Information Fusion 58 (2020) 24–39

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Multi-level information fusion for learning a blood pressure predictive model using sensor data

Monika Simjanoska [a,*], Stefan Kochev [a], Jovan Tanevski [b], Ana Madevska Bogdanova [a], Gregor Papa [c], Tome Eftimov [c]

[a] Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University, Rugjer Boshkovikj 16, Skopje 1000, Macedonia
[b] Department of Knowledge Technologies, Jožef Stefan Institute, Jamova cesta 39, Ljubljana 1000, Slovenia
[c] Computer Systems Department, Jožef Stefan Institute, Jamova cesta 39, Ljubljana 1000, Slovenia

ARTICLE INFO

Keywords:
Sensor fusion
Data fusion
Information fusion
Performance metric fusion
Blood pressure prediction

ABSTRACT

The availability of commercial wearable bio-sensors provides an opportunity for developing smart phone applications for real-time diagnosis that can be used to improve the health of the user. We propose a multi-level information fusion approach for learning a predictive model for blood pressure (BP) using electrocardiogram (ECG) sensor data. The approach fuses the information on five different levels: i) data collection, where data from multiple ECG sensors is collected; ii) feature extraction, where features are extracted from the collected data by different preprocessing methods; iii) information fusion, fusing the evaluation information from different classifiers; iv) information fusion using the information from multi-target regression models for each BP class; and v) information fusion using the information from multi-target regression models from all configurations as a single model. This is used for predicting the blood pressure values (systolic BP (SBP), diastolic BP (DBP), and mean arterial pressure (MAP)). Evaluating the methodology by using a separate test set indicates that the multi-level information fusion provides promising results, which are acceptable and comparable to the state-of-the-art results obtained for blood pressure prediction.

1. Introduction

According to the World Health Organization (WHO) statistics, nearly 63% of the global deaths are considered to be preventable [1]. Up to 30% of them are related to cardiovascular disease. High blood pressure (BP) is one of the factors that increases the risk of cardiovascular diseases and strokes [2,3]. Therefore, hypertension detection is crucial for appropriate and on-time treatments [4].

Recently, the expansion of commercial wearable bio-sensors enables real-time monitoring of multiple physiological signals and other vital signs (heart rate, respiratory rate, and oxygen saturation), with the exception of blood pressure, aim to bring the possibility for developing smart phone applications for real-time diagnosis [5]. For monitoring blood pressure, either standalone devices are developed for non-invasive monitoring [6–12], or, blood pressure is estimated by combining multiple physiological signals [13–19]. However, no strong evidence of the reliability of those approaches exists [20–22].

Since electrocardiography (ECG) is most common physiological signal that is emitted by the commercially available wearable bio-sensors, we consider the task of predicting blood pressure solely from ECG signals. Being able to make such predictions would be of great advantage to preventive medicine, especially in ambulatory/clinical situations [23,24] and military environments [25].

The task of BP prediction from ECG signals has been approached from the point of analyzing the relationship between morphological changes in ECG and BP at a very coarse level [26,27]. In our previous work, we approached the problem from another side, relying on complexity analysis (whose reliability is previously proved for other medical conditions [28–30]) and machine learning. Relying on the literature, we applied a band-pass Butterworth filter, allowing frequencies from 0.30 Hz up to 50 Hz. We chose the lowest frequency to be 0.30 Hz since it assures removing the baseline wander without deforming the ECG signal [31]. We obtained promising results when predicting the systolic (SBP), the diastolic (DBP), and the mean arterial pressure (MAP) [32].

In this paper, we follow an information fusion paradigm, combining it with a specific data-driven preference approach [33] aimed at predicting blood pressure from ECG signals. The idea is to extract features from the ECG signal by applying different settings of parameters used for preprocessing (i.e. sampling length and baseline removal cut-off fre-

* Corresponding author.
E-mail addresses: monika.simjanoska@finki.ukim.mk (M. Simjanoska), stefan.kocev@students.finki.ukim.mk (S. Kochev), jovan.tanevski@ijs.si (J. Tanevski), ana.madevska.bogdanova@finki.ukim.mk (A.M. Bogdanova), gregor.papa@ijs.si (G. Papa), tome.eftimov@ijs.si (T. Eftimov).

https://doi.org/10.1016/j.inffus.2019.12.008
Received 1 November 2018; Received in revised form 6 November 2019; Accepted 22 December 2019
Available online 23 December 2019

quencies), in order to produce different datasets, which will be further processed and used in order to learn SBP, DBP, and MAP predictive models. Since the lowest ECG component is known to occur at 0.05 Hz [34], by setting the lowest frequency to start from 0.05 Hz, up to 0.50 Hz, with the step of 0.05 Hz, we decided to fuse the information from multiple predictive models to achieve as accurate and reliable prediction as possible. Furthermore, in our previous work [32], we relied on the statement that 30 s long ECG segments are proved to carry enough information [35–37] for creating robust estimations. In this paper, we inspect multiple signal length 10 s, 20 s and 30 s. The upper limit is chosen as the maximum affordable time not transcending the minimum time needed to perform the traditional cuff-based BP measurements.

The contributions of the paper are as follows:

- A methodology for fusing multi-level information for learning a predictive model for blood pressure using ECG sensor-fused data.
- Models (i.e. configurations) evaluation, using ECG signals with different lengths, 10 s, 20 s, and 30 s, filtered with different cut-off frequencies starting from 0.05 Hz, up to 0.50 Hz, with the step of 0.05 Hz.
- A contribution to the arguments published in the literature discussing the optimal ECG sample lengths needed for building predictive models [35–37], as well as the lower frequencies where the ECG components overlap with the baseline wander noise [31,38–40].
- We identify that fusing models learned for different configurations can improve the results.

In the reminder of the paper, we first explain the problem (Section 2). In Section 3, we present an overview of the related work. In Section 4, we present the multi-level information fusion methodology for learning a BP predictive model using sensor data. We present the experimental results obtained from the evaluation in Section 5 followed by a discussion in Section 6. Finally, we present the conclusions and the directions for future work in Section 7.

2. Problem definition

A biomedical signal processing system encompasses the biological system of interest, the sensors used to capture the activity of the biomedical system and the methodology developed to analyze the signals and extract the desired information [41]. In our study, the analyzed biological system is the heart system whose electrical activity is presented through the ECG signals. There are two phases for managing the blood flow,a systole phase known as pumping phase, and a diastole phase known as filling phase. The pressure that the blood flow causes on the blood vessels walls is referred to as BP and is measured in millimeters of mercury (mmHg). The maximum pressure during one heart beat is the SBP and the minimum pressure in between two heart beats is DBP.

Having proved the relation between the complexity features of the ECG signals and the BP [32], we intent to discover the most reliable ECG settings in order to produce as accurate predictions as possible. The work is inspired by real-life civil and military emergency situations where we are not always guaranteed of placing desired ECG sensors, or that there will be enough time to measure long ECG samples. The applicability of this research is mostly seen in specific occasions where there is a lack of medical equipment, and thus it can be combined with additional mobile technology in order to assure the transfer of the medical data to the remote hospital centers [42–44].

3. Related work

An overview of techniques and approaches based on the photoplethysmographic (PPG) signal for non invasive and continuous measurement of blood pressure is presented at [45]. Gravina et al. [46] provide a comprehensive review of the state-of-the-art techniques on multi-sensor fusion in the area of Body Sensor Networks. The advantages of such approach are particularly found in the physical activity recognition field [47], emotion recognition[48,49], and health in general [50]. However, multi-sensor data fusion requires sophisticated signal processing techniques, since the wearable sensor devices are characterized with limited computational power and storage capacity. Even though, fusing data from various and diverse sensors assures the inference of high quality information, it is still a demanding task for the developed monitoring applications as more complex methods are applied for signal preprocessing, features extraction, and machine learning.

Working with ML models, at the end, the learned predictive models should be evaluated. For this reason different performance metrics can be applied on a test set and regarding them the best predictive model for the specific application scenario is selected [51]. These performance metrics are related to different properties that describe the quality of solutions. There are studies in which researchers select the best model based only on one performance metric or they report the results for several performance metrics and then explain them for each metric separately using statistical analyses [52,53]. Recently, researchers are not only interested in results obtained for each performance metric separately, and this is not only a case in ML, but they want to make a general conclusion in which the results for all of them will be fused. One way to treat the problem of evaluation regarding a set of performance metrics is to treat them as a multiple criteria for which a set of predictive models (i.e., algorithms, alternatives) are evaluated. This is a well-known problem in decision making, for which PROMETHEE methods [54] are used to solve a decision problem in which a set of alternatives are evaluated according to a set of criteria that are often conflicting. The idea behind these methods is that an evaluation matrix is constructed, in which each alternative is estimated for each criteria. They perform pairwise comparisons between all the alternatives for each criteria to provide either a complete or partial rankings of the alternatives. They are applicable in different domains such as, business, chemistry, manufacturing, social sciences, agriculture and medicine [54–56]. Recently, they were also used in a data-driven approach for evaluating multi-objective optimization algorithms regarding different performance metrics [33]. In this study, all involved performance metrics are equal, but the influence of each performance metric is fused, based on its estimated performance according to its entropy.

4. Methodology

Using the idea of information fusion, we propose a multi-level information fusion methodology aimed to predict SBP, DBP, and MAP from single ECG signal. Fig. 1 explains the complete methodology, which consists of five different levels:

- *Level 1.* Since we measure a single phenomena (ECG) with multiple sensors, the data is collected (i.e. fused) into a single pool used for knowledge discovery.
- *Level 2.* Next, feature datasets are extracted from multiple data sources (multiple sensors), representing signals sampled at different lengths and filtered with different cut-off frequencies for noise removal. This step is essential since it assures that none of the ECG information will be lost when filters for baseline removal are applied, and at the same time it assures that the baseline information will be removed and will not affect the prediction models. The features for each dataset are extracted following the same procedure. The number of features, and their interpretation is the same across datasets, however, the values differ depending on the length of the signal and the cut-off frequency that are used. Later in the text, these feature sets referred to as “different configurations”, explaining the diversity in sample lengths and cut-off frequencies.
- *Level 3.* Further, each of the different datasets is used as input to a stacking procedure for BP classification. Since we apply a stacking based classification, the outputs from all classifiers are fused and used as an input into a single META classifier [57] (**features-level**

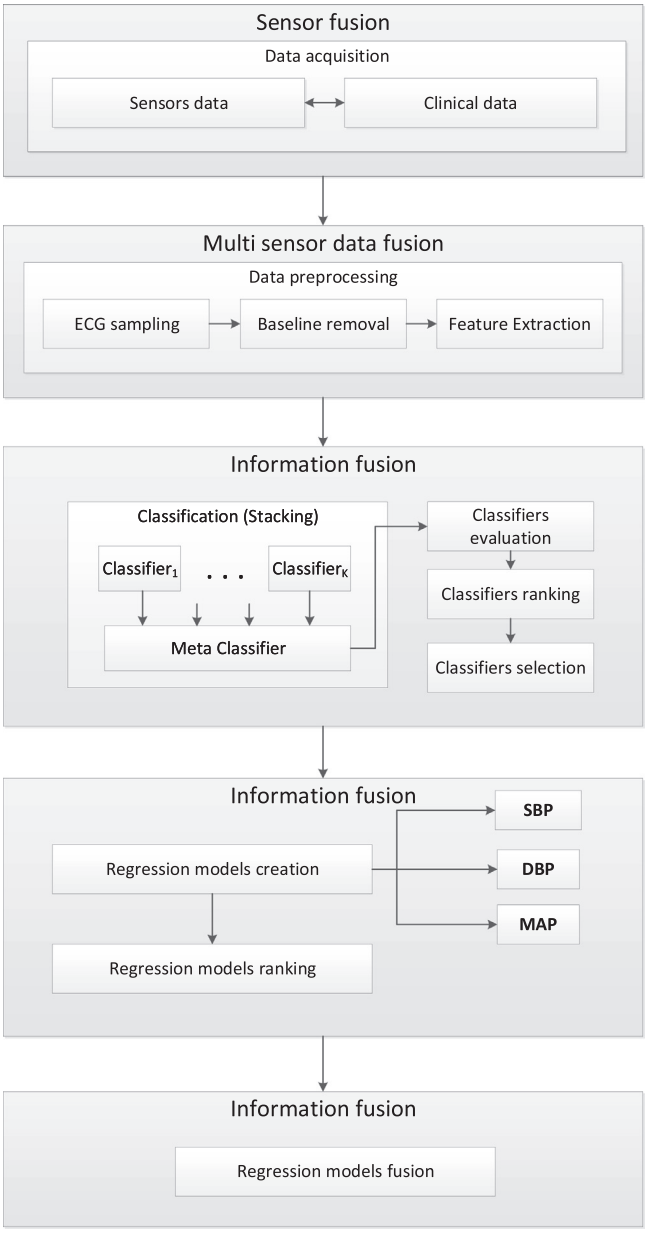

Fig. 1. The methodology for SBP, DBP and MAP prediction.

fusion). Due to the nature of the samples, we use two distinct splitting procedures for evaluation, from which we obtain multiple performance metrics. Next, a data-driven preference approach based on the PROMETHEE method [54] is applied to select the best classification model for each feature dataset. The selection itself can also be considered as **performance metrics fusion**, since it selects the best training and validation set for each configuration. At this level, we have **information fusion**, since information covering by different performance metrics is merged to select the best training and validation set for each configuration.

- *Level 4.* Having chosen the most representative feature dataset for each different configuration, regression models are built for each feature dataset, considering each of the BP classes, with the aim to predict the real SBP, DBP and MAP values. At this level, a **decision-level fusion** is applied since the final prediction on the SBP, DBP, and MAP values is made by using the probabilities produced in the previous Level 3. Next, these models are ranked by using the same ranking approach as in Level 3; the difference is that the performance metrics used are those for evaluation of regression models. At this level, we have **information fusion**, since information provided by the regression model for each BP class is merged together to produce a single prediction for SBP, DBP, and MAP for each configuration.
- *Level 5.* At this level the information obtained from the regression models for all configuration is merged together to obtain a single prediction for SBP, DBP, and MAP. The information is fused by using the influence of each regression model, which is estimated by using the rankings of the training and the validation sets on which the regression models are trained and evaluated, correspondingly. These rankings come from the evaluation of the classifiers used for each configuration since from this evaluation the selection of the training and the validation set for each configuration is made. Here, **information fusion** is performed, since information from different models is fused to obtain a single value prediction for SBP, DBP, and MAE.

Table 1
Application of data fusion for BP prediction in each methodology step.

Step	Fusion level
ECG data gathering	Multi-sensor fusion
Data preprocessing	Multi-sensor fusion
ML (Classification)	Information fusion
ML (Regression)	Information fusion
Regression models fusion	Information fusion

Table 2
Sensors and datasets summary information.

Dataset	Reliability	Participants	Age
Cooking hacks [62]	[65–67]	16	16–72
180° eMotion FAROS [61]	[68–70]	3	25–27
Zephyr Bioharness module [63]	[71–74]	14	20–73
Savvy sensor platform [60]	[75,76]	21	15–54
Charis Physionet database [64]	[77]	7	20–74

Table 1 provides a summary of data fusion levels recognized in each step of the BP prediction methodology.

Fig. 1 provides a description of the steps of the methodology proposed for SBP, DBP, and MAP prediction, focusing on the modules for ranking the developed ML models according to a set of performance metrics and selecting the most suitable accordingly. A comprehensive description of the ML approach for developing the classification is presented in [32]. The regression problem is tackled as a multi-target regression task [58,59], in contrast to building separate independent regression models, as was done in [32]. The following subsections describe each of the modules presented in Fig. 1.

4.1. ECG Signals acquisition

To create reliable and robust models, we used five distinct sources of ECG data, four of which we obtained from commercially available ECG sensors [60–63] and one from the online available Physionet database [64]. Each ECG signal is accompanied with reference SBP and DBP values measured in parallel by using electronic sphygmomanometer.

Table 2 provides summary information for each sensor, as well as its reliability considered within the available literature, the number of participants measured with the sensor, and their age. All the human subjects involved in the research with commercial sensors have signed an agreement allowing their ECG data to be used for research goals. The participants are of diverse age, health status, personal estimation of the current stress level, and were measured in both moving (normal walking) and sitting positions. The measurements obtained from the Physionet database are explicitly reported to be form patients suffering from brain injuries.

More sensors details as well as datasets public availability information can be read in [32].

4.2. ECG preprocessing

ECG sampling refers to the length of the ECG sample that is processed in the feature extraction module. The baseline removal refers to removal of noise caused by respiration.

Taking into account the different signal lengths 10 s, 20 s, and 30 s, and the different cut-off frequencies starting from 0.05 Hz, up to 0.50 Hz, with the step of 0.05 Hz, 30 different datasets (configurations) were created, which were processed in the feature extraction module.

4.3. Feature extraction

Recently published work relies on multiple physiological signal analysis for BP prediction, taking into account morphological features of the signal [17,20–22,78–80]. We propose approaching the problem from another perspective: representing the ECG information via complexity metrics: signal mobility, signal complexity, fractal dimension, autocorrelation and entropy [32].

Therefore, each of the 30 datasets, created in the previous module explained in Section 4.2, has gone through the process of feature extraction, calculating the metrics as follows.

4.3.1. Signal mobility

Given $x_i, i = 1, N$ is the ECG signal of length N and $d_j = x_{j+1} - x_j$ is the first-order variations in the signal, then the first-order factors, S_0 and S_1, are calculated as:

$$S_0 = \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N}} \tag{1}$$

$$S_1 = \sqrt{\frac{\sum_{j=2}^{N-1} d_j^2}{N-1}} \tag{2}$$

The signal mobility quantitatively measures the level of variation in the signal. It is calculated as a ratio between the factors S_1 and S_0:

$$Mobility = \frac{S_1}{S_0}, \tag{3}$$

4.3.2. Signal complexity

Given the first-order variation of the ECG signal d_j, $j = 1, N-1$, the second-order variation of the signal is presented by $g_k = d_{k+1} - d_k$. Then, the second-order factor is calculated as:

$$S_2 = \sqrt{\frac{\sum_{k=3}^{N-2} g_k^2}{N-2}}, \tag{4}$$

Hereupon, the complexity is calculated by using the first-order and the second-order factors:

$$Complexity = \sqrt{\frac{{S_2}^2}{{S_1}^2} - \frac{{S_1}^2}{{S_0}^2}} \tag{5}$$

Both the signal mobility and signal complexity were computed by using the Hjorth parameters method [81].

4.3.3. Fractal dimension

Fractal dimension measures the self-similarity of the signal. By zooming and comparing different portions, it describes fundamental patterns hidden in the signal. Higuchi algorithm [82,83] has been used to calculate the fractal dimension. The method works with a set of k subseries with different resolutions, creating a new time series X_k, for $m = 1, k$:

$$X_k^m : x(m), x(m+k), x(m+2k), \ldots, x(m + \lfloor \frac{N-m}{k} \rfloor k) \tag{6}$$

The length of the curve X_k^m, $l(k)$ is calculated as:

$$l(k) = \frac{(\sum_{i=1}^{\lfloor N-m/k \rfloor} |x(m+ik) - x(m+(i-1)k)|(N-1))}{(\lfloor \frac{N-m}{k} \rfloor)k} \tag{7}$$

Then, for each k in range 1 to $k_{\max}$, the average length is calculated as the mean of the k lengths $l(k)$ for $m = 1, \ldots k$. The fractal dimension is the estimation of the slope of the plot $ln(l(k))$ vs. $ln(1/k)$.

4.3.4. Autocorrelation

The similarity between the signal and its shifted version is measured via autocorrelation. Let τ be the amount of shift, and then the autocorrelation is calculated as:

$$r_{xx}(\tau) = \int_{-\inf}^{+\inf} x(t)x(t-\tau)p_{xx}(x(t), x(t-\tau))dt, \tag{8}$$

where $p_{xx}(x(t), x(t-\tau))$ presents the joint probability density function of $x(t)$ and $x(t-\tau)$.

4.3.5. Entropy

Entropy expresses the randomness of the signal. The decrease of entropy often indicates a disease [84]. The amount of information is expressed through the concept of probability. Let p_i denote the probability of each outcome x_i within the ECG signal X for $i = 1, N-1$. Then, entropy is calculated as:

$$Entropy = \sum_{i=0}^{N-1} p_i \log(\frac{1}{p_i}) \tag{9}$$

4.4. Classification

Having created 30 datasets (configurations) described by the five features for each of them we extracted the feature values. The first step for the SBP, DBP and MAP prediction according to our methodology is to predict the BP class. Each classifier predicts the probability of a given feature vector to belong in each of the k BP classes.

Classification itself follows a stacking design, in which we model the problem by using seven different classifiers: Bagging [85], Boosting [86], SVM [87], K-means [88], Random Forest [89], Naive Bayes [90], and J48 [91]. The output probabilities from each of them for all the feature vectors to belong in each class, comprise a new feature vectors upon which a META classifier (Random Forest) is trained. Details of the design are provided in [32].

4.5. Ranking of classifiers

For each configuration, a classifier is trained by using a training set and it is evaluated on a corresponding validation set. The procedure is repeated m times on different choice training and validation sets. Thus, for each configuration, m classifiers are trained.

To evaluate the performance of the classifiers trained for each configuration, the focus is not only on one performance metric, but on a set of 17 performance metrics, including:

- **Accuracy (ACC)** - represents the fraction between the number of correct predictions by the total number of predictions made.
- **Cohen's Kappa** - compares the observed accuracy (the number of correctly classified instances) with the expected accuracy (taking into account the number of instances in each class, along with the number of instances that the classifier agreed with the ground truth label).
- **Precision (PR)** - captures the effect of the large number of negative examples on the classifier's performance, by comparing false positives (FP) to true positives (TP) rather than true negatives (TN), i.e. it measures how many of the predicted instances were TP.
- **Recall (RC)** - to measure how many of the TP instances were predicted.

- **F-Measure** - measures the trade-off between recall and precision giving equal importance to recall and precision.
- **Area Under PRC** - precision-recall curve does not account for TN, since TN is not a component of either Precision or Recall. Given there are more negatives (normal) than positives (hypertension), PRC might be very informative evaluation metric of the performance of the classifier. Having a model at the upper right corner, means that the classifier gets only the TP with no FP and no false negatives (FN) at all and thus is perfect classifier.
- **Area Under ROC** - the ROC curve plots the True Positive Rate (TPR) vs. False Positive Rate (FPR). Therefore achieving a model at the upper left corner, means that the classifiers is getting no FP at all and thus is perfect classifier.
- **Matthews Correlation (COR)** - takes the advantage from all four metrics TP, FP, TN and FN to calculate the correlation coefficient between the observed and predicted classifications [92].
- **Relative absolute error (RAE)** - is an error measure relative to a simple predictor, meaning it takes the total absolute error and normalizes it by dividing it by the one of the simple predictor.
- **Root relative squared error (RRSE)** - takes the total squared error and normalizes it by dividing by the total squared error of the simple predictor. The square root reduces the error to the same dimensions as the problem at hand.
- **Root mean squared error (RMSE)** - explains the standard deviation of the prediction error.;
- **Informedness (INF)** - quantifies how informed a classifier is for the specified class. It provides insight into how consistently the classifier predicts the class [93].
- **Markedness (MAR)** - quantifies how marked a class is for the specified classifier, i.e. how consistently the class has the classifier as a marker by combining measures about correct classifications [93].
- **Micro F-measure (MF)** - aggregates the contributions of all classes to compute the average F-measure by considering the total TPs, FNs and FPs.
- **Log likelihood (LL)** - presents the probability of the observed prediction given the real class.
- **Mutual information (MI)** - measures whether the real and the predicted labels are statistically depended.
- **Pearson's chi-squared test (PRS)** - to measure whether the observed and expected frequencies are the same by comparing the predicted contingency table to an expected table.

4.5.1. The PROMETHEE II

The idea behind this process is to select the best model that can be learned for a particular "length-frequency" configuration (as explained above) and for this reason we need to rank the classifiers, such that the contribution of each performance metric (PM) is taken into account. Let us consider that the number of PMs used is p.

A comparison needs to be made between m classifiers (i.e., alternatives) regarding p performance evaluation metrics (i.e., criteria). One way to do this is to use the PROMETHEE methods. They are used in decision making to solve a decision problem in which a set of alternatives are evaluated according to a set of criteria that are often conflicting. For the method, an evaluation matrix is constructed, in which each alternative is estimated for each criteria. The method performs pairwise comparisons between all the alternatives for each criteria to provide either a complete or partial rankings of the alternatives.

Let $C = \{C_1, C_2, \dots, C_m\}$ be the set of classifiers we want to compare regarding the set of performance metrics $PM = \{pm_1, pm_2, \dots, pm_n\}$. The decision matrix is a $m \times p$ matrix (see Table 3) that contains the values obtained for the classifiers for each PM separately.

The appropriate method in our case is PROMETHEE II. It is based on pairwise comparisons that need to be made between all classifiers for each performance metric. The differences between values for each pair of classifiers according to a specified performance metric are taken into consideration. For larger differences the decision maker might consider larger preferences. The preference function of a performance metric for two classifiers is defined as the degree of preference of classifier C_1 over classifier C_2 as seen in the following equation:

Table 3
Decision matrix.

	pm_1	pm_2	...	pm_p
C_1	$pm_1(C_1)$	$pm_2(C_1)$	...	$pm_p(C_1)$
C_2	$pm_1(C_2)$	$pm_2(C_2)$	...	$pm_p(C_2)$
⋮	⋮	⋮	⋮	
C_m	$pm_1(C_m)$	$pm_2(C_m)$	...	$pm_p(C_m)$

$$P_j(C_1, C_2) = \begin{cases} p_j(d_j(C_1, C_2)), & if\ maximizing\ the\ performance\ metric \\ p_j(-d_j(C_1, C_2)), & if\ minimizing\ the\ performance\ metric \end{cases}, \tag{10}$$

where $d_j(C_1, C_2) = pm_j(C_1) - pm_j(C_2)$ is the difference between the values of the classifiers for the performance metric pm_j and $p_j(\cdot)$ is a generalized preference function assigned to the performance metric. Six types of generalized preference functions are known to exist [94]. Some of them require certain preferential parameters to be defined, such as the preference and indifference thresholds. The preference threshold is the smallest amount that is assumed as preference, while the indifference threshold is the greatest amount of difference that is insignificant. In our case, a V-shape generalized preference function is used for each performance metric, in which the threshold of strict preference, q, is set to the maximum difference that exists for each preference metric for each configuration. The V-shape preference function is presented in Eq. 11. Using this preference function, all difference values are take into account using a linear function.

$$p(x) = \begin{cases} 0, & x \le 0 \\ \frac{x}{q}, & 0 \le x \le q \\ 1, & x > q \end{cases}, \tag{11}$$

After selecting the preference function for each PM, the next step is to define the average preference index and outranking (preference and net) flows. The average preference index for each pair of classifiers gives information of global comparison between them using all PMs. The average preference index can be calculated as:

$$\pi(C_1, C_2) = \frac{1}{p} \sum_{j=1}^{p} w_j P_j(C_1, C_2), \tag{12}$$

where w_j represents the relative significance (weight) of the j^{th} PM. The higher the weight value of a given PM the higher its relative significance. The selection of the weights is a crucial step in the PROMETHEE II method because it defines the priorities used by the decision-maker. In our case, we used the Shannon entropy weight method [33]. For the average preference index, we need to point out that it is not a symmetric function, so $\pi(C_1, C_2) \neq \pi(C_2, C_1)$.

To rank the classifiers, a net flow for each classifier needs to be calculated. It is the difference between the positive preference flow, $\phi(C_i^+)$, and the negative preference flow of the classifier, $\phi(C_i^-)$. The positive preference flow gives information how a given classifier is globally better than the other classifiers, while the negative preference flow gives the information about how a given classifier is outranked by all the other classifiers. The positive preference flow is defined as:

$$\phi(C_i^+) = \frac{1}{(m-1)} \sum_{x \in C} \pi(C_i, x), \tag{13}$$

while the negative preference flow is defined as:

$$\phi(C_i^-) = \frac{1}{(m-1)} \sum_{x \in C} \pi(x, C_i). \tag{14}$$

The net flow of a classifier is defined as:

$$\phi(C_i) = \phi(C_i^+) - \phi(C_i^-). \tag{15}$$

The PROMETHEE II method ranks the classifiers by ordering them according to decreasing values of net flows.

4.6. Regression

Upon ranking the classifiers, the training sets that were used for creating the classification models that produced best results for each configuration, were used to train regression models in order to determine approximate SBP, DBP, and MAP values. MAP is calculated by using the following equation [95]:

$$MAP = \frac{SBP + 2 \times DBP}{3} \tag{16}$$

At this phase, we train three multi-target regression models, where the targets are the SBP, DBP, and MAP values. The three multi-target regressions refer to the three distinct BP classes, which will be further explained, and therefore we obtained MR_{ij} models, for $i = 0, 1, 2$ denoting the classes and $j = 0, 1, 2$ denoting SBP, DBP and MAP, correspondingly. The training set is split into three training sets according to the BP class. Each multi-target regression model is an ensemble (random forest) of predictive clustering trees, which generalize classification trees [96].

The idea in this phase is to fuse the probabilities produced by the META classifier with the values predicted by the MR models. Hereupon, the final BP predictions are obtained as follows:

$$SBP = \sum_{i=0}^{2} p_i MR_{i0}; \tag{17}$$

$$DBP = \sum_{i=0}^{2} p_i MR_{i1}; \tag{18}$$

$$MAP = \sum_{i=0}^{2} p_i MR_{i2}; \tag{19}$$

where p_0, p_1, and p_2 are the probabilities for the current feature vector to belong in each of the BP classes.

4.7. Ranking of regression models

After we obtain regression models, the methodology used for classifiers ranking is used again with the difference the performance metrics that are used are those for regression evaluation. The method of ranking the regression models includes training on separate subset, followed by performance estimation using unseen data. The three regression models were evaluated by using the correlation between the real and the predicted values, the Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). This is made with the purpose to select one best configuration that can be used for SBP, DBP, and MAP prediction.

- **Correlation coefficient (COR)** - presents the correlation between the real and the predicted values, for which the Pearson correlation coefficient is calculated.
- **Mean squared error (MSE)** - is the average of squared differences between the predicted and the actual BP values and shows how concentrated the data is around the line of best fit.
- **Mean Absolute Error (MAE)** - is the average error obtained from the absolute differences between the real and the predicted values.
- **Root Mean Squared Error (RMSE)** - is used to represent a higher weight for the large errors, which is especially important for the BP problem, meaning the differences between the real absolute and the predicted values are first squared, then averaged, and afterwards a square root of the average is taken.

Using the previously explained methodology, the regression models are ranked considering the information gained from each performance metric.

4.8. Regression models fusion

The main goal of the proposed methodology is to fuse the information of each configuration in order to predict the SBP, DBP, and MAP values. For this reason, an information fusion is performed by merging the information of all regression models (i.e. configurations) using their separate influences. The influence is estimated using the rankings for the training and the validation sets on which the regression models are trained and evaluated. These rankings come from the evaluation of the classifiers used for each configuration because from this evaluation the selection of the training and the validation set for each configuration is made.

The obtained rankings, that actually provide an information of how good is the particular configuration comparing to the others, are normalized for their sum to be equal to 1. Thus, once a regression model predicts a value, that value is multiplied by the ranking value of the configuration for that particular regression model. Having the predictions of all regression models multiplied by their rankings, they are fused (summed up) to produce a single reliable SBP, DBP and MAP prediction.

Formally, let P be the patient's ECG for which SBP, DBP and MAP need to be predicted. Denoting the rank for the i-th configuration with R_i and the total number of models with T, the normalized rank, NR_i is calculated as follows:

$$NR_i = \frac{\frac{(T-R_i)}{T}}{\sum_{i=1}^{T} \frac{(T-R_i)}{T}} \tag{20}$$

Eventually, the predicted SBP, DBP or MAP values are obtained as a sum of the contributions from all the models. Let FP denote the fused predicted value (i.e. it can be the value for SBP, DBP, or MAP), NR is the normalized rank and the p is the predicted value. Then the prediction can be calculated as expected value from all configurations, which is the sum of all predictions multiplied by the corresponding ranks of the particular configuration, as explained in the following equation:

$$FP = \sum_{i=1}^{T} p_i NR_i \tag{21}$$

5. Evaluation

5.1. Level 1: Data collection

The data collection explained in Section 2 was crucial for achieving appropriate design of the study. Each human subject was measured only by one ECG sensor (i.e. Cooking hacks, 180 eMotion Faros, Zephyr Bioharness module, and Savvy sensor platform) and the ground truth for the blood pressure SBP and DBP reference values were measured in parallel using a medically certified electronic sphygmomanometer. The electronic sphygmomanometer (MyTech) is a cuff-based device certified by Conformité Européenne (CE) and US Food and Drug Administration (FDA) for blood pressure monitoring and has been previously used as a reliable device in other studies [97,98]. For a given time, the device provides two values, one for SBP and one for DBP. The human subjects that had the ECG sensors attached to their chest were in parallel periodically measured for the blood pressure with the sphygmomanometer. Since our goal was to develop a methodology that can be used in occasion where there is a lack of standard medical equipment, we measured short ECG signals and therefore, there were not big deviations for the SBP and DBP values that were measured periodically. In the case of the Charis Physionet database, each ECG signal is associated with the continuous reference SBP and DBP values. In this case to each signal, average values of the SBP and DBP values were assigned.

Given that each measurement is accompanied with SBP and DBP values, we enriched this information by including the calculated of MAP and the BP class. To obtain the BP class and map the ECG measurements accordingly, we used the publicly available scheme presented in Table 4 [99,100]. Even though, the BP classes seem not to be mutually exclusive,

Table 4
Rules and categorization.

Joined class	Class	SBP (mmHg)	Logical	DBP (mmHg)
Normal	HPTN	<=90	OR	<=60
	N	90–119	AND	60–79
Prehypertension	PHTN	120–139	OR	80–89
Hypertension	S1HTN	140–159	OR	90–99
	S2HTN	>=160	OR	>=100
	ISHTN	>=140	AND	<90
	HTNC	>=180	OR	>=110

Table 5
Descriptive statistics of the targets.

Target	Mean	Median	Mode	SD	Var	Min	Max
SBP	128.31	126	115	20.89	436.42	84	208
DBP	71.97	71	68	11.76	138.41	37	134
MAP	90.75	89	85	11.24	126.37	53	152

we solved the problem of giving priority to the more severe BP conditions by checking the "AND" conditions at the end of the ECG samples mapping procedure. To solve the imbalanced-class data, we grouped BP classes into three main categories: hypotension (HPTN) and normal (N) as Normal class (denoted as 0); prehypertension (PHTN) as Prehypertension class (denoted as 1); and stage 1 hypertension (S1HTN), stage 2 hypertension (S2HTN), isolated systolic hypertension (ISHTN), and hypertensive crisis (HTNC) as Hypertension class (denoted as 2).

5.2. Level 2: Feature extraction

Collecting the data, by acquisition of measurements from different sources measured at different sampling rates and under various conditions, we organized them in three datasets. Each dataset contains one x seconds long random sample per measurement, where x=10 s, 20 s, and 30 s. Table 5 provides descriptive statistics for the targets (SBP, DBP and MAP) in general, no matter if 10, 20 or 30 s sample lengths are taken from the signals.

As mentioned in Section 4.2, the baseline wander and the ECG information might overlap, and if not cut-off appropriately, either important information might be lost, or, a sensor-biased predictor may be developed. Considering the literature, we decided to find the proper cut-off frequency experimentally. Each of the three datasets was preprocessed with 10 different cut-off frequencies, starting from 0.05 Hz up to 0.50 Hz by step of 0.05 Hz. Calculating the complexity features for each, we obtained 30 different configurations, or, 30 feature datasets used for training classification models for BP prediction and multi-target regression models for predicting the SBP, DBP, and MAE values.

5.3. Level 3: Information fusion using the evaluation information from the classifiers for each configuration

The datasets should be split on an appropriate way in terms of finding the best training, validation and testing sets, which will be used for learning classification models for BP prediction and multi-target regression models for SBP, DBP, and MAE prediction. Taking into account that we have multiple **independent measurement** for each subject (i.e. participant) and as the **number of measurements varies for each subject**, we proposed and evaluated two approaches as follows:

1. **Traditional split**: In the first approach, each measurement is considered to be independent. Before start of the experiment, chosen randomly, 15% of all measurements has been taken as test set and are removed from the pool. Hereafter, for each of the 30 configurations, the rest of the pool is 30 times split by using 75% for training and 25% for validation. For each split, a classifier is created as described in Section 4.4. As an output, we have obtained 30 performance matrices with dimension 30x17, where 30 is the number of splits and 17 is the number of performance metrics (described in Section 4.5) obtained by using the corresponding validation set.
2. **Custom split**: In the second approach, we follow the rule that if a subject is included in one set, none of its measurements may occur in another set. Similarly as in the previous approach, 15% random subject with all their measurements have been taken as distinct test set and are removed from the pool. Following the same rule, for each of the 30 configurations, the rest of the pool is 30 times split by using 75% of the subjects for training and 25% of the subjects for validation. For each split, a classifier is created, and again, we have obtained 30 performance matrices of same dimensions as in the previous case.

Hereupon, the ranking method described in Section 4.5.1 is applied to find the best performance train/validation split in each configuration, considering not one performance metric, but a set of 17, of which 14 (ACC, KAPPA, PRC, ROC, F, COR, PR, RC, INF, MAR, MF, LL, MI, and PRS) require maximization, and 3 (RAE, RRSE, and RMSE) require minimization. The chosen best train/validation split in each configuration has produced the results presented in Tables 6 and 8, for the traditional and custom split correspondingly. The rows represent the configurations, starting from 10 s with 0.05 Hz, up to 30 s with 0.50 Hz.

Taking into account the best representative from each configuration, we applied the same ranking method to find the best configuration. The last column in tables 6 and 8 presents the results from the ranking method. In the case of the traditional split, the best classifier is obtained from the dataset containing features calculated with signal length of 30 s and filtered by using cut-off frequency of 0.35 Hz. In the case of the custom split, the best classifier is obtained for the configuration in which the signal length is 10 s and the cut-off frequency of 0.30 Hz.

To consider the influence of all trials, not only the best, we provided an insight into the ranking of the mean performance values by configuration. For this reason, the results of each performance measure was averaged across all trials for each configuration. The results for both splitting procedures are presented in Tables 7 and 9. Using them, in the case of the traditional split, the best configuration is obtained using signal length of 30 s filtered with cut-off frequency of 0.50 Hz. On the other side, in the case of the custom split, the configuration with a signal length of 30 s and cut-off frequency of 0.10 Hz is ranked as the best.

However, if we consider all four tables, it can be noted that the ranks from 1 to 10, presenting the degradation of the classifier in ascending order, are mainly focused in the 30 s-cluster of the feature datasets.

5.4. Level 4: Information fusion using the information from multi-target regression model for each BP class

In the next step, each training set was split into three subsets according to the BP class (0, 1, and 2). The three training subsets are used to train three multi-target regression models able to predict the SBP, DBP, and MAP values.

Repeating the procedure for the selected representatives for all configurations, we have obtained a classifier that produces probabilities for a given feature vector to belong in each of the three BP classes, and three multi-target predicting machines able to produce SBP, DBP and MAP values. The final SBP, DBP and MAP values are computed by using the Eqs. (17)–(19), as described in Section 4.6. By this step, we aim to introduce the level of uncertainty by the classifier's prediction over the classes.

The methodology is tested on the external testing datasets we created at the beginning, one per each spiting procedure. Tables 10 and 11 present the regression results for the traditional and the custom case, respectively. The rankings in both Tables 10 and 11 favor 30 s to be most informative with variable frequency of 0.40 Hz and 0.50 Hz in both cases, respectively. We need to mention that for applying the rank-

Table 6
Best classification results for each configuration following the traditional approach.

Conf	ACC	KAPPA	PRC	ROC	F	COR	PR	RC	RAE	RRSE	RMSE	INF	MAR	MF	LL	MI	PRS	Ranking
10/0.05	62.64	0.44	0.58	0.75	0.62	0.44	0.62	0.63	60.32	99.99	0.47	0.44	0.45	0.63	58.49	0.22	108.80	30
10/0.10	63.02	0.45	0.59	0.75	0.63	0.44	0.63	0.63	60.70	98.98	0.47	0.44	0.45	0.63	62.04	0.23	113.45	29
10/0.15	66.04	0.49	0.66	0.80	0.66	0.49	0.66	0.66	53.36	91.39	0.43	0.49	0.50	0.66	67.58	0.26	128.29	24
10/0.20	66.79	0.50	0.66	0.80	0.66	0.50	0.66	0.67	52.74	91.80	0.43	0.50	0.52	0.67	76.05	0.29	139.57	21
10/0.25	66.04	0.49	0.63	0.79	0.66	0.49	0.66	0.66	54.48	94.55	0.45	0.49	0.50	0.66	70.22	0.26	131.07	26
10/0.30	67.17	0.51	0.67	0.81	0.67	0.51	0.67	0.67	52.88	93.77	0.44	0.51	0.51	0.67	74.61	0.28	140.49	20
10/0.35	66.42	0.50	0.68	0.81	0.66	0.49	0.66	0.66	54.56	91.93	0.43	0.49	0.51	0.66	72.72	0.27	134.24	22
10/0.40	69.43	0.54	0.66	0.81	0.69	0.54	0.69	0.69	49.32	90.30	0.43	0.54	0.56	0.69	92.99	0.35	164.39	14
10/0.45	68.68	0.53	0.66	0.81	0.68	0.53	0.68	0.69	51.07	92.01	0.43	0.53	0.54	0.69	84.97	0.32	153.19	16
10/0.50	70.57	0.56	0.67	0.81	0.71	0.56	0.71	0.71	47.80	91.31	0.43	0.56	0.55	0.71	89.11	0.34	169.56	8
20/0.05	64.53	0.47	0.64	0.79	0.64	0.47	0.65	0.65	56.09	96.77	0.46	0.47	0.47	0.65	63.29	0.24	119.38	28
20/0.10	66.04	0.49	0.64	0.78	0.66	0.49	0.66	0.66	56.55	96.33	0.45	0.49	0.49	0.66	69.30	0.26	130.68	27
20/0.15	71.32	0.57	0.70	0.83	0.71	0.57	0.72	0.71	48.77	86.82	0.41	0.57	0.57	0.71	93.12	0.35	175.72	7
20/0.20	67.92	0.52	0.69	0.82	0.68	0.52	0.68	0.68	50.09	87.66	0.41	0.52	0.52	0.68	75.98	0.29	143.74	17
20/0.25	69.06	0.54	0.67	0.81	0.69	0.54	0.69	0.69	51.28	90.79	0.43	0.54	0.53	0.69	76.84	0.29	149.91	15
20/0.30	66.04	0.49	0.69	0.82	0.66	0.49	0.66	0.66	52.90	92.99	0.44	0.49	0.49	0.66	71.55	0.27	133.50	23
20/0.35	67.17	0.51	0.66	0.80	0.67	0.51	0.68	0.67	53.25	92.09	0.43	0.51	0.50	0.67	73.06	0.28	138.99	19
20/0.40	69.06	0.54	0.69	0.82	0.69	0.53	0.69	0.69	48.28	87.86	0.41	0.53	0.54	0.69	93.23	0.35	161.51	12
20/0.45	69.43	0.54	0.68	0.82	0.69	0.54	0.69	0.69	48.56	89.90	0.42	0.54	0.54	0.69	90.15	0.34	162.24	13
20/0.50	72.45	0.59	0.72	0.84	0.72	0.59	0.72	0.72	44.67	85.49	0.40	0.58	0.59	0.72	99.74	0.38	184.55	3
30/0.05	67.55	0.51	0.65	0.80	0.67	0.51	0.67	0.68	52.21	91.82	0.43	0.51	0.52	0.68	73.56	0.28	140.16	18
30/0.10	66.04	0.49	0.63	0.79	0.66	0.49	0.66	0.66	52.52	95.07	0.45	0.49	0.50	0.66	73.00	0.28	132.83	25
30/0.15	69.81	0.55	0.69	0.82	0.70	0.55	0.70	0.70	50.11	89.51	0.42	0.55	0.55	0.70	94.26	0.36	168.34	10
30/0.20	72.08	0.58	0.68	0.82	0.72	0.58	0.72	0.72	47.05	89.65	0.42	0.58	0.59	0.72	94.15	0.36	177.48	6
30/0.25	69.81	0.55	0.70	0.84	0.69	0.55	0.69	0.70	46.96	88.40	0.42	0.55	0.56	0.70	88.02	0.33	162.42	9
30/0.30	69.81	0.55	0.71	0.83	0.69	0.55	0.69	0.70	47.57	89.27	0.42	0.55	0.57	0.70	86.39	0.33	160.76	11
30/0.35	**72.83**	**0.59**	**0.75**	**0.86**	**0.72**	**0.59**	**0.72**	**0.73**	**42.69**	**82.24**	**0.39**	**0.59**	**0.60**	**0.73**	**98.87**	**0.37**	**185.14**	**1**
30/0.40	73.58	0.60	0.69	0.83	0.73	0.60	0.73	0.74	42.64	86.07	0.41	0.60	0.61	0.74	102.81	0.39	194.63	2
30/0.45	71.32	0.57	0.72	0.84	0.71	0.57	0.72	0.71	43.56	85.65	0.40	0.57	0.57	0.71	91.83	0.35	173.94	4
30/0.50	71.32	0.57	0.74	0.85	0.71	0.57	0.71	0.71	46.81	86.16	0.41	0.57	0.58	0.71	89.51	0.34	170.30	5

Table 7
Mean classification results for each configuration following the traditional approach.

Conf	ACC	KAPPA	PRC	ROC	F	COR	PR	RC	RAE	RRSE	RMSE	INF	MAR	MF	LL	MI	PRS	Ranking
10/0.05	55.90	0.34	0.55	0.72	0.56	0.34	0.56	0.56	67.66	106.06	0.50	0.34	0.34	0.56	41.21	0.16	73.51	30
10/0.10	57.14	0.36	0.55	0.72	0.57	0.36	0.57	0.57	65.98	104.75	0.49	0.36	0.36	0.57	44.73	0.17	79.91	29
10/0.15	60.31	0.40	0.59	0.75	0.60	0.40	0.60	0.60	61.56	101.06	0.48	0.41	0.41	0.60	52.70	0.20	96.16	26
10/0.20	59.58	0.39	0.59	0.75	0.59	0.39	0.60	0.60	61.48	101.11	0.48	0.39	0.40	0.60	52.82	0.20	94.83	27
10/0.25	61.71	0.43	0.60	0.76	0.62	0.43	0.62	0.62	59.50	99.47	0.47	0.43	0.43	0.62	57.58	0.22	104.93	24
10/0.30	61.72	0.43	0.61	0.76	0.62	0.43	0.62	0.62	59.39	99.42	0.47	0.43	0.42	0.62	59.62	0.22	107.92	23
10/0.35	62.44	0.44	0.62	0.77	0.62	0.44	0.62	0.62	58.41	98.32	0.46	0.44	0.44	0.62	59.65	0.23	109.37	19
10/0.40	62.36	0.44	0.62	0.77	0.62	0.44	0.62	0.62	58.14	98.21	0.46	0.44	0.44	0.62	59.76	0.23	109.30	20
10/0.45	63.41	0.45	0.63	0.78	0.63	0.45	0.64	0.63	56.76	97.20	0.46	0.45	0.45	0.63	62.30	0.24	115.22	12
10/0.50	64.54	0.47	0.64	0.79	0.64	0.47	0.65	0.65	55.09	95.66	0.45	0.47	0.47	0.65	66.12	0.25	122.79	9
20/0.05	58.62	0.38	0.59	0.74	0.59	0.38	0.59	0.59	63.53	102.87	0.48	0.38	0.38	0.59	49.30	0.19	88.73	28
20/0.10	59.99	0.40	0.60	0.75	0.60	0.40	0.60	0.60	61.16	100.92	0.48	0.40	0.40	0.60	54.65	0.21	98.10	25
20/0.15	62.43	0.44	0.62	0.77	0.62	0.44	0.62	0.62	58.09	98.66	0.47	0.44	0.44	0.62	61.82	0.23	112.15	16
20/0.20	62.15	0.43	0.62	0.77	0.62	0.43	0.62	0.62	57.95	98.19	0.46	0.43	0.43	0.62	61.28	0.23	110.71	18
20/0.25	62.28	0.43	0.62	0.77	0.62	0.43	0.62	0.62	58.44	98.69	0.47	0.44	0.43	0.62	60.43	0.23	109.98	22
20/0.30	62.30	0.43	0.62	0.77	0.62	0.43	0.62	0.62	57.57	98.31	0.46	0.44	0.44	0.62	60.86	0.23	110.67	17
20/0.35	63.08	0.45	0.63	0.78	0.63	0.45	0.63	0.63	56.79	97.37	0.46	0.45	0.45	0.63	63.93	0.24	116.04	13
20/0.40	63.03	0.45	0.63	0.78	0.63	0.45	0.63	0.63	57.01	97.40	0.46	0.45	0.44	0.63	64.69	0.24	116.94	14
20/0.45	63.45	0.45	0.63	0.78	0.63	0.45	0.63	0.63	56.34	96.95	0.46	0.45	0.46	0.63	65.98	0.25	118.96	11
20/0.50	64.13	0.46	0.64	0.79	0.64	0.46	0.64	0.64	55.59	95.93	0.45	0.46	0.46	0.64	67.94	0.26	123.58	10
30/0.05	62.96	0.44	0.62	0.77	0.63	0.44	0.63	0.63	57.95	97.14	0.46	0.44	0.45	0.63	60.71	0.23	111.83	15
30/0.10	62.20	0.43	0.62	0.77	0.62	0.43	0.62	0.62	58.17	98.56	0.46	0.43	0.43	0.62	61.49	0.23	111.13	21
30/0.15	64.47	0.47	0.65	0.79	0.64	0.47	0.64	0.64	55.17	95.38	0.45	0.47	0.47	0.64	67.96	0.26	124.11	8
30/0.20	65.53	0.48	0.65	0.80	0.65	0.48	0.65	0.66	53.85	94.20	0.44	0.48	0.49	0.66	71.30	0.27	131.13	6
30/0.25	66.33	0.49	0.66	0.80	0.66	0.49	0.66	0.66	52.30	92.80	0.44	0.50	0.50	0.66	74.72	0.28	137.65	4
30/0.30	64.97	0.47	0.65	0.79	0.65	0.47	0.65	0.65	54.30	95.02	0.45	0.48	0.48	0.65	69.85	0.26	128.06	7
30/0.35	66.50	0.50	0.67	0.80	0.66	0.50	0.66	0.67	52.03	92.87	0.44	0.50	0.50	0.67	73.57	0.28	136.63	3
30/0.40	66.64	0.50	0.67	0.81	0.67	0.50	0.67	0.67	51.48	92.09	0.43	0.50	0.50	0.67	75.69	0.29	140.08	2
30/0.45	66.34	0.50	0.67	0.81	0.66	0.49	0.66	0.66	52.03	92.94	0.44	0.50	0.50	0.66	73.10	0.28	135.94	5
30/0.50	**66.83**	**0.50**	**0.68**	**0.81**	**0.67**	**0.50**	**0.67**	**0.67**	**51.14**	**91.97**	**0.43**	**0.50**	**0.51**	**0.67**	**74.14**	**0.28**	**138.77**	**1**

ing method, the MAE, MSE, and RMSE, require minimization, while the COR requires maximization.

Given the winners and taking a look of their performance in tables 6 and 8, we can perceive that those configurations were ranked as 2nd and 6th in the best classification analysis, and, as 2nd and 9th in the mean classification analysis provided in Tables 7 and 9. On the other hand, the configurations that showed best performance in the classification step (Tables 6 and 8), are now ranked as 12th and 20th (Tables 10 and 11). When the ranking was done on the mean performance values instead of selecting the best for each configuration, the best configura-

Table 8
Best classification results for each configuration following the custom approach.

Conf	ACC	KAPPA	PRC	ROC	F	COR	PR	RC	RAE	RRSE	RMSE	INF	MAR	MF	LL	MI	PRS	Ranking
10/0.05	55.16	0.33	0.54	0.71	0.53	0.32	0.54	0.55	68.23	108.14	0.51	0.31	0.37	0.55	56.59	0.18	96.75	14
10/0.10	46.99	0.24	0.53	0.68	0.45	0.25	0.54	0.47	77.09	109.25	0.54	0.27	0.29	0.47	32.67	0.12	58.12	27
10/0.15	55.30	0.30	0.53	0.70	0.55	0.33	0.56	0.55	66.71	104.72	0.50	0.32	0.34	0.55	40.72	0.19	61.14	16
10/0.20	51.46	0.22	0.51	0.67	0.49	0.23	0.50	0.51	73.99	110.49	0.52	0.24	0.28	0.51	19.38	0.09	29.71	28
10/0.25	54.01	0.31	0.52	0.69	0.54	0.31	0.55	0.54	71.47	110.71	0.52	0.32	0.30	0.54	33.26	0.14	57.64	22
10/0.30	**63.14**	**0.44**	**0.62**	**0.78**	**0.64**	**0.47**	**0.68**	**0.63**	**58.84**	**95.76**	**0.46**	**0.47**	**0.45**	**0.63**	**52.12**	**0.22**	**93.10**	**1**
10/0.35	56.38	0.28	0.60	0.72	0.56	0.32	0.56	0.56	67.75	101.45	0.49	0.32	0.32	0.56	25.18	0.13	40.82	18
10/0.40	50.19	0.24	0.48	0.64	0.50	0.24	0.51	0.50	80.58	116.81	0.55	0.23	0.27	0.50	26.72	0.10	45.47	30
10/0.45	52.68	0.30	0.48	0.66	0.51	0.28	0.52	0.53	75.96	112.93	0.53	0.27	0.34	0.53	32.58	0.15	54.57	24
10/0.50	52.68	0.29	0.49	0.67	0.50	0.28	0.50	0.53	73.16	112.33	0.53	0.27	0.34	0.53	29.21	0.13	49.54	25
20/0.05	56.88	0.35	0.55	0.73	0.54	0.34	0.53	0.57	63.21	106.86	0.51	0.33	0.42	0.57	45.39	0.21	73.83	11
20/0.10	56.07	0.34	0.55	0.72	0.52	0.33	0.53	0.56	69.25	106.40	0.50	0.33	0.42	0.56	48.09	0.20	77.33	12
20/0.15	57.24	0.24	0.57	0.65	0.57	0.23	0.58	0.57	66.19	103.87	0.50	0.23	0.24	0.57	11.33	0.07	20.14	23
20/0.20	57.14	0.32	0.55	0.70	0.57	0.33	0.57	0.57	67.37	104.82	0.50	0.34	0.34	0.57	29.84	0.15	47.90	15
20/0.25	51.64	0.25	0.51	0.66	0.51	0.25	0.52	0.52	75.33	109.22	0.51	0.26	0.23	0.52	26.49	0.11	44.91	26
20/0.30	55.56	0.28	0.58	0.70	0.56	0.30	0.57	0.56	66.80	101.72	0.48	0.30	0.29	0.56	12.89	0.08	20.42	21
20/0.35	65.56	0.37	0.63	0.73	0.65	0.39	0.65	0.66	56.85	94.44	0.46	0.39	0.39	0.66	19.70	0.13	34.84	5
20/0.40	51.03	0.24	0.45	0.61	0.51	0.23	0.51	0.51	78.26	115.49	0.55	0.24	0.22	0.51	39.74	0.10	68.30	29
20/0.45	61.03	0.41	0.58	0.76	0.59	0.41	0.59	0.61	65.11	93.24	0.44	0.41	0.45	0.61	53.41	0.25	86.30	2
20/0.50	54.87	0.33	0.51	0.67	0.55	0.33	0.56	0.55	77.03	103.46	0.49	0.34	0.35	0.55	35.22	0.16	60.30	20
30/0.05	58.64	0.36	0.58	0.75	0.57	0.37	0.56	0.59	61.19	101.95	0.49	0.36	0.40	0.59	70.00	0.24	115.57	7
30/0.10	54.15	0.32	0.53	0.68	0.55	0.34	0.59	0.54	72.85	109.31	0.52	0.35	0.30	0.54	68.93	0.23	115.65	13
30/0.15	60.96	0.41	0.60	0.75	0.62	0.43	0.64	0.61	62.35	97.53	0.47	0.43	0.42	0.61	45.74	0.20	80.81	3
30/0.20	58.48	0.38	0.53	0.69	0.57	0.37	0.58	0.58	67.72	105.90	0.50	0.37	0.41	0.58	47.47	0.21	77.22	10
30/0.25	52.73	0.30	0.54	0.69	0.52	0.34	0.59	0.53	74.38	108.59	0.52	0.36	0.32	0.53	52.02	0.19	88.54	19
30/0.30	56.07	0.33	0.52	0.68	0.56	0.33	0.56	0.56	70.54	105.38	0.50	0.32	0.33	0.56	43.67	0.16	79.52	17
30/0.35	58.37	0.37	0.58	0.74	0.57	0.37	0.57	0.58	64.60	101.66	0.48	0.37	0.40	0.58	49.26	0.22	79.46	9
30/0.40	62.73	0.36	0.59	0.74	0.60	0.43	0.65	0.63	58.09	100.81	0.48	0.48	0.42	0.63	27.00	0.25	40.09	4
30/0.45	59.41	0.38	0.59	0.73	0.59	0.38	0.59	0.59	68.63	100.51	0.47	0.39	0.38	0.59	45.21	0.15	88.38	8
30/0.50	60.48	0.40	0.59	0.74	0.60	0.41	0.60	0.60	68.74	95.42	0.45	0.41	0.41	0.60	37.36	0.18	67.44	6

Table 9
Mean classification results for each configuration following the custom approach.

Conf	ACC	KAPPA	PRC	ROC	F	COR	PR	RC	RAE	RRSE	RMSE	INF	MAR	MF	LL	MI	PRS	Ranking
10/0.05	44.32	0.14	0.45	0.60	0.43	0.14	0.46	0.44	83.80	117.68	0.56	0.14	0.16	0.44	14.64	0.06	24.40	16
10/0.10	43.05	0.12	0.46	0.58	0.43	0.13	0.48	0.43	85.39	119.59	0.57	0.14	0.12	0.43	15.59	0.06	26.19	21
10/0.15	44.02	0.13	0.45	0.59	0.43	0.13	0.46	0.44	84.55	119.00	0.57	0.14	0.14	0.44	14.43	0.06	23.65	19
10/0.20	44.28	0.15	0.44	0.59	0.43	0.15	0.47	0.44	85.16	118.83	0.57	0.16	0.16	0.44	18.22	0.07	30.27	14
10/0.25	45.89	0.15	0.47	0.60	0.45	0.15	0.48	0.46	81.93	116.21	0.55	0.16	0.15	0.46	14.88	0.06	24.46	12
10/0.30	43.78	0.14	0.46	0.60	0.43	0.14	0.47	0.44	83.49	117.40	0.56	0.15	0.14	0.44	15.84	0.06	26.07	15
10/0.35	44.07	0.13	0.46	0.59	0.43	0.14	0.46	0.44	84.54	117.58	0.56	0.13	0.14	0.44	11.94	0.05	18.98	18
10/0.40	42.43	0.11	0.43	0.58	0.41	0.11	0.44	0.42	86.35	120.03	0.57	0.11	0.11	0.42	12.52	0.04	21.03	26
10/0.45	44.51	0.13	0.46	0.59	0.43	0.13	0.45	0.45	83.31	116.29	0.55	0.13	0.15	0.45	10.94	0.05	17.06	17
10/0.50	41.80	0.10	0.44	0.58	0.40	0.10	0.43	0.42	86.91	118.43	0.57	0.10	0.12	0.42	10.41	0.04	16.55	29
20/0.05	42.80	0.13	0.45	0.59	0.42	0.13	0.46	0.43	85.39	118.85	0.57	0.13	0.15	0.43	16.57	0.06	28.14	20
20/0.10	46.08	0.16	0.46	0.59	0.45	0.16	0.49	0.46	81.75	117.28	0.56	0.17	0.17	0.46	18.36	0.06	31.33	11
20/0.15	41.65	0.10	0.44	0.57	0.41	0.11	0.45	0.42	87.05	121.09	0.58	0.12	0.11	0.42	11.59	0.05	18.65	28
20/0.20	42.90	0.12	0.45	0.58	0.42	0.13	0.47	0.43	86.06	118.77	0.57	0.13	0.13	0.43	15.25	0.06	25.31	22
20/0.25	41.97	0.09	0.44	0.57	0.41	0.10	0.45	0.42	86.71	120.46	0.58	0.10	0.10	0.42	11.93	0.05	19.48	30
20/0.30	42.65	0.12	0.44	0.57	0.41	0.11	0.44	0.43	86.93	119.17	0.57	0.12	0.13	0.43	14.81	0.05	24.20	23
20/0.35	41.61	0.10	0.44	0.57	0.40	0.10	0.44	0.42	87.53	120.54	0.58	0.10	0.11	0.42	14.66	0.06	23.60	27
20/0.40	41.76	0.11	0.44	0.57	0.40	0.11	0.45	0.42	87.62	120.57	0.58	0.12	0.12	0.42	15.25	0.06	23.65	25
20/0.45	44.69	0.13	0.47	0.60	0.44	0.14	0.47	0.45	83.58	116.39	0.56	0.15	0.14	0.45	15.72	0.07	24.58	13
20/0.50	42.97	0.10	0.45	0.58	0.42	0.11	0.44	0.43	86.02	118.99	0.57	0.11	0.11	0.43	11.88	0.05	18.69	24
30/0.05	47.96	0.19	0.49	0.62	0.47	0.19	0.50	0.48	78.77	115.69	0.55	0.20	0.21	0.48	23.61	0.09	40.29	5
30/0.10	**49.21**	**0.19**	**0.49**	**0.62**	**0.49**	**0.20**	**0.52**	**0.49**	**78.07**	**114.34**	**0.54**	**0.21**	**0.20**	**0.49**	**21.45**	**0.08**	**36.08**	**1**
30/0.15	47.88	0.20	0.49	0.63	0.47	0.21	0.52	0.48	78.95	114.58	0.55	0.22	0.21	0.48	21.63	0.09	37.32	2
30/0.20	48.21	0.20	0.48	0.62	0.47	0.21	0.51	0.48	79.81	115.27	0.55	0.22	0.22	0.48	24.05	0.10	40.22	3
30/0.25	46.82	0.18	0.48	0.62	0.46	0.18	0.49	0.47	80.50	115.71	0.55	0.19	0.19	0.47	24.06	0.09	40.49	7
30/0.30	45.69	0.16	0.46	0.61	0.44	0.16	0.48	0.46	82.11	116.52	0.56	0.17	0.17	0.46	19.97	0.08	33.25	10
30/0.35	47.53	0.16	0.48	0.60	0.46	0.17	0.50	0.48	79.27	114.75	0.55	0.18	0.16	0.48	19.43	0.08	32.22	8
30/0.40	48.41	0.18	0.48	0.60	0.47	0.18	0.50	0.48	78.94	114.78	0.55	0.19	0.19	0.48	19.29	0.07	32.74	6
30/0.45	48.31	0.18	0.47	0.62	0.47	0.19	0.49	0.48	79.17	114.26	0.54	0.19	0.19	0.48	25.97	0.10	43.66	4
30/0.50	45.79	0.18	0.48	0.63	0.44	0.19	0.50	0.46	82.13	115.70	0.56	0.20	0.21	0.46	19.59	0.07	32.44	9

tion for the classification (Tables 7 and 9) are now ranked as 2nd and 6th (Tables 10 and 11), respectively.

This leads to a conclusion that if we demand an indicator for developing good predicting models (regression step) by taking the best configuration from the mean values ranking, might be a good approach.

Fig. 2 shows the BP predictions for the best configurations, depicting the BP in mmHg on the y-axis, and the particular instance on the x-axis. Blue lines are the actual BP values and the red lines are the predicted BP values. Black lines present the mean values of the SBP, DBP, and MAP, obtained from the training sets. Those lines are used as a reference

Table 10
Testing set results following the traditional approach.

Conf	SBP				DBP				MAE				Ranking
	MAE	MSE	RMSE	COR	MAE	MSE	RMSE	COR	MAE	MSE	RMSE	COR	
10/0.05	10.88	250.55	15.83	0.68	8.43	151.08	12.29	0.24	7.73	133.84	11.57	0.34	30
10/0.10	10.31	212.43	14.58	0.71	7.82	138.20	11.76	0.34	7.27	120.64	10.98	0.39	29
10/0.15	10.56	223.38	14.95	0.70	7.66	133.35	11.55	0.37	7.21	118.93	10.91	0.41	28
10/0.20	10.37	229.38	15.15	0.70	7.53	120.54	10.98	0.45	7.03	111.73	10.57	0.45	25
10/0.25	10.23	219.76	14.82	0.71	7.46	118.58	10.89	0.46	6.92	106.72	10.33	0.48	22
10/0.30	9.52	192.45	13.87	0.74	7.58	120.04	10.96	0.47	6.95	103.45	10.17	0.50	15
10/0.35	10.32	219.49	14.82	0.71	7.15	114.52	10.70	0.49	6.81	105.07	10.25	0.49	17
10/0.40	10.33	222.90	14.93	0.70	7.61	129.55	11.38	0.41	7.18	117.85	10.86	0.41	27
10/0.45	10.41	225.00	15.00	0.70	7.19	119.17	10.92	0.47	6.87	109.66	10.47	0.46	23
10/0.50	9.56	204.58	14.30	0.72	7.53	129.81	11.39	0.42	6.93	114.05	10.68	0.44	24
20/0.05	10.41	209.18	14.46	0.74	7.85	123.51	11.11	0.44	7.58	115.19	10.73	0.45	26
20/0.10	9.39	181.81	13.48	0.76	7.86	130.69	11.43	0.41	7.30	107.93	10.39	0.48	19
20/0.15	9.51	193.94	13.93	0.75	7.44	124.69	11.17	0.44	6.84	102.35	10.12	0.51	14
20/0.20	9.37	181.94	13.49	0.77	7.69	126.99	11.27	0.42	7.05	104.47	10.22	0.50	16
20/0.25	9.87	191.28	13.83	0.75	7.49	127.50	11.29	0.41	7.07	110.53	10.51	0.46	21
20/0.30	9.44	170.15	13.04	0.78	7.38	117.57	10.84	0.48	6.90	98.22	9.91	0.53	9
20/0.35	9.33	187.44	13.69	0.76	7.05	114.71	10.71	0.49	6.74	99.55	9.98	0.52	10
20/0.40	9.67	178.49	13.36	0.77	7.17	116.80	10.81	0.49	6.91	100.11	10.01	0.53	11
20/0.45	9.56	180.75	13.44	0.76	7.31	121.34	11.02	0.45	6.84	104.33	10.21	0.49	13
20/0.50	9.53	181.52	13.47	0.78	6.96	115.83	10.76	0.49	6.57	99.61	9.98	0.53	8
30/0.05	9.85	187.32	13.69	0.75	7.55	129.22	11.37	0.41	7.02	108.26	10.40	0.48	18
30/0.10	10.31	201.68	14.20	0.74	7.14	124.47	11.16	0.46	6.84	111.77	10.57	0.49	20
30/0.15	8.49	142.39	11.93	0.82	6.93	112.97	10.63	0.52	6.29	91.00	9.54	0.58	6
30/0.20	8.66	152.72	12.36	0.80	6.60	103.80	10.19	0.56	6.09	88.64	9.41	0.58	4
30/0.25	8.43	150.76	12.28	0.80	6.67	107.95	10.39	0.54	5.97	86.15	9.28	0.59	3
30/0.30	8.90	168.70	12.99	0.78	6.49	101.65	10.08	0.57	5.94	88.21	9.39	0.58	5
30/0.35	9.56	184.66	13.59	0.76	7.12	119.68	10.94	0.48	6.72	105.06	10.25	0.50	12
30/0.40	**8.61**	**156.25**	**12.50**	**0.80**	**6.15**	**95.62**	**9.78**	**0.60**	**5.77**	**82.75**	**9.10**	**0.61**	**1**
30/0.45	8.51	163.70	12.79	0.79	6.80	111.27	10.55	0.53	6.08	94.23	9.71	0.56	7
30/0.50	8.54	168.42	12.98	0.78	6.28	99.43	9.97	0.58	5.85	87.19	9.34	0.59	2

Table 11
Testing set results following the custom approach.

Conf	SBP				DBP				MAE				Ranking
	MAE	MSE	RMSE	COR	MAE	MSE	RMSE	COR	MAE	MSE	RMSE	COR	
10/0.05	15.49	401.55	20.04	0.63	10.87	197.58	14.06	-0.22	11.37	208.01	14.42	0.15	15
10/0.10	19.22	564.06	23.75	0.51	9.35	171.16	13.08	0.16	12.02	222.90	14.93	0.03	19
10/0.15	15.73	400.63	20.02	0.63	10.72	184.99	13.60	-0.15	11.79	217.59	14.75	0.08	17
10/0.20	17.01	507.84	22.54	0.56	9.82	167.25	12.93	0.00	11.70	212.95	14.59	0.20	13
10/0.25	15.95	418.39	20.45	0.61	10.15	175.42	13.24	-0.04	11.43	204.27	14.29	0.13	9
10/0.30	15.90	428.25	20.69	0.61	11.26	189.50	13.77	-0.13	12.30	223.57	14.95	0.08	20
10/0.35	14.96	383.98	19.60	0.64	11.19	195.60	13.99	-0.15	11.81	214.62	14.65	0.14	16
10/0.40	15.23	411.75	20.29	0.64	11.83	220.68	14.86	-0.31	12.43	238.22	15.43	0.10	27
10/0.45	16.47	448.62	21.18	0.57	11.17	197.05	14.04	-0.10	12.60	236.76	15.39	0.03	26
10/0.50	15.53	412.91	20.32	0.63	11.49	212.08	14.56	-0.21	12.28	231.66	15.22	0.07	23
20/0.05	18.57	526.36	22.94	0.53	9.53	166.24	12.89	0.04	11.15	199.71	14.13	0.19	11
20/0.10	20.44	639.12	25.28	0.46	8.79	148.61	12.19	0.17	11.30	199.46	14.12	0.23	12
20/0.15	17.88	487.77	22.09	0.59	11.57	216.64	14.72	-0.17	12.31	231.04	15.20	0.08	28
20/0.20	18.11	514.65	22.69	0.54	9.39	157.87	12.56	0.14	10.98	198.92	14.10	0.17	5
20/0.25	18.71	546.87	23.39	0.52	8.96	159.18	12.62	0.10	10.97	200.21	14.15	0.15	8
20/0.30	16.17	412.49	20.31	0.62	12.04	223.79	14.96	-0.27	12.02	222.56	14.92	0.08	25
20/0.35	17.00	436.39	20.89	0.64	9.94	178.64	13.37	0.02	10.87	196.44	14.02	0.20	4
20/0.40	15.64	361.26	19.01	0.70	10.57	190.56	13.80	-0.14	10.19	187.34	13.69	0.18	3
20/0.45	16.29	413.83	20.34	0.67	10.25	173.99	13.19	-0.06	10.42	186.67	13.66	0.22	2
20/0.50	19.31	531.20	23.05	0.50	9.50	158.78	12.60	0.10	11.40	201.99	14.21	0.01	14
30/0.05	21.49	649.85	25.49	0.40	9.89	175.00	13.23	0.16	12.54	233.38	15.28	0.04	29
30/0.10	22.02	685.54	26.18	0.31	7.86	124.11	11.14	0.32	11.30	187.46	13.69	0.19	6
30/0.15	19.43	556.33	23.59	0.46	10.37	176.74	13.29	-0.03	11.58	215.40	14.68	0.02	22
30/0.20	21.60	650.16	25.50	0.49	9.21	167.23	12.93	0.05	12.31	224.92	15.00	0.13	24
30/0.25	19.42	525.79	22.93	0.51	11.21	197.13	14.04	0.01	12.09	233.00	15.26	-0.10	30
30/0.30	15.68	415.29	20.38	0.65	10.86	183.53	13.55	-0.11	11.34	203.42	14.26	0.22	10
30/0.35	21.30	663.79	25.76	0.44	8.93	148.61	12.19	0.25	12.19	218.75	14.79	0.12	18
30/0.40	16.72	438.41	20.94	0.60	10.09	170.14	13.04	0.10	11.22	202.91	14.24	0.09	7
30/0.45	17.20	450.34	21.22	0.57	11.57	207.80	14.42	-0.09	11.73	209.57	14.48	0.03	21
30/0.50	**17.39**	**462.46**	**21.50**	**0.59**	**9.23**	**145.64**	**12.07**	**0.19**	**10.54**	**181.21**	**13.46**	**0.19**	**1**

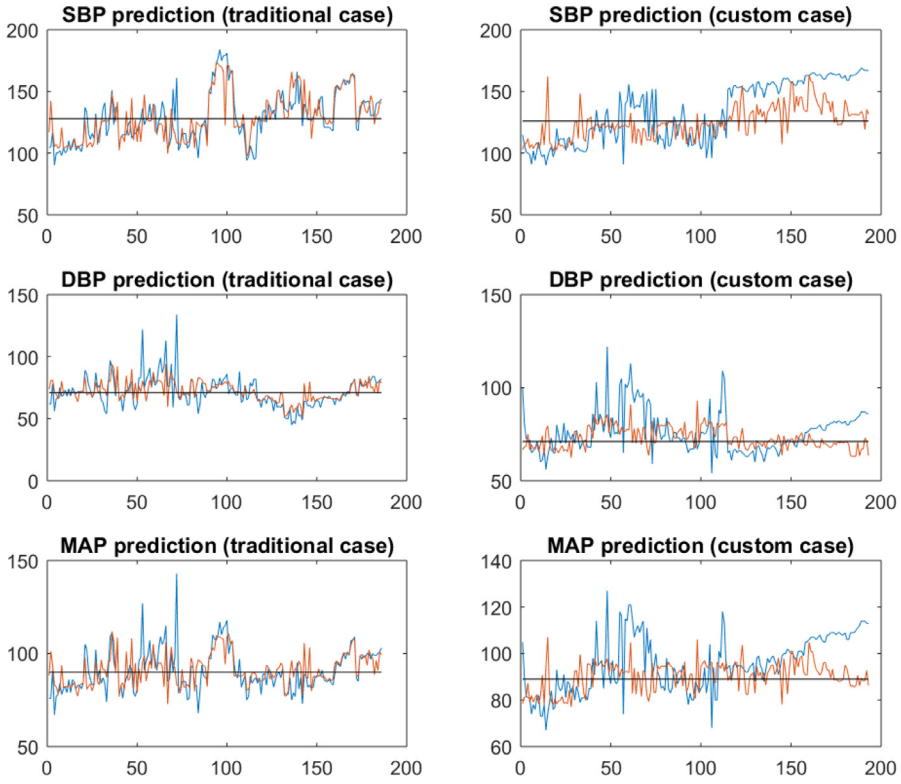

Fig. 2. Best predictions in traditional (left) and custom (right) approach.

for the success of our developed predictors, showing that they perform better than simple predictors that would always predict the mean values. From the figure it can be perceived that the prediction line clearly follows the actual line in the traditional approach, whereas in the custom approach the prediction line cannot catch all the trends in the BP changes. We believe that this phenomena is due to the fact that in the custom case, no instances from a single patient are involved in the training, validation and the testing phase. Thus, there are some training sets that miss hypertensive patients to learn from, and therefore, the testing results are not reliable.

5.5. *Level 5: Information fusion using the information from multi-target regression models from all configurations*

At the end, we performed an information fusion to merge the information obtained from the multi-target regression models from all configurations. Because the training set, which is used for training multi-target regression model for each configuration, was selected using the classifiers ranking, this ranking was used as the influence of each configuration multi-target regression model to the end prediction of the SBP, DBP, and MAE. Therefore, as soon as the regression models are created from the best training sets in the classification analysis, the produced values are multiplied by their final ranking. Those rankings are scaled in a way that all the ranks sum is one. As they are multiplied by the corresponding prediction, all the predictions are summed into a single prediction for each SBP, DBP, and MAP.

The procedure is performed considering both, the best classification rankings from each configuration (Tables 6 and 8) and the mean results rankings of each configuration (Tables 7 and 9), for both splitting approaches. Table 12 presents the evaluation of the end results, where “Best” are the best results for the traditional (T) and custom (C) approach, according to the ranking in Tables 10 and 11; “Fused 1” is the case where we used the rankings from the best classifiers, and “Fused 2” is the case where we used the rankings from the mean performance results. All results in this table are presented in mmHg. Considering the table, in both cases we can conclude that the prediction fusion led to better prediction. The SBP prediction is significantly improved and there are minor improvements in the MAP prediction. For the DBP prediction, the results are almost the same. These results, encouraged us to follow the proposed multi-level information fusion methodology. The predictions for the “Fused 1” and “Fused 2” cases are shown in Figs. 3 and 4, correspondingly.

Comparing the left and the right part of Figs. 2–4, we can see greater dissimilarity between blue and orange lines in the right part. This happens since the lines on the left side are for the traditional split, where all measurements were assumed as independent, so part of the measurements from the same subject are in the training set and the rest in the test set. However, following the studies that involve humans subjects, we also performed a custom split, where measurements from one subject occurs only in one set (training or test set). This is the reason why greater dissimilarities are detected on the right parts of these figures.

6. Discussion

Fig. 5 depicts the influence of the prediction errors on the BP class. In both approaches (traditional and custom), we inspected whether the error caused by the multi-target regression leads to a change of the actual BP class, as medically determined according to the Table 4, or to a change of the joined class (as presented in Table 4).

From the Fig. 5, it can be concluded that configuration number 26, which refers to 30 s long signals and cut-off frequency of 0.30 Hz, can be considered as best trade-off between the first and the second approach (traditional and custom) with least changes given in percentage.

Fig. 6 provides an insight into the severity of the misses - meaning we consider the BP class miss to be a severe one if the difference between the actual and the predicted class is more than one class. Adding the severity analysis to the class changes, we can see that most promising configurations are the 3rd and the 26th. Comparing to the total number of misses in the Fig. 5, where we selected 26th configuration to be the best, according to the severity analysis, we can expect that 4% included

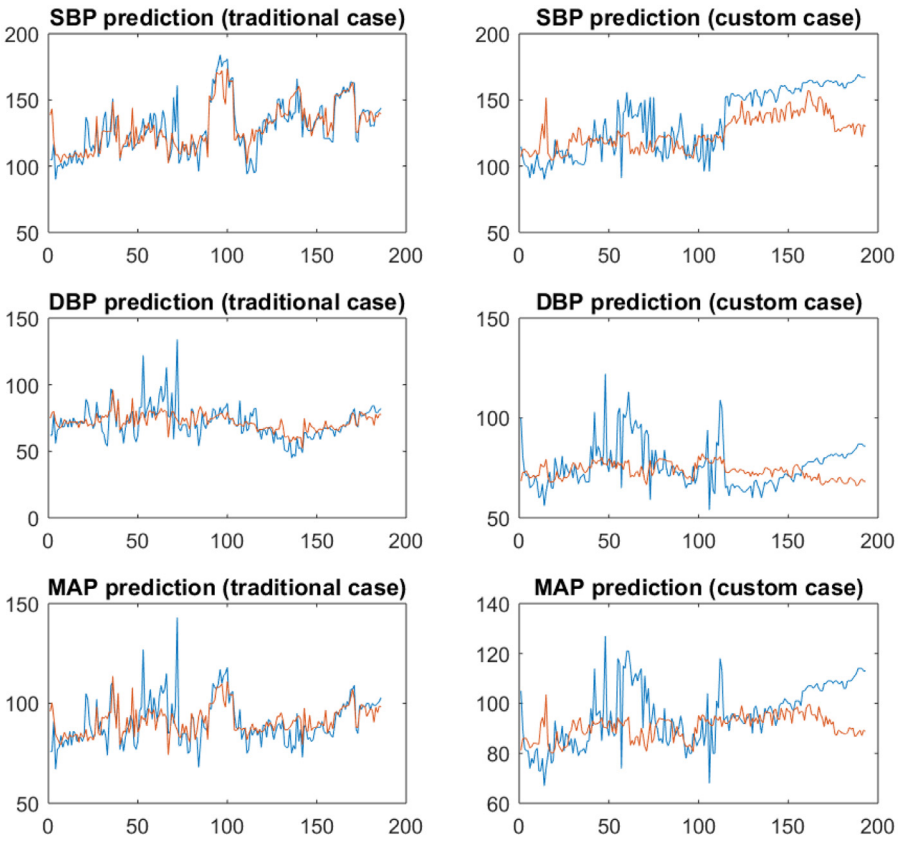

Fig. 3. Fused predictions in traditional (left) and custom (right) approach by using best classifiers rankings.

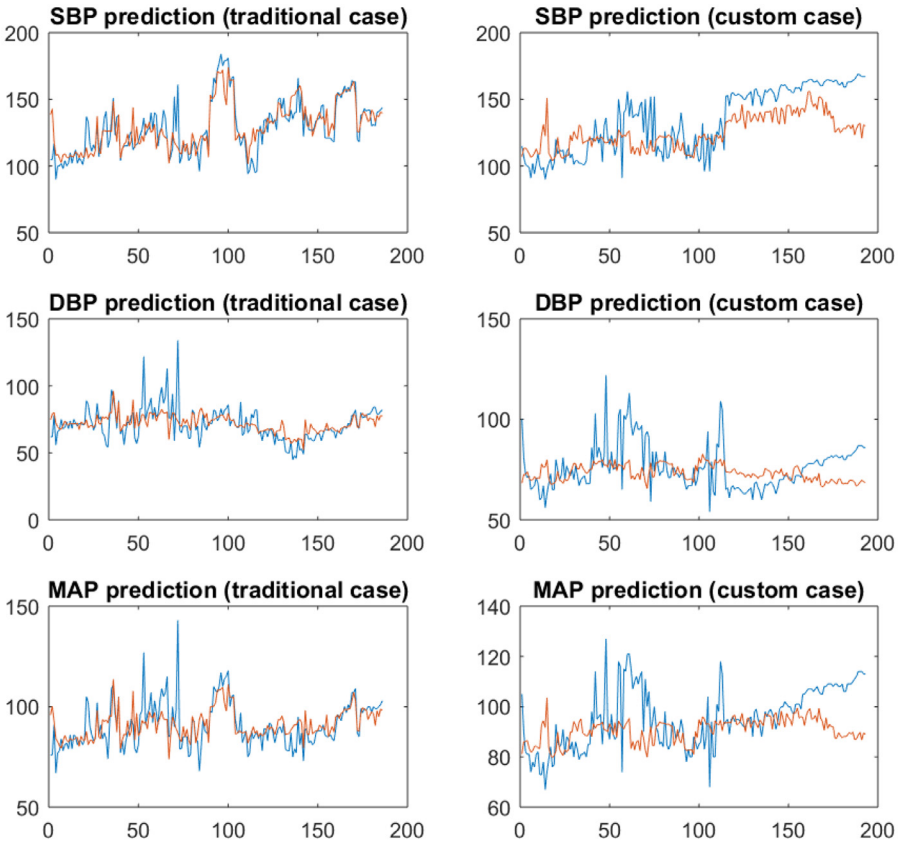

Fig. 4. Fused predictions in traditional (left) and custom (right) approach by using mean classifications rankings.

Table 12
Regression results from the fusion-prediction procedure.

T	Best		Fused 1		Fused 2		C	Best		Fused 1		Fused 2	
	MAE	RMSE	MAE	RMSE	MAE	RMSE		MAE	RMSE	MAE	RMSE	MAE	RMSE
SBP	8.61	12.50	7.94	11.40	**7.93**	**11.36**	SBP	17.39	21.50	**16.40**	**19.69**	16.60	19.92
DBP	**6.15**	**9.78**	6.43	9.88	6.41	9.85	DBP	**9.23**	**12.07**	9.26	12.15	9.24	12.11
MAP	5.77	9.10	5.73	8.82	**5.72**	**8.79**	MAP	10.54	13.46	**9.76**	**12.92**	9.80	12.98

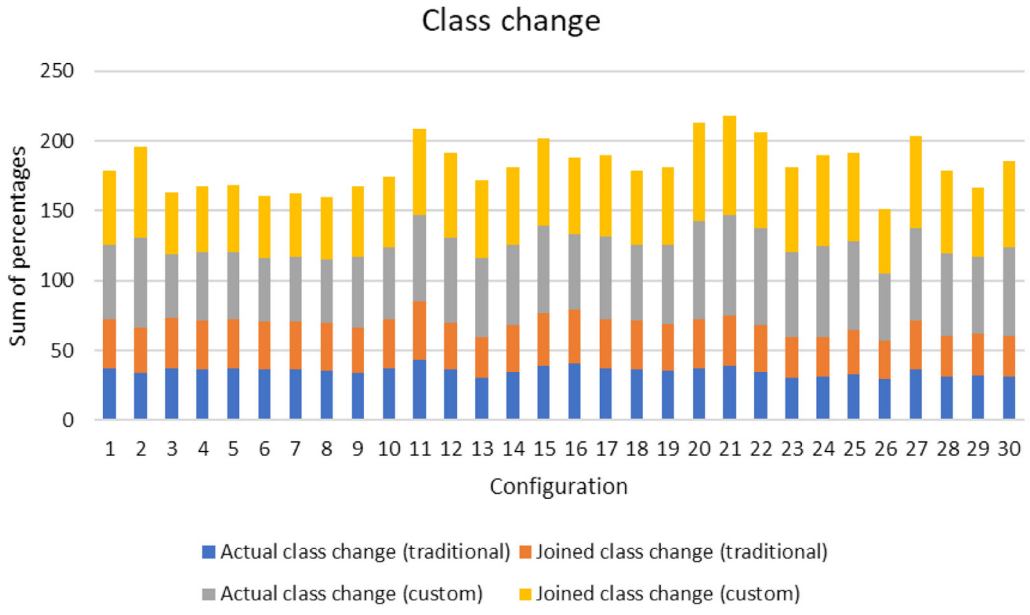

Fig. 5. Percentage of misses in classes caused by prediction error.

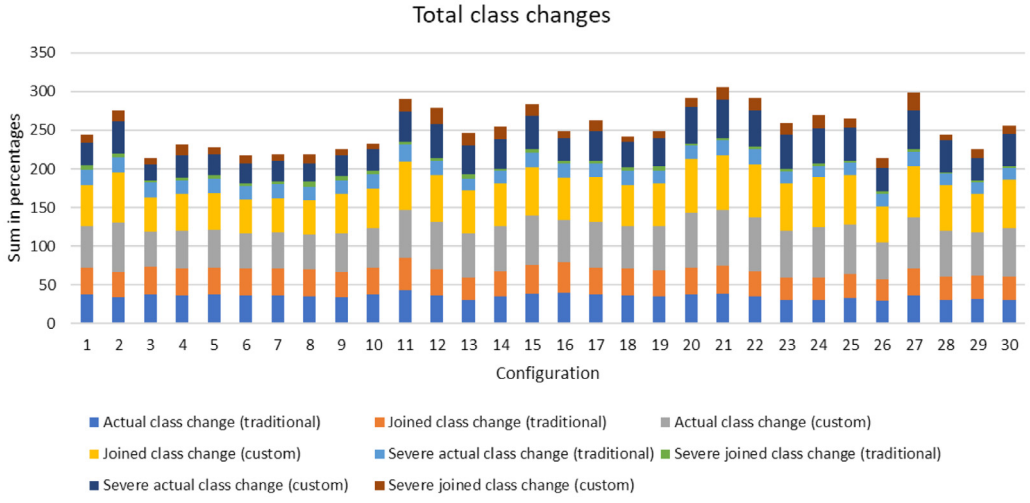

Fig. 6. Percentage of severe misses in classes caused by prediction error.

in all misses (27%) in the traditional approach, and 13% included in all misses (47%) in the custom approach, to be severe.

Going back to the tables showing the regressions (Tables 10 and 11) it can be observed that the 26th configuration has been ranked as 5th and 10th. Since we cannot make a conclusion whether this configuration might be the best, we decided to include the percentage of class misses and the percentage of severe class misses in the pool with the other performance metrics. Therefore, we repeated the ranking analysis, however, no significant changes were detected for the best and the worst configuration performances. That means that there was no significant difference between the new metrics that would beat the influence of the other metrics.

Furthermore, we inspected the severe misses distributions by classes, in both real and joined class cases. As it can be perceived in Fig. 7, class 4 (S2HTN) is most problematic in both traditional and custom approach. This phenomena is due to the fact that most of the hypertension representatives are patients with brain injuries that suffer isolated systolic hypertension characterized with high SBP and low DBP (given in Table 3). This is also reflected in the joined classes case since class 4 is joined with the rest of the hypertension classes into a single class labeled as class 2.

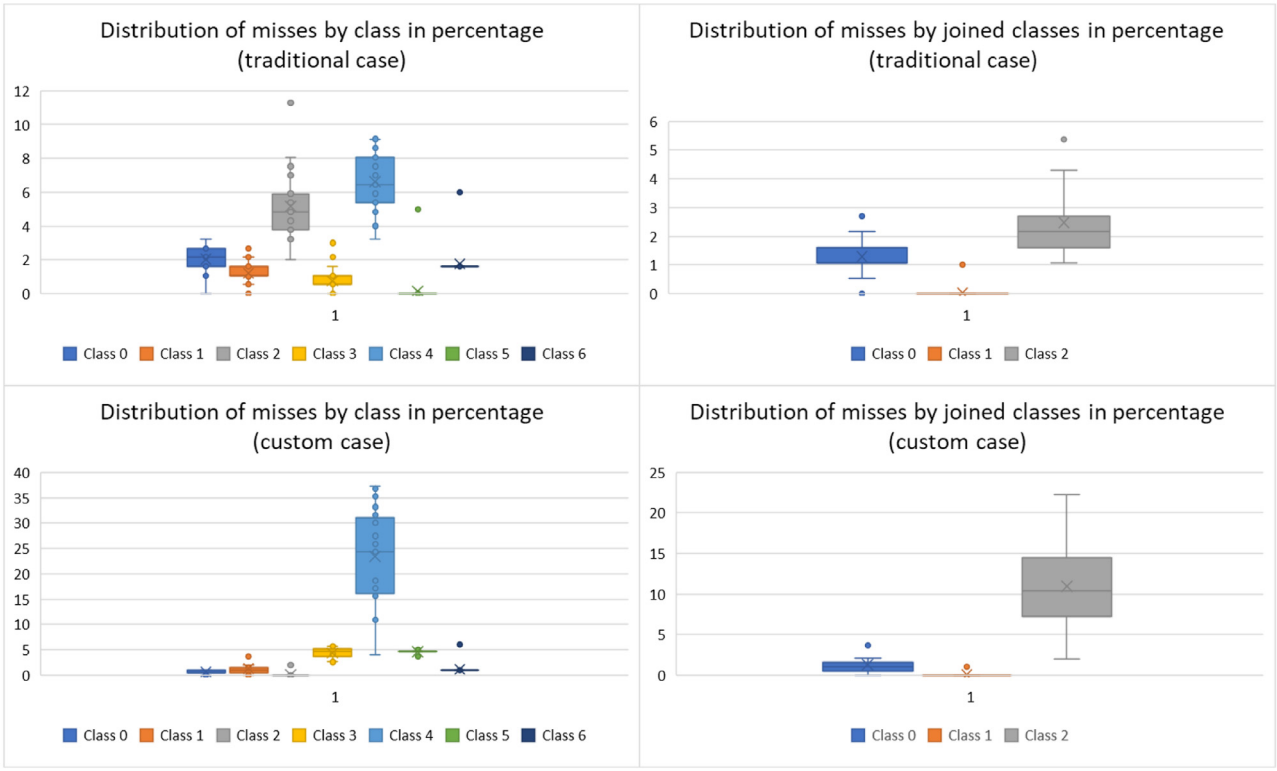

Fig. 7. Percentage of severe misses by class.

Table 13
State-of-the-art literature results published in 2018.

Ref.	Source	Features	Methods	SBP (mmHg)	DBP (mmHg)
[101]	ECG + PPG	Morphological	Deep RNN	[3.84–5.81] RMSE	[1.80–5.21] RMSE
[102]	FBG sensor	Morphological	PLSR, ANN	Unknown: 12 ± 17 MAE ± SD	
[103]	Utrasound images	Levenberg-Marquardt	Regression, Bland-Altman	10.21 MAE	8.23 MAE
[104]	PPG	Morphological, time-frequency	DT,SVR, ABR,RFR	[4.17–7.51] SD	[8.90–18.54] SD
[105]	ECG + PPG	Waveform, artificial and personal	DNN	3.63 MAE	2.45 MAE
[106]	rPPG	PTT estimation	Gaussian model	8.42 ± 8.81 MAE ± SD	12.34 ± 7.10 MAE ± SD

The same results are also confirmed to be valid for the prediction-fusion approach.

In our previous work [32], the best results obtained for SBP, DBP, and MAP are 8.64, 18.20, and 13.52 MAE in mmHg without calibration, and 7.72, 9.45, and 8.13 MAE in mmHg with calibration.

In the multi-level information fusion methodology design as presented in this paper, we we have achieved 7.93 ± 8.16, 6.41 ± 7.5, and 5.72 ± 6.69 (*MAE* ± *SD* in mmHg), for SBP, DBP, and MAP when we have patient's data available for training (the traditional approach). In the case of completely unknown patient's data (the custom approach), the (*MAE* ± *SD* in mmHg) achieved for SBP, DBP, and MAP is 16.60 ± 11.05, 9.24 ± 7.85, and 9.80 ± 8.53. The presented results are **with no calibration method applied**.

Table 13, provides insight into the state-of-the-art literature published in 2018. According to the information provided, all of the methodologies rely either on a combination of multiple physiological signals, or completely new techniques involving ultrasound images and RGB cameras. None of the reported relies on ECG signal only.

To provide results that are more acceptable for medical purposes, the goal is to achieve an error as close as possible to that of a certified medical device for BP estimation (5 mmHg, and SD within 8 mmHg according to BHS and AAMI standards [107]. Our results are appropriately compared with latest literature [108] and confirm the relation between the ECG signal and the blood pressure. However, there is still an effort that should be done to make these models more close to be acceptable for medical purposes.

7. Conclusion

In this paper, we propose a multi-level information fusion methodology for learning a blood pressure predictive model using ECG sensor data. The methodology estimates systolic blood pressure, diastolic blood pressure, and the mean arterial pressure from ECG sensor data. We evaluate models (i.e. configurations) when the ECG signals were processed using different signal lengths, 10 s, 20 s, and 30 s, filtered with different cut-off frequencies starting from 0.05 Hz, up to 0.50 Hz, with the step of 0.05 Hz. Using them, we identify that the models trained for a signal length of 30 s with a variable cut-off frequency of 0.40 Hz and 0.50 Hz to be most informative. The novelty of the methodology is in fusing the models built for different configurations, which improves the results obtained by the best configuration only. In this study, a design relationship between BP and ECG is presented in the form of fused ML models. Since the trend of ECG sensor usage indicates a continuous increase in demand, we believe that our proposed solution has promising real-world applications in civilian and military environments.

For our future work, we will focus on a comparison analysis to show the impact of different physical activities to the blood pressure measure-

ment, since intuitively, results with sitting subjects would be higher due to the lack of motion artifacts in the collected signals.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the Slovenian Research Agency, Programme P2-0098 and the projects Z2-1867 and J1-8155. Also it received funding from the project "ECG2BP: ECG Signal Processing for Blood Pressure Estimation" within Faculty of Computer Science and Engineering at the Ss. Cyril and Methodius University in Skopje.

References

[1] A. Alwan, et al., Global Status Report on Noncommunicable Diseases 2010., World Health Organization, 2011.
[2] C. Rosendorff, D.T. Lackland, M. Allison, W.S. Aronow, H.R. Black, R.S. Blumenthal, C.P. Cannon, J.A. De Lemos, W.J. Elliott, L. Findeiss, et al., Treatment of hypertension in patients with coronary artery disease, Hypertension 65 (6) (2015) 1372–1407.
[3] G.F. Mitchell, Arterial stiffness and hypertension, Hypertension 64 (1) (2014) 13–18.
[4] W.H. Organization, I.S. of Hypertension Writing Group, et al., 2003 World Health Organization (WHO)/International Society of Hypertension (ISH) statement on management of hypertension, J. Hypertens. 21 (11) (2003) 1983–1992.
[5] G. Cosoli, L. Casacanditella, F. Pietroni, A. Calvaresi, G.M. Revel, L. Scalise, A novel approach for features extraction in physiological signals, in: Medical Measurements and Applications (MeMeA), 2015 IEEE International Symposium on, IEEE, 2015, pp. 380–385.
[6] J. Canning, K. Helbert, G. Iashin, J. Matthews, J. Yang, M.K. Delano, C.G. Sodini, Q. Zhang, Noninvasive and continuous blood pressure measurement via superficial temporal artery tonometry, in: Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the, IEEE, 2016, pp. 3382–3385.
[7] V. Mouradian, A. Poghosyan, L. Hovhannisyan, Noninvasive continuous mobile blood pressure monitoring using novel PPG optical sensor, in: Biomedical Wireless Technologies, Networks, and Sensing Systems (BioWireleSS), 2015 IEEE Topical Conference on, IEEE, 2015, pp. 1–3.
[8] A.K. Sahani, V. Ravi, M. Sivaprakasam, Automatic estimation of carotid arterial pressure in artsens, in: India Conference (INDICON), 2014 Annual IEEE, IEEE, 2014, pp. 1–6.
[9] R. Marani, A.G. Perri, An intelligent system for continuous blood pressure monitoring on remote multi-patients in real time, arXiv:1212.06514 (2012).
[10] S. Tanaka, M. Nogawa, T. Yamakoshi, K. Yamakoshi, Accuracy assessment of a noninvasive device for monitoring beat-by-beat blood pressure in the radial artery using the volume-compensation method, IEEE Trans. Biomed. Eng. 54 (10) (2007) 1892–1895.
[11] Y. Li, Y. Gao, N. Deng, Mechanism of cuff-less blood pressure measurement using mmsb, Engineering 5 (10) (2013) 123.
[12] B. Ilie, Portable equipment for monitoring human functional parameters, in: Roedunet International Conference (RoEduNet), 2010 9th, IEEE, 2010, pp. 299–302.
[13] A. Sahoo, P. Manimegalai, K. Thanushkodi, Wavelet based pulse rate and blood pressure estimation system from ECG and PPG signals, in: Computer, Communication and Electrical Technology (ICCCET), 2011 International Conference on, IEEE, 2011, pp. 285–289.
[14] S. Ilango, P. Sridhar, A non-invasive blood pressure measurement using android smart phones, IOSR J. Dent. Med. Sci. 13 (1) (2014) 28–31.
[15] S.S. Thomas, V. Nathan, C. Zong, K. Soundarapandian, X. Shi, R. Jafari, Biowatch: a noninvasive wrist-based blood pressure monitor that incorporates training techniques for posture and subject variability, IEEE J. Biomed. Health Inform. 20 (5) (2016) 1291–1300.
[16] R. Nye, Z. Zhang, Q. Fang, Continuous non-invasive blood pressure monitoring using photoplethysmography: a review, in: Bioelectronics and Bioinformatics (ISBB), 2015 International Symposium on, IEEE, 2015, pp. 176–179.
[17] S. Goli, T. Jayanthi, Cuff less continuous non-invasive blood pressure measurement using pulse transit time measurement, Int. J. Recent Dev.Eng. Technol. 2 (1) (2014) 87.
[18] Y. Choi, Q. Zhang, S. Ko, Noninvasive cuffless blood pressure estimation using pulse transit time and Hilbert–Huang transform, Comput. Electr. Eng. 39 (1) (2013) 103–111.
[19] J. Seo, S.J. Pietrangelo, H.-S. Lee, C.G. Sodini, Noninvasive arterial blood pressure waveform monitoring using two-element ultrasound system, IEEE Trans. Ultrason. Ferroelectr. Freq. Control 62 (4) (2015) 776–784.
[20] J.M.S. i Carós, Continuous Non-Invasive Blood Pressure Estimation, ETH, 2011 Ph.D. thesis.
[21] M.Y.-M. Wong, C.C.-Y. Poon, Y.-T. Zhang, An evaluation of the cuffless blood pressure estimation based on pulse transit time technique: a half year study on normotensive subjects, Cardiovasc. Eng. 9 (1) (2009) 32–38.
[22] R. Payne, C. Symeonides, D. Webb, S. Maxwell, Pulse transit time measured from the ECG: an unreliable marker of beat-to-beat blood pressure, J. Appl. Physiol. 100 (1) (2006) 136–141.
[23] R. Trobec, I. Tomašić, A. Rashkovska, M. Depolli, V. Avbelj, ECG pilot studies, in: Body Sensors and Electrocardiography, Springer, 2018, pp. 61–75.
[24] The ATLS SubcommitteeAmerican College of Surgeons Committee on Trauma, and the International ATLS working group, Advanced trauma life support (ATLS®): the ninth edition, J. Trauma Acute Care Surg. 74 (5) (2013) 1363–1366.
[25] A. Strahovnik, N. Koceska, R. Komadina, J. Franc Tasic, The use of new technology to lessen the death-rate of soldiers injured in combat (2016).
[26] E.B. Schroeder, D. Liao, L.E. Chambless, R.J. Prineas, G.W. Evans, G. Heiss, Hypertension, blood pressure, and heart rate variability, Hypertension 42 (6) (2003) 1106–1111.
[27] M.K.B.A. Hassan, M. Mashor, N.M. Nasir, S. Mohamed, Measuring of systolic blood pressure based on heart rate, in: 4th Kuala Lumpur International Conference on Biomedical Engineering 2008, Springer, 2008, pp. 595–598.
[28] C. Gómez, R. Hornero, D. Abásolo, A. Fernández, M. López, Complexity analysis of the magnetoencephalogram background activity in Alzheimer's disease patients, Med. Eng. Phys. 28 (9) (2006) 851–859.
[29] M. Costa, A.L. Goldberger, C.-K. Peng, Multiscale entropy analysis of biological signals, Phys. Rev. E 71 (2) (2005) 21906.
[30] M.R. Raoufy, T. Ghafari, A.R. Mani, Complexity analysis of respiratory dynamics, Am. J. Respir. Crit. Care Med. (ja) (2017).
[31] Y. Luo, R.H. Hargraves, A. Belle, O. Bai, X. Qi, K.R. Ward, M.P. Pfaffenberger, K. Najarian, A hierarchical method for removal of baseline drift from biomedical signals: application in ECG analysis, Sci. World J. 2013 (2013).
[32] M. Simjanoska, M. Gjoreski, M. Gams, A. Madevska Bogdanova, Non-invasive blood pressure estimation from ECG using machine learning techniques, Sensors 18 (4) (2018) 1160.
[33] T. Eftimov, P. Korošec, B.K. Seljak, Data-driven preference-based deep statistical ranking for comparing multi-objective optimization algorithms, in: International Conference on Bioinspired Methods and Their Applications, Springer, 2018, pp. 138–150.
[34] R.M. Rangayyan, Biomedical Signal Analysis, 33, John Wiley & Sons, 2015.
[35] X.L. Aubert, A. Brauers, Estimation of vital signs in bed from a single unobtrusive mechanical sensor: algorithms and real-life evaluation, in: Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, IEEE, 2008, pp. 4744–4747.
[36] N. Takahashi, A. Kuriyama, H. Kanazawa, Y. Takahashi, T. Nakayama, Validity of spectral analysis based on heart rate variability from 1-min or less ECG recordings, Pacing Clin. Electrophysiol. (2017).
[37] A.Y. Shdefat, M.-I. Joo, S.-H. Choi, H.-C. Kim, Utilizing ECG waveform features as new biometric authentication method, Int. J. Electr. Comput.Eng. 8 (2) (2018).
[38] M. Blanco-Velasco, B. Weng, K.E. Barner, Ecg signal denoising and baseline wander correction based on the empirical mode decomposition, Comput. Biol. Med. 38 (1) (2008) 1–13.
[39] Y. Xu, M. Luo, T. Li, G. Song, Ecg signal de-noising and baseline wander correction based on Ceemdan and wavelet threshold, Sensors 17 (12) (2017) 2754.
[40] B.M.M.A. Tinati, Ecg baseline wander elimination using wavelet packets, World Acad. Sci. Eng. Technol. 3 (2005) (2005) 14–16.
[41] R.M. Rangayyan, Biomedical Signal Analysis, 33, John Wiley & Sons, 2015.
[42] M. Simjanoska, B. Koteska, A.M. Bogdanova, N. Ackovska, V. Trajkovik, M. Kostoska, Automated triage parameters estimation from ecg, Technol. Health Care (Preprint) (2018) 1–4.
[43] A. Madevska Bogdanova, M. Simjanoska, N. Ackovska, M. Kostoska, B. Koteska, M. Tashkoski, Biosensors technology in massive civil disasters, in: 23th International Symposium on Emergency Medicine, Slovenian Society for Emergency Medicine, 2016, pp. 355–359.
[44] M. Kostoska, M. Simjanoska, B. Koteska, A.M. Bogdanova, Real-time smart advisory health system, in: Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics, ACM, 2018, p. 47.
[45] G. Fortino, V. Giampà, PPG-based methods for non invasive and continuous blood pressure measurement: an overview and development issues in body sensor networks, in: 2010 IEEE International Workshop on Medical Measurements and Applications, IEEE, 2010, pp. 10–13.
[46] R. Gravina, P. Alinia, H. Ghasemzadeh, G. Fortino, Multi-sensor fusion in body sensor networks: state-of-the-art and research challenges, Inf. Fusion 35 (2017) 68–80.
[47] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, I. Korhonen, Activity classification using realistic data from wearable sensors, IEEE Trans. Inf. Technol. Biomed. 10 (1) (2006) 119–128.
[48] A. Haag, S. Goronzy, P. Schaich, J. Williams, Emotion recognition using bio-sensors: first steps towards an automatic system, in: Tutorial and Research Workshop on Affective Dialogue Systems, Springer, 2004, pp. 36–48.
[49] R. Subramanian, J. Wache, M.K. Abadi, R.L. Vieriu, S. Winkler, N. Sebe, Ascertain: emotion and personality recognition using commercial sensors, IEEE Trans. Affect. Comput. (2) (2018) 147–160.
[50] D. Seo, B. Yoo, H. Ko, Information fusion of heterogeneous sensors for enriched personal healthcare activity logging, Int. J. Ad Hoc Ubiquitous Comput. 27 (4) (2018) 256–269.
[51] J. Davis, M. Goadrich, The relationship between precision-recall and ROC curves, in: Proceedings of the 23rd International Conference on Machine Learning, ACM, 2006, pp. 233–240.

[52] A. Benavoli, G. Corani, J. Demšar, M. Zaffalon, Time for a change: a tutorial for comparing multiple classifiers through Bayesian analysis, J. Mach. Learn. Res. 18 (1) (2017) 2653–2688.
[53] J. Demšar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (2006) 1–30.
[54] Z. Andreopoulou, C. Koliouska, E. Galariotis, C. Zopounidis, Renewable energy sources: using promethee ii for ranking websites to support market opportunities, Technol. Forecast. Soc. Change 131 (2018) 31–37.
[55] A. Ishizaka, P. Nemery, Selecting the best statistical distribution with promethee and gaia, Comput. Ind. Eng. 61 (4) (2011) 958–969.
[56] M.A. Nikouei, M. Oroujzadeh, S. Mehdipour-Ataei, The promethee multiple criteria decision making analysis for selecting the best membrane prepared from sulfonated poly (ether ketone) s and poly (ether sulfone) s for proton exchange membrane fuel cell, Energy 119 (2017) 77–85.
[57] P. Brazdil, C.G. Carrier, C. Soares, R. Vilalta, Metalearning: Applications to Data Mining, Springer Science & Business Media, 2008.
[58] X. Zhen, M. Yu, X. He, S. Li, Multi-target regression via robust low-rank learning, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2) (2018) 497–504.
[59] M. Breskvar, D. Kocev, S. Džeroski, Ensembles for multi-target regression with random output selections, Mach. Learn. 107 (11) (2018) 1673–1709.
[60] Savvy ECG, http://www.savvy.si/Savvy/.
[61] Bittium Biosignals, Emotion faros, 2016,. http://www.megaemg.com/products/faros/.
[62] C. Hacks., e-health sensor platform v2.0 for arduino and raspberry pi., https://www.cooking-hacks.com/documentation/tutorials/ehealth-biometric-sensor-platform-arduino-raspberry-pi-medical.
[63] Z. Technology., Zephyr bioharness 3.0 user manual. 2017, https://www.zephyranywhere.com/media/download/bioharness3-user-manual.pdf.
[64] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, Physiobank, physiotoolkit, and physionet, Circulation 101 (23) (2000) e215–e220.
[65] P. Winderbank-Scott, P. Barnaghi, A non-invasive wireless monitoring device for children and infants in pre-hospital and acute hospital environments (2017).
[66] M.A. Bereksi-Reguid, F. Bereksi-Reguig, A. Nait-Ali, A new system for measurement of the pulse transit time, the pulse wave velocity and its analysis, J. Mech. Med. Biol. 17 (1) (2017) 1750010.
[67] J. Morales, C. Díaz-Piedra, L.L. Di Stasi, S. Romero, et al., Low-cost remote monitoring of biomedical signals, in: International Work-Conference on the Interplay Between Natural and Artificial Computation, Springer, 2015, pp. 288–295.
[68] L. Ahonen, B. Cowley, J. Torniainen, A. Ukkonen, A. Vihavainen, K. Puolamäki, Cognitive collaboration found in cardiac physiology: study in classroom environment, PLoS One 11 (7) (2016) e0159178.
[69] T. Miettinen, K. Myllymaa, S. Westeren-Punnonen, J. Ahlberg, T. Hukkanen, J. Töyräs, R. Lappalainen, E. Mervaala, K. Sipilä, S. Myllymaa, Success rate and technical quality of home polysomnography with self-applicable electrode set in subjects with possible sleep bruxism, IEEE J. Biomed. Health Inform. (2017).
[70] D.P. Cliff, J. McNeill, S. Vella, S.J. Howard, M.A. Kelly, D.J. Angus, I.M. Wright, R. Santos, M. Batterham, E. Melhuish, et al., The preschool activity, technology, health, adiposity, behaviour and cognition (path-abc) cohort study: rationale and design, BMC Pediatr. 17 (1) (2017) 95.
[71] J.A. Johnstone, P.A. Ford, G. Hughes, T. Watson, A.T. Garrett, Bioharness multivariable monitoring device: part. I: validity, J. Sports Sci. Med. 11 (3) (2012) 400.
[72] H. Ding, A. Sarela, R. Helmer, M. Mestrovic, M. Karunanithi, Evaluation of ambulatory ECG sensors for a clinical trial on outpatient cardiac rehabilitation, in: Complex Medical Engineering (CME), 2010 IEEE/ICME International Conference on, IEEE, 2010, pp. 240–243.
[73] J.A. Johnstone, P.A. Ford, G. Hughes, T. Watson, A.C. Mitchell, A.T. Garrett, Field based reliability and validity of the bioharness multivariable monitoring device, J. Sports Sci. Med. 11 (4) (2012) 643.
[74] J. Hailstone, A.E. Kilding, Reliability and validity of the zephyr bioharness to measure respiratory responses to exercise, Meas. Phys. Educ. Exerc. Sci. 15 (4) (2011) 293–300.
[75] R. Trobec, I. Tomašić, A. Rashkovska, M. Depolli, V. Avbelj, Commercial ECG systems, in: Body Sensors and Electrocardiography, Springer, 2018, pp. 101–114.
[76] A. Rashkovska, V. Avbelj, Three-year experience with a wireless ECG sensor, 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), IEEE, 2018.
[77] N. Kim, A. Krasner, C. Kosinski, M. Wininger, M. Qadri, Z. Kappus, S. Danish, W. Craelius, Trending autoregulatory indices during treatment for traumatic brain injury, J. Clin. Monit. Comput. 30 (6) (2016) 821–831.
[78] M. Wong, C. Poon, Y. Zhang, Can the timing-characteristics of phonocardiographic signal be used for cuffless systolic blood pressure estimation? in: Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE, IEEE, 2006, pp. 2878–2879.
[79] X.-Y. Zhang, Y.-T. Zhang, A model-based study of relationship between timing of second heart sound and systolic blood pressure, in: Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE, IEEE, 2006, pp. 1387–1390.
[80] M. Nitzan, Automatic noninvasive measurement of arterial blood pressure, IEEE Instrum. Meas. Mag. 14 (1) (2011).
[81] D. Kugiumtzis, A. Tsimpiris, Measures of analysis of time series (mats): a matlab toolkit for computation of multiple measures on time series data bases, arXiv:1002.1940 (2010).
[82] J. Monge-lvarez, Higuchi and Katz fractal dimension measures, 2015, https://www.mathworks.com/matlabcentral/fileexchange/50290-higuchi-and-katz-fractal-dimension-measures/content/FractaldimensionmeasuresHiguchiFD.m.
[83] T.L. Doyle, E.L. Dugan, B. Humphries, R.U. Newton, Discriminating between elderly and young using a fractal dimension analysis of centre of pressure, Int. J. Med. Sci. 1 (1) (2004) 11.
[84] P. Zhou, M. Zong-Xia, H. Chun-Lan, Y.-X. Huang, Power spectral entropy in the ECG of patients suffered from nocturnal frontal lobe epilepsy, J. Pharm. Biomed. Sci. 7 (3) (2017).
[85] L. Breiman, Bagging predictors, Mach. Learn. 24 (2) (1996) 123–140.
[86] Y. Freund, R.E. Schapire, et al., Experiments with a new boosting algorithm, in: ICML, 96, Citeseer, 1996, pp. 148–156.
[87] M.A. Hearst, S.T. Dumais, E. Osuna, J. Platt, B. Scholkopf, Support vector machines, IEEE Intelli. Syst. Appl. 13 (4) (1998) 18–28.
[88] M.J. Islam, Q.J. Wu, M. Ahmadi, M.A. Sid-Ahmed, Investigating the performance of naive-Bayes classifiers and k-nearest neighbor classifiers, in: Convergence Information Technology, 2007. International Conference on, IEEE, 2007, pp. 1541–1546.
[89] A. Liaw, M. Wiener, et al., Classification and regression by randomforest, R news 2 (3) (2002) 18–22.
[90] I. Rish, et al., An empirical study of the naive Bayes classifier, in: IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, 3, IBM New York, 2001, pp. 41–46.
[91] T.R. Patil, S. Sherekar, Performance analysis of naive Bayes and j48 classification algorithm for data classification, Int. J. Comput. Sci.Appl. 6 (2) (2013) 256–261.
[92] M. Vihinen, How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis, in: BMC Genomics, 13, BioMed Central, 2012, p. S2.
[93] D.M. Powers, Evaluation: from precision, recall and f-measure to ROC, informedness, markedness and correlation (2011).
[94] J.-P. Brans, B. Mareschal, Promethee methods, in: Multiple Criteria Decision Analysis: State of the Art Surveys, Springer, 2005, pp. 163–186.
[95] S. Magder, Mean arterial pressure calculator, 2018, https://www.mdcalc.com/mean-arterial-pressure-map.
[96] D. Kocev, C. Vens, J. Struyf, S. Džeroski, Tree ensembles for predicting structured outputs, Pattern Recognit. 46 (3) (2013) 817–833.
[97] N. Koceska, R. Komadina, M. Simjanoska, B. Koteska, A. Strahovnik, A. Jošt, A. Madevska-Bogdanova, V. Trajkovik, J.F. Tasič, J. Trontelj, et al., Mobile wireless monitoring system for prehospital emergency care, Eur. J. Trauma Emerg. Surg. (2019) 1–8.
[98] M.A. Serhani, M. Al Hemairy, S. Amin, M. Alahmad, System and method for remote healthcare, 2018, US Patent App. 15/395,121.
[99] N.H.B.P.E. Program, et al., The seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure (2004).
[100] AHA, Understanding blood pressure readings, 2016, http://www.heart.org/HEARTORG/Conditions/HighBloodPressure/AboutHighBloodPressure.
[101] P. Su, X.-R. Ding, Y.-T. Zhang, J. Liu, F. Miao, N. Zhao, Long-term blood pressure prediction with deep recurrent neural networks, in: 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), IEEE, 2018, pp. 323–328.
[102] K. Katayama, H. Ishizawa, S. Koyama, K. Fujimoto, Improvement of blood pressure prediction using artificial neural network, in: 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA), IEEE, 2018, pp. 1–5.
[103] A.M. Zakrzewski, B.W. Anthony, Noninvasive blood pressure estimation using ultrasound and simple finite element models, IEEE Trans. Biomed. Eng. 65 (9) (2018) 2011–2022.
[104] S.S. Mousavi, M. Firouzmand, M. Charmi, M. Hemmati, M. Moghadam, Y. Ghorbani, Blood pressure estimation from appropriate and inappropriate PPG signals using a whole-based method, Biomed. Signal Process. Control 47 (2019) 196–206.
[105] D. Wu, L. Xu, R. Zhang, H. Zhang, L. Ren, Y.-T. Zhang, Continuous cuff-less blood pressure estimation based on combined information using deep learning approach, J. Med. Imaging Health Inform. 8 (6) (2018) 1290–1299.
[106] X. Fan, Q. Ye, X. Yang, S.D. Choudhury, Robust blood pressure estimation using an RGBcamera, J. Ambient Intell. Humaniz. Comput. (2018) 1–8.
[107] D.W. Jones, J.E. Hall, The national high blood pressure education program: thirty years and counting, 2002.
[108] S.S. Mousavi, M. Charmi, M. Firouzmand, M. Hemmati, M. Moghadam, A new approach based on dynamical model of the ECG signal to blood pressure estimation, in: 2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA), IEEE, 2019, pp. 210–215.