URL: https://hype-submit.aha-journals.org/

Manuscript Number: HYPE/2023/22529-T2

Title: Artificial intelligence derived risk prediction: a novel risk calculator using office and ambulatory blood pressure

Authors:
Pedro Guimarães (Saarland University)
Andreas Keller (Saarland University)
Michael Bohm (Saarland University Medical Center)
Lucas Lauder (Universitätsklinikum des Saarlandes)
Tobias Fehlmann (Saarland University)
Luis Ruilope (Hospital Universitario 12 de Octubre)
Ernest Vinyoles (La Mina Primary Care Center. University of Barcelona)
Manuel Gorostidi (Hospital Universitario Central de Asturias)
Julián Segura (Hospital Universitario 12 de Octubre)
Gema Ruiz-Hurtado (Hospital Universitario 12 de Octubre)
Natalie Staplin (University of Oxford)
Bryan Williams (University College London)
Alejandro de la Sierra (University of Barcelona)
Felix Mahfoud (Saarland University Hospital)

Artificial intelligence derived risk prediction: a novel risk calculator using office and ambulatory blood pressure

Pedro Guimarães[1,2], Andreas Keller[1,3], Michael Böhm[4], Lucas Lauder[4], Tobias Fehlmann[1], Luis M. Ruilope[5,6,7], Ernest Vinyoles[8], Manuel Gorostidi[9], Julián Segura[5], Gema Ruiz-Hurtado[5,6], Natalie Staplin[10], Bryan Williams[11], Alejandro de la Sierra[12], Felix Mahfoud[4,13]

1 Chair for Clinical Bioinformatics, Saarland University, Saarbrücken, Germany

2 University of Coimbra, Coimbra Institute for Biomedical Imaging and Translational Research (CIBIT), Institute for Nuclear Sciences Applied to Health (ICNAS), Coimbra, Portugal

3 Department of Neurology and Neurological Sciences, Stanford University, Stanford, CA, USA

4 Department of Internal Medicine III, Cardiology, Angiology, Intensive Care Medicine, Universitätsklinikum des Saarlandes, Saarland University, Homburg/Saar, Germany

5 Cardiorenal Translational Laboratory & Hypertension Unit, Institute of Research i+12, Hospital Universitario 12 de Octubre, Madrid, Spain

6 CIBER-CV, Hospital Universitario 12 de Octubre, Madrid, Spain

7 Faculty of Sport Sciences, European University of Madrid, Madrid, Spain

8 La Mina Primary Care Center, University of Barcelona; IDIAP Jordi Gol, Barcelona, Spain

9 Department of Nephrology, Hospital Universitario Central de Asturias, RedinRen, Oviedo, Spain

10 Medical Research Council Population Health Research Unit, Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK

11 University College London (UCL), Institute of Cardiovascular Science, National Institute for Health Research, UCL Hospitals Biomedical Research Centre, London, UK

12 Department of Internal Medicine, Hospital Universitario Mútua Terrasa, Universidad de Barcelona, Terrasa, Barcelona, Spain

13 Institute for Medical Engineering and Science, MIT, Cambridge, MA, USA

X: @FelixMahfoud; @Lucas Lauder

Short title: Artificial intelligence derived risk prediction

Address for correspondence:
Professor Dr. Felix Mahfoud, MD, MA
Klinik für Innere Medizin III
Universitätsklinikum des Saarlandes/Saarland University
Kirrberger Str. 100, IMED, Geb. 41.1
66421 Homburg/Saar, Germany
Telephone: +49 6841 16 15911
Fax: +49 6841 16 15910
E-Mail: felix.mahfoud@uks.eu

Abstract

Background

Quantification of total CV risk is essential for individualizing hypertension treatment. This study aimed to develop and validate a novel, machine learning-derived model to predict cardiovascular (CV) mortality risk using office (OBP) and ambulatory blood pressure (ABP).

Methods

The performance of the novel risk score was compared with existing risk scores, and the possibility of predicting ABP phenotypes utilizing clinical variables was assessed. Using data from 59,124 patients enrolled in the Spanish ABP monitoring registry, machine-learning approaches (logistic regression, gradient-boosted decision trees, and deep neural networks) and stepwise forward feature selection were employed.

Results

For the prediction of CV mortality, deep neural networks yielded the highest clinical performance. The novel mortality prediction models using OBP (CV-Mortality$_{OBP}$) and ABP (CV-Mortality$_{ABP}$) outperformed other risk scores. The area under the curve (AUC) achieved by the novel approach, already when using OBP variables, was significantly higher when compared with the AUC of the Framingham, SCORE-2, and ASCVD risk scores. However, the prediction of CV mortality with ABP instead of OBP data significantly increased the AUC (0.870 vs. 0.865, $p = 3.61\times10^{-28}$), accuracy, and specificity, respectively. The prediction of ABP phenotypes (i.e., white-coat, ambulatory, and masked hypertension) using clinical characteristics was limited.

Conclusions

The ROC curves for CV-Mortality$_{OBP}$ and CV-Mortality$_{ABP}$ with deep neural networks models outperformed all other risk metrics, indicating the potential for improving current risk

scores by applying state-of-the-art machine learning approaches. The prediction of CV mortality using ABP data led to a significant increase in AUC and performance metrics.

Key Words

Risk prediction, ambulatory blood pressure, office blood pressure, machine learning, artificial intelligence, neural networks.

Abbreviations

ABP – ambulatory blood pressure

AH – ambulatory hypertension

ASCVD – Atherosclerotic Cardiovascular Disease

AUC – area under the curve

BP – blood pressure

CV – cardiovascular

DBP – diastolic blood pressure

ESH – European Society of Hypertension

ESC – European Society of Cardiology

FRS – Framingham Risk Score

MH – masked hypertension

OBP – office blood pressure

SBP – systolic blood pressure

SCORE – Systemic Coronary Risk Estimation

SH – sustained hypertension

WH – white-coat hypertension

Introduction

Hypertension continues to be the leading risk factor for premature death worldwide[1] and typically clusters with other cardiovascular (CV) risk factors, hence quantification of total CV risk is essential for individualizing treatment. For this purpose, several risk assessment systems exist, such as the high-risk Systemic Coronary Risk Estimation (SCORE-2), the Atherosclerotic Cardiovascular Disease (ASCVD) score, or the Framingham Risk Score.[2] Office blood pressure (OBP) is integral to these risk scores. Ambulatory BP (ABP) is more informative about all-cause and CV mortality than OBP values.[3,4] Further, ABP monitoring (ABPM) allows for the differentiation of various hypertension phenotypes (i.e., white-coat and masked hypertension) and provides prognostic information for the risk of CV and renal disease.[5-7] However, the value of integrating ABP as part of risk prediction scores has not been systematically analyzed. The 2018 European Society of Cardiology (ESC)/European Society of Hypertension (ESH) and the 2023 ESH guidelines for the management of hypertension[7] emphasize the importance of ABP recordings in selected patients, particularly in individuals with suspected white-coat and masked hypertension. Given the limited resources and inability to perform ABPM in all hypertensive individuals, identifying clinical characteristics associated with either condition is warranted to potentially reduce the number of unnecessarily performed ABP recordings.

Machine learning approaches set a new standard for processing and analyzing data. These methods allow us to explore linear and nonlinear multi-dimensional relationships in large datasets to extract relevant patterns for improved classification and regression models. Several studies have shown the potential benefits of machine learning in multiple medical fields such as neurology[8,9], gastroenterology[10,11], and cardiology.[12,13] The present study utilized a completely new database comprising 59,124 patients included in the Spanish

Ambulatory Blood Pressure Monitoring Registry and used machine learning to i) develop a risk score for the prediction of CV mortality with and without ABP data and compare its performance to established approaches and ii) assess whether clinical variables including OBP can predict ABP phenotypes (namely white-coat, ambulatory and masked hypertension).

Methods

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Data source

This study evaluates data from the Spanish ABP Study, a national study performed at 223 primary care centers within the Spanish National Health System. Baseline data, protocols, inclusion and exclusion criteria were described elsewhere and recently updated.[4,14-16] Patients were selected by their physicians within the Spanish National Health system, had to be 18 years or older, and had guideline-recommended indications for ABPM, including suspected white-coat, high-risk, borderline and labile hypertension, as well as the assessment of BP control.[17] The study was sponsored by the Spanish Society of Hypertension, Lacer Laboratories, and European government agencies. The sponsors had no role in the study's design, data analyses, and interpretation. All patients provided written informed consent and institutional review boards approved the protocols.

Office blood pressure (OBP) was measured with validated oscillometric devices (in 85% of patients) or calibrated mercury sphygmomanometers (15%) after 5 minutes of rest while seated, using standardized procedures. The mean of two OBP readings was calculated and used for analysis. ABPM was performed using validated oscillometric devices (Spacelabs

90207, Snoqualmie, WA, USA) with appropriate cuff sizes, programmed to record blood pressure at 20-minute intervals for the day and 30-minute intervals for the night. The mean of all valid readings (based on pre-defined quality criteria, including successful recording of ≥70% of systolic and diastolic pressure readings for 24 hours) was used. Day and night periods were defined according to the patient´s self-reported data on sleeping and waking times. Other variables were obtained from personal interviews and physical examinations at study visits or from clinical records. The full list of variables considered in this study, along with additional details, are presented in the Table S1.

The date and cause of death were ascertained by a computerized search of the vital registry of the Spanish National Institute of Statistics (contract 20534 between the University of Barcelona and the National Institute of Statistics), which has been shown to be accurate and reliable with complete coverage.[18] Individuals were designated as having died if their status in the vital registry was marked as deceased. The cause of death was determined from the death certificate and was coded according to the 10th Revision of the International Statistical Classification of Diseases. Deaths with codes in the range I00-I99 were classified as of cardiovascular origin. For each study participant, follow-up was from the date of their recruitment to the date of death or December 31, 2019, whichever occurred first.

A total of 59,746 subjects, aged 18 or older were included, out of which 622 (1%) were excluded because of incomplete information. This analysis comprised 59,124 patients. Along with OBP and ABP, 60 clinically relevant variables were obtained from interviews and physical examinations during visits and from patient records. The data analysis workflow is shown in Figure 1.

Statistical analysis and performance metrics

Continuous data were presented as mean ± standard deviation (SD), and categorical data as frequencies and/or percentages. Before classification, all variables were normalized to have zero mean and unit variance. Missing values were replaced by the mean of all available values for the same variable. The percentage of missing data for each of the selected variables was computed per class and feature and are shown in Figure S1. Variables with more than 50% missing for any class were excluded. There was no set limit for discrepancy (missing percentage difference cardiac death vs alive) put in place since upon checking no feature presented a higher difference than 2% (maximum 1.85% difference). Furthermore, sensitivity analyses excluding subjects with missing values were performed for each of the missing variables. Two scenarios were tested: extreme positive – the variable was replaced with the mean value for all subjects with a positive outcome, and extreme negative – the variable was replaced with the mean value for all subjects with a negative outcome (Figure S2). To compare the prediction performance of the models tested, the receiver operating characteristic (ROC) curve, and the area under it (AUC) were used. Accuracy, balanced accuracy, sensitivity, and specificity metrics were computed for each model. The Kolmogorov-Smirnov normality test was used to assess the normal distribution of the data at a 10% significance level. Statistical differences between metrics were evaluated using the independent samples t-test when normality was not rejected, and the nonparametric alternative Mann-Whitney U test when at least one of the metrics did not follow a normal distribution. Significance levels of 5%, 1%, and 0.1% were considered.

Outcomes

The following models were built to predict 4 unique outcomes defined according to the 2018 ESC/ESH guidelines for the management of hypertension[7]:

- Ambulatory hypertension (AH): 24-hour ABP ≥130/80 mmHg;
- Masked hypertension (MH) vs Normotension (or controlled hypertension): OBP <140/90 mmHg and 24-hour ABP ≥130/80 mmHg vs OBP <140/90 mmHg and 24-hour ABP <130/80 mmHg;
- White-coat hypertension (WH) vs Sustained hypertension (SH): OBP ≥140/90 mmHg and 24-hour ABP <130/80 mmHg vs OBP ≥140/90 mmHg and 24-hour ABP ≥130/80 mmHg;
- CV mortality: only patients who died during follow-up for CV causes were considered; all deceased subjects with unrelated causes were excluded.

Machine learning approaches

Various classifiers, namely logistic regression, gradient-boosted decision trees, and deep neural networks were used for the classification of the data to have a comprehensive spectrum of models and/or approaches. Logistic regression uses a logistic sigmoid function to model the classification problem to a probability that can be mapped to discrete classes. Herein, we used L2 regularized logistic regression. Gradient boosting approaches use a collection of weak classifiers to generate a final decision on the data. The maximum tree depth, number of leaves, and minimum amount of data in each leave were optimized to avoid under- and overfitting (grid-search evaluated with the tuning set). The deep neural network is a multi-layer perceptron with 3 hidden layers, each followed by batch normalization and optimized using stochastic gradient descent.[19] Typically, artificial neurons are organized into layers and a deep neural network contains multiple layers in between the input and the output layers, called hidden layers. Multiple architectures were evaluated and empirically decided upon. Rectified linear unit activation was used for all hidden layers. Dropout regularization was also added to reduce overfitting. The random deletion of network nodes approximates the behavior of an ensemble of multiple architectures. Learning rate, momentum, and dropout

retaining probability hyperparameters were optimized in the tuning set, with early stopping to decide on the number of epochs. Variational dropout was used to estimate uncertainty.[20] Dropout was kept during 50 different prediction calls approximating the output of an ensemble of multiple models. For the final classification, all output probabilities were averaged, and uncertainty was estimated as their standard deviation value. All classifiers were implemented in Python v3.5. All implementations considered the possible imbalance between the number of subjects in different classes, weighting them accordingly. Logistic regression was implemented using the open-source machine learning library scikit-learn[21] v0.22.1. For the multi-layer perceptron, the open-source neural network library Tensorflow[22] v2.0.0 backend was used. Finally, for gradient-boosted decision trees, we resorted to the LightGBM[23] v2.3.1 gradient-boosting framework. The github repository contains the neural networks model and code to exemplify how the models were created, and how they can be used to classify new data. The repository is publicly available here: https://github.com/pedrogsc/CVMortality.

k-fold cross-validation

We performed 5-fold stratified cross-validation to train, tune, and evaluate all the models. The dataset was randomly split into 5 groups of equal size while maintaining the ratio of samples from each class in every group. The classifier was then trained and tested 5 times. Each time, one group was used for testing, one for tuning, and the remainder groups were used for training. Every group has been selected for testing only once. The test set was classified using the best-performing hyperparameter combination, as assessed by grid-search evaluated with the tuning set, to ensure that the test set was always independent of the training and tuning processes.

Feature selection

The Spanish ABP Registry dataset collected 60 clinical variables. Limiting the number of features to a minimum was mandatory to increase clinical applicability and enhance the practicability of any machine learning-based tool, as each clinical feature comes with extra time and potential cost. However, the number of features included had to be high enough to maintain adequate performance. Forward stepwise selection was applied to each machine learning approach (see section 'Machine Learning Approaches') to select the best subset of features. First, each variable was evaluated independently using the AUC, to select which one can better discriminate the data. Thereafter, we recursively tested the addition of all other variables, adding the combination that performed best (higher AUC). This process was repeated until all variables were tested. The feasibility of each variable in daily clinical practice was also evaluated. Some variables were deemed unfeasible or redundant and were not considered for further analyses. These included variables, which are uncommonly available in daily clinical practice, such as the albumin-creatinine ratio. All variables that were not considered are highlighted in the Table S1. The risk calculator is available online at https://www.ccb.uni-saarland.de/cvm-risk.

Risk scores

Several risk scores have been proposed. The widely used Framingham 10-year risk score (FRS)[24], the European CV disease risk assessment score (SCORE-2)[25], and the American College of Cardiology/American Heart Association (ACC/AHA) 10-year risk score (ASCVD)[26] were computed for the Spanish ABP registry dataset for validation reasons. Subjects with missing values that prevented the computation of the respective risk score were excluded from the corresponding analysis. Of note, the evaluated risk scores use different age restraints leading to different subsets (FRS: 30-74 years; the SCORE-2: 45-64 years;

ASCVD: 40-75 years). All analyses were tested in each of these subsets and then independently compared.

Results

Baseline characteristics for the 59,124 patients included in the present analysis are summarized in Table 1 and have been published previously.[4] Data are also presented per hypertensive phenotype and vital status at the end of the study. During a median follow-up of 9.7 years, 7,174 deaths (12.1%) occurred, from which 2,361 were from CV causes, including 685 coronary heart disease, 503 stroke, and 302 heart failure deaths.

Feature selection and machine learning approaches

Three different classifiers were used, i.e., logistic regression, deep neural networks, and gradient-boosted decision trees. The data analysis workflow is depicted in figure 1. Figure 2 shows the impact of varying numbers of selected features for each classifier as % of maximum AUC (upper panel) and performance of each model (lower panel) for predicting CV mortality using office blood pressure (Figure 2a. and Figure 2d.), predicting CV mortality using ambulatory blood pressure (Figure 2b. and Figure 2e.), and ambulatory hypertension prediction (Figure 2c. and Figure 2e.). The performance for all tested classifiers increased steadily with the addition of the first features and plateaued thereafter. The combination of multiple features performed considerably better than any single feature. Already after the first 10 features, classifier performance achieved >99% of its maximum value. After 20 features, ~100% of top performance was attained for all classifiers. Limiting the number of features while maintaining satisfactory relative classification performance was aimed to enhance applicability. For this reason, we established 10 features as the cut-off point for both classification tasks. Age, history of CV disease, number of anti-hypertensive drugs, smoking status, diabetes, and sex were common to all classification approaches for predicting CV

mortality. Office systolic and diastolic BP, office pulse pressure, sex, smoking status, body mass index, and treatment with alpha-blockers were common in the selection for all 3 classification approaches for the prediction of ambulatory hypertension. For the prediction of CV mortality, there were no significant differences between the three classification approaches (AUC). Gradient-boosted decision trees and deep neural networks performed significantly better than logistic regression to detect ambulatory hypertension (AUC) (Figure 2f.). Hence, deep neural networks were selected as the approach of choice for all classification tasks. The best 10 features for each condition are shown in Table 2. Their pairwise Pearson correlations are shown in Figure S3.

Prediction of cardiovascular mortality

Two deep neural network models were trained to predict CV mortality using OBP (CV-Mortality$_{OBP}$) and ABP (CV-Mortality$_{ABP}$) data, respectively. The goal was to evaluate whether ABP could improve risk assessment compared with OBP. Figure 3a shows the ROC curves for both models. Obtained performance metrics are depicted in Table 3. Predicting CV mortality with ABP vs. OBP data led to a significant increase in AUC (0.870 vs 0.865, $p = 3.61\times10^{-28}$), accuracy, and specificity.

Comparison with established risk scores

The FRS, the SCORE-2, and ASCVD risk scores were computed with the data of the Spanish ABPM Study. Due to different exclusion criteria for each of these scores, 3 different subsets were created. For each subset, the ROC curves for the proposed mortality prediction with deep neural networks models for CV-Mortality$_{OBP}$ and CV-Mortality$_{ABP}$ were computed and compared with the respective established risk score. The resulting ROC curves are illustrated in Figures 3c, 3d, and 3e. As shown, both novel mortality prediction models (CV-Mortality$_{OBP}$ and CV-Mortality$_{ABP}$) outperformed all other risk metrics. Of note, the AUC

achieved by the novel approach, using OBP variables only, was significantly higher when compared with the AUC of FRS (0.828 vs. 0.785, $p = 6.09\times10^{-5}$), the SCORE-2 (0.805 vs. 0.778, $p = 4.48\times10^{-3}$), and ASCVD (0.789 vs. 0.709, $p = 3.13\times10^{-3}$).

Prediction of hypertension phenotypes

A second analysis evaluated whether it was possible to detect hypertension phenotypes solely from clinical features. The ROC curves for the detection of ambulatory hypertension, masked hypertension, and white-coat hypertension are illustrated in Figure 3b. Performance metrics are shown in Table 3. The AUC for predicting the ABP from clinically available features was 0.704 ± 0.004. The distinction between different hypertension phenotypes was particularly difficult: normotensive vs. masked hypertension achieved an AUC of about 0.633 ± 0.006. Sustained hypertension (ABP >130/80 mmHg and OBP >140/90 mmHg) vs. white-coat hypertension achieved an AUC of about 0.661 ± 0.004, respectively.

Discussion

The present analysis from a large cohort of 59,124 individuals with 9.7 years follow-up indicates that existing risk scores (i.e., FRS, the SCORE-2, and ASCVD) underperform when compared with a novel CV mortality prediction score derived by artificial intelligence. This novel risk assessment tool relies on 10 clinical features which can be collected relatively easily in clinical practice (Table 2). When adding ABP data to the score, performance metrics improved even further. These results have implications for individual risk assessment and patient management. The prediction of white-coat, masked, and ambulatory hypertension using clinical variables, including OBP, was modest, indicating that ABP measurements provide supplementary information that cannot be reproduced reliably from OBP variables alone.

Artificial intelligence and machine learning have shown increased performance in classifying medical imaging and patient data.[27-29] Various classifiers, namely logistic regression, gradient-boosted decision trees, and deep neural networks are available for data classification. From those machine learning algorithms analyzed herein, the deep neural network approach had the highest accuracy for both prediction tasks and outperformed all other techniques, followed by gradient-boosted decision trees and logistic regression. Hence, we selected deep neural networks as the approach of choice. Algorithmic approaches, such as deep learning[30], whose performance improves as more data become available, offer entirely new opportunities to analyze large datasets. Deep neural networks are composed of a collection of interconnected individual units, roughly modeled after the brain neurons. Each unit can transmit a signal to connected units modeled by weights and biases that are adjusted in the training process. The architecture of the entire network and how these artificial neurons are connected influences its response. To evaluate the machine learning models, cross-validation has been used to infer how well the models generalize to independent datasets. This can only be done reliably with the use of large enough datasets such as the current database, which included 59,124 individuals with more than 3.5 million data points and 7,174 deaths (out of which 2,361 were from CV causes). Both novel risk scores proposed herein, CV-Mortality$_{OBP}$ and CV-Mortality$_{ABP}$, outperformed the established FRS, the SCORE-2, and ASCVD, which were also computed using the data of the Spanish ABPM Study. When ambulatory BP was added to the model, AUC for prediction of CV mortality significantly increased (0.870 vs 0.865, $p = 3.61\times10^{-28}$), indicating that the addition of ambulatory BP features enhanced the discriminative power of the classifier even further. With an AUC in the range of 0.865-0.870, the chances of correctly predicting CV death in a pair of individuals are higher than those reported for other established scores, such as the widely used CHA_2DS_2-VASc score to predict the risk of stroke in patients with atrial fibrillation. In the context of

this study, this would indicate that if 1 individual were picked from the group who died because of CV causes and 1 were selected from the group alive, the novel models (CV-Mortality$_{OBP}$ and CV-Mortality$_{ABP}$) would correctly assign higher risk to the individual deceased for CV causes in approximately 4 out of 5 times. The boxplots in Figure S4 show the distribution of the normalized values of the newly proposed and established risk scores for prediction of CV mortality. Studies from Denmark[31] and Italy[32] were unable or found only modest improvements in the prognostic accuracy of cardiovascular risk by addition of ABP. However, these studies used standard statistics, such as cox-regression, did not utilize novel artificial intelligence and machine learning approaches, and had smaller sample sizes, all of which may, at least in part, explain the differences observed herein.

Ambulatory BP readings are unanimously recommended by the current North American[26] and European[7,33] guidelines for the management of hypertension. Recent data indicate that ambulatory BP monitoring can indeed improve risk stratification in patients with hypertension and those with suspected hypertension over and beyond OBP readings.[3] Particularly, nighttime BP was more informative about the risk of death than standard office BP readings. Ambulatory BP monitoring is mandatory to detect white-coat and masked hypertension; the latter has been associated with increased CV risk compared with normotension or white-coat hypertension.[4] However, top-ranked barriers to ambulatory BP monitoring are challenges in accessing devices, costs of testing, and concerns about the willingness or ability of patients to successfully complete tests.[34] Given the limited resources and inability to perform ABPM in all hypertensive individuals, the present study also aimed at identifying clinical characteristics associated with white-coat and masked hypertension. However, even with the help of sophisticated artificial intelligence techniques applied to this large cohort of patients, one must realize that predicting ambulatory BP with clinical features alone is limited. In line, office and ambulatory BP showed only a moderate correlation

(correlation coefficient for SBP: 0.43 correlation coefficient for DBP: 0.52). These results indicate that clinical and ambulatory features provide very different information and one can barely predict ambulatory output using clinical characteristics, including OBP. Particularly for the distinction between different hypertension phenotypes and subsequent treatment decisions, ABPM appears irreplaceable.[6]

Some limitations must be considered. In most patients, ABPM was performed at a single time point, thus limiting its prognostic power. Data on medication during the follow-up, except in patients who had two ABPM sessions, is limited. The indication for ABPM was left at physicians' discretion, some selection bias can, therefore, not be excluded entirely. Only Caucasian patients were included; hence, the results may not apply to people of other races. For the computation of the established FRS, the SCORE-2, and ASCVD, subjects with missing values that prevented the calculation were subsequently excluded from the analysis. However, all analyses were tested in each of these subsets and then independently compared. The number of missing values was different across variables which could have introduced some bias. The present study did not consider the association of BP with non-fatal events, as only data for deaths were available from the Spanish National Institute of Statistics. Examination of model performance in different datasets (i.e. external validation) should be considered to further examine its validity. This novel risk calculator was developed and validated in the same cohort, which could have contributed to its excellence performance.

In conclusion, we developed a novel risk calculator for the prediction of CV mortality using artificial intelligence based on a large cohort of patients included in the Spanish ABPM registry. The ROC curves for CV-Mortality$_{OBP}$ and CV-Mortality$_{ABP}$ with deep neural networks models outperformed all other risk metrics, indicating the potential for improving current risk scores by applying state-of-the-art machine learning approaches. Further, the prediction of CV mortality using ABP data led to a significant increase in AUC and

performance metrics. Additional studies are warranted to validate this artificial intelligence-based approach and explore its ability in risk prediction.

Perspectives

Several studies have reported that ABP, and in particular nighttime BP, associates more closely with outcomes than conventional office BP. The value of integrating ABP as part of risk prediction scores has not been systematically analyzed. Using machine learning approaches, which set a new standard for processing and analyzing large dataset, we developed and validated a novel risk score (https://www.ccb.uni-saarland.de/cvm-risk). This novel risk assessment tool relies on 10 clinical features which can be collected relatively easily in clinical practice. Already when using OBP parameters, this risk score outperformed existing risk scores and improved further when ABP was used instead of OBP. Further studies are required to examine and validate this novel risk score in independent datasets.

Funding

Spanish Society of Hypertension, Lacer Laboratories, UK Medical Research Council, Health Data Research UK, National Institute for Health and Care Research Biomedical Research Centres (Oxford and University College London Hospitals), and British Heart Foundation Centre for Research Excellence.

Disclosure

Conflicts of interest: MB is supported by Deutsche Gesellschaft für Kardiologie (DGK), and Deutsche Forschungsgemeinschaft (SFB TRR219, Project-ID 322900939). MB has received consulting fees from Abbott Vascular, Bayer AG, Amgen, AstraZeneca, Servier, Medtronic, Vifor, and Boehringer Ingelheim. LL has received speaker honoraria from Medtronic and ReCor Medical. BW has received research grant support from Omron and speaker honoraria

from Omron, Pfizer, Vifor, Sevier and Menarini. FM is supported by Deutsche Gesellschaft für Kardiologie (DGK), Deutsche Forschungsgemeinschaft (SFB TRR219, Project-ID 322900939), and Deutsche Herzstiftung. He has received scientific support from Ablative Solutions, Medtronic and ReCor Medical and speaker honoraria/consulting fees from Ablative Solutions, Amgen, Astra-Zeneca, Bayer, Boehringer Ingelheim, Inari, Medtronic, Merck, ReCor Medical, Servier, and Terumo. NS reports institutional grant support from Boehringer Ingelheim, Lilly, and Novo Nordisk for investigator led/designed research unrelated to this project. AdlS reports lecture fees from Viatris, and support for travel to attend meetings from Menarini and Lacer. LMR reports consulting fees from Novartis, Bayer, Pfizer, Medtronic, ReCor, Sanofi, and Sandoz and lecture fees from Novartis, Bayer, Pfizer, Medtronic, ReCor, and Sanofi. BW reports lecture fees from Daichi Sankyo, Servier, and Pfizer, and institutional grant support from Omron for investigator designed/led research unrelated to this work. The Clinical Trial Service Unit and Epidemiological Studies Unit has a staff policy of not accepting honoraria or other payments from the pharmaceutical industry, except for the reimbursement of costs to participate in scientific meetings. BW on behalf of University College London has received an investigator-led grant award from Omron, Japan and honoraria for lectures on blood pressure measurement, and is supported by the National Institute for Health and Care Research University College London Hospitals Biomedical Research Centre. All other authors declare no competing interests.

References

1. Olsen MH, Angell SY, Asma S, et al. A call to action and a lifecourse strategy to address the global burden of raised blood pressure on current and future generations: the Lancet Commission on hypertension. The Lancet. Nov 26 2016;388(10060):2665-2712. doi:10.1016/s0140-6736(16)31134-5

2. Goff DC, Jr., Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. Jun 24 2014;129(25 Suppl 2):S49-73. doi:10.1161/01.cir.0000437741.48606.98

3. Yang WY, Melgarejo JD, Thijs L, et al. Association of Office and Ambulatory Blood Pressure With Mortality and Cardiovascular Outcomes. JAMA. Aug 6 2019;322(5):409-420. doi:10.1001/jama.2019.9811

4. Staplin N, de la Sierra A, Ruilope LM, et al. Relationship between clinic and ambulatory blood pressure and mortality: an observational cohort study in 59 124 patients. Lancet. Jun 17 2023;401(10393):2041-2050. doi:10.1016/S0140-6736(23)00733-X

5. Niiranen TJ, Maki J, Puukka P, Karanko H, Jula AM. Office, home, and ambulatory blood pressures as predictors of cardiovascular risk. Hypertension. Aug 2014;64(2):281-6. doi:10.1161/HYPERTENSIONAHA.114.03292

6. Minutolo R, Agarwal R, Borrelli S, et al. Prognostic role of ambulatory blood pressure measurement in patients with nondialysis chronic kidney disease. Arch Intern Med. Jun 27 2011;171(12):1090-8. doi:10.1001/archinternmed.2011.230

7. Williams B, Mancia G, Spiering W, et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension. Eur Heart J. Sep 1 2018;39(33):3021-3104. doi:10.1093/eurheartj/ehy339

8. Engemann DA, Raimondo F, King JR, et al. Robust EEG-based cross-site and cross-protocol classification of states of consciousness. Brain. Nov 1 2018;141(11):3179-3192. doi:10.1093/brain/awy251

9. Yang G, Yu S, Dong H, et al. DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction. IEEE Trans Med Imaging. Jun 2018;37(6):1310-1321. doi:10.1109/TMI.2017.2785879

10. Yamada M, Saito Y, Imaoka H, et al. Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy. Sci Rep. Oct 8 2019;9(1):14465. doi:10.1038/s41598-019-50567-5

11. Guimaraes P, Keller A, Fehlmann T, Lammert F, Casper M. Deep-learning based detection of gastric precancerous conditions. Gut. Jan 2020;69(1):4-6. doi:10.1136/gutjnl-2019-319347

12. Mortazavi BJ, Downing NS, Bucholz EM, et al. Analysis of Machine Learning Techniques for Heart Failure Readmissions. Circ Cardiovasc Qual Outcomes. Nov 2016;9(6):629-640. doi:10.1161/CIRCOUTCOMES.116.003039

13. Johnson KW, Torres Soto J, Glicksberg BS, et al. Artificial Intelligence in Cardiology. J Am Coll Cardiol. Jun 12 2018;71(23):2668-2679. doi:10.1016/j.jacc.2018.03.521

14. Banegas JR, Ruilope LM, de la Sierra A, et al. High prevalence of masked uncontrolled hypertension in people with treated hypertension. Eur Heart J. Dec 7 2014;35(46):3304-12. doi:10.1093/eurheartj/ehu016

15. Gorostidi M, Banegas JR, de la Sierra A, Vinyoles E, Segura J, Ruilope LM. Ambulatory blood pressure monitoring in daily clinical practice - the Spanish ABPM Registry experience. Eur J Clin Invest. Jan 2016;46(1):92-8. doi:10.1111/eci.12565

16. de la Sierra A, Armario P, Oliveras A, et al. Antihypertensive drug use in resistant and nonresistant hypertension and in controlled and uncontrolled resistant hypertension. J Hypertens. Jul 2018;36(7):1563-1570. doi:10.1097/HJH.0000000000001729

17. Parati G, Stergiou G, O'Brien E, et al. European Society of Hypertension practice guidelines for ambulatory blood pressure monitoring. J Hypertens. Jul 2014;32(7):1359-66. doi:10.1097/HJH.0000000000000221

18. Instituto Nacional de Estadística. Informes metodológicos estandarizados. Accessed 21 February 2017, http://www.ine.es/dynt3/metadatos/es/RespuestaPrint.html?oper=23

19. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. presented at: Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37; 2015; Lille, France.

20. Gal Y, Ghahramani Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. PMLR; 2016:1050-1059.

21. Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F, Mueller A. Scikit-learn: Machine Learning Without Learning the Machinery. GetMobile Mob Comput Commun. 2015;19:29-33.

22. Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 2016:arXiv:1603.04467. doi:10.48550/arXiv.1603.04467 Accessed March 01, 2016. https://ui.adsabs.harvard.edu/abs/2016arXiv160304467A

23. Ke G, Meng Q, Finley T, et al. LightGBM: A highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst. 2017;30:3146–3154.

24. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation. May 12 1998;97(18):1837-47. doi:10.1161/01.cir.97.18.1837

25. Score working group and ESC Cardiovascular risk collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. Eur Heart J. Jul 1 2021;42(25):2439-2454. doi:10.1093/eurheartj/ehab309

26. Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension. Jun 2018;71(6):1269-1324. doi:10.1161/HYP.0000000000000066

27. Howard JP, Cook CM, van de Hoef TP, et al. Artificial Intelligence for Aortic Pressure Waveform Analysis During Coronary Angiography: Machine Learning for Patient

Safety. JACC Cardiovasc Interv. Oct 28 2019;12(20):2093-2101. doi:10.1016/j.jcin.2019.06.036

28. Howard JP, Fisher L, Shun-Shin MJ, et al. Cardiac Rhythm Device Identification Using Neural Networks. JACC Clin Electrophysiol. May 2019;5(5):576-586. doi:10.1016/j.jacep.2019.02.003

29. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med. Jan 2019;25(1):65-69. doi:10.1038/s41591-018-0268-3

30. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. May 28 2015;521(7553):436-44. doi:10.1038/nature14539

31. Mortensen RN, Gerds TA, Jeppesen JL, Torp-Pedersen C. Office blood pressure or ambulatory blood pressure for the prediction of cardiovascular events. Eur Heart J. Nov 21 2017;38(44):3296-3304. doi:10.1093/eurheartj/ehx464

32. Mancia G, Facchetti R, Seravalle G, Cuspidi C, Corrao G, Grassi G. Adding Home and/or Ambulatory Blood Pressure to Office Blood Pressure for Cardiovascular Risk Prediction. Hypertension. Feb 2021;77(2):640-649. doi:10.1161/HYPERTENSIONAHA.120.16303

33. Mancia G, Kreutz R, Brunstrom M, et al. 2023 ESH Guidelines for the management of arterial hypertension The Task Force for the management of arterial hypertension of the European Society of Hypertension Endorsed by the European Renal Association (ERA) and the International Society of Hypertension (ISH). J Hypertens. Jun 21 2023;doi:10.1097/HJH.0000000000003480

34. Kronish IM, Kent S, Moise N, et al. Barriers to conducting ambulatory and home blood pressure monitoring during hypertension screening in the United States. J Am Soc Hypertens. Sep 2017;11(9):573-580. doi:10.1016/j.jash.2017.06.012

Pathophysiological Novelty and Relevance

What is new?

By applying machine learning approaches to a large dataset of patients undergoing ambulatory blood pressure monitoring, a novel risk novel risk score was developed, which outperformed existing risk scores (i.e. Framingham Risk Score, the SCORE-2, and ASCVD) for prediction of CV mortality. Using ambulatory instead of office BP led to a significant increase in AUC and performance metrics.

What is relevant?

With this novel risk score, CV mortality can be predicted more precisely than with other established risk scores, particularly when ABP data is used. Given the limited resources and inability to perform ABPM in all hypertensive individuals, we also aimed at identifying clinical characteristics associated with white-coat and masked hypertension. However, predicting ABP with clinical features alone was limited, indicating that clinical and ambulatory BP features provide very different information. Particularly for the distinction between different hypertension phenotypes (i.e. white-coat hypertension, masked hypertension) and subsequent treatment decisions, performing ABPM appears irreplaceable.

Clinical Implications?

This novel risk assessment tool relies on 10 clinical features which can be collected relatively easily in clinical practice (https://www.ccb.uni-saarland.de/cvm-risk). Data derived from ABPM may be used to further improve risk assessment.

Figure Legends

Figure 1. Data analysis workflow. Stepwise forward feature selection was used with four different machine-learning approaches to select the best feature set and classifier for each outcome. Cross validation guaranteed that the test set was completely independent of the training and tuning sets. Variational dropout was applied for uncertainty estimation. Final average results and variability were computed. The workflow is further detailed in the Methods section.

Figure 2. At the top is the evolution of the area under the curve (AUC) with feature selection for each classifier. The percentage of the maximum AUC versus the number of features is shown. At the bottom, the performance of each model with 10 features. Accuracy, sensitivity, specificity, balanced accuracy, and AUC box plots are shown. Median, first and third quartiles are shown. Whiskers at each quartile plus 1.5 times the interquartile range. Outliers are shown as diamonds. (a) and (d) for cardiovascular mortality prediction using office blood pressure, (b) and (e) for cardiovascular mortality prediction using ambulatory blood pressure, and (c) and (f) for ambulatory hypertension prediction. * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

Figure 3. On the top, receiving operating characteristics (ROC) curves for (a) cardiovascular (CV) mortality prediction using office blood pressure (CV-MortalityOBP) and ambulatory blood pressure (CV-MortalityABP) data, and (b) detection of ambulatory hypertension, masked hypertension and white-coat hypertension. On the bottom, comparison of the ROC curves obtained with the proposed approaches, using office blood pressure (CV-MortalityOBP) and ambulatory blood pressure (CV-MortalityABP) data, and (c) Framingham (FRS) risk score, (d) the SCORE-2 and (e) ASCVD. The uncertainty range for each curve is also represented. The area under each curve (AUC) ± standard deviation values (if applicable) are showed in the graphic's legend. * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

Table 1. Baseline characteristics

	All subjects	Normotension	Ambulatory Hypertension	Sustained Hypertension	Masked Hypertension	White-coat Hypertension	Patients Alive at the End of the Study	Patients Who Died during Follow-up for CV causes
Number of Subjects	59124	10006	31937	26938	4999	17181	51950	2361
Female	27787 (47.0)	4593 (45.9)	13184 (41.3)	11164 (41.4)	2020 (40.4)	7991 (46.5)	24672 (47.5)	1132 (47.9)
Age, years	58.7 ± 14.1	58.1 ± 14.9	58.3 ± 13.9	58.5 ± 13.7	56.9 ± 14.5	59.7 ± 14.1	56.8 ± 13.6	73.3 ± 10.5
Body mass index, kg/m^2	29.0 ± 4.8	28.6 ± 5.0	28.9 ± 4.7	29.0 ± 4.7	28.4 ± 4.6	29.3 ± 5.0	29.0 ± 4.8	29.4 ± 4.9
Risk factors and history								
Smoker	9294 (15.7)	9294 (15.7)	5703 (17.9)	8342 (16.1)	924 (18.5)	2128 (12.4)	8342 (16.1)	293 (12.4)
Diabetic	11391 (19.3)	11391 (19.3)	6454 (20.2)	8913 (17.2)	903 (18.1)	3255 (18.9)	8913 (17.2)	888 (37.6)
Glucose level (mg/dL)	105.5 ± 32.0	103.1 ± 27.3	106.5 ± 33.7	106.9 ± 33.9	103.9 ± 32.7	105.0 ± 31.2	104.2 ± 30.9	119.2 ± 41.4
History of CV disease	7574 (12.8)	7574 (12.8)	4007 (12.5)	5363 (10.3)	706 (14.1)	2044 (11.9)	5363 (10.3)	888 (37.6)
Target organ damage	6245 (10.6)	6245 (10.6)	3587 (11.2)	4518 (8.7)	565 (11.3)	1638 (9.5)	4518 (8.7)	646 (27.4)

Chronic kidney disease	2768 (4.7)	2768 (4.7)	1714 (5.4)	1847 (3.6)	265 (5.3)	670 (3.9)	1847 (3.6)	331 (14.0)
High-density lipoprotein (mg/dL)	53.6 ± 15.8	54.3 ± 15.7	52.9 ± 15.7	52.9 ± 15.6	53.0 ± 16.2	54.6 ± 16.0	53.7 ± 15.8	50.9 ± 14.6
Blood pressure (mmHg)								
Ambulatory SBP	128.77 ± 13.65	116.32 ± 7.86	137.75 ± 10.97	138.58 ± 11.13	133.30 ± 8.83	119.33 ± 7.26	128.12 ± 13.16	134.77 ± 16.22
Office SBP	147.96 ± 18.78	125.77 ± 10.14	152.52 ± 18.33	156.80 ± 16.39	129.45 ± 8.22	152.40 ± 13.54	147.31 ± 18.21	153.60 ± 21.89
Daytime SBP	131.91 ± 13.98	119.09 ± 8.45	140.97 ± 11.15	141.87 ± 11.29	136.15 ± 8.94	122.53 ± 7.95	131.43 ± 13.55	136.46 ± 16.59
Nighttime SBP	119.97 ± 15.73	108.90 ± 10.40	128.35 ± 14.68	129.02 ± 14.81	124.74 ± 13.37	110.84 ± 10.11	118.79 ± 14.89	130.42 ± 18.81
Ambulatory DBP	76.23 ± 10.11	69.51 ± 6.65	81.61 ± 9.35	81.82 ± 9.58	80.51 ± 7.93	70.13 ± 6.80	76.84 ± 9.92	71.23 ± 10.90
Office DBP	86.49 ± 11.50	76.21 ± 8.40	88.98 ± 11.48	90.87 ± 11.01	78.77 ± 8.09	87.85 ± 9.69	87.15 ± 11.19	81.31 ± 12.66
Daytime DBP	79.13 ± 10.68	72.25 ± 7.33	84.54 ± 10.02	84.77 ± 10.24	83.27 ± 8.66	73.07 ± 7.47	79.85 ± 10.46	73.22 ± 11.31
Nighttime DBP	68.20 ± 10.18	62.23 ± 7.25	73.19 ± 9.57	73.34 ± 9.74	72.39 ± 8.53	62.38 ± 7.44	68.43 ± 10.05	66.32 ± 11.59

Ambulatory PP	52.54 ± 11.86	46.81 ± 7.97	56.14 ± 13.25	56.76 ± 13.39	52.79 ± 11.89	49.20 ± 8.15	51.28 ± 10.87	63.54 ± 14.73
Office PP	61.47 ± 16.97	49.56 ± 9.75	63.55 ± 17.93	65.93 ± 18.12	50.68 ± 9.43	64.55 ± 15.44	60.16 ± 16.12	72.29 ± 19.80
Daytime PP	52.78 ± 12.02	46.84 ± 8.12	56.44 ± 13.37	57.10 ± 13.53	52.88 ± 11.88	49.46 ± 8.38	51.58 ± 11.10	63.24 ± 15.02
Nighttime PP	51.78 ± 12.46	46.67 ± 8.77	55.16 ± 13.99	55.68 ± 14.12	52.35 ± 12.92	48.46 ± 8.85	50.36 ± 11.30	64.10 ± 15.52
No. of antihypertensive drugs								
0	23996 (40.6)	3901 (39.0)	13559 (42.5)	11432 (42.4)	2127 (42.5)	6536 (38.0)	22357 (43.0)	606 (25.7)
1	12543 (21.2)	2364 (23.6)	6430 (20.1)	5291 (19.6)	1139 (22.8)	3749 (21.8)	11183 (21.5)	523 (22.2)
≥2	22585 (38.2)	3741 (37.4)	11948 (37.4)	10215 (37.9)	1733 (34.7)	6896 (40.1)	18410 (35.4)	1232 (52.2)
Treatment with								
Alpha blocker	2656 (4.5)	318 (3.2)	1708 (5.3)	1482 (5.5)	226 (4.5)	630 (3.7)	2003 (3.9)	231 (9.8)

CV – cardiovascular; BP – blood pressure; OBP – office BP; ABP – ambulatory BP, SBP – systolic BP; DBP – diastolic BP, PP – pulse pressure.

Target organ damage: left ventricular hypertrophy, chronic kidney disease, evidence of atherosclerotic cardiovascular disease. Normotension: office BP <140/90 mmHg, ABP <130/80 mmHg; ambulatory hypertension: ABP ≥130/80 mmHg; sustained hypertension: ABP ≥130/80 mmHg; masked hypertension: OBP <140/90 mmHg, ABP≥130/80 mmHg; white-coat hypertension: OBP ≥140/90 mmHg, ABP <130/80 mmHg; Data are numbers (%), or mean ± SD

Table 2. Feature selection

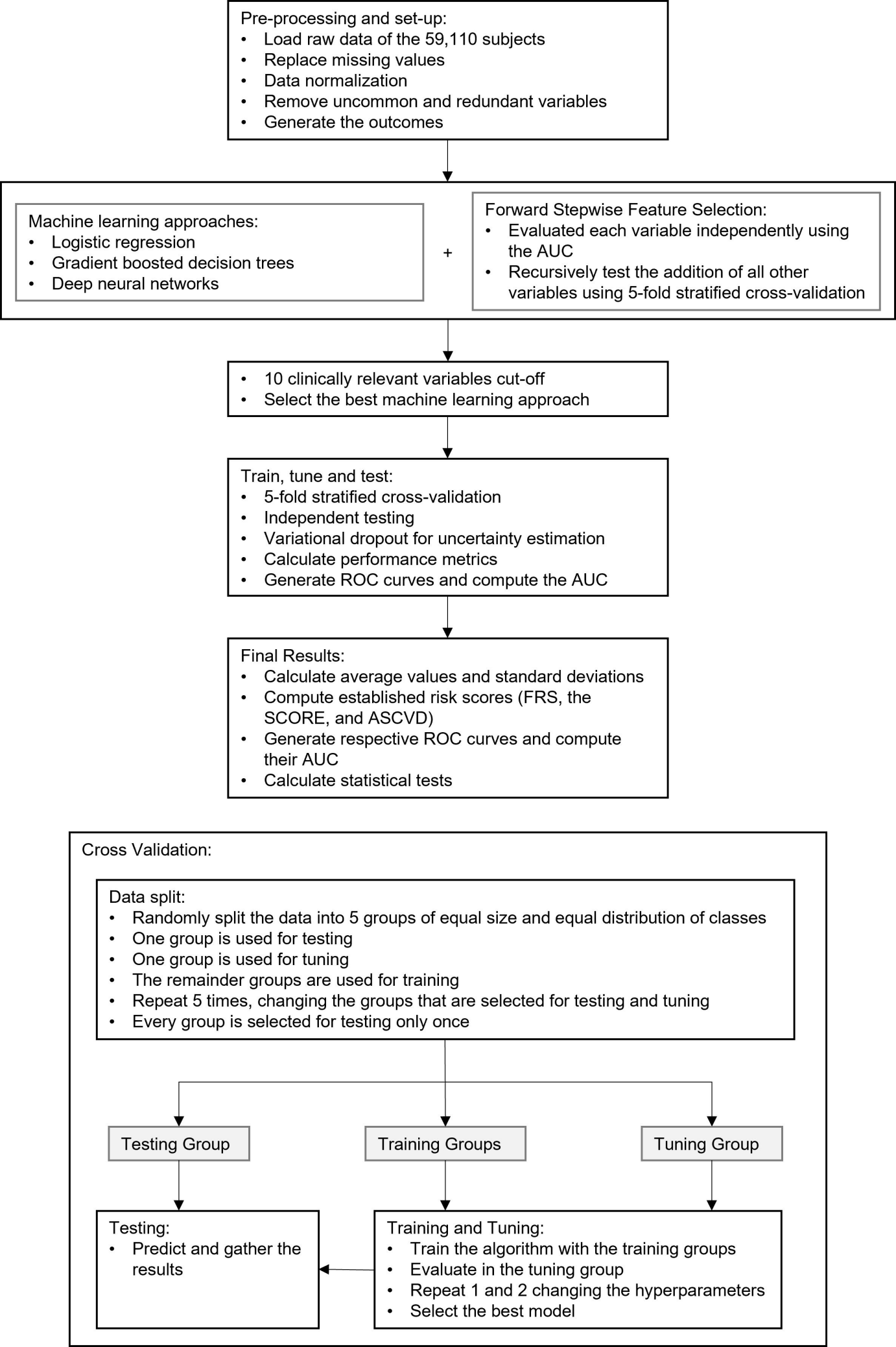
	CV-Mortality$_{OBP}$	**CV-Mortality$_{ABP}$**	**Ambulatory Hypertension**
1	Age (years)	Age (years)	Office systolic BP (mmHg)
2	History of CV disease (yes/no)	History of CV disease (yes/no)	Office diastolic BP (mmHg)
3	Number of anti-hypertensive drugs (n)	Nighttime systolic BP (mmHg)	Sex (male/female)
4	Smoker (yes/no)	Number of anti-hypertensive drugs (n)	Age (years)
5	Diabetic (yes/no)	Smoker (yes/no)	Body Mass Index (kg/m^2)
6	Triglycerides (mg/dL)	Diabetic (yes/no)	Number of anti-hypertensive drugs (n)
7	Office diastolic BP (mmHg)	Daytime diastolic BP (mmHg)	Smoker (yes/no)
8	Target organ damage (yes/no)	Daytime HR (bpm)	Diabetic (yes/no)
9	Sex (male/female)	Abdominal circumference (cm)	Office PP (mmHg)
10	Office systolic BP (mmHg)	Sex (male/female)	Alpha blocker treatment (yes/no)

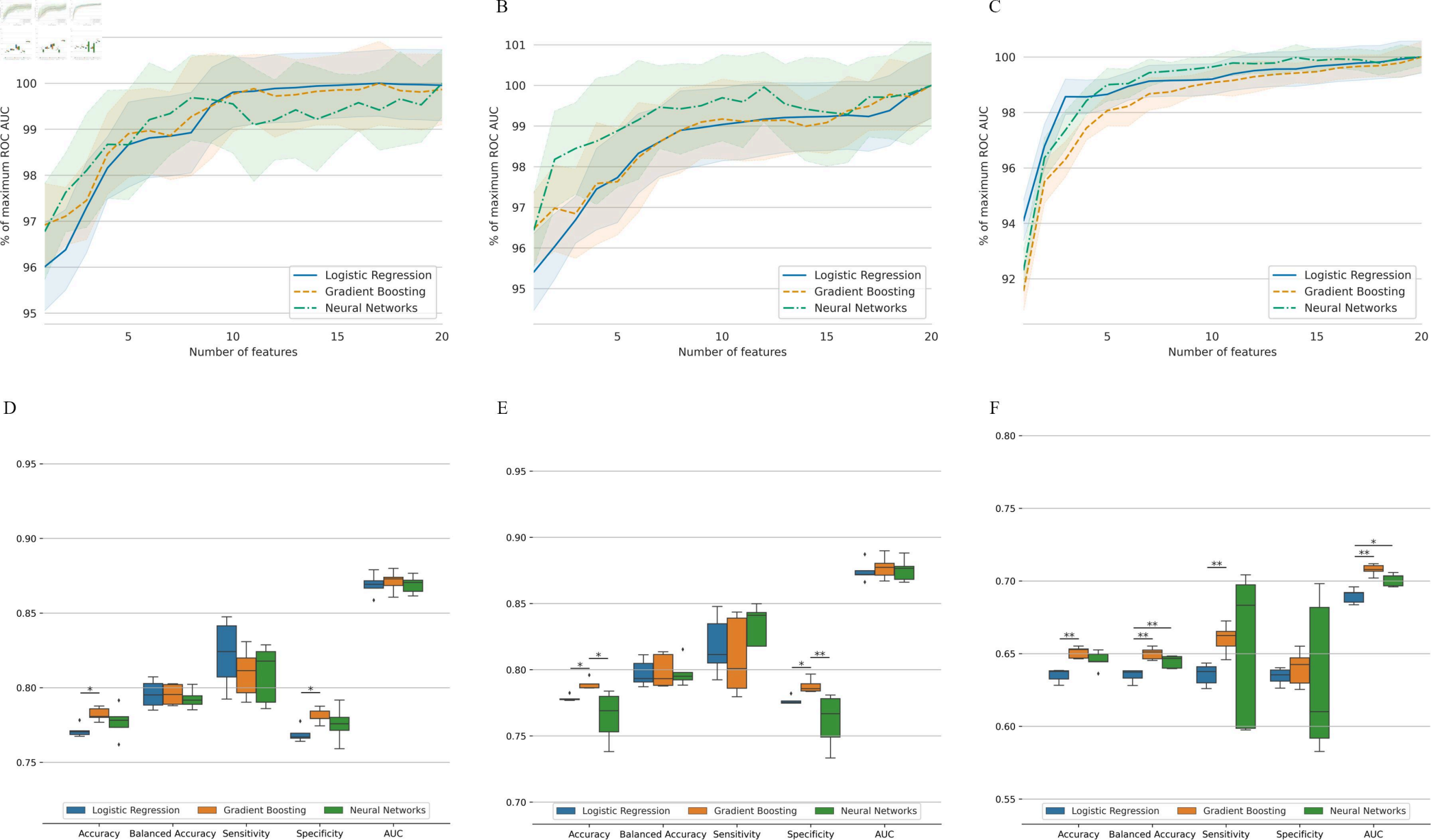
CV – cardiovascular; BP – blood pressure, PP – pulse pressure, HR – heart rate.
Target organ damage: left ventricular hypertrophy, chronic kidney disease, evidence of atherosclerotic cardiovascular disease

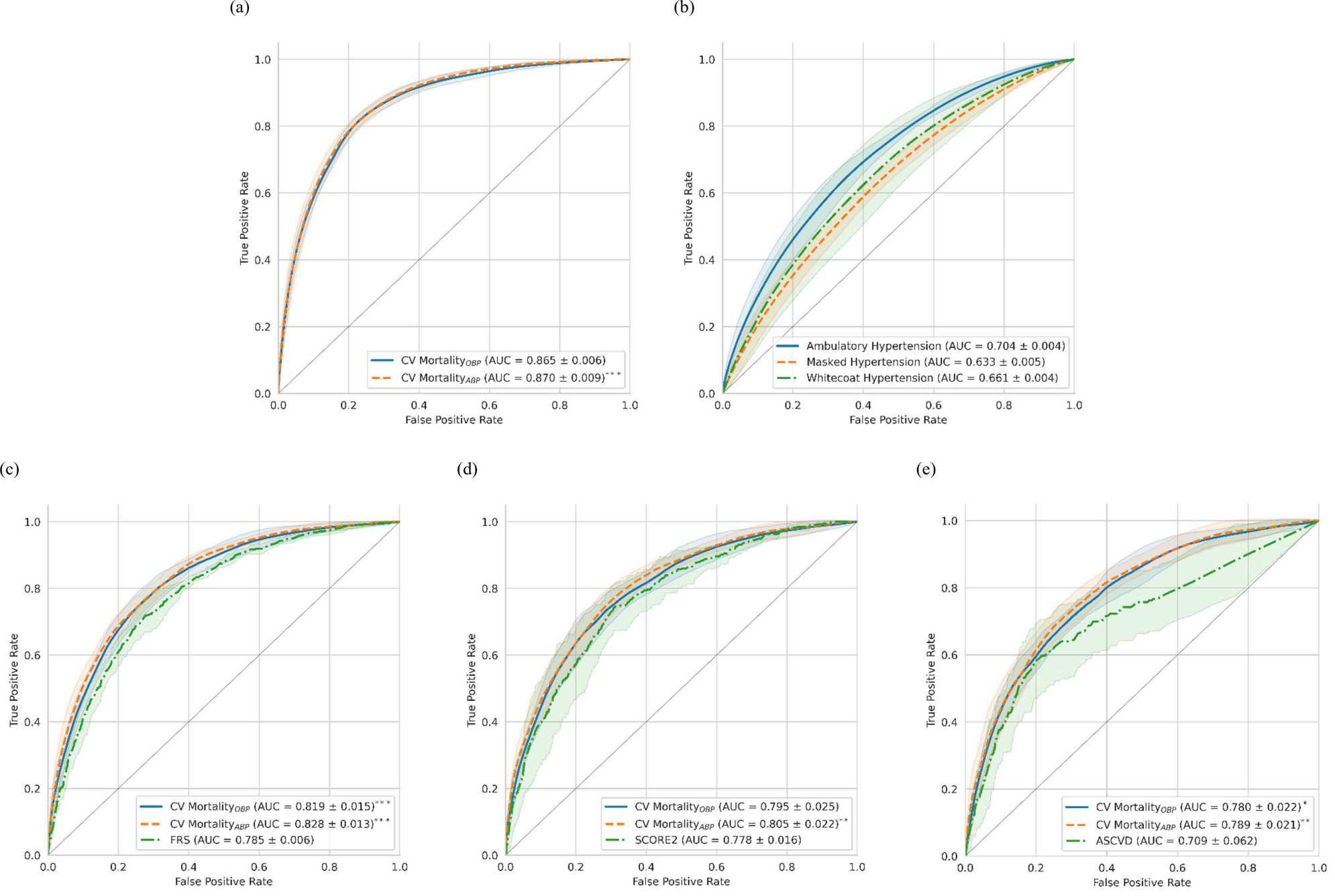
Table 3. Classification performance metrics

	AUC	Accuracy	Balanced Accuracy	Sensitivity	Specificity
CV Mortality$_{OBP}$	0.865 ± 0.006	0.774 ± 0.012	0.794 ± 0.006	0.815 ± 0.019	0.772 ± 0.013
CV Mortality$_{ABP}$	0.870 ± 0.010	0.781 ± 0.020	0.793 ± 0.009	0.806 ± 0.025	0.780 ± 0.022
Ambulatory Hypertension	0.704 ± 0.004	0.650 ± 0.004	0.648 ± 0.003	0.682 ± 0.028	0.614 ± 0.027
Masked Hypertension	0.633 ± 0.006	0.596 ± 0.010	0.594 ± 0.006	0.591 ± 0.020	0.598 ± 0.021
White-coat Hypertension	0.661 ± 0.004	0.620 ± 0.015	0.614 ± 0.005	0.590 ± 0.079	0.639 ± 0.073

AUC – Area under the curve; CV – Cardiovascular; OBP – Office Blood Pressure; ABP – Ambulatory Blood Pressure.


Pre-processing and set-up:
• Load raw data of the 59,110 subjects
• Replace missing values
• Data normalization
• Remove uncommon and redundant variables
• Generate the outcomes
Machine learning approaches:
• Logistic regression
• Gradient boosted decision trees
• Deep neural networks
+
Forward Stepwise Feature Selection:
• Evaluated each variable independently using the AUC
• Recursively test the addition of all other variables using 5-fold stratified cross-validation
• 10 clinically relevant variables cut-off
• Select the best machine learning approach
Train, tune and test:
• 5-fold stratified cross-validation
• Independent testing
• Variational dropout for uncertainty estimation
• Calculate performance metrics
• Generate ROC curves and compute the AUC
Final Results:
• Calculate average values and standard deviations
• Compute established risk scores (FRS, the SCORE, and ASCVD)
• Generate respective ROC curves and compute their AUC
• Calculate statistical tests
Cross Validation:
Data split:
• Randomly split the data into 5 groups of equal size and equal distribution of classes
• One group is used for testing
• One group is used for tuning
• The remainder groups are used for training
• Repeat 5 times, changing the groups that are selected for testing and tuning
• Every group is selected for testing only once
Testing Group
Training Groups
Tuning Group
Testing:
• Predict and gather the results
Training and Tuning:
• Train the algorithm with the training groups
• Evaluate in the tuning group
• Repeat 1 and 2 changing the hyperparameters
• Select the best model


A
B
C
D
E
F
% of maximum ROC AUC
Number of features
Logistic Regression
Gradient Boosting
Neural Networks
Accuracy
Balanced Accuracy
Sensitivity
Specificity
AUC


(a)
True Positive Rate
False Positive Rate
CV Mortality$_{OBP}$ (AUC = 0.865 ± 0.006)
CV Mortality$_{ABP}$ (AUC = 0.870 ± 0.009)***
(b)
True Positive Rate
False Positive Rate
Ambulatory Hypertension (AUC = 0.704 ± 0.004)
Masked Hypertension (AUC = 0.633 ± 0.005)
Whitecoat Hypertension (AUC = 0.661 ± 0.004)
(c)
True Positive Rate
False Positive Rate
CV Mortality$_{OBP}$ (AUC = 0.819 ± 0.015)***
CV Mortality$_{ABP}$ (AUC = 0.828 ± 0.013)***
FRS (AUC = 0.785 ± 0.006)
(d)
True Positive Rate
False Positive Rate
CV Mortality$_{OBP}$ (AUC = 0.795 ± 0.025)
CV Mortality$_{ABP}$ (AUC = 0.805 ± 0.022)**
SCORE2 (AUC = 0.778 ± 0.016)
(e)
True Positive Rate
False Positive Rate
CV Mortality$_{OBP}$ (AUC = 0.780 ± 0.022)*
CV Mortality$_{ABP}$ (AUC = 0.789 ± 0.021)**
ASCVD (AUC = 0.709 ± 0.062)