfying conclusion to the work of this thesis. Further research is, of course, required before intent sensing will be truly ready for implementation in assistive medical devices, but the potential next steps are clear.

Chapter	Contribution
1	Introduction to the idea of intent to be discussed in the thesis.
2	Establishment of a more holistic definition of intent, and a novel systematic literature review of the current state-of-the-art according to this definition.
3	Proposal and experimental verification of a new model for the performance of a multi-sensor probabilistic sensor network (PSN).
4	Development of a novel, modular algorithm for combining sensors into a PSN to predict short-term intent of a reach/grasp activity. This is shown to

	be robust to sensor drop-in/drop-out, and is theoretically scalable to large numbers of sensors which may vary over time.
5	Enhancement of the intent algorithm to apply in real-time, with verification on a new electromyography data set. This application attempts to predict intent up to half a second before activity onset, exploring the Activity Prediction phase of intent.
6	Development of a novel method of applying deep learning Long Short-Term Memory neural networks to time-series data without requiring a prohibitively large data set by time-segmenting the data, making individual intent predictions for each time step, and re-combining these predictions as part of a PSN. This is tested against a new data set including Parkinson's disease patients and controls.
7	Combination of the algorithms from Chapters 5 and 6 to produce a multimodal, holistic, deep learning-driven intent prediction algorithm, tested on an expanded version of the Parkinson's disease data set and shown to provide much higher performance than comparison methods. This particularly showcases the scalability of the modular approach, as sensors can be added without increasing risk of overfitting.
8	Summary of the work done in the thesis, with detailed suggestions for future research.

Table 8.1. *A chapter-by-chapter breakdown of the key contributions of this thesis.*

8.3 Suggestions for Future Research

8.3.1 Incorporating Classifier Confidence

One area that was explored to an extent as part of the work for this DPhil, but considered beyond the scope of this thesis, was the use of sensor confidence to further weight the predictions made by each sensor at each time step. Deep-learning classifiers such as LSTMs do not only output a prediction of class, but rather a likelihood, or confidence, of each possible class, with the most likely class selected for the final prediction.

It is possible to use these likelihoods in the Bayes' rule-based method proposed in Chapter 3 by introducing fractional weightings for each sensor, where the weighting used is the likelihood of each class as given by the LSTM classifier. Bayes' rule has been used in this thesis in the form:

$$P(E|V) = \frac{P(V|E) \cdot P(E)}{P(V|E) \cdot P(E) + P(V|E') \cdot P(E')} \tag{8.1}$$

Where:

$$P(V|E) = \prod_{s=1}^{n} \big(P(V_s|E)\big) \tag{8.2}$$

Where s is the sensor number and n is the total number of sensors.

If the same sensor was to be included twice in this system, i.e. given a double weighting, then it is clear that in Equation 8.2, its probability would be effectively squared. It

therefore seems intuitive that classifier confidence weightings should be applied as powers. With the proposed weightings, this equation now becomes:

$$P(V|E) = \prod_{s=1}^{n} \left(P(V_s|E)\right)^{k_s} \tag{8.3}$$

Where k_s is the likelihood of intent event E according to sensor s.

This allows sensors to make fractional contributions to the overall classifier depending on their confidence. In this case, the confidence of each class is always within the range of 0-1, so all weightings will be fractional, with each sensor contributing some of its "vote" in favour of each class according to how likely it determines each class is.

It is hypothesised that this will increase overall classifier accuracy, provided that the accuracy of the individual classifiers themselves was already high. If only a small training set is used, and the individual classifiers for each sensor have largely wrong estimations of the likelihood of each intent event, then this inaccuracy will be amplified by the weighting, reducing overall classifier accuracy.

This should be properly investigated in a further study, to quantify the potential benefits of weighting according to sensor prediction confidence and determine the limitations of this method.

8.3.2 Human-In-The-Loop

This thesis has focused on accurately determining user intent, which could theoretically be used as an input for an assistive device. However, the scope of the thesis has not included any physical output. A priority for future research should be to apply the

techniques developed in this thesis to a control scheme for an assistive system such as a prosthetic device or a smart-home environment in a useability study.

This would require the participant to complete activities both with and without the assistive system, and to have their performance measured. Metrics assessed could include objective factors such as time and accuracy but should also include subjective factors reported by the user such as comfort and ease of use.

This would provide a real-world measure of the level of benefit offered by an intent sensing system over alternative control methods, and would be a vital step in bringing intent into use in commercially available products.

8.3.3 Sensor Co-Dependency

The models used in this thesis assume probabilistic independence between each sensor. While this assumption holds true to a certain extent, as the sensors used have all been on different sensing sites, it is not always an absolute. For example, a system composed only of EMG sensors would not be able to detect an intent before the idea to act progresses beyond the brain and nervous system to the muscles themselves, regardless of how many EMG sensors are used.

Further research should modify the mathematical models used to incorporate sensor co-dependency in some way. This could be done by measuring correlations in sensor error during training (i.e. when sensors A and B are wrong, are they wrong in the same way?), and then reducing the weighting of predictions made by sensors with strong correlations. It is important to note that correlation is not the same thing as dependency,

and lack of correlation merely supports the possibility of independence, rather than proving it [156].

Care should be taken to only measure correlation in sensor errors, rather than overall sensor correlation, as it is natural that if the sensors are accurate, their measurements will be highly correlated, while still being independent, as they are measuring the same event.

This proposed correlation weighting should be implemented such that if n sensors are entirely dependent on one another, their total weighting should be 1, and if they are entirely independent, their total weighting should be n.

There are many possible models that could satisfy these constraints, and determining which is the correct one will require further analysis and experimentation. A study could be conducted by simulating a controllable degree of sensor co-dependence, varying this with a set of sensors of known accuracy and measuring how closely the accuracy of the combined system matches that predicted by this new, modified model.

Improving the models in this way would reduce the need for care to be taken in sensor selection, allowing any and all sensors to be included without risking a reduction in overall system accuracy. This would improve the system's versatility and generalisability in real-world scenarios.

8.3.4 More Possible Intents

As mentioned in Section 8.2, the studies in this thesis worked with a limited number of possible intents, and these were primarily considered after the inception of the idea to

act. This was useful for constraining the system and establishing the methodology for intent sensing, but in the real world, users will have hundreds, if not thousands of possible intents at a given moment. The scale of this problem can be reduced using contextual and environmental factors, which were not used in the studies performed here, but it will certainly still be a much larger number of classes than were tested in this thesis.

It is therefore important that future work tests intent sensing with a more realistic number of classes. With a fixed amount of training data, this will reduce the accuracy of the system, and so to achieve a sufficiently high level of accuracy for practical use, much more training data will be required.

As such, this would have to be a large-scale study using many participants to collect as large a training data set as possible. Data collection for each participant would also have to be extensive, taking place over a long enough period of time for them to complete the required number of activities.

This could be achieved by asking participants to wear sensors throughout their normal working day. This would allow users' typical patterns of behaviour to be learned, enabling contextual and environmental factors to be incorporated into the intent sensing system. However, if this were unsupervised, the question arises of task labelling, i.e. how would the training system know which activity is which? It may be possible to overcome this by exploiting the fact that training does not have to take place in real-time, and so each activity can be classified after-the-fact by applying a pre-trained

activity recognition classifier to provide the labels, which can then be used to train the much more difficult intent classifier.

8.3.5 Self-Learning

The fact that activity classification is an easier problem than intent sensing could also be used in real-time. It is demonstrated by the fact that classification accuracy does increase monotonically over time after activity inception. This could mean that an initial prediction of intent could be made at the point of activity inception, with this then improving in accuracy as the activity takes place – and then at the activity's conclusion, the final prediction could be used to label the recorded data, which could then be added to the training set, and used to update the classifier.

This would result in a continuously-improving intent classifier, gradually learning from the user's specific habits and personalising the system for them. The precise nature of motor control has been shown to vary from person to person [10], and so it is hypothesised that personalisation will result in an improved level of accuracy, and/or a faster speed to reach a fixed required level of accuracy.

Even this after-the-fact activity classification will not be 100% accurate, however, meaning this self-learned data will be imperfectly labelled. The effect of imperfect labelling has not been widely explored in literature, but it is hypothesised that, as long as the proportion of correctly labelled points is higher than random labelling, it will improve the classifier's accuracy as more data is collected.

9

Conclusion

Intent sensing is an issue of huge scope, and there remains a great amount of work to do to further develop it. This thesis has, however, established a strong foundation for intent research and has showcased its great potential for accurate and rapid device control.

It is my firm belief that, whether or not through the methods proposed in this thesis, intent will one-day be a ubiquitous aspect of technology. The devices of the future will anticipate our needs and respond to our intentions before even we ourselves know what they are. From prosthetics, to orthotics, to virtual reality and human-robot cooperation, intent sensing has the potential to bring about a step-change impact in many developing and exciting fields.

The work in this thesis has taken major, novel steps towards achieving these goals, and I look forward to the day that intent technology becomes available for all.

10

Appendices

10.1 Appendix A: Feature Reduction

In Chapter 4, the high dimensionality of the features used could create an issue of overfitting in the classifiers trained. This risk is higher in the NMM (Combined KNN) than in the MM (Bayesian Fusion), as the same amount of data is used to train both, but the dimensionality is 12 times greater in the former than in the latter.

To mitigate the risk of overfitting, a visual inspection was performed (see Figure 10.1) to check how the measured accuracy with both algorithms varied with the number of features, N, increasing from 250 to 4500 (the number of features present in the full one second of data). For this test, the number of features used was constrained through feature reduction, performed using a Principal Component Analysis (PCA) [157]. The features were ranked in order of "importance" as estimated by the PCA, and the top N features were retained.

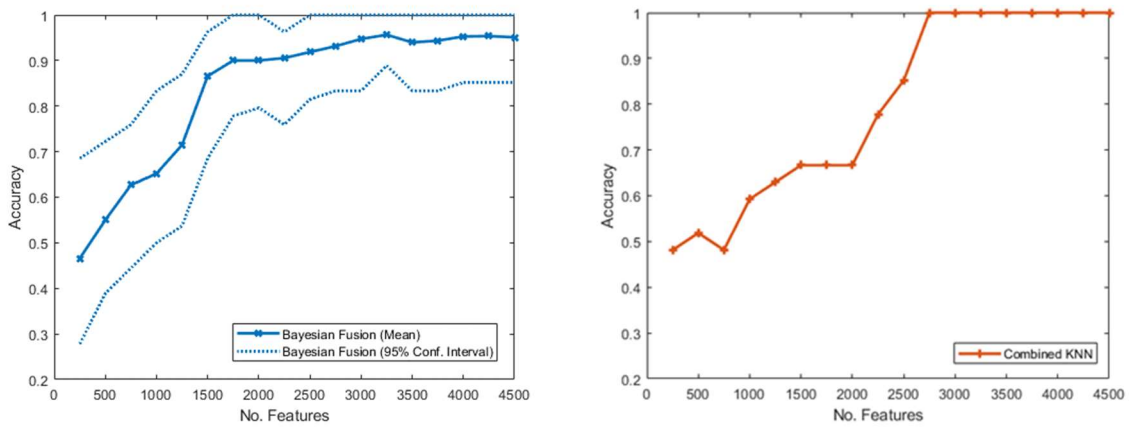

Figure 10.1. *Example graphs of randomly selected volunteers to show the accuracy across the number of features selected in the Bayesian Fusion method with all sensors included. The Bayesian Fusion algorithm has confidence intervals, as its performance depends on how the training data is subdivided. The Combined KNN method does not have confidence intervals, as all the sensors are included and there is no subdivision of the training data, so its performance is entirely reproducible.*

This test demonstrated that, allowing for random fluctuation, the accuracy increased approximately monotonically in both algorithms as more features were included. This suggested that overfitting was not occurring and that the features should not be reduced in the algorithm, and as such all 4500 features were included. This number is generally consistent with the number of features used with similar amounts of data in similar studies [92].

10.2 Appendix B: Justification of Number of Time Segments

The intent sensing algorithms used in Chapter 5 made continuously updating predictions of the users' intent, with a new prediction being made every 50 ms. For each

prediction, the features from the previous ten 200 ms wide segments (650 ms) were used to allow the full pattern of the changing EMG signal to be recognised. This does not introduce input lag, as each segment was labelled according to its delay, and as such, the earlier segments simply provided context for the later segments. To further justify this, a test was performed by repeating the study, where varying numbers of time windows, nW, were retained. $nW = 1$ indicates that only the most recent time window was used as a source of features for training and testing. $nW = 10$ indicates all ten preceding time windows were used. The results of this are shown in Figure 10.2.

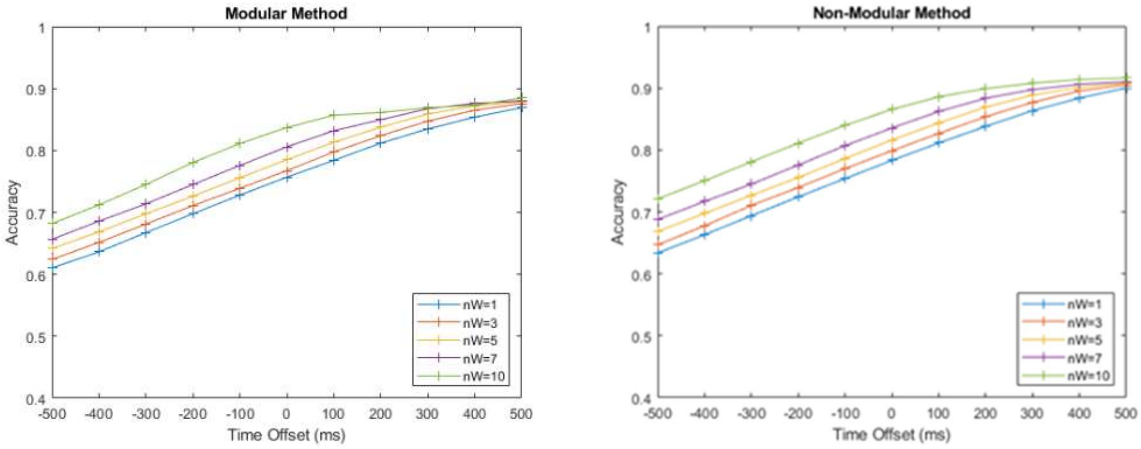

Figure 10.2. *Accuracy vs time for the two weighted methods for five different values of R, the proportion of the training set that is held back as the probability learning subset.*

If the inclusion of these preceding time windows caused input lag, the accuracy of earlier time offsets would decrease as nW increases, as the algorithm would not be able to correctly predict the intent so far in advance. This was not seen in the test. Instead, the graphs show a clear improvement in the performance of both algorithms with a larger number of time windows at all values of time offset. This justifies the inclusion

of the preceding ten time windows and shows that doing so does not introduce input lag.

10.3 Appendix C: Effect of Varying Training Set Subdivision Ratio

In Chapter 6, the variation of accuracy with time after activity inception was also calculated for different values of R, the proportion of the training set held back to learn the confusion matrices for the two weighted methods. Results from this are shown in Figure 10.3.

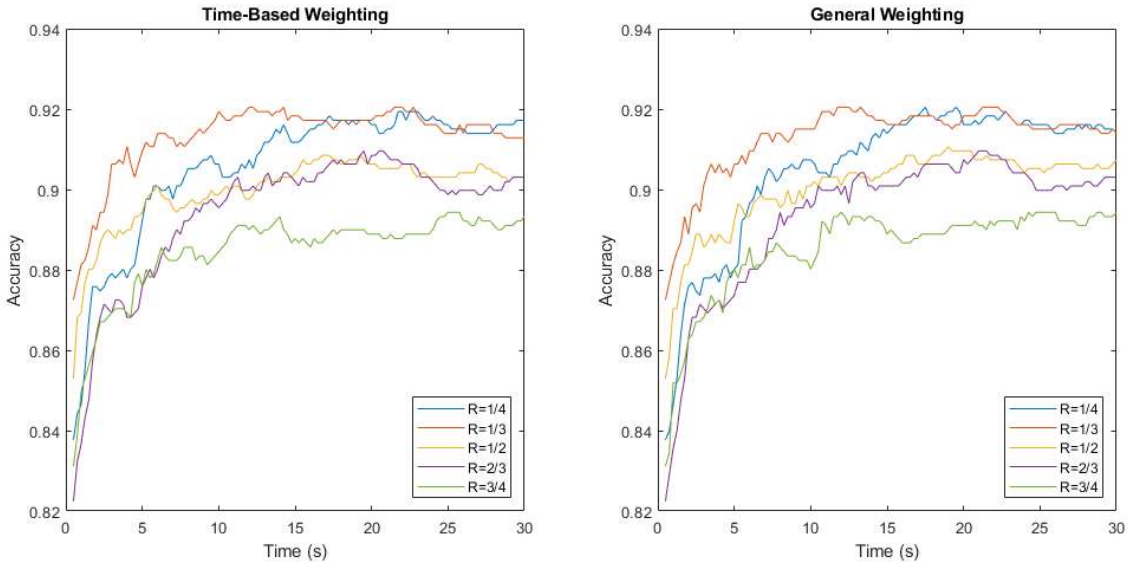

Figure 10.3. *Accuracy vs time for the two weighted methods for five different values of R, the proportion of the training set that is held back as the probability learning subset.*

Figure 10.3 shows the effect on the two weighted methods of varying R, the proportion of the training set that is held back as the probability learning subset. While $R = 1/2$

was used for the analysis in the study, it is interesting to note that $R = 1/3$ appears to be optimal for this data set – with two thirds of the data being used to train the classifier, and one third of the data being withheld to learn the confusion matrix. This implies that a larger data set is more important than a more accurate confusion matrix for intent sensing, but only up to a point – the $R = 1/4$ test produced lower accuracy performance, maintaining that a balance must still be found between the two.

10.4 Appendix D: Comparison Between Support Vector Machine and Deep Learning

In Chapter 6, a novel method was proposed for applying time-segmented deep learning to intent sensing in patients with Parkinson's disease and controls. The method was developed with deep learning in mind, as the time-series nature of the data lent itself well to the pattern-recognising capabilities of the LSTM classifier (though other time-series algorithms may be just as effective). However, for the sake of comparison, the time-segmenting algorithm was also tested with a Support Vector Machine (SVM), applied with appropriate feature reduction using Principle Component Analysis (PCA). This was implemented using the Classification Learner tool from MATLAB's Statistics and Machine Learning Toolbox v1.4, with default settings. The PCA was set to retain only the features needed to explain 95% of the variance in the training data.

A comparison for the two methods, based on a single sensor, both using the Majority Voting time segmentation method described in Chapter 6, is shown in Figure 10.4.

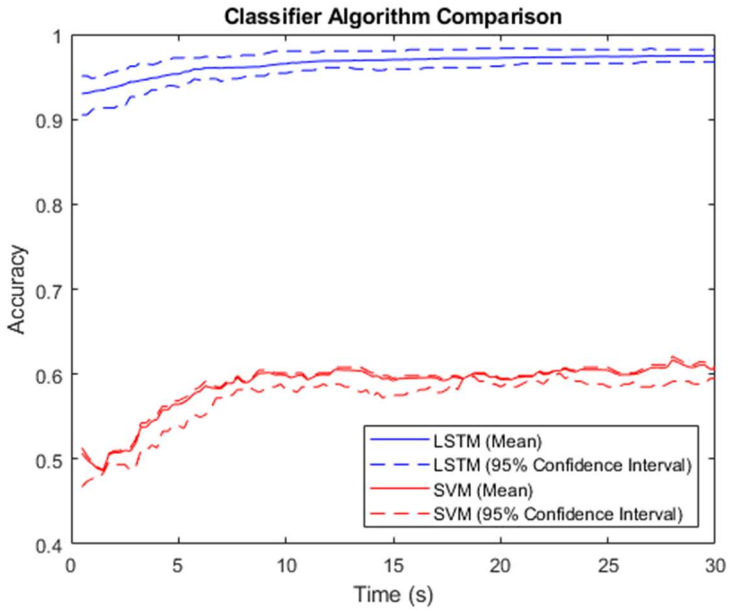

Figure 10.4. *Accuracy vs time for the two classifier methods, both using the Majority Voting time segmentation method on a single sensor, with the central 95% confidence intervals shown for both.*

The accuracy of the LSTM method was far higher at all time steps than the SVM method. This was an expected result, as the LSTM method is far better suited to time series, pattern-based data such as was collected in this study. It should be noted, however, that the time-segmentation method produces an approximately monotonic increase in accuracy with time in the SVM method as well as with the LSTM, suggesting that this benefit of time segmentation is not dependent on the specific classifier used.

10.5 Appendix E: Six-Class Version of the Parkinson's Disease Study

The study performed in Chapter 7 classified three different action intents for a combined group of patients with Parkinson's disease and controls. In order to further

challenge the algorithm, it was proposed that the number of classes could be (somewhat artificially) inflated to six, by simultaneously classifying the action intent and whether or not the user was a Parkinson's disease patient. Results of this are shown in Figure 10.5.

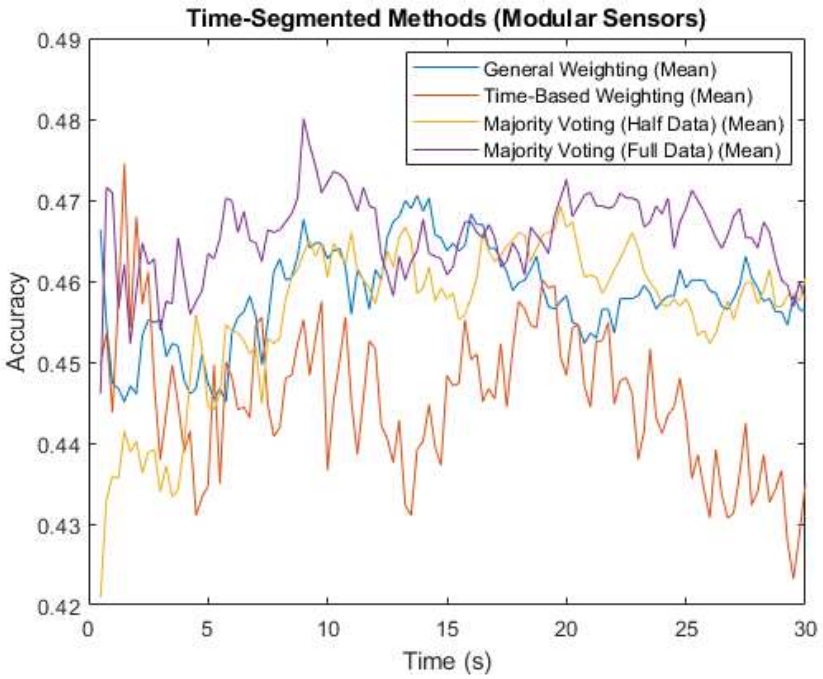

Figure 10.5. *Accuracy vs time for the time-segmented methods treating sensors as modular, with the number of classes inflated to 6 by simultaneously classifying both whether the patient has Parkinson's disease and which of the three actions they are intending to take.*

The accuracy of the 6-class classifier was very poor, never surpassing 50% in any of the tested classifiers. This was expected, as the algorithms were designed to predict user intent, not to determine whether or not the user was a Parkinson's disease patient. Additionally, the size of the training data set for each class had been halved relative to the 3-class problem, as the same number of samples was now shared between twice as many classes.

This is a failure case for the algorithm, as not only is the accuracy very low, but the approximately monotonically increasing accuracy over time has been lost. This was also expected, as the algorithm's inability to accurately classify whether or not the user has Parkinson's disease is a systematic error, not one which is independent between time samples, so making repeated predictions over time should not significantly improve the accuracy.

There was one flaw to this approach that was potentially mitigable – the fact that it treated the Parkinson's and non-Parkinson's variants of each action intent as entirely separate classes, rather than exploiting the similarity between the two. This had the effect of both increasing the complexity of the classifier and halving the number of available training samples for each class. An alternative, two-tier method was proposed to overcome this problem (though the performance of the overall classifier was still expected to be low).

Two separate classifiers were trained using the same data – one to determine the action intent, and one to determine whether or not the user had Parkinson's disease. These decisions from these two classifiers were then combined with a logical "AND" operator, placing each trial into one of the six proposed classes. Results from this are shown in Figure 10.6.

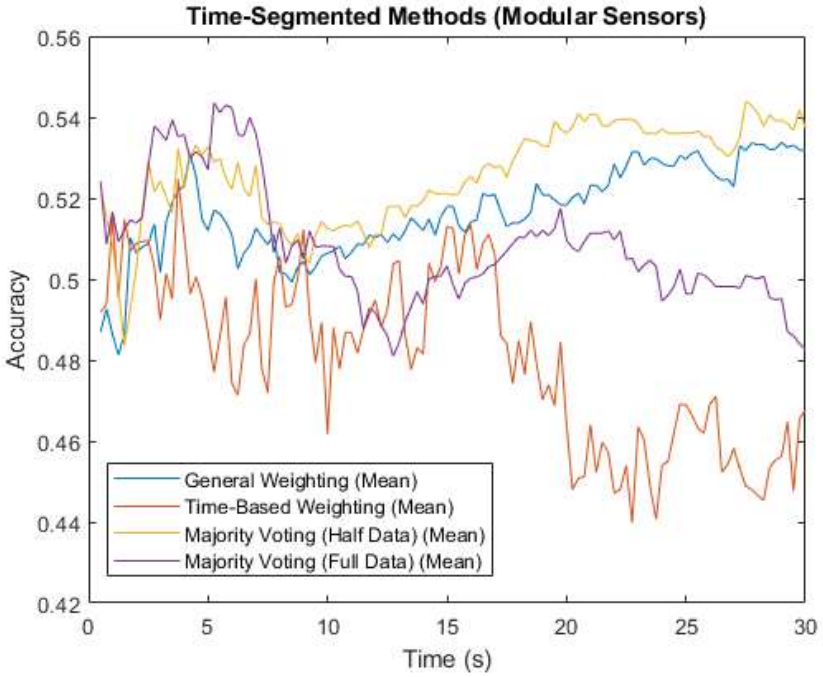

Figure 10.6. *Accuracy vs time for the time-segmented methods treating sensors as modular, with the number of classes inflated to 6 by classifying first whether the patient has Parkinson's disease and secondly which of the three actions they are intending to take, and then combining these with a logical "AND".*

It can be seen that in this case, the accuracy of the methods is improved, reaching around 54% in the best case. The fact that this surpassed 50% showed that in the two-tier system, the Parkinson's disease component of the classifier was at least better than random, though the overall performance is still poor when this intent-prediction algorithm is applied outside of its intended scope.

10.6 Appendix F: Shortened Time Window

The 500ms time windowing used for the deep learning methods in Chapters 6 and 7 was selected to be representative of human reaction time, and therefore long enough to

capture any anticipatory response between inception of the idea to act and activity onset. However, the algorithm was further challenged by reducing this time window to 200ms, with results shown in Figure 10.7.

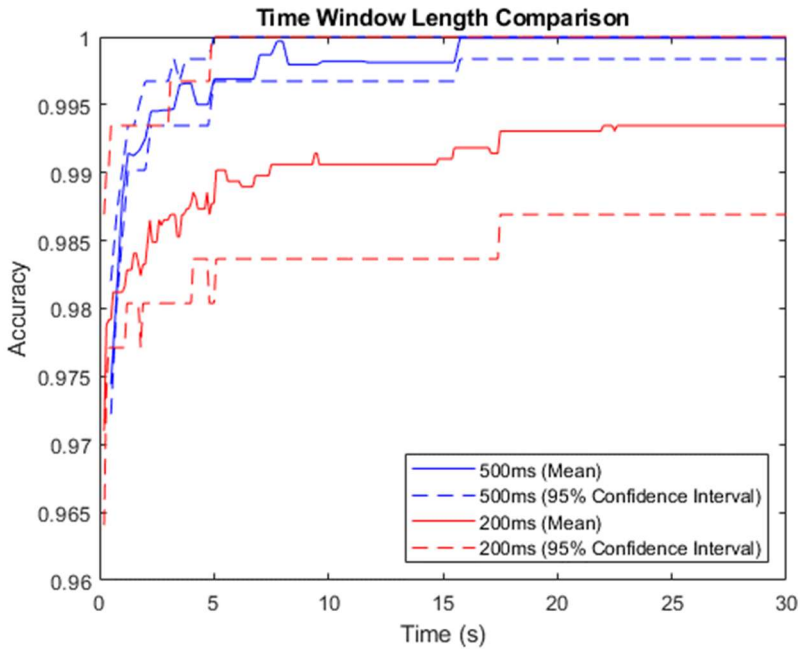

Figure 10.7. *Accuracy vs time for the two different time window lengths, both using the Majority Voting time segmentation method, with the central 95% confidence intervals shown for both.*

In this case, it can be seen that while the accuracy of the 200ms system is still high, it was only able to reach a mean accuracy of ~99.3% over the first thirty seconds of the activity, lower than the performance of the 500ms system. This was expected, as the narrower time window was unable to learn any patterns longer than 200ms that may have emerged in the user's response, and therefore did not have as much information to work with as the 500ms system. A narrower time window also increases the number of the intent predictions calculated per unit time, thereby increasing the computational requirements of the system.

This suggests that a practical intent-sensing system should carefully select a time window length that is optimal for its particular implementation. Future studies could explore whether it is possible to learn this automatically, as it will vary from user to user and from application to application. This kind of intelligent optimisation will be an essential component of personal assistive technology of the future, reducing need for clinical oversight and improving user independence and quality of life.

Glossary of Terms and Abbreviations

Activities of Daily Living (ADLs) – a term used to refer to a user's regular, routine activities essential to their wellbeing.

Bayes' Rule – a law giving the probability of event A happening given that event B has happened, as a function of the probabilities of A, B and the probability of event B happening given that event A has happened.

BITalino – a low-cost body sensing board developed by PLUX Wireless Biosignals, used to record electromyography data.

Classifier Training Set – a subset of the training set in the Modular Method, used to train the machine learning classifier. The other subset of the training set is the Probability Learning Set.

Confusion Matrix – a common probabilistic representation of sensor performance, summarising the probability of each true class being classified as each predicted class.

Convolutional Neural Network (CNN) – a kind of deep learning classifier algorithm.

Electrocardiography (ECG) – a technique for measuring heartrate from electrical activity on the surface of the skin.

Electroencephalography (EEG) – the practice of inferring activity in the brain through electrical signals measured on the surface of the scalp.

Electromyography (EMG) – the practice of measuring the electrical activity in muscles.

Inertial Measurement Unit (IMU) – a kinematic sensing unit consisting of an accelerometer, gyroscope and magnetometer, used to measure acceleration, angular velocity and orientation in a magnetic field.

Force Myography (FMG) – a technique for measuring muscle activity through volumetric changes in the muscle cross-sectional area.

Intent – the measurement and prediction of what it is that a user wants to make happen. Divided into three aspects, as stated in Chapter 2: Prediction of Future Actions, Activity Transition Recognition and Inference of Task Goal.

K-Nearest Neighbours (KNN) – A classifier method that assigns a class to a data point based on the class of a number (*k)* of data points nearest to it.

Long Short-Term Memory (LSTM) – a recurrent neural network used to classify time series data.

Mann-Kendall Test – a statistical test for trend analysis, used to determine whether a time series is monotonically increasing or decreasing.

MATLAB – a programming and mathematical computing platform developed by MathWorks, used for analysis in this thesis.

Mechanomyography (MMG) – a technique for detecting activity in muscles through vibrations on the surface of the skin.

Mel-Frequency Cepstral Coefficients (MFCCs) – representations of the power spectrum of a sound, used in Chapter 7 as features for intent classification from a microphone audio source.

Modular Method (MM) – the technique used for combining sensors in a modular fashion using Bayes' rule, introduced in Chapter 4.

Naïve Model – a probabilistic model introduced in Chapter 3 which assumes that if any one sensor in a network detected an event, then it is detected by the whole network.

Non-Modular Method (NMM) – a comparison technique that combines sensors into a single classifier, rather than treating them in a modular fashion.

Parkinson's Disease – a neurodegenerative illness limiting mobility. The focus of the dataset used in Chapters 6 and 7.

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) – a quality checklist used in systematic reviews.

Principle Component Analysis (PCA) – a feature reduction method frequently used in the training of classifiers such as Support Vector Machines (SVM).

Probability Learning Set – A subset of the training set in the Modular Method, used to learn the sensitivity and specificity of the classifier. The other subset of the training set is the Classifier Training Set.

Probabilistic Sensor Network (PSN) – a mathematical model introduced in Chapter 3, used to represent the probability of a signal passing through the four stages of detection. May include multiple sensors and sensing environments.

Revised Model – a probabilistic model introduced in Chapter 3 giving the probability of a majority voting-based sensor network correctly detecting an event.

Root Mean Square Error (RMSE) – a standard statistical measure of deviation, calculated as the square root of the average of the squared values of the errors.

Spearman's Rank – a common statistical test used to quantify to what extent a correlation between two variables is monotonic.

Support Vector Machine (SVM) – a standard classifier, trained by fitting an N-dimensional plane to best separate data points with N features.

Surface Electromyography (sEMG) – the practice of measuring the electrical activity in muscles through the surface of the skin.

Surface Electromyography for the Non-Invasive Assessment of Muscles (SENIAM) – a European project focused on standardising best practices for electromyography.

Xsens – A supplier of 3D motion capture products, including the Inertial Measurement Units used in this project.

References

1. Gambon, T.M.; Schmiedeler, J.P.; Wensing, P.M. Effects of User Intent Changes on Onboard Sensor Measurements During Exoskeleton-Assisted Walking. *IEEE Access* **2020**, *8*, 224071–224082, doi:10.1109/ACCESS.2020.3044255.

2. Lee Chang, M.; Gutierrez, R.A.; Khante, P.; Schaertl Short, E.; Lockerd Thomaz, A. Effects of Integrated Intent Recognition and Communication on Human-Robot Collaboration. *IEEE Int. Conf. Intell. Robot. Syst.* **2018**, 3381–3386, doi:10.1109/IROS.2018.8593359.

3. Fotteler, M.L.; Mühlbauer, V.; Brefka, S.; Mayer, S.; Kohn, B.; Holl, F.; Swoboda, W.; Gaugisch, P.; Risch, B.; Denkinger, M.; et al. The Effectiveness of Assistive Technologies for Older Adults and the Influence of Frailty: Systematic Literature Review of Randomized Controlled Trials. *JMIR aging* **2022**, *5*, e31916, doi:10.2196/31916.

4. Bates, T.J.; Fergason, J.R.; Pierrie, S.N. Technological Advances in Prosthesis Design and Rehabilitation Following Upper Extremity Limb Loss. *Curr. Rev. Musculoskelet. Med.* **2020**, *13*, 485–493, doi:10.1007/s12178-020-09656-6.

5. García-Bustillo, Á.; Valiñas-Sieiro, F.; Allende-Río, M.; González-Santos, J.; Cubo, E. Assistive Devices for Personal Mobility in Parkinson's Disease: A Systematic Review of the Literature. *Mov. Disord. Clin. Pract.* **2022**, *9*, 1040–1046,

doi:10.1002/mdc3.13568.

6. Pozaic, T.; Lindemann, U.; Grebe, A.-K.; Stork, W. Sit-to-Stand Transition Reveals Acute Fall Risk in Activities of Daily Living. *IEEE J. Transl. Eng. Heal. Med.* **2016**, *4*, 2700211, doi:10.1109/JTEHM.2016.2620177.

7. Össur i-Limb Ultra Available online: https://www.ossur.com/en-us/prosthetics/arms/i-limb-ultra (accessed on Jul 12, 2020).

8. Ribeiro, J.; Mota, F.; Cavalcante, T.; Nogueira, I.; Gondim, V.; Albuquerque, V.; Alexandria, A. Analysis of man-machine interfaces in upper-limb prosthesis: A review. *Robotics* 2019, *8*, 1–17.

9. Össur i-Limb Quantum Available online: https://www.ossur.com/en-gb/prosthetics/arms/i-limb-quantum (accessed on Aug 17, 2020).

10. Bergmann, J.H.M.; Langdon, P.M.; Mayagoitia, R.E.; Howard, N. Exploring the use of sensors to measure behavioral interactions: An experimental evaluation of using hand trajectories. *PLoS One* **2014**, *9*, 1–10, doi:10.1371/journal.pone.0088080.

11. Ghazaei, G.; Alameer, A.; Degenaar, P.; Morgan, G.; Nazarpour, K. Deep learning-based artificial vision for grasp classification in myoelectric hands. *J. Neural Eng.* **2017**, *14*, doi:10.1088/1741-2552/aa6802.

12. Espinosa, M.; Nathan-Roberts, D. Understanding Prosthetic Abandonment. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2019**, *63*, 1644–1648, doi:10.1177/1071181319631508.

13. Harmon, S.H.E.; Haddow, G.; Gilman, L. New risks inadequately managed: the case of smart implants and medical device regulation. *Law, Innov. Technol.* **2015**, *7*, 231–252, doi:10.1080/17579961.2015.1106107.

14. Navarathna, P.; Bequette, B.W.; Cameron, F. Wearable Device Based Activity Recognition and Prediction for Improved Feedforward Control. In Proceedings of the 2018 Annual American Control Conference (ACC); AACC, 2018; pp. 3571–3576.

15. Brush, A.J.; Albrecht, J.; Miller, R.; Albrecht, J.; Brush, A.J.; Hazas, M. Smart Homes. *IEEE Pervasive Comput.* **2020**, *19*, 69–73, doi:10.1109/MPRV.2020.2977739.

16. Raeiszadeh, M.; Tahayori, H. A novel method for detecting and predicting resident's behavior in smart home. In Proceedings of the 2018 6th Iranian Joint Congress on Fuzzy and Intelligent Systems, CFIS 2018; IEEE, 2018; pp. 71–74.

17. Lew, E.Y.L.; Chavarriaga, R.; Silvoni, S.; Millán, J. del R. Single trial prediction of self-paced reaching directions from EEG signals. *Front. Neurosci.* **2014**, *8*, 1–13, doi:10.3389/fnins.2014.00222.

18. Engel, W.; Ding, W. Reliable Human Fall Prediction with Practical Alert Time and Robot Prediction Algorithm. *2017 Int. Conf. Green Informatics* **2017**, 1–5.

19. Saini, R.; Maan, V. Human Activity and Gesture Recognition: A Review. *2020 Int. Conf. Emerg. Trends Commun. Control Comput.* **2020**, 1–2, doi:10.1109/iconc345789.2020.9117535.

20. Perumal, T.; Chui, Y.L.; Ahmadon, M.A. Bin; Yamaguchi, S. IoT based activity

recognition among smart home residents. *2017 IEEE 6th Glob. Conf. Consum. Electron. GCCE 2017* **2017**, *2017-Janua*, 1–2, doi:10.1109/GCCE.2017.8229478.

21. Alam, M.A.U.; Roy, N.; Misra, A. Tracking and Behavior Augmented Activity Recognition for Multiple Inhabitants. *IEEE Trans. Mob. Comput.* **2019**, 1–1, doi:10.1109/tmc.2019.2936382.

22. Mukhopadhyay Integrating a Bed Sensor in a Smart Home Monitoring System. *I2MTC 2008 - IEEE Int. Instrum. Meas. Technol. Conf. Victoria, Vancouver Island, Canada, May 12-15* **2008**, 15–18.

23. Wang, Z.; Huo, Y. A multi-Attribute fusion acceleration feature selection algorithm for activity recognition on smart phones. *Proc. - 2014 Int. Conf. Inf. Sci. Electron. Electr. Eng. ISEEE 2014* **2014**, *1*, 145–148, doi:10.1109/InfoSEEE.2014.6948085.

24. Haggard, P.; Wing, A. Coordinated responses following mechanical perturbation of the arm during prehension. *Exp. Brain Res.* **1995**, *102*, 483–494, doi:10.1007/BF00230652.

25. Koochaki, F.; Najafizadeh, L. Predicting Intention Through Eye Gaze Patterns. *2018 IEEE Biomed. Circuits Syst. Conf. BioCAS 2018 - Proc.* **2018**, 2018–2021, doi:10.1109/BIOCAS.2018.8584665.

26. Castellanos, J.L.; Gomez, M.F.; Adams, K.D. Using machine learning based on eye gaze to predict targets: An exploratory study. *2017 IEEE Symp. Ser. Comput. Intell. SSCI 2017 - Proc.* **2018**, *2018-Janua*, 1–7, doi:10.1109/SSCI.2017.8285207.

27. Lee, S.; Khan, M.Q.; Husen, M.N. Continuous Car Driving Intent Detection Using Structural Pattern Recognition. *IEEE Trans. Intell. Transp. Syst.* **2020**, 1–13, doi:10.1109/tits.2019.2961928.

28. Argall, B.D. Modular and adaptive wheelchair automation. *Springer Tracts Adv. Robot.* **2016**, *109*, 835–848, doi:10.1007/978-3-319-23778-7_55.

29. Lee, S.; Yoo, J.; Han, G. Gaze-assisted user intention prediction for initial delay reduction in web video access. *Sensors (Switzerland)* **2015**, *15*, 14679–14700, doi:10.3390/s150614679.

30. Henry B, M. Nonparametric Tests Against Trend. *Econometrica* **1945**, *13*, 245–259.

31. PRISMA Available online: http://www.prisma-statement.org/ (accessed on Aug 3, 2020).

32. Kwong, M.T.; Colopy, G.W.; Weber, A.M.; Ercole, A.; Bergmann, J.H.M. The efficacy and effectiveness of machine learning for weaning in mechanically ventilated patients at the intensive care unit: a systematic review. *Bio-Design Manuf.* **2019**, *2*, 31–40, doi:10.1007/s42242-018-0030-1.

33. Joanna Briggs Institute Critical Appraisal Tool Available online: http://joannabriggs.org/research/critical-appraisal-tools.html (accessed on Sep 20, 2018).

34. Zhang, F.; Huang, H. Source selection for real-time user intent recognition toward volitional control of artificial legs. *IEEE J. Biomed. Heal. Informatics* **2013**, *17*, 907–914, doi:10.1109/JBHI.2012.2236563.

35. Planelles, D.; Hortal, E.; Costa, Á.; Úbeda, A.; Iáñez, E.; Azorín, J.M. Evaluating Classifiers to Detect Arm Movement Intention from. *Sensors* **2014**, *14*, 18172–18186, doi:10.3390/s141018172.

36. Chen, B.; Zheng, E.; Wang, Q. A locomotion intent prediction system based on multi-sensor fusion. *Sensors (Switzerland)* **2014**, *14*, 12349–12369, doi:10.3390/s140712349.

37. Zhang, X.; Liu, Y.; Zhang, F.; Ren, J.; Sun, Y.L.; Yang, Q.; Huang, H. On design and implementation of neural-machine interface for artificial legs. *IEEE Trans. Ind. Informatics* **2012**, *8*, 418–429, doi:10.1109/TII.2011.2166770.

38. McGibbon, C.A. A biomechanical model for encoding joint dynamics: Applications to transfemoral prosthesis control. *J. Appl. Physiol.* **2012**, *112*, 1600–1611, doi:10.1152/japplphysiol.01251.2011.

39. Ridi, A.; Zarkadis, N.; Gisler, C.; Hennebert, J. Duration models for activity recognition and prediction in buildings using Hidden Markov Models. *Proc. 2015 IEEE Int. Conf. Data Sci. Adv. Anal. DSAA 2015* **2015**, doi:10.1109/DSAA.2015.7344784.

40. Park, K.; Han, Y.; Lee, Y.K. Activity graph feature selection for activity pattern classification. *Int. J. Distrib. Sens. Networks* **2014**, *2014*, doi:10.1155/2014/254256.

41. Lethaus, F.; Baumann, M.R.K.; Köster, F.; Lemmer, K. A comparison of selected simple supervised learning algorithms to predict driver intent based on gaze data. *Neurocomputing* **2013**, *121*, 108–130, doi:10.1016/j.neucom.2013.04.035.

42. Choi, J.S.; Bang, J.W.; Park, K.R.; Whang, M. Enhanced perception of user intention by combining EEG and Gaze-tracking for brain-computer interfaces (BCIs). *Sensors (Switzerland)* **2013**, *13*, 3454–3472, doi:10.3390/s130303454.

43. Yan, W.; Gao, Y.; Liu, Q. Human-object Interaction Recognition Using Multitask Neural Network. *3rd Int. Symp. Auton. Syst. ISAS 2019* **2019**, 323–328, doi:10.1109/ISASS.2019.8757767.

44. Zhang, K.; Zhang, W.; Xiao, W.; Liu, H.; de Silva, C.W.; Fu, C. Sequential decision fusion for environmental classification in assistive walking. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 1780–1790, doi:10.1109/TNSRE.2019.2935765.

45. Krausz, N.E.; Lenzi, T.; Hargrove, L.J. Depth sensing for improved control of lower limb prostheses. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 2576–2587, doi:10.1109/TBME.2015.2448457.

46. Massalin, Y.; Abdrakhmanova, M.; Varol, H.A. User-independent intent recognition for lower limb prostheses using depth sensing. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 1759–1770, doi:10.1109/TBME.2017.2776157.

47. Guo, W.; Sheng, X.; Liu, J.; Hua, L.; Zhang, D.; Zhu, X. Towards zero training for myoelectric control based on a wearable wireless sEMG armband. *IEEE/ASME Int. Conf. Adv. Intell. Mechatronics, AIM* **2015**, *2015-Augus*, 196–201, doi:10.1109/AIM.2015.7222531.

48. Hiyama, T.; Kato, Y.; Inoue, T. Sit-to-stand assistance system based on using EMG to predict movement. *RO-MAN 2017 - 26th IEEE Int. Symp. Robot Hum. Interact.*

Commun. **2017**, *2017-Janua*, 81–87, doi:10.1109/ROMAN.2017.8172284.

49. Paek, A.Y.; Gailey, A.; Parikh, P.; Santello, M.; Contreras-Vidal, J. Predicting hand forces from scalp electroencephalography during isometric force production and object grasping. *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS* **2015**, *2015-Novem*, 7570–7573, doi:10.1109/EMBC.2015.7320144.

50. Schreiber, M.A.; Trkov, M.; Merryweather, A. Influence of Frequency Bands in EEG Signal to Predict User Intent. *Int. IEEE/EMBS Conf. Neural Eng. NER* **2019**, *2019-March*, 1126–1129, doi:10.1109/NER.2019.8716947.

51. Wilson, S.; Vaidyanathan, R. Upper-limb prosthetic control using wearable multichannel mechanomyography. *IEEE Int. Conf. Rehabil. Robot.* **2017**, 1293–1298, doi:10.1109/ICORR.2017.8009427.

52. Geyik, C.S.; Dutta, A.; Ogras, U.Y.; Bliss, D.W. Decoding human intent using a wearable system and multi-modal sensor data. *Conf. Rec. - Asilomar Conf. Signals, Syst. Comput.* **2017**, 846–850, doi:10.1109/ACSSC.2016.7869168.

53. Xiao, Z.G.; Menon, C. Towards the development of a wearable feedback system for monitoring the activities of the upper-extremities. *J. Neuroeng. Rehabil.* **2014**, *11*, 1–13, doi:10.1186/1743-0003-11-2.

54. Yap, H.K.; Mao, A.; Goh, J.C.H.; Yeow, C.H. Design of a wearable FMG sensing system for user intent detection during hand rehabilitation with a soft robotic glove. *Proc. IEEE RAS EMBS Int. Conf. Biomed. Robot. Biomechatronics* **2016**, *2016-July*, 781–786, doi:10.1109/BIOROB.2016.7523722.

55. Shen, B.; Li, J.; Bai, F.; Chew, C.M. Motion intent recognition for control of a lower extremity assistive device (LEAD). *2013 IEEE Int. Conf. Mechatronics Autom. IEEE ICMA 2013* **2013**, 926–931, doi:10.1109/ICMA.2013.6618039.

56. Zhang, H.; Parker, L.E. Bio-inspired predictive orientation decomposition of skeleton trajectories for real-time human activity prediction. *Proc. - IEEE Int. Conf. Robot. Autom.* **2015**, *2015-June*, 3053–3060, doi:10.1109/ICRA.2015.7139618.

57. Muangsrinoon, S.; Boonbrahm, P. Using ensemble algorithms for physical activity recognition prediction. *AIP Conf. Proc.* **2018**, *2016*, doi:10.1063/1.5055504.

58. Aipperspach, R.; Cohen, E.; Canny, J. Modeling human behavior from simple sensors in the home. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **2006**, *3968 LNCS*, 337–348, doi:10.1007/11748625_21.

59. Wang, C.; De, D.; Song, W.Z. Trajectory mining from anonymous binary motion sensors in Smart Environment. *Knowledge-Based Syst.* **2013**, *37*, 346–356, doi:10.1016/j.knosys.2012.08.020.

60. Stefanou, T.; Turton, A.; Lenz, A.; Dogramadzi, S. Upper limb motion intent recognition using tactile sensing. *IEEE Int. Conf. Intell. Robot. Syst.* **2017**, *2017-Septe*, 6601–6608, doi:10.1109/IROS.2017.8206573.

61. De Chazal, P.; O'Dwyer, M.; Reilly, R.B. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 1196–1206, doi:10.1109/TBME.2004.827359.

62. Jennings, J.R. Beat-by-beat vascular responses during anticipatory heart rate deceleration. *Physiol. Psychol.* **1982**, *10*, 422–430, doi:10.3758/BF03332976.

63. Pantelopoulos, A.; Bourbakis, N.G. A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis. *IEEE Trans. Syst. Man Cybern. - Part C Appl. Rev.* **2010**, *40*, 1–12.

64. Mills, K.R. The Basics of Electromyography. *J. Neurol. Neurosurg. Psychiatry* **2005**, *76*, 32–35, doi:10.1136/jnnp.2005.069211.

65. Olson, J.S.; Redkar, S. A survey of wearable sensor networks in health and entertainment. *MOJ Appl. Bionics Biomech.* **2018**, *2*, 280–287.

66. Wang, G.G. Definition and Review of Virtual Prototyping. *J. Comput. Inf. Sci. Eng.* **2003**, *2*, 232–236, doi:10.1115/1.1526508.

67. Tilak, S.; Abu-ghazaleh, N.B.; Heinzelman, W. A Taxonomy of Wireless Micro-Sensor Network Models. *ACM SIGMOBILE Mob. Comput. Commun. Rev.* **2002**, *6*, 28–36, doi:10.1145/565702.565708.

68. Bergmann, J.; Noble, A.; Thompson, M. Probabilistic sensor network design. In Proceedings of the 2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN); San Francisco, CA, 2016; pp. 66–70.

69. Stokes, I.A.F.; Henry, S.M.; Single, R.M. Surface EMG electrodes do not accurately record from lumbar multifidus muscles. *Clin. Biomech. (Bristol, Avon)* **2003**, *18*, 9–13.

70. Winter, D.A.; Fuglevand, A.J.; Archer, S.E. Crosstalk in Surface Electromyography: Theoretical and Practical Estimates. *J. Electromyogr. Kinesiol.* **1994**, *4*, 15–26.

71. Trevethan, R. Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. *Front. Public Heal.* **2017**, *5*, 307.

72. Philips, K.; van Roermund, A. *Sigma Delta A/D Conversion for Signal Conditioning*; 2006 editi.; Springer: New York, 2006; ISBN 978-1402046797.

73. Giannakis, G.B. Statistical Signal Processing. In *Digital Signal Processing Handbook*; Madisetti, V.K., Williams, D.B., Eds.; CRC Press: Boca Raton, 1999; pp. V–1 ISBN 978-0849385728.

74. Dawid, A.P. Conditional Independence in Statistical Theory. *J. R. Stat. Soc. Ser. B* **1979**, *41*, 1–31.

75. Ragan, A. Taking the Confusion Out of Confusion Matrices Available online: https://towardsdatascience.com/taking-the-confusion-out-of-confusion-matrices-c1ce054b3d3e (accessed on Jan 1, 2019).

76. Bai, Y. Enhancement of the Reliability of an Embedded Surveillance System by Multiple Sensors Using a Majority Voting Mechanism. In Proceedings of the 2009 IEEE Instrumentation and Measurement Technology Conference; IEEE: Singapore, 2009; pp. 1157–1162.

77. PLUX Wireless Biosignals S.A. BITalino Plugged Kit Data Sheet Available online: http://bitalino.com/datasheets/BITalino_Plugged_Datasheet.pdf (accessed on Jan

4, 2019).

78. SENIAM Recommendations for Sensor Locations on Individual Muscles Available online: http://www.seniam.org (accessed on Aug 12, 2022).

79. imusic-school Online Metronome Available online: https://www.imusic-school.com/en/tools/online-metronome/ (accessed on Dec 4, 2018).

80. De Luca, C.J. Surface Electromyography: Detection and Recording. *Delsys Inc.* **2002**.

81. Fahidy, T.Z. Some Applications of Bayes' Rule in Probability Theory to Electrocatalytic Reaction Engineering. *Int. J. Electrochem.* **2011**, *2011*, 1–6, doi:10.4061/2011/404605.

82. Khrennikov, A. Quantum Bayesianism as the basis of general theory of decision-making. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*.

83. Barnston, A.G. Correspondence among the Correlation, RMSE, and Heidke Forecast Verification Measures; Refinement of the Heidke Score. *Weather Forecast.* **1992**, *7*, 699–709.

84. Özgünen, K.T.; Çelik, U.; Kurdak, S.S. Determination of an Optimal Threshold Value for Muscle Activity Detection in EMG Analysis. *J. Sport. Sci. Med.* **2010**, *9*, 620–628.

85. Martens, J.; Daly, D.; Deschamps, K.; Jorge, R.; Fernandes, P. Intra-Individual Variability of Surface Electromyography in Front Crawl Swimming. *PLoS One*

2015, *10*, 1–13, doi:10.1371/journal.pone.0144998.

86. Karam, A.; Alnajjar, F.; Gochoo, M. Assistive and rehabilitation robotics for upper limb impairments in post-stroke patients: Evaluation criteria for the design and functionality. *2020 Adv. Sci. Eng. Technol. Int. Conf. ASET 2020* **2020**, 14–17, doi:10.1109/ASET48392.2020.9118320.

87. McDonald, C.L.; Westcott-McCoy, S.; Weaver, M.R.; Haagsma, J.; Kartin, D. Global prevalence of traumatic non-fatal limb amputation. *Prosthet. Orthot. Int.* **2020**, doi:10.1177/0309364620972258.

88. Chadwell, A.; Kenney, L.; Thies, S.; Galpin, A.; Head, J. The Reality of Myoelectric Prostheses: Understanding What Makes These Devices Difficult for Some Users to Control. *Front. Neurorobot.* **2016**, *10*, 7, doi:10.3389/fnbot.2016.00007.

89. Childress, D. Historical aspects of powered limb prostheses. *Clin. Prosthetics Orthot.* **1985**, *9*, 2–13.

90. Bennett, D.A.; Goldfarb, M. IMU-Based Wrist Rotation Control of a Transradial Myoelectric Prosthesis. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 419–427, doi:10.1109/TNSRE.2017.2682642.

91. Comito, C.; Falcone, D.; Forestiero, A. Current Trends and Practices in Smart Health Monitoring and Clinical Decision Support. *Proc. - 2020 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2020* **2020**, 2577–2584, doi:10.1109/BIBM49941.2020.9313449.

92. Liu, R.; Liu, M. Recognizing human activities based on multi-sensors fusion. *2010*

4th Int. Conf. Bioinforma. Biomed. Eng. iCBBE 2010 **2010**, 2–5, doi:10.1109/ICBBE.2010.5514802.

93. Kim, K.; Yun, G.; Park, S.K.; Kim, D.H. Fall Detection for the Elderly Based on 3-Axis Accelerometer and Depth Sensor Fusion with Random Forest Classifier. *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS* **2019**, 4611–4614, doi:10.1109/EMBC.2019.8856698.

94. Vasuhi, S.; Vaidehi, V.; Midhunkrishna, P.R. Multiple target tracking using Support Vector Machine and data fusion. *3rd Int. Conf. Adv. Comput. ICoAC 2011* **2011**, 407–411, doi:10.1109/ICoAC.2011.6165210.

95. Nagarja, V..; Cheng, R.; Kwong, E.; Bergmann, J.H.; Anderson, M.S.; Thompson, M.S. Marker-based vs Intertial-based Motion CaptureL Musculoskeletaal Modelling of Upper Extremity Kinetics. *Trent Int. Prosthetics Symp.* **2019**, 1–3.

96. XSens Mtw Awinda Available online: https://www.xsens.com/products/mtw-awinda (accessed on Aug 10, 2021).

97. Xsens MVN Analyze Available online: https://www.xsens.com/products/mvn-analyze (accessed on Dec 24, 2021).

98. Vicon Nexus Available online: https://www.vicon.com/software/nexus/ (accessed on Aug 10, 2021).

99. Xsens Syncronising Xsens systems with Vicon Nexus Available online: https://www.xsens.com/hubfs/Downloads/plugins tools/SynchronisingXsenswithVicon.pdf.

100. Nazmi, N.; Rahman, M.A.A.; Yamamoto, S.I.; Ahmad, S.A.; Zamzuri, H.; Mazlan, S.A. A review of classification techniques of EMG signals during isotonic and isometric contractions. *Sensors (Switzerland)* **2016**, *16*, 1–28, doi:10.3390/s16081304.

101. Spiewak, C. A Comprehensive Study on EMG Feature Extraction and Classifiers. *Open Access J. Biomed. Eng. Biosci.* **2018**, *1*, 17–26, doi:10.32474/oajbeb.2018.01.000104.

102. Osman, H.; Ghafari, M.; Nierstrasz, O. Hyperparameter optimization to improve bug prediction accuracy. *MaLTeSQuE 2017 - IEEE Int. Work. Mach. Learn. Tech. Softw. Qual. Eval. co-located with SANER 2017* **2017**, 33–38, doi:10.1109/MALTESQUE.2017.7882014.

103. Ruano, P.; Delgado, L.L.; Picco, S.; Villegas, L.; Tonelli, F.; Merlo, M.; Rigau, J.; Diaz, D.; Masuelli, M. Artificial Human Arm Driven by EMG Signal. *Intech* **2016**, 13.

104. Daniel, W.W. Spearman rank correlation coefficient. In *Applied Nonparametric Statistics*; PWS-Kent: Boston, 1990; pp. 358–365 ISBN 978-0-534-91976-4.

105. Kose, M.; Incel, O.D.; Ersoy, C. Online Human Activity Recognition on Smart Phones. *Perform. Eval.* **2012**, 0–4.

106. Medical Center Orthotics & Prosthetics Arm & Hand Prosthetics Available online: https://mcopro.com/blog/resources/arm-hand-prosthetics (accessed on Feb 13, 2022).

107. Marino, M.; Pattni, S.; Greenberg, M.; Miller, A.; Hocker, E.; Ritter, S.; Mehta, K.

Access to prosthetic devices in developing countries: Pathways and challenges. *Proc. 5th IEEE Glob. Humanit. Technol. Conf. GHTC 2015* **2015**, 45–51, doi:10.1109/GHTC.2015.7343953.

108. Wu, W.; Dasgupta, S.; Ramirez, E.E.; Peterson, C.; Norman, G.J. Classification accuracies of physical activities using smartphone motion sensors. *J. Med. Internet Res.* **2012**, *14*, 1–9, doi:10.2196/jmir.2208.

109. Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A.J. Machine learning algorithm validation with a limited sample size. *PLoS One* **2019**, *14*, 1–20, doi:10.1371/journal.pone.0224365.

110. Chadwell, A.; Kenney, L.; Granat, M.; Thies, S.; Galpin, A.; Head, J. Upper limb activity of twenty myoelectric prosthesis users and twenty healthy anatomically intact adults. *Sci. Data* **2019**, *6*, 1–11, doi:10.1038/s41597-019-0211-6.

111. Wei, S.J.; Zhang, B.; Tan, X.W.; Zhao, X.G.; Ye, D. A Real-time Human Activity Recognition Approach with Generalization Performance. *Chinese Control Conf. CCC* **2020**, *2020-July*, 6334–6339, doi:10.23919/CCC50068.2020.9188860.

112. Cloutier, A.; Yang, J. Control of Hand Prostheses: A Literature Review. In Proceedings of the Proceedings of the ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference; Portland, 2013; Vol. 6A.

113. Schultz, A.E.; Kuiken, T.A. Neural Interfaces for Control of Upper Limb Prostheses: The State of the Art and Future Possibilities. *PM&R* **2011**, *3*, 55–67,

doi:https://doi.org/10.1016/j.pmrj.2010.06.016.

114. Lovely, D.F. Signals and Signal Processing for Myoelectric Control BT - Powered Upper Limb Prostheses: Control, Implementation and Clinical Application. In; Muzumdar, A., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2004; pp. 35–54 ISBN 978-3-642-18812-1.

115. Boxtel, A.; Damen, E.J.P.; Brunia, C.H.M. Anticipatory EMG responses of pericranial muscles in relation to heart rate during a warned simple reaction time task. *Psychophysiology* **1996**, *33*, 576–583, doi:10.1111/j.1469-8986.1996.tb02434.x.

116. Hameed, H.K.; Hassan, W.Z.W.; Shafie, S.; Ahmad, S.A.; Jaafar, H. A Review on Surface Electromyography-Controlled Hand Robotic Devices Used for Rehabilitation and Assistance in Activities of Daily Living. *JPO J. Prosthetics Orthot.* **2020**, *32*.

117. Zhang, X.; Huang, H. A real-time, practical sensor fault-tolerant module for robust EMG pattern recognition. *J. Neuroeng. Rehabil.* **2015**, *12*, 18, doi:10.1186/s12984-015-0011-y.

118. PLUX Wireless Biosignals MuscleBIT BT Available online: https://plux.info/bundles/426-bitalino-revolution-musclebit-bt.html.

119. ProHands Gripmaster Available online: https://prohands.net/collections/frontpage/products/gripmaster (accessed on Jan 5, 2019).

120. Farrell, T.R.; Weir, R.F. The Optimal Controller Delay for Myoelectric Prostheses.

IEEE Trans. Neural Syst. Rehabil. Eng. **2007**, *15*, 111–118, doi:10.1109/tnsre.2007.891391.

121. Woods, D.L.; Wyma, J.M.; Yund, E.W.; Herron, T.J.; Reed, B. Factors influencing the latency of simple reaction time. *Front. Hum. Neurosci.* **2015**, *9*.

122. John Clarkson, P.; Coleman, R. History of Inclusive Design in the UK. *Appl. Ergon.* **2015**, *46*, 235–247, doi:https://doi.org/10.1016/j.apergo.2013.03.002.

123. Kadir, S.A.; Jamaludin, M. Universal Design as a Significant Component for Sustainable Life and Social Development. *Procedia - Soc. Behav. Sci.* **2013**, *85*, 179–190, doi:https://doi.org/10.1016/j.sbspro.2013.08.349.

124. Willis, A.W.; Roberts, E.; Beck, J.C.; Fiske, B.; Ross, W.; Savica, R.; Van Den Eeden, S.K.; Tanner, C.M.; Marras, C.; Alcalay, R.; et al. Incidence of Parkinson disease in North America. *npj Park. Dis.* **2022**, *8*, 170, doi:10.1038/s41531-022-00410-y.

125. Balestrino, R.; Schapira, A.H. V Parkinson disease. *Eur. J. Neurol.* **2020**, *27*, 27–42, doi:10.1111/ene.14108.

126. Yuan, H.; Zhang, Z.-W.; Liang, L.-W.; Shen, Q.; Wang, X.-D.; Ren, S.-M.; Ma, H.-J.; Jiao, S.-J.; Liu, P. Treatment strategies for Parkinson's disease. *Neurosci. Bull.* **2010**, *26*, 66–76, doi:10.1007/s12264-010-0302-z.

127. Lora-Millan, J.S.; Delgado-Oleas, G.; Benito-León, J.; Rocon, E. A Review on Wearable Technologies for Tremor Suppression. *Front. Neurol.* 2021, *12*.

128. Kim, J.; Lee, G.; Heimgartner, R.; Arumukhom Revi, D.; Karavas, N.; Nathanson,

D.; Galiana, I.; Eckert-Erdheim, A.; Murphy, P.; Perry, D.; et al. Reducing the metabolic rate of walking and running with a versatile, portable exosuit. *Science.* **2019**, *365*, 668–672, doi:10.1126/science.aav7536.

129. Verbaan, D.; Marinus, J.; Visser, M.; van Rooden, S.M.; Stiggelbout, A.M.; Middelkoop, H.A.M.; van Hilten, J.J. Cognitive impairment in Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* **2007**, *78*, 1182–1187, doi:10.1136/jnnp.2006.112367.

130. Staunton, H.; Kelly, K.; Newton, L.; Leddin, M.; Rodriguez-Esteban, R.; Chaudhuri, K.R.; Weintraub, D.; Postuma, R.B.; Martinez-Martin, P. A Patient-Centered Conceptual Model of Symptoms and Their Impact in Early Parkinson's Disease: A Qualitative Study. *J. Parkinsons. Dis.* **2022**, *12*, 137–151, doi:10.3233/JPD-202457.

131. Cybart, N. Apple Watch Is Now Worn on 100 Million Wrists Available online: https://www.aboveavalon.com/notes/2021/2/11/apple-watch-is-now-worn-on-100-million-wrists (accessed on Jan 17, 2023).

132. Laricchia, F. Fitbit - statistics & facts Available online: https://www.statista.com/topics/2595/fitbit (accessed on Jan 17, 2023).

133. Blatchford Linx Available online: https://www.blatchfordmobility.com/en-us/products/limb-systems/linx/ (accessed on Feb 8, 2023).

134. StackCare UK Ltd StackCare Available online: https://www.stackcare.co.uk/ (accessed on Feb 8, 2023).

135. Yapıcı, M.M.; Tekerek, A.; Topaloğlu, N. Literature Review of Deep Learning Research Areas. *Gazi J. Eng. Sci.* **2019**, *5*, 188–215.

136. Kamath, C.N.; Bukhari, S.S.; Dengel, A. Comparative Study between Traditional Machine Learning and Deep Learning Approaches for Text Classification. In Proceedings of the Proceedings of the ACM Symposium on Document Engineering 2018; Association for Computing Machinery: New York, NY, USA, 2018.

137. Shaheen, F.; Verma, B.; Asafuddoula, M. Impact of Automatic Feature Extraction in Deep Learning Architecture. In Proceedings of the 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA); 2016; pp. 1–8.

138. Jain, A.; Bansal, R.; Kumar, A.; Singh, K. A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. *Int. J. Appl. Basic Med. Res.* **2015**, *5*, 124, doi:10.4103/2229-516X.157168.

139. Graves, A.; Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM networks. In Proceedings of the Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.; 2005; Vol. 4, pp. 2047–2052 vol. 4.

140. Rajendran, G.; Poornachandran, P.; Chitturi, B. Deep learning model on stance classification. In Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI); 2017; pp. 2407–2409.

141. Bhaskar, N.; Suchetha, M.; Philip, N.Y. Time Series Classification-Based Correlational Neural Network With Bidirectional LSTM for Automated Detection of Kidney Disease. *IEEE Sens. J.* **2021**, *21*, 4811–4818, doi:10.1109/JSEN.2020.3028738.

142. Bergmann, J.H.M.; Chandaria, V.; McGregor, A. Wearable and Implantable Sensors: The Patient's Perspective. *Sensors* **2012**, *12*, 16695–16709, doi:10.3390/s121216695.

143. Kulkarni, P.; Kirkham, R.; McNaney, R. Opportunities for Smartphone Sensing in E-Health Research: A Narrative Review. *Sensors* **2022**, *22*, doi:10.3390/s22103893.

144. Chandrabhatla, A.S.; Pomeraniec, I.J.; Ksendzovsky, A. Co-evolution of machine learning and digital technologies to improve monitoring of Parkinson's disease motor symptoms. *NPJ Digit. Med.* **2022**, *5*, 32, doi:10.1038/s41746-022-00568-y.

145. Firmansyah, M.R.; Hidayat, R.; Bejo, A. Comparison of Windowing Function on Feature Extraction Using MFCC for Speaker Identification. In Proceedings of the 2021 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA); 2021; pp. 1–5.

146. Tiwari, V. MFCC and its applications in speaker recognition. *Int. J. Emerg. Technol.* **2010**, *1*, 19–22.

147. Hasan, M.R.; Hasan, M.M.; Hossain, M.Z. How many Mel-frequency cepstral coefficients to be utilized in speech recognition? A study with the Bengali language. *J. Eng.* **2021**, *2021*, 817–827, doi:https://doi.org/10.1049/tje2.12082.

148. Uddin, M.Z.; Soylu, A. Human activity recognition using wearable sensors, discriminant analysis, and long short-term memory-based neural structured learning. *Sci. Rep.* **2021**, *11*, 16455, doi:10.1038/s41598-021-95947-y.

149. Preatoni, E.; Nodari, S.; Lopomo, N.F. Supervised Machine Learning Applied to Wearable Sensor Data Can Accurately Classify Functional Fitness Exercises Within a Continuous Workout. *Front. Bioeng. Biotechnol.* 2020, *8*.

150. Mohamad, S.; Sayed-Mouchaweh, M.; Bouchachia, A. Online active learning for human activity recognition from sensory data streams. *Neurocomputing* **2020**, *390*, 341–358, doi:10.1016/j.neucom.2019.08.092.

151. von Rosen, P.; Hagströmer, M.; Franzén, E.; Leavy, B. Physical activity profiles in Parkinson's disease. *BMC Neurol.* **2021**, *21*, 71, doi:10.1186/s12883-021-02101-2.

152. Waller, S.; Bradley, M.; Hosking, I.; Clarkson, P.J. Making the case for inclusive design. *Appl. Ergon.* **2015**, *46*, 297–303, doi:https://doi.org/10.1016/j.apergo.2013.03.012.

153. Össur Össur and Alfred Mann Foundation Sign New Agreement to Extend Exploration of Mind-Controlled Prosthetics Available online: https://www.ossur.com/en-us/about-ossur/newsroom/ossur-and-alfred-mann-foundation-sign-new-agreement (accessed on Jul 21, 2023).

154. Musk, E. An Integrated Brain-Machine Interface Platform With Thousands of Channels. *J Med Internet Res* **2019**, *21*, e16194, doi:10.2196/16194.

155. Al-Faiz, M.Z.; Ali, A.A.; Miry, A.H. A k-nearest neighbor based algorithm for

human arm movements recognition using EMG signals. In Proceedings of the EPC-IQ01 2010 - 2010 1st International Conference on Energy, Power and Control; 2010; Vol. 6, pp. 159–167.

156. Mukaka, M.M. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Med. J.* **2012**, *24*, 69–71.

157. Jolliffe, I.T. Definition and Derivation of Principal Components. In *Principal Component Analysis*; Springer: Berlin, 2002; pp. 1–6 ISBN 9780080448947.

Acknowledgements

Firstly, I must thank my supervisor, Prof. Jeroen Bergmann, for his advice and support throughout the project. Without his encouragement and guidance, I would not have begun this DPhil, let alone finished it, and so in some sense, all of this is his fault.

Thanks go to Dr Vikranth Nagaraja for providing the dataset used in Chapter 4, and to Jemma Inches and Prof. Camille Carroll for their work collecting the data used in Chapters 6 and 7, and for their feedback on the papers themselves.

I would also like to thank my college advisors at Worcester, Prof. Ton van der Bremer, Dr Brian Tang and Prof. Wouter Mostert, for a number of supportive Zoom chats and excellent free lunches in the SCR.

I would like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out the analysis in Chapters 6 and 7 (http://dx.doi.org/10.5281/zenodo.22558).

Thanks go to Prof. Perla Maiolino and Prof. David Clifton for their feedback on my Transfer of Status and Confirmation of Status reports, and to Prof. Maiolino (again) and Prof. Ravi Vaidyanathan for examining the final thesis. Fingers crossed it passed muster.

Finally, this DPhil is dedicated to my father, Iain, who passed away in 2023. I would like to thank all my colleagues, friends, family and in particular my mother Judi and brother James for continuing to support me during this time. Miss you, Dad.