medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Risk prediction with office and ambulatory blood pressure using artificial intelligence

Pedro Guimarães[1], Andreas Keller[1,2], Michael Böhm[3], Lucas Lauder[3], José L. Ayala[4]
José R. Banegas[5], Alejandro de la Sierra[6], Ernest Vinyoles[7], Manuel Gorostidi[8],
Julián Segura[9], Gema Ruiz-Hurtado[9,10] Luis M. Ruilope[9,10,11], Felix Mahfoud[2,12]

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

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

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

4 Department of Computer Architecture, Complutense University of Madrid, Spain

5 Department of Preventive Medicine and Public Health, School of Medicine, Universidad Autónoma de Madrid/IdiPAZ and CIBERESP, Madrid, Spain

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

7 La Mina Primary Care Center, University of Barcelona; Cardiology Service, Barcelona, Spain

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

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

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

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

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

Address for correspondence

Professor Dr. Felix Mahfoud, MD FESC

Klinik für Innere Medizin III

Universitätsklinikum des Saarlandes/Saarland University

Kirrberger Str. 1, IMED, Geb. 41.1

66421 Homburg/Saar, Germany

Telephone: +49 6841 16 15911

Fax: +49 6841 16 15910

E-Mail: Felix.Mahfoud@uks.eu

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Abstract

Background

To develop and validate a novel, machine learning-derived model for prediction of cardiovascular (CV) mortality risk using office (OBP) and ambulatory blood pressure (ABP), to compare its performance with existing risk scores, and to assess the possibility of predicting ABP phenotypes (i.e. white-coat, ambulatory and masked hypertension) utilizing clinical variables.

Methods

Using data from 63,910 patients enrolled in the Spanish ABP monitoring registry, machine-learning approaches (logistic regression, support vector machine, gradient boosted decision trees, and deep neural networks) and stepwise forward feature selection were used for the classification of the data.

Results

Over a median follow-up of 4.7 years, 3,808 deaths occurred from which 1,295 were from CV causes. The performance for all tested classifiers increased while adding up to 10 features and converged thereafter. 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 all other risk scores. The area under the curve (AUC) achieved by the novel approach, using OBP variables only, was already significantly higher when compared with the AUC of Framingham score (0.685 vs 0.659, $p = 1.97\times10^{-22}$), the SCORE (0.679 vs 0.613, $p = 6.21\times10^{-22}$), and ASCVD (0.722 vs 0.639, $p = 8.03\times10^{-30}$) risk score. However, prediction of CV mortality with ABP instead of OBP data led to a significant increase in AUC (0.781 vs 0.752, $p = 1.73\times10^{-42}$), accuracy, balanced

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

accuracy and sensitivity. The sensitivity and specificity for detection of ambulatory, masked, and white-coat hypertension ranged between 0.653-0.661 and 0.573-0.651, respectively.

Conclusion

We developed a novel risk calculator for CV death using artificial intelligence based on a large cohort of patients included in the Spanish ABP monitoring registry. The receiver operating characteristic curves for CV-Mortality$_{OBP}$ and CV-Mortality$_{ABP}$ with deep neural networks models outperformed all other risk metrics. Prediction of CV mortality using ABP data led to a significant increase in performance metrics. The prediction of ambulatory phenotypes using clinical characteristics, including OBP, was limited.

Key words

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

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

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), the Atherosclerotic Cardiovascular Disease (ASCVD) score, or the Framingham Risk Score.[2] Office blood pressure (OBP) is an integral part of these risk score. Recently, ambulatory BP (ABP) has been shown to be a stronger predictor of all-cause and CV mortality than OBP values.[3,4] Further, ABP monitoring (ABPM) allows to differentiate various phenotypes of hypertension (i.e. white-coat and masked hypertension) and provides prognostic information for risk of CV and renal disease.[5–7] However, the value of integrating ABP as part of risk predictions scores has not been systematically analyzed. The 2018 European Society of Cardiology (ESC)/European Society of Hypertension (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. For the diagnosis of both conditions, which have been associated with increased CV risk compared with normotensive individuals, ABPM is mandatory. Given the limited resources and inability to perform ABPM in all hypertensive individuals, identification of clinical characteristics associated with either condition is warranted to potentially reduce the number of unnecessarily performed ABP recordings.

Machine learning approaches have revolutionized the way data can be processed and analyzed. 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

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

present study utilized data of 63,910 patients included in the Spanish Ambulatory Blood Pressure Monitoring Registry[3] and used machine learning to i) develop a risk score for 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

Data source

This study evaluates data from the Spanish ABP Registry, 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.[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 the European government agencies. The sponsors had no role in the design, data analyses, and interpretation of the data of the study. All patients provided written informed consent and institutional review boards approved the protocols. All patients provided written informed consent and institutional review boards approved the protocols.

A total of 63,910 subjects, aged 18 or older, were enrolled until 2014. Along with OBP and ABP, an additional 52 clinically relevant variables were obtained from interviews and physical examinations during visits and from patient's records. These variables ranged from physical characteristics such as age, height, or weight, to clinical history and medication. The full list of variables considered in this study, along with additional details, is presented in the

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Supplementary Appendix. Office BP was the average result of 2 consecutive readings. Ambulatory BP was monitored using oscillometric (Spacelabs model 90207, Spacelabs Healthcare) devices programmed to record every 20 minutes during the day and every 30 minutes during the night. Day and night cycles were defined based on patient information. Mortality data were recorded during the study period. Date and cause of death were confirmed by the vital registry of the Spanish National Institute of Statistics. Causes of death were determined from certificates coded according to the International Statistical Classification of Diseases. Follow-up started from the first recruitment visit until the date of death or end of recruitment (December 31, 2014). 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. 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. Statistical differences were assessed with the t-test.

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: OBP <140/90 mmHg and 24-hour ABP ≥130/80 mmHg vs OBP <140/90 mmHg and 24-hour ABP <130/80 mmHg;

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

- 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, support vector machine, 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. The support vector machine classifier is a widely used supervised-learning algorithm for classification tasks. Support vector machines transform the data implicitly to a higher dimensional space and attempt to separate it with a linear hyperplane in the higher dimensional space. Only a subset of training points is used together with a decision function to determine the position of the hyperplane. This subset is also known as support vectors. The classifier attempts to maximize the distance between the support vectors and the hyperplane, i.e., the margin. A support vector classification with a linear-function kernel was used. The hyperparameter that controls the separability margin was optimized in the tuning set. We used the signed distance to the hyperplane to compute the resulting ROC curve. 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 2 hidden layers, optimized using stochastic gradient descent. Typically, artificial neurons are organized into layers and a deep neural network contains

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

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 added to reduce overfitting. Learning rate, momentum, and dropout retaining probability hyperparameters were optimized in the tuning set, with early stopping to decide on the number of epochs. All classifiers were implemented in Python. All implementations considered the possible imbalance between the number of subjects in different classes, weighting them accordingly. Logistic regression and support vector machine were implemented using the open-source machine learning library scikit-learn.[18] For the multi-layer perceptron, the open-source neural network library Keras[19] with Tensorflow[20] backend was used. Finally, for gradient boosted decision trees, we resorted to the LightGBM[21] gradient boosting framework.

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. The cross-validation was repeated 20 times with different splits. Performance metrics were averaged, and the standard deviation computed.

Feature selection

The Spanish ABP registry dataset collected 56 clinical variables. Limiting the number of features to a minimum was mandatory to increase clinical applicability and enhance the

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

practicability of any machine learning-based tool, as each clinical feature comes with an extra time and potential cost. However, the number of features included had to be sufficient 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 as 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 Supplementary Appendix. The risk calculator is online available at https://ccb-test.cs.uni-saarland.de/cvm-risk.

Risk scores

Several risk scores have been proposed in the literature. The widely used Framingham 10-year risk score (FRS)[22], the European CV disease risk assessment score (SCORE)[23], and the American College of Cardiology/American Heart Association (ACC/AHA) 10-year risk score (ASCVD)[24] 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: 45-64 years; ASCVD: 40-75 years). All analyses were tested in each of these subsets and then independently compared.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Results

Baseline characteristics for the 63,910 patients included in the present analysis are summarized in Table 1. Data are also presented per hypertensive phenotype and vital status at the end of the study. During a median follow-up of 4.7 years, 3,808 deaths occurred from which 1,295 (34 %) were from CV causes.

Feature selection and machine learning approaches

Four different classifiers were used, i.e. logistic regression, support vector machine, deep neural networks, and gradient boosted decision trees. Figure 2 depicts the impact of varying numbers of selected features for each classifier as AUC (left panel) and % of maximum AUC (right panel) for prediction of CV mortality and ambulatory hypertension, top and bottom panels, respectively. 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 almost 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, office systolic BP, number of anti-hypertensive drugs, glucose level, smoking status, and the history of CV disease were common to all classification approaches for the prediction of CV mortality. Office systolic and diastolic BP, sex, smoking status, and treatment with diuretic and statins were common in the selection for all 4 classification approaches for the prediction of ambulatory hypertension. For the prediction of CV mortality, deep neural networks outperformed all other approaches. Feature selection for CV mortality with ambulatory BP data was performed for deep neural networks only. Gradient boosted decision trees and deep neural networks performed better than support

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

vector machines and logistic regression for the detection of ambulatory hypertension. At the cut-off point, deep neural networks outperformed gradient boosting. 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 3.

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 instead of OBP could improve risk assessment. Figure 4 shows the ROC curves for both models. Obtained performance metrics are depicted in Table 3. Prediction of CV mortality with ABP data led to a significant increase in AUC (0.781 vs 0.752, $p = 1.73\times10^{-42}$), accuracy, balanced accuracy, and specificity.

Comparison with established risk scores

The FRS, the SCORE, and ASCVD risk scores were computed with the Spanish ABPM registry data. Due to different exclusion criteria for each of these scores, 3 different subsets had to be 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 Figure 5. 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 (0.685 vs 0.659, $p = 1.97\times10^{-22}$), the SCORE (0.679 vs 0.613, $p = 6.21\times10^{-22}$), and ASCVD (0.722 vs 0.639, $p = 8.03\times10^{-30}$). The boxplots in Figure 6 shows the distribution of the normalized values of the different risk scores for prediction of CV mortality.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Prediction of hypertension phenotypes

A second analysis aimed at evaluating 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 7. Performance metrics are shown in Table 3. The AUC for prediction of the ABP from clinically available features was 0.715. The distinction between different hypertension phenotypes was particularly difficult: normotensive vs masked hypertension achieved an AUC of about 0.659, sustained hypertension (ABP >130/80 mmHg and OBP >140/90 mmHg) vs white-coat hypertension achieved an AUC of about 0.667, respectively.

Discussion

The present analysis from a large cohort of 63,910 individuals indicates that existing risk scores (i.e. FRS, the SCORE, 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, including OBP, which can be collected with relative ease 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 the classification of medical imaging and patient data.[25–27] Various classifiers, namely logistic

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

regression, support vector machine, 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, support vector machines and logistic regression. Hence, we selected deep neural networks as the approach of choice. Algorithmic approaches, such as deep learning[28], whose performance improves as more data become available, offer entirely new opportunities in analyzing 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 63,910 individuals with more than 3.5 million data points and 3,808 deaths (out of which 1,295 were from CV causes). Both novel risk scores proposed herein, CV-Mortality$_{OBP}$ and CV-Mortality$_{ABP}$, outperformed the established FRS, the SCORE, and ASCVD, which were also computed using the data of the Spanish ABPM Registry. When ambulatory BP was added to the model, AUC for prediction of CV mortality significantly increased (0.752 vs. 0.781, $p = 1.73\times10^{-42}$), indicating that the addition of ambulatory BP features enhanced the discriminative power of the classifier. With an AUC in the range of 0.752-0.781, 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 picked

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

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 6 show the distribution of the normalized values of the newly proposed and established risk scores for prediction of CV mortality and illustrates the largest separation between boxes, resulting in the best discrimination (alive vs deceased) for CV-Mortality$_{ABP}$.

Ambulatory BP readings are unanimously recommended by the current North American[24] and European[7] 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.[3,4] Ambulatory BP monitoring is mandatory to detect white-coat and masked hypertension, both of which have been associated with increased CV risk compared with normotensive individuals.[7,29] However, top-ranked barriers to ambulatory BP monitoring are challenges in accessing devices, costs of testing, concerns about the willingness or ability of patients to successfully complete tests.[30] 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 has to realize that predicting ambulatory BP with clinical features alone is rather limited. The sensitivity and specificity for the detection of ambulatory, masked, and white-coat hypertension ranged between 0.657-0.661 and 0.573-0.651, respectively. In line, office and ambulatory BP showed only a moderate correlation (intraclass correlation coefficient for SBP: 0.57; intraclass correlation coefficient for DBP: 0.70). These results indicate that

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

clinical and ambulatory features provide very different information and one can barely predict ambulatory output using clinical characteristics, including OBP.

Limitations

In most patients, ABPM was performed at a single time point, thus limiting its prognostic power. Data on medication during the follow-up period 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 completed excluded. Only Caucasian patients were included; hence, the results may not be applicable to people of other races. For the computation of the established FRS, the SCORE, 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.

Conclusion

We developed a novel risk calculator for 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 that there is potential for improvement of 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 to explore its ability in risk prediction.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Conflict of interest

FM received speaker honoraria from Medtronic, Recor, Berlin Chemie and Boehringer Ingelheim, and is supported by Deutsche Hochdruckliga, Deutsche Gesellschaft für Kardiologie and Deutsche Forschungsgemeinschaft (SFB TRR219). All other authors report no conflict of interest.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

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. *Lancet*. 2016;388(10060):2665-2712. doi:10.1016/S0140-6736(16)31134-5

2. Goff DC, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk. *Circulation*. 2014;129(25 suppl 2):S49-S73. doi:10.1161/01.cir.0000437741.48606.98

3. Banegas JR, Ruilope LM, de la Sierra A, et al. Relationship between Clinic and Ambulatory Blood-Pressure Measurements and Mortality. *N Engl J Med*. 2018;378(16):1509-1520. doi:10.1056/NEJMoa1712231

4. Yang W-Y, Melgarejo JD, Thijs L, et al. Association of Office and Ambulatory Blood Pressure With Mortality and Cardiovascular Outcomes. *JAMA*. 2019;322(5):409. doi:10.1001/jama.2019.9811

5. Niiranen TJ, Mäki J, Puukka P, Karanko H, Jula AM. Office, Home, and Ambulatory Blood Pressures as Predictors of Cardiovascular Risk. *Hypertension*. 2014;64(2):281-286. 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*. 2011;171(12):1090-1098. 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*. 2018;39(33):3021-3104. doi:10.1093/eurheartj/ehy339

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

8. Engemann DA, Raimondo F, King J-R, et al. Robust EEG-based cross-site and cross-protocol classification of states of consciousness. *Brain*. 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*. 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*. 2019;9(1):14465. doi:10.1038/s41598-019-50567-5

11. Guimarães P, Keller A, Fehlmann T, Lammert F, Casper M. Deep-learning based detection of gastric precancerous conditions. *Gut*. 2019;2:gutjnl-2019-319347. 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*. 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*. 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*. 2014;35(46):3304-3312. 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

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Registry experience. *Eur J Clin Invest*. 2016;46(1):92-98. 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*. 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*. 2014;32(7):1359-1366. doi:10.1097/HJH.0000000000000221

18. Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F, Mueller A. Scikit-learn. *GetMobile Mob Comput Commun*. 2015;19(1):29-33. doi:10.1145/2786984.2786995

19. Chollet F. Keras. https://keras.io. Published 2015.

20. Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. March 2016. http://arxiv.org/abs/1603.04467.

21. Ke G, Meng Q, Finley T, et al. LightGBM: A highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst*. 2017;2017-Decem(Nips):3147-3155.

22. Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of Coronary Heart Disease Using Risk Factor Categories. *Circulation*. 1998;97(18):1837-1847. doi:10.1161/01.CIR.97.18.1837

23. Piepoli MF, Hoes AW, Agewall S, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice. *Eur Heart J*. 2016;37(29):2315-2381. doi:10.1093/eurheartj/ehw106

24. Whelton PK, Carey RM, Aronow WS, et al. *2017*

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

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 a Report of the American College of Cardiology/American Heart Association Task Force on Clinical Pr. Vol 71.; 2018. doi:10.1161/HYP.0000000000000065

25. Howard JP, Cook CM, van de Hoef TP, et al. Artificial Intelligence for Aortic Pressure Waveform Analysis During Coronary Angiography. *JACC Cardiovasc Interv*. 2019;12(20):2093-2101. doi:10.1016/j.jcin.2019.06.036

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

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

28. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444. doi:10.1038/nature14539

29. Pickering TG, Shimbo D, Haas D. Ambulatory Blood-Pressure Monitoring. *N Engl J Med*. 2006;354(22):2368-2374. doi:10.1056/NEJMra060433

30. 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*. 2017;11(9):573-580. doi:10.1016/j.jash.2017.06.012

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	63910	10913	35327	29957	5370	17670	60102	1295
Female	26860 (42.0)	5329 (48.8)	12952 (36.7)	11037 (36.8)	1915 (35.7)	8579 (48.6)	25127 (41.8)	639 (49.3)
Age, years	58.4 ± 14.2	57.6 ± 14.9	58.1 ± 13.9	58.4 ± 13.8	56.8 ± 14.4	59.6 ± 14.2	57.9 ± 14.0	68.5 ± 13.7
Height, cm	164.2 ± 11.7	163.6 ± 11.6	165.0 ± 11.8	164.8 ± 11.7	165.7 ± 11.8	163.1 ± 11.6	164.4 ± 11.7	160.0 ± 12.4
Weight, kg	78.7 ± 14.7	77.0 ± 14.8	79.4 ± 14.6	79.6 ± 14.6	78.5 ± 14.5	78.3 ± 14.9	78.8 ± 14.7	75.5 ± 14.9
Risk factors and history								
Smoker	10141 (15.9)	1512 (13.9)	6374 (18.0)	5411 (18.1)	963 (17.9)	2255 (12.8)	9588 (16.0)	172 (13.3)
Diabetic	12510 (19.6)	1906 (17.5)	7268 (20.6)	6261 (20.9)	1007 (18.8)	3336 (18.9)	11271 (18.8)	444 (34.3)
Target organ damage	7266 (11.4)	1227 (11.2)	4265 (12.1)	3597 (12.0)	668 (12.4)	1774 (10.0)	6410 (10.7)	324 (25.0)
History of CV disease	8347 (13.1)	1772 (16.2)	4452 (12.6)	3666 (12.2)	786 (14.6)	2123 (12.0)	7378 (12.3)	390 (30.1)
Glucose level (mg/dl)	105.5 ± 29.7	102.9 ± 27.1	106.4 ± 30.9	106.9 ± 31.3	103.6 ± 28.1	105.2 ± 28.6	104.9 ± 29.0	120.3 ± 43.9
Blood pressure (mmHg)								
Ambulatory systolic BP	129.2 ± 13.7	116.5 ± 7.8	137.9 ± 11.1	138.8 ± 11.3	133.2 ± 8.9	119.5 ± 7.1	128.6 ± 13.1	138.5 ± 18.4

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

Office systolic BP	147.9 ± 18.8	125.7 ± 10.1	152.6 ± 18.2	156.7 ± 16.3	129.4 ± 8.0	152.2 ± 13.6	147.3 ± 18.4	156.0 ± 21.9
Ambulatory diastolic BP	76.5 ± 10.1	69.7 ± 6.6	81.8 ± 9.4	82.0 ± 9.6	80.6 ± 7.9	70.3 ± 6.7	76.6 ± 10.0	74.7 ± 12.2
Office diastolic BP	86.7 ± 11.6	76.3 ± 8.3	89.2 ± 11.5	91.1 ± 11.1	78.9 ± 8.0	87.9 ± 9.7	86.8 ± 11.4	84.3 ± 13.1
No. of antihypertensive drugs								
0	25682 (40.2)	4221 (38.7)	14833 (42.0)	12555 (41.9)	2278 (42.4)	6628 (37.5)	24589 (40.9)	320 (24.7)
1	13791 (21.6)	2666 (24.4)	7211 (20.4)	5981 (20.0)	1230 (22.9)	3914 (22.2)	13081 (21.8)	218 (16.8)
≥2	24437 (38.2)	4026 (36.9)	13283 (37.6)	11421 (38.1)	1862 (34.7)	7128 (40.3)	22432 (37.3)	757 (58.5)
Treatment with								
Statin	9721 (15.2)	1838 (16.8)	4940 (14.0)	4141 (13.8)	799 (14.9)	2943 (16.7)	9015 (15.0)	258 (19.9)
Alpha blocker	3189 (5.0)	365 (3.3)	2073 (5.9)	1809 (6.0)	264 (4.9)	751 (4.3)	2842 (4.7)	127 (9.8)

CV – Cardiovascular; BP – Blood Pressure; OBP – Office BP; ABP – Ambulatory BP. Normotension: office BP <140/90 mmHg, ABP <130/80 mmHg; ambulatory hypertension: ABP ≥130/80 mmHg; sustained hypertension: ABP ≥130/80mmHg; masked hypertension: office BP <140/90 mmHg, ABP≥130/80 mmHg; white-coat hypertension: office BP ≥140/90 mmHg, ABP <130/80 mmHg; Data are numbers (%), or mean ± SD

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

Table 2. Feature selection

	CV-Mortality$_{OBP}$	CV-Mortality$_{ABP}$	Ambulatory Hypertension
1	Age (years)	Age (years)	Office systolic BP (mmHg)
2	Office systolic BP (mmHg)	Ambulatory systolic BP (mmHg)	Office diastolic BP (mmHg)
3	Number of anti-hypertensive drugs (n)	Number of anti-hypertensive drugs (n)	Sex (male/female)
4	Glucose level (mg/dl)	Glucose level (mg/dl)	Age (years)
5	Target organ damage (yes/no)	Target organ damage (yes/no)	Weight (kg)
6	History of CV disease (yes/no)	History of CV disease (yes/no)	Smoker (yes/no)
7	Smoker (yes/no)	Smoker (yes/no)	Diabetic (yes/no)
8	Diabetic (yes/no)	Diabetic (yes/no)	Treatment with diuretic (yes/no)
9	Treatment with statins (yes/no)	Sex (male/female)	Treatment with alpha blocker (yes/no)
10	Height (cm)	Office systolic BP (mmHg)	Treatment with statins (yes/no)

CV – cardiovascular; BP – blood pressure.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

Table 3. Classification performance metrics

	AUC	Accuracy	Balanced Accuracy	Sensitivity	Specificity
CV Mortality$_{OBP}$	$0.752 \pm 9.5\times10^{-4}$	$0.723 \pm 4.8\times10^{-3}$	$0.691 \pm 2.3\times10^{-3}$	$0.657 \pm 7.0\times10^{-3}$	$0.724 \pm 5.0\times10^{-3}$
CV Mortality$_{ABP}$	$0.781 \pm 1.4\times10^{-3}$	$0.772 \pm 3.7\times10^{-3}$	$0.710 \pm 2.1\times10^{-3}$	$0.645 \pm 6.2\times10^{-3}$	$0.774 \pm 3.9\times10^{-3}$
Ambulatory Hypertension	$0.715 \pm 3.6\times10^{-4}$	$0.656 \pm 1.3\times10^{-3}$	$0.656 \pm 8.0\times10^{-4}$	$0.661 \pm 7.9\times10^{-3}$	$0.651 \pm 7.3\times10^{-3}$
Masked Hypertension	$0.659 \pm 1.7\times10^{-3}$	$0.602 \pm 2.8\times10^{-3}$	$0.617 \pm 2.0\times10^{-4}$	$0.661 \pm 9.1\times10^{-3}$	$0.573 \pm 7.7\times10^{-3}$
White-coat hypertension	$0.667 \pm 4.8\times10^{-4}$	$0.614 \pm 1.9\times10^{-3}$	$0.622 \pm 9.5\times10^{-3}$	$0.653 \pm 5.8\times10^{-3}$	$0.590 \pm 6.1\times10^{-3}$

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

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

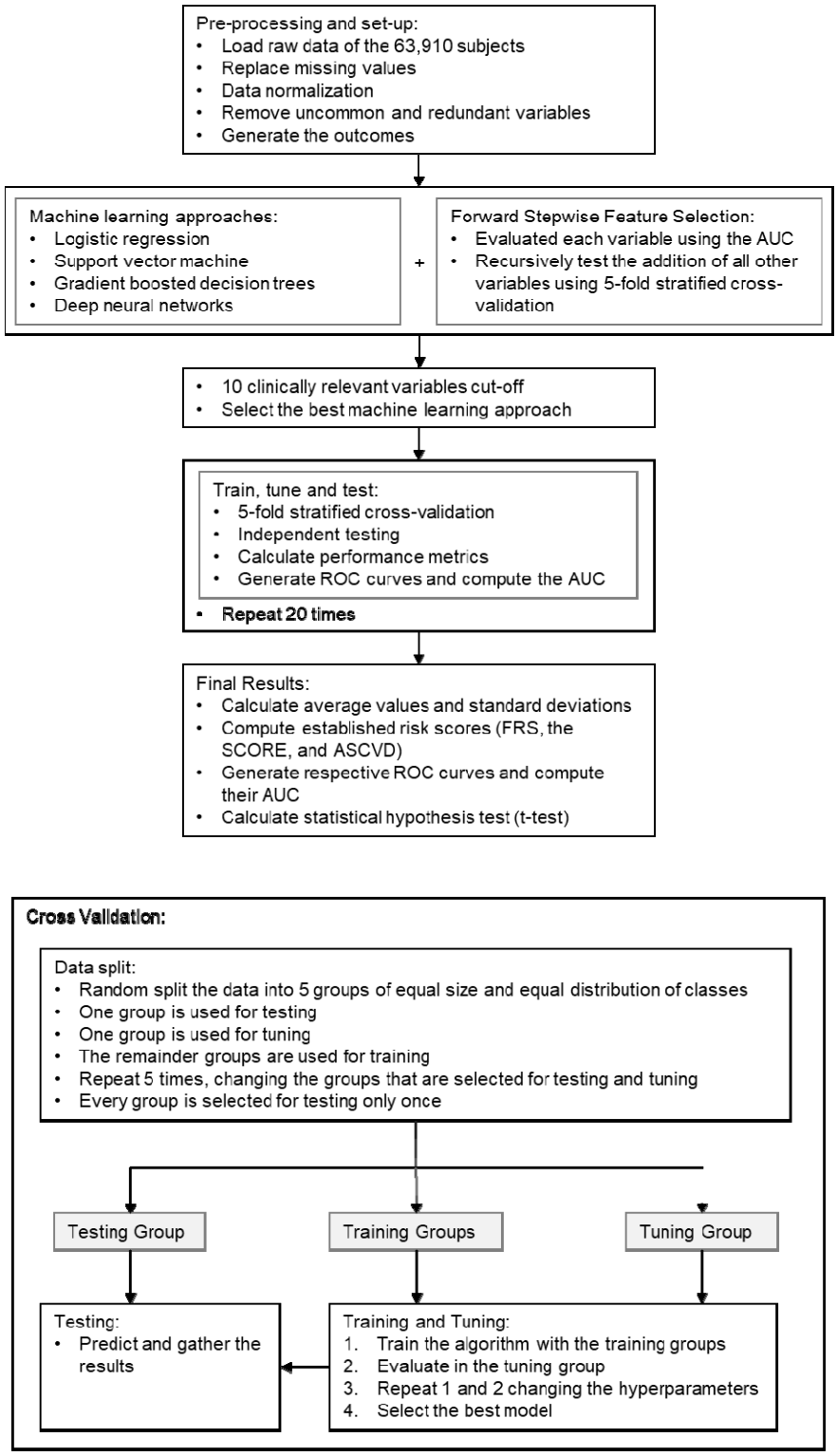

Figure 1. Data analysis workflow. Stepwise forward feature selection was used with four different machine-learning approaches to select the best feature set and classifier

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

for each outcome. Cross validation guaranteed that the test set was completely independent of the training and tuning sets. Training, tuning and testing was repeated 20 times to gather final average results and variability. The workflow is further detailed in the Methods section.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

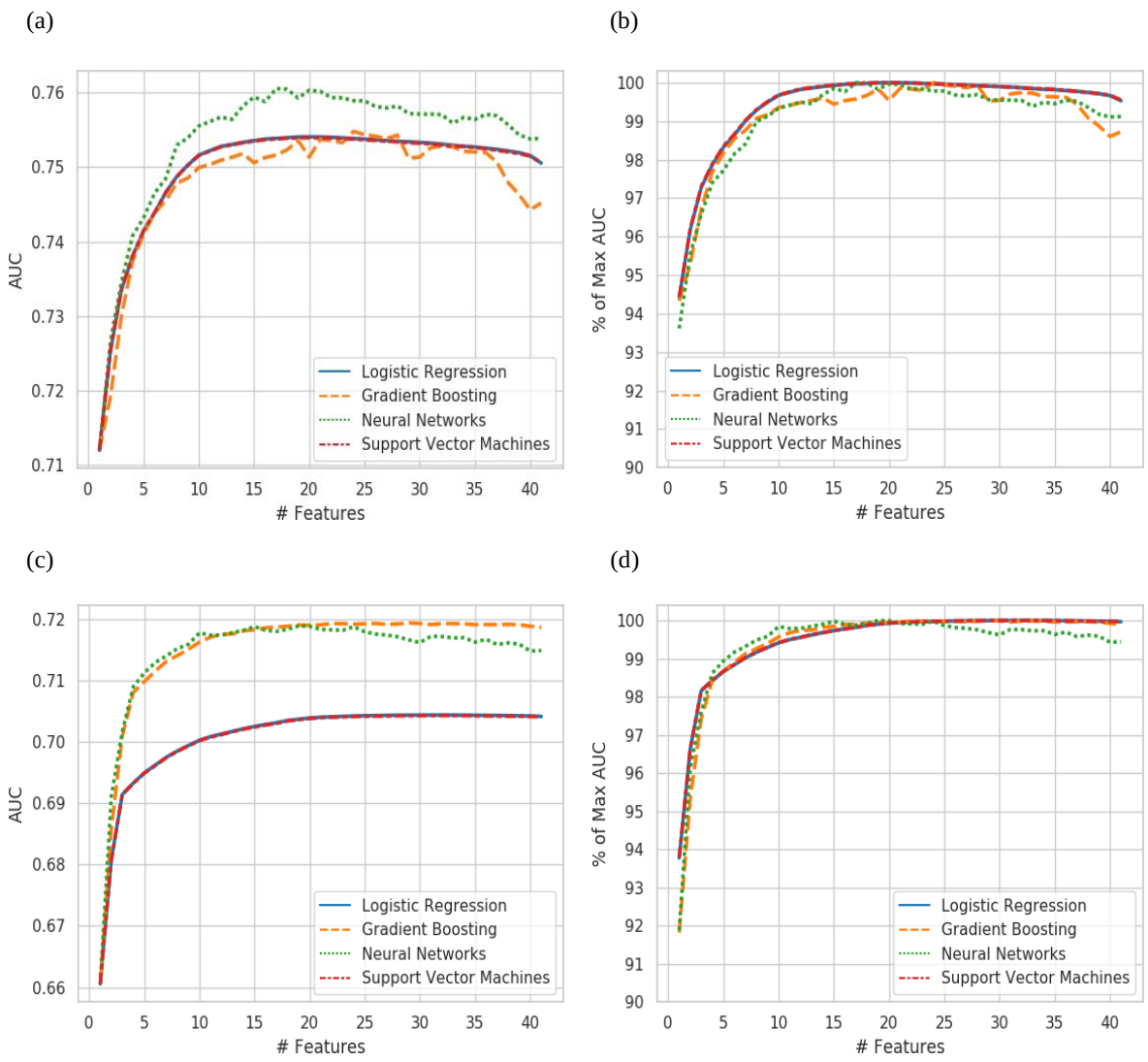

Figure 2. On the left, the area under the curve (AUC) is shown versus the number (#) of features, for each classifier. On the right the same information is presented but as a percentage of the maximum AUC value. (a) and (b) for cardiovascular mortality prediction using office BP. (c) and (d) for prediction of ambulatory hypertension.

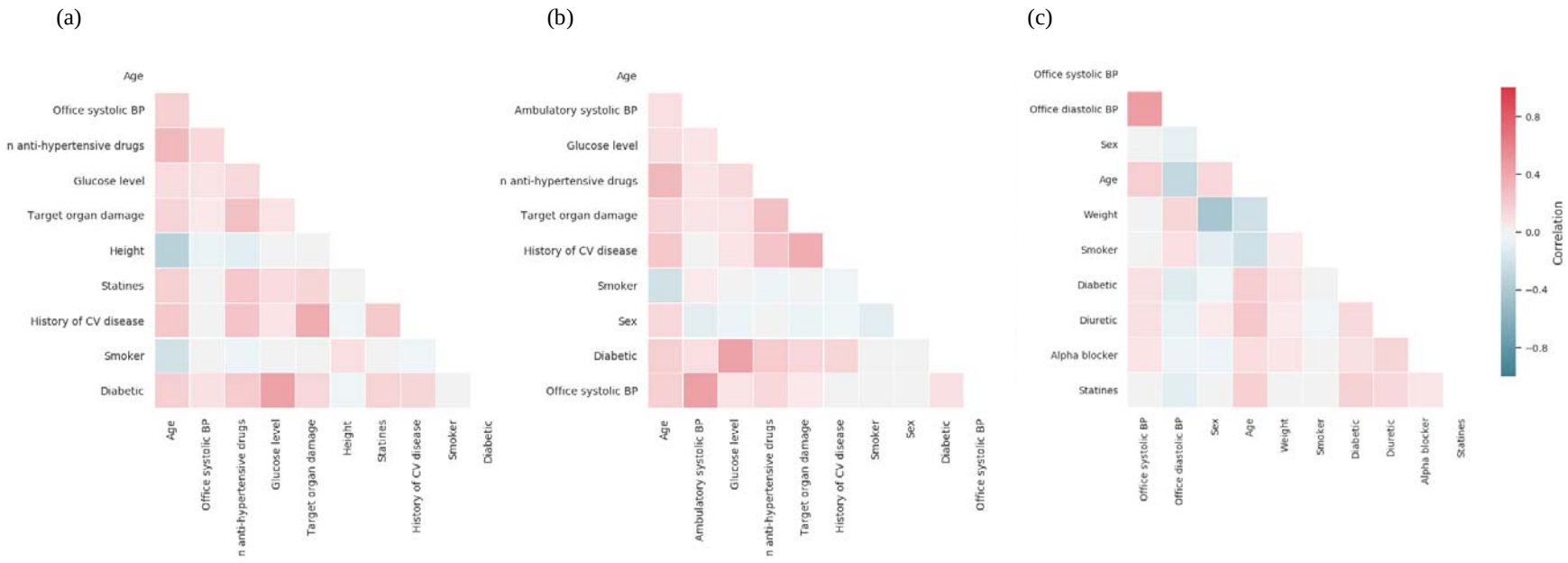

Figure 3. Pairwise correlations for the sorted best 10 features as determined by feature selection for the prediction of: (a) cardiovascular (CV) mortality using office blood pressure data (CV-MortalityOBP), (b) CV mortality using ambulatory blood pressure (CV-MortalityABP) data, and (c) ambulatory hypertension. BP – Blood Pressure; n – number, CV – Cardiovascular.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

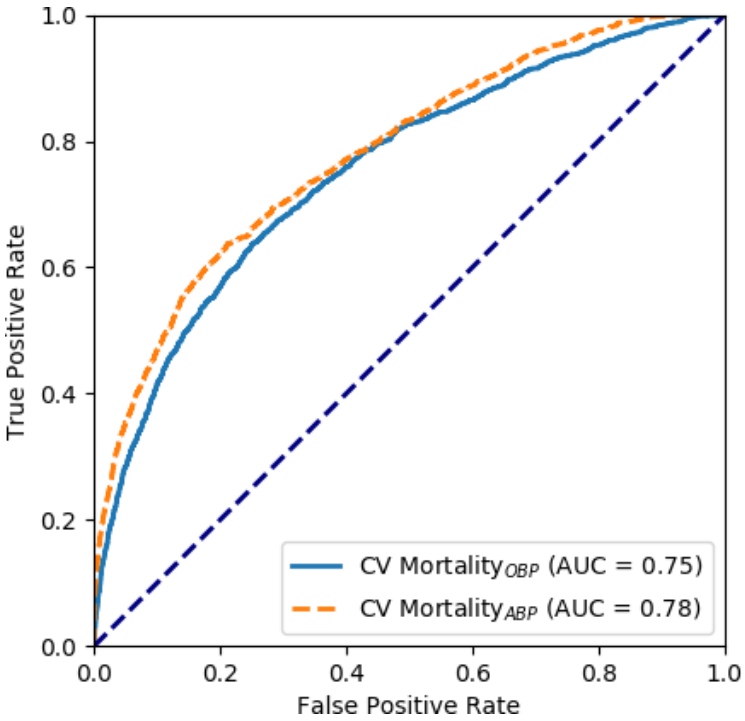

Figure 4. Receiving operating characteristics (ROC) curves for cardiovascular (CV) mortality prediction using office blood pressure (CV-Mortality$_{OBP}$) and ambulatory blood pressure (CV-Mortality$_{ABP}$) data.

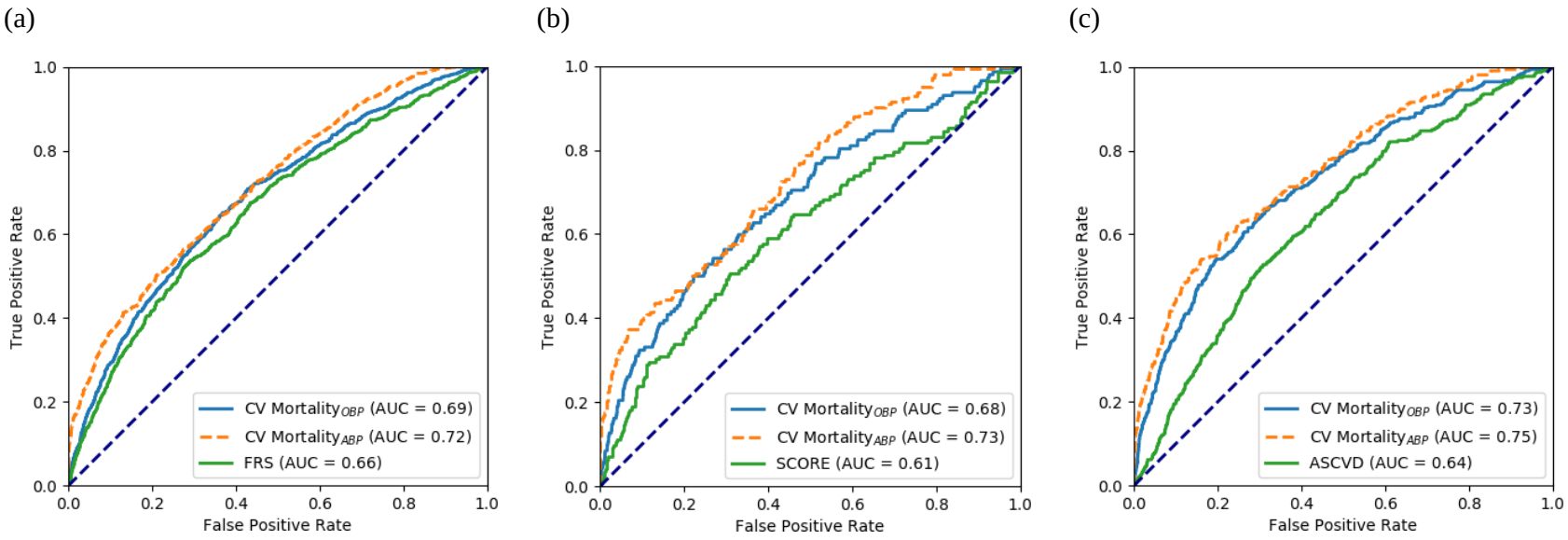

Figure 5. Comparison of receiving operating characteristics (ROC) curves between the proposed approaches, using office blood pressure (CV-Mortality$_{OBP}$) and ambulatory blood pressure (CV-Mortality$_{ABP}$) data, and (a) Framingham (FRS) risk score, (b) the SCORE and (c) ASCVD.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

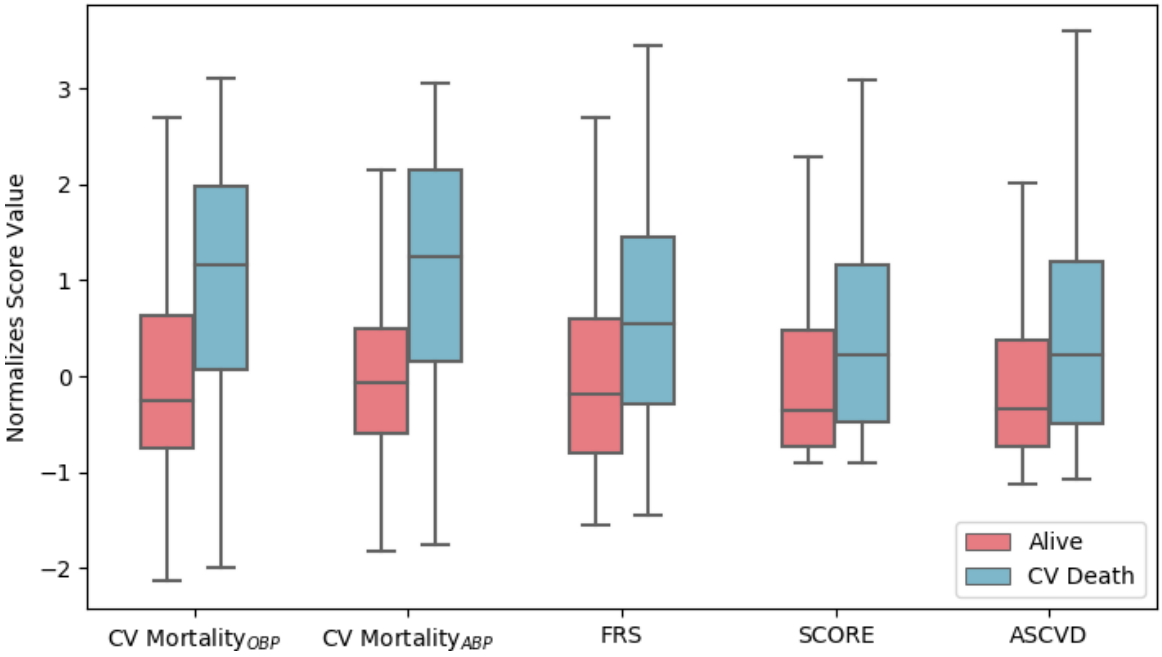

Figure 6. Distribution of the normalized values of the risk scores using office blood pressure (CV-Mortality$_{OBP}$), ambulatory blood pressure (CV-Mortality$_{ABP}$) data, Framingham (FRS) risk score, the SCORE and ASCVD. Red – alive; Blue – Cardiovascular (CV) death.

medRxiv preprint doi: https://doi.org/10.1101/2020.01.17.20017798; this version posted January 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

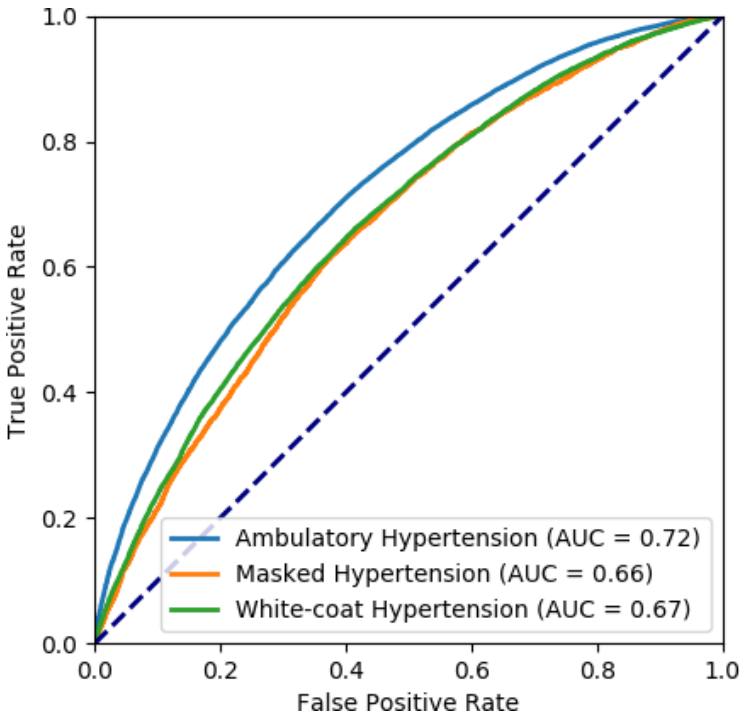

Figure 7. Receiving operating characteristics (ROC) curves for the detection of ambulatory hypertension, masked hypertension and white-coat hypertension.