# Novel Data Processing Approach for Deriving Blood Pressure from ECG only

Monika Simjanoska[1], Martin Gjoreski[2], Matjaž Gams[2], and Ana Madevska Bogdanova[1]

[1] Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University, Rugjer Boshkovikj, 16, 1000, Skopje, Macedonia
{monika.simjanoska, ana.madevska.bogdanova}@finki.ukim.mk
[2] Department of Intelligent Systems, Jožef Stefan Institute, Jožef Stefan International Postgraduate School, Jamova cesta 39, 1000, Ljubljana, Slovenia
{martin.gjoreski,matjaz.gams}@ijs.si

**Abstract.** Blood pressure is one of the most valuable vital signs. Recently, the use of bio-sensors has expanded, however, the blood pressure estimation still requires additional devices. We proposed a method based on complexity analysis and machine learning techniques for blood pressure estimation using only ECG signals. Using ECG recordings from 51 different subjects by using three commercial bio-sensors and clinical equipment, we evaluated the proposed methodology by using leave-one-subject-out evaluation. The method achieves mean absolute error (MAE) of 8.2 mmHg for SBP, 8.7 mmHg for DBP and 7.9 mmHg for the MAP prediction. When models are calibrated using person-specific labelled data, the MAE decreases to 7.1 mmHg for SBP, 6.3 mmHg for DBP and 5.4 mmHg for MAP. The experimental results indicate that when a person-specific calibration data is used, the proposed method can achieve results close to a certified medical device for BP estimation.

**Keywords:** Blood pressure · ECG · Machine learning · Complexity analysis · Classification · Regression · Stacking.

## 1 Introduction

Blood pressure (BP) increase, hypertension, is one of the key factors for cardiovascular diseases [22, 19]. The recent advances in bio-sensors technology has brought the opportunity to continuously monitor physiological signals (e.g., ECG, PPG, EMG, etc.) and consequently calculate or estimate the vital parameters: heart rate, respiratory rate, peripheral capillary oxygen saturation (SpO2) and blood pressure. BP estimation is considered to be a great challenge since the methodologies reported in the literature [21, 23, 13, 26, 12] usually require multiple physiological signals and devices for its estimation. However, in our previous research we proved that the systolic BP (SBP), diastolic BP (DBP) and mean arterial pressure (MAP) can be estimated by using only the ECG signal as a single source of information [24]. The methodology proposed relied on

a combination of complexity analysis and machine learning (ML) techniques to build regression models that are able to predict the actual SBP, DBP and MAP values. By using a train-validation-test evaluation, we achieved a mean absolute error (MAE) of 8.6 mmHg for SBP, 18.2 mmHg for DBP, and 13.5 mmHg for the MAP prediction. By applying a probability distribution-based calibration, the MAE decreases to 7.7 mmHg for SBP, 9.4 mmHg for DBP and 8.1 mmHg for MAP.

In this paper, we consider a different evaluation of the methodology by performing leave-one-subject-out instead of the traditional train-validation-test evaluation and allowing a person-specific calibration to adapt the models to a particular user. The results obtained show significant improvement, especially for the DBP and MAP, decreasing the MAE error $\sim$ -10 for non-calibrated DBP case and $\sim$ -6 for the non-calibrated MAP case. When using a person-specific calibration, the improvements obtained are $\sim$ -3 for DBP and MAP.

The rest of the paper is organized as follows. The proposed method is briefly described in Section 2. The experimental results are presented in Section 3, followed by a discussion in Section 4 and the conclusions of the study given in Section 5.

## 2 Methods and Materials

### 2.1 Methods

The complete methodology is comprehensively explained in [24] and is briefly depicted in Figure 1. Raw ECG signals are divided into 30-seconds segments, each accompanied with SBP and DBP values. Those values pass through the preprocessing method, labelling the segments into the appropriate BP class and applying the low-pass filter. Hereupon, the signals are forwarded to the module for complexity analysis and feature extraction. Having computed the complexity metrics (signal mobility, signal complexity, fractal dimension, entropy and autocorrelation), the feature vectors are inputted to the classification module, which implements a stacking ML approach. The output of the classification module, in a combination with the extracted features, is inputted to a regression module which outputs the actual SBP and DBP estimation. The last module is a calibration module which allows for person-specific calibration. The calibration is performed by considering the mean error of the predictions for five randomly selected instances (measurements) of each subject, compared to the actual absolute SBP and DBP values. The error is either added or subtracted from the predicted values, depending on the models tendency to predict higher or lower values.

### 2.2 Materials

The database we created for this research (publicly available online [1]) is built by using three different commercial ECG sensors (whose reliability is proven in

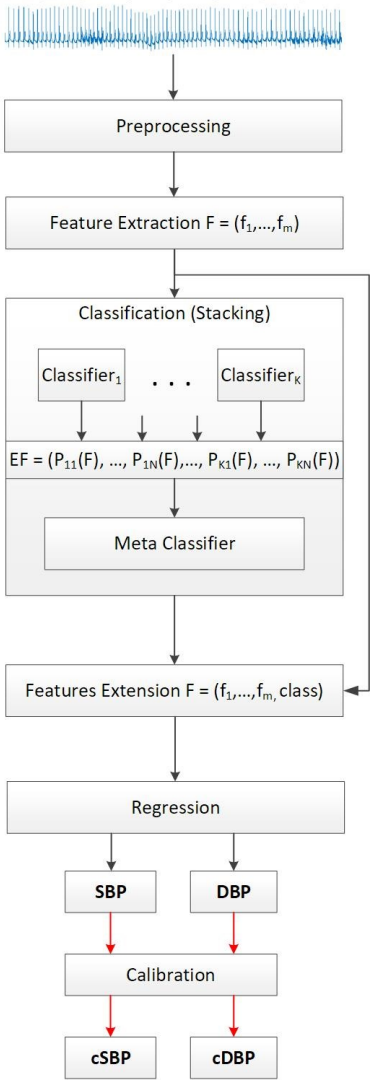


**Fig. 1.** Proposed methodology for blood pressure estimation.

previous studies [27], [3], [20], [2], [18], [5], [14], [6], [15], [11]) and the reference SBP and DBP values measured by using an electronic sphygmomanometer. The second database that we use to additionally evaluate the methodology is obtained from the Physionet database and is created by using clinical equipment [9]. A summarized information of the datasets is provided in the following Table 1. Most of the participants are healthy (33). The rest 18 unhealthy participants were measured in hospital conditions, 11 of which are with cardiovascular problems and 7 are with brain injuries.

**Table 1.** Datasets summary information.

| Source | Num. part | Age | Status |
|---|---|---|---|
| Cooking hacks sensor [10] | 16 | 16 - 72 | healthy |
| 180◦ eMotion FAROS [4] | 3 | 25 - 27 | healthy |
| Zephyr Bioharness module [25] | 25 | 20 - 73 | 14 healthy, 11 unhealthy |
| Charis Physionet database [17] | 7 | 20 - 74 | brain injuries |

## 3 Results

Four types of experiments are performed: classification, regression, feature analysis and devices evaluation. The classification experiments were performed to measure the ability of the classification algorithms to estimate BP class (hypotension, normal or hypertension - described in details in [24]) and were needed as an additional input for the regression algorithms. The regression experiments provide the error measurement for predicting the actual BP values of the method. Here we present three types of experiments, leave-one-subject-out (LOSO) inter-dataset, LOSO within dataset and leave-one-dataset-out (LODO). The Feature analysis experiments were performed to analyze the quality of the chosen features for the specific tasks of BP estimation. Eventually, the evaluation experiments were performed to provide an insight into devices evaluation, i.e. the performance of the wearable technology vs. the validated and reliable clinical monitors. The details for each experiment are in the following subsections.

### 3.1 Classification experiments

The preprocessing and the feature extraction phase produced a total number of 3129 feature vectors mapped into three BP classes (hypotension - 0, normal - 1 and hypertension - 2). When developing both the classification and the regression models, we used LOSO cross-validation, meaning that we trained 51 models by including 50 subjects in the training set and leaving 1 subject out for testing. The performance of the stacking ML solution used for the classification was evaluated through the F-measure as a balanced mean between precision and recall for each class, and the overall accuracy of the classifier. The recall shows the proportion of the given class cases correctly predicted among all the instances that belong in the given class:

$$Recall = \frac{True\_positives}{Real\_positives} \tag{1}$$

Precision is a measure showing the proportion of the given class cases correctly predicted among all instances predicted to belong in the given class:

$$Precision = \frac{True\_positives}{Predicted\_positives} \tag{2}$$

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \tag{3}$$

The stacking design produced the results presented in Table 2.

**Table 2.** Stacking approach results.

| Class/Metric | Precision | Recall | F-measure | Accuracy |
|---|---|---|---|---|
| 0 | 0.71 | 0.67 | 0.69 | |
| 1 | 0.58 | 0.89 | 0.71 | 0.73 |
| 2 | 0.94 | 0.63 | 0.76 | |

### 3.2 Regression experiments

The predicted BP classes were used to extend the initial feature vectors and prepare the data for regression, as depicted in Figure 1. Following the same principle for LOSO cross-validation, we evaluated three distinct models for predicting the SBP, DBP and MAP. To improve the prediction, we applied a calibration procedure as described in the Section 2. The results are sublimated in Table 3. The regression models were evaluated by using the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). MAE is the average error obtained from the absolute differences between the actual, $a_i$, and the predicted values $p_i$, for $i = 1, , n$, where $n$ is the number of instances within a subject. MAE weights all the differences equally and is calculated as:

$$MAE = \frac{\sum_1^n |p_i - a_i|}{n} \tag{4}$$

To obtain higher weight for the large errors, which is important for the BP problem, the differences between the actual absolute and the predicted values are first squared, then averaged, and afterwards a square root of the average is performed. The RMSE is calculated according to the following equation:

$$RMSE = \sqrt{\frac{\sum_1^n |p_i - a_i|^2}{n}} \tag{5}$$

Table 3 presents the MAE and RMSE evaluation for SBP, DBP and MAP. For each subject, $ID = 1, ..., 51$, and the errors from both the prediction and the calibration are presented for all three cases. The results show that the calibration goes in favor of the prediction by reducing the overall MAE from 8.24 $\pm$ 5.34 ($\mu$ $\pm$ $\sigma$) to 7.11 $\pm$ 5.29 for the SBP, from 8.75 $\pm$ 7.90 to 6.28 $\pm$ 5.02 for the DBP case, and from 7.92 $\pm$ 9.66 to 5.35 $\pm$ 4.16 for the MAP case.

Given the four datasets used in the experiments, two more experiments were performed using LOSO within a dataset and LODO evaluations. The results for the LOSO within dataset are presented in Table 4. The different datasets are labelled 1-4 (according to the device used for the measurements). Considering the obtained errors, for the SBP the MAE of 8.05 is close to the mean MAE obtained from the LOSO testing in Table 3. For the DBP case, the most critical datasets are number 2 and 4. Perhaps, this is due to the reduced number of participants available in those datasets. However, the calibration method is still able to perform well and provides a mean MAE of 6.5 $\pm$ 0.99.

**Table 3.** MAE and RMSE evaluation for SBP, DBP and MAP.

| ID | SBP | | | | DBP | | | | MAP | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Prediction | | Calibration | | Prediction | | Calibration | | Prediction | | Calibration | |
| | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| 1 | 10.8 | 11.3 | 3.1 | 3.5 | 17.0 | 15.5 | 2.9 | 3.2 | 15.8 | 14.4 | 3.1 | 3.2 |
| 2 | 9.7 | 10.6 | 10.1 | 10.1 | 4.9 | 6.5 | 4.2 | 4.6 | 9.3 | 9.9 | 4.2 | 4.7 |
| 3 | 5.8 | 6.8 | 6.8 | 9.0 | 6.3 | 7.0 | 2.8 | 3.3 | 7.5 | 7.7 | 2.3 | 2.5 |
| 4 | 5.3 | 5.3 | 5.3 | 5.3 | 4.9 | 5.0 | 4.9 | 5.0 | 5.6 | 5.7 | 5.6 | 5.7 |
| 5 | 10.2 | 13.0 | 7.8 | 10.0 | 5.6 | 7.3 | 4.6 | 5.9 | 5.7 | 7.1 | 6.1 | 6.9 |
| 6 | 12.5 | 13.0 | 2.8 | 3.8 | 6.2 | 8.0 | 4.0 | 5.1 | 9.3 | 10.5 | 4.7 | 5.3 |
| 7 | 6.3 | 6.7 | 2.6 | 3.7 | 4.5 | 5.6 | 3.6 | 4.5 | 5.3 | 5.8 | 2.5 | 3.2 |
| 8 | 6.7 | 7.5 | 6.3 | 7.5 | 5.5 | 7.7 | 5.2 | 6.3 | 4.5 | 5.6 | 3.4 | 4.0 |
| 9 | 5.0 | 6.4 | 4.6 | 6.2 | 4.5 | 5.3 | 3.1 | 3.5 | 4.0 | 4.5 | 4.6 | 5.5 |
| 10 | 5.1 | 4.9 | 5.0 | 5.0 | 7.6 | 7.4 | 7.4 | 7.7 | 6.2 | 7.8 | 6.5 | 8.6 |
| 11 | 20.9 | 19.8 | 4.8 | 7.0 | 15.5 | 18.4 | 5.2 | 10.2 | 20.8 | 18.2 | 7.1 | 7.3 |
| 12 | 6.1 | 7.2 | 6.2 | 7.1 | 2.1 | 2.4 | 2.4 | 2.3 | 3.4 | 3.3 | 3.7 | 3.5 |
| 13 | 10.2 | 9.2 | 3.4 | 3.3 | 8.2 | 7.2 | 4.0 | 5.1 | 10.9 | 10.5 | 5.2 | 5.2 |
| 14 | 10.2 | 13.7 | 16.9 | 19.0 | 8.6 | 12.2 | 9.9 | 11.9 | 6.5 | 9.1 | 14.4 | 16.1 |
| 15 | 12.3 | 14.8 | 9.2 | 11.6 | 15.9 | 18.0 | 10.5 | 10.9 | 11.1 | 14.8 | 8.0 | 10.5 |
| 16 | 19.0 | 21.2 | 18.1 | 21.4 | 11.7 | 10.7 | 11.3 | 10.1 | 7.2 | 9.5 | 7.1 | 9.7 |
| 17 | 8.6 | 9.8 | 10.9 | 11.0 | 15.4 | 15.6 | 4.1 | 4.8 | 7.7 | 7.5 | 3.1 | 3.5 |
| 18 | 29.3 | 17.9 | 31.3 | 20.0 | 52.6 | 33.6 | 35.4 | 24.8 | 68.5 | 32.5 | 25.6 | 12.6 |
| 19 | 7.1 | 7.7 | 7.1 | 7.9 | 20.2 | 21.5 | 6.1 | 7.5 | 15.0 | 14.6 | 3.8 | 4.8 |
| 20 | 7.1 | 7.2 | 5.2 | 5.8 | 6.7 | 7.5 | 5.7 | 6.2 | 7.6 | 7.5 | 3.4 | 3.3 |
| 21 | 10.5 | 11.4 | 11.4 | 10.1 | 14.4 | 12.3 | 13.5 | 12.0 | 7.6 | 9.1 | 11.9 | 9.5 |
| 22 | 10.9 | 11.8 | 3.4 | 4.6 | 4.5 | 4.8 | 6.2 | 6.1 | 11.8 | 10.9 | 4.1 | 4.8 |
| 23 | 4.9 | 5.6 | 5.5 | 7.5 | 4.1 | 5.5 | 3.5 | 4.7 | 3.3 | 4.4 | 3.9 | 5.0 |
| 24 | 5.8 | 7.1 | 5.8 | 7.1 | 8.6 | 9.5 | 8.6 | 9.5 | 2.3 | 4.0 | 2.3 | 4.0 |
| 25 | 7.0 | 7.0 | 7.0 | 7.0 | 3.1 | 3.1 | 3.1 | 3.1 | 3.0 | 3.0 | 3.0 | 3.0 |
| 26 | 8.7 | 9.8 | 8.7 | 9.8 | 9.5 | 10.1 | 9.5 | 10.1 | 2.9 | 5.0 | 2.9 | 5.0 |
| 27 | 17.6 | 17.6 | 17.6 | 17.6 | 7.2 | 7.2 | 7.2 | 7.2 | 0.7 | 0.7 | 0.7 | 0.7 |
| 28 | 2.3 | 2.7 | 2.3 | 2.7 | 8.3 | 8.4 | 8.3 | 8.4 | 2.6 | 4.5 | 2.6 | 4.5 |
| 29 | 4.6 | 5.3 | 4.6 | 5.3 | 6.3 | 7.2 | 6.3 | 7.2 | 4.7 | 7.7 | 4.7 | 7.7 |
| 30 | 4.2 | 4.8 | 4.2 | 4.8 | 3.6 | 5.0 | 3.6 | 5.0 | 1.6 | 2.4 | 1.6 | 2.4 |
| 31 | 4.0 | 4.6 | 4.0 | 4.6 | 2.6 | 3.6 | 2.6 | 3.6 | 2.7 | 3.6 | 2.7 | 3.6 |
| 32 | 8.3 | 9.9 | 8.3 | 9.9 | 8.2 | 10.4 | 8.2 | 10.4 | 7.0 | 9.3 | 7.0 | 9.3 |
| 33 | 2.4 | 3.2 | 2.4 | 3.2 | 7.4 | 8.8 | 7.4 | 8.8 | 4.8 | 6.9 | 4.8 | 6.9 |
| 34 | 6.8 | 6.8 | 6.8 | 6.8 | 4.4 | 4.4 | 4.4 | 4.4 | 6.6 | 6.6 | 6.6 | 6.6 |
| 35 | 13.3 | 13.3 | 13.3 | 13.3 | 11.9 | 11.9 | 11.9 | 11.9 | 12.1 | 12.1 | 12.1 | 12.1 |
| 36 | 7.9 | 7.9 | 7.9 | 7.9 | 0.4 | 0.4 | 0.4 | 0.4 | 1.2 | 1.2 | 1.2 | 1.2 |
| 37 | 6.6 | 6.6 | 6.6 | 6.6 | 2.1 | 2.1 | 2.1 | 2.1 | 2.7 | 2.7 | 2.7 | 2.7 |
| 38 | 0.5 | 0.5 | 0.5 | 0.5 | 6.1 | 6.1 | 6.1 | 6.1 | 4.7 | 4.7 | 4.7 | 4.7 |
| 39 | 2.3 | 2.3 | 2.3 | 2.3 | 5.3 | 5.3 | 5.3 | 5.3 | 3.2 | 3.2 | 3.2 | 3.2 |
| 40 | 4.8 | 4.8 | 4.8 | 4.8 | 6.5 | 6.5 | 6.5 | 6.5 | 4.2 | 4.2 | 4.2 | 4.2 |
| 41 | 4.9 | 4.9 | 4.9 | 4.9 | 5.3 | 5.3 | 5.3 | 5.3 | 5.8 | 5.8 | 5.8 | 5.8 |
| 42 | 3.2 | 3.2 | 3.2 | 3.2 | 6.1 | 6.1 | 6.1 | 6.1 | 7.1 | 7.1 | 7.1 | 7.1 |
| 43 | 0.7 | 0.7 | 0.7 | 0.7 | 4.0 | 4.0 | 4.0 | 4.0 | 1.6 | 1.6 | 1.6 | 1.6 |
| 44 | 11.4 | 11.4 | 11.4 | 11.4 | 2.0 | 2.0 | 2.0 | 2.0 | 2.5 | 2.5 | 2.5 | 2.5 |
| 45 | 16.0 | 20.3 | 11.4 | 13.2 | 12.1 | 15.2 | 9.6 | 11.2 | 12.7 | 15.8 | 13.4 | 16.3 |
| 46 | 10.2 | 12.6 | 6.7 | 13.2 | 3.6 | 4.1 | 4.9 | 5.3 | 7.0 | 7.2 | 4.2 | 5.4 |
| 47 | 3.0 | 3.6 | 4.5 | 5.2 | 6.2 | 6.7 | 2.2 | 2.5 | 2.4 | 3.0 | 2.2 | 2.7 |
| 48 | 4.4 | 5.1 | 4.3 | 5.8 | 17.2 | 17.8 | 3.9 | 4.7 | 11.8 | 12.6 | 4.1 | 5.2 |
| 49 | 11.6 | 16.2 | 11.4 | 15.8 | 20.3 | 21.0 | 5.6 | 6.8 | 15.2 | 16.8 | 7.0 | 7.8 |
| 50 | 8.2 | 9.8 | 5.0 | 6.1 | 8.5 | 10.4 | 8.8 | 10.3 | 8.7 | 11.0 | 8.9 | 11.0 |
| 51 | 5.1 | 6.4 | 4.2 | 5.4 | 12.2 | 14.3 | 6.2 | 7.7 | 7.4 | 8.7 | 4.5 | 5.5 |
| Mean | 8.2 | 8.8 | 7.1 | 7.8 | 8.7 | 9.1 | 6.3 | 6.7 | 7.9 | 7.9 | 5.4 | 5.8 |
| SD | 5.3 | 5.0 | 5.3 | 4.8 | 7.9 | 6.1 | 5.0 | 3.9 | 9.7 | 5.5 | 4.2 | 3.5 |

The results for the LODO are presented in Table 5. Given the MAE and RMSE, it can be perceived that if the datasets are completely unknown to the classifier, then the model performs worse even in the calibration case.

**Table 4.** Leave-one-subject-out within dataset results.

| BP | Dataset | Prediction | | Calibration | |
|---|---|---|---|---|---|
| | | MAE | RMSE | MAE | RMSE |
| SBP | 1 | 8.6 | 9.8 | 5.8 | 7.1 |
| | 2 | 8.0 | 8.9 | 6.1 | 6.6 |
| | 3 | 7.5 | 7.8 | 7.3 | 7.7 |
| | 4 | 8.1 | 10.3 | 7.6 | 10.1 |
| DBP | 1 | 7.4 | 8.5 | 5.5 | 6.6 |
| | 2 | 14.9 | 15.5 | 6.5 | 7.4 |
| | 3 | 6.4 | 7.0 | 6.4 | 6.8 |
| | 4 | 11.8 | 13.4 | 7.9 | 9.7 |

**Table 5.** Leave-one-dataset-out results.

| BP | Prediction | | Calibration | |
|---|---|---|---|---|
| | MAE | RMSE | MAE | RMSE |
| SBP | 13.0 | 15.7 | 10.2 | 12.6 |
| DBP | 12.3 | 14.9 | 14.3 | 17.0 |

Finally, Table 6 presents a summarization of the results for the three different evaluations: leave-one-subject-out inter-dataset, leave-one-subject-out within dataset and leave-one-dataset-out. The main three observations are:

1. For both the SBP and DBP prediction, the LODO evaluation results are worse compared to the LOSO inter-dataset and within dataset.
2. Regarding the LOSO evaluations, for the SBP prediction the within dataset models have slightly better MAE, but worse RMSE, meaning that the models (within dataset and inter dataset) perform similarly.
3. However, for the DBP prediction, the inter-dataset models yield the best results.

**Table 6.** Comparison of the evaluation metrics (MAE and RMSE) for the three evaluations: LOSO inter-dataset, LOSO within dataset and LODO.

| BP | Evaluation type | Prediction | | Calibration | |
|---|---|---|---|---|---|
| | | $\mu$(MAE) | $\mu$(RMSE) | $\mu$(MAE) | $\mu$(RMSE) |
| SBP | LOSO inter-dataset | 8.2 | **8.8** | 7.1 | **7.8** |
| | LOSO within dataset | **8.0** | 9.2 | **6.7** | 7.9 |
| | LODO | 13.0 | 9.2 | 6.7 | 7.9 |
| DBP | LOSO inter-dataset | **8.7** | **9.1** | **6.3** | **6.7** |
| | LOSO within dataset | 10.1 | 11.1 | 6.7 | 7.6 |
| | LODO | 12.3 | 14.9 | 6.6 | 7.6 |

The differences between the LOSO and LODO are visualized in Figure 2 and Figure 3, correspondingly. Both figures present the SBP absolute values,

rather than the absolute errors as provided in Table 3, of a patient referred to as patient X, from the Charis/Physionet database. The patient is chosen to be suitable since there is high variability in the BP values which is appropriate to visually represent whether the predictions follow the trend of the actual absolute values. The real BP are marked with a black line. The blue line represents the stacking approach predictions, and the green line represents the predictions after the calibration. The red line represents the performance of a simple classifier the one that always predicts the mean value from the training set. The x-axis shows the continuous instances (samples) for the particular patient and the y-axis presents the absolute values of the SBP in mmHg. In Figure 2 it can be perceived that in case of LOSO, the stacking classifier (before and after the calibration) follows the tendency of the actual BP values. On the contrary, when LODO, the stacking classifier (the blue line) is unable to predict better than the simple classifier predicting the mean BP values (the red line). Considering the results, it can be concluded that the stacking classifier needs to be provided with a training instance from the particular dataset before accurate predictions can be made.

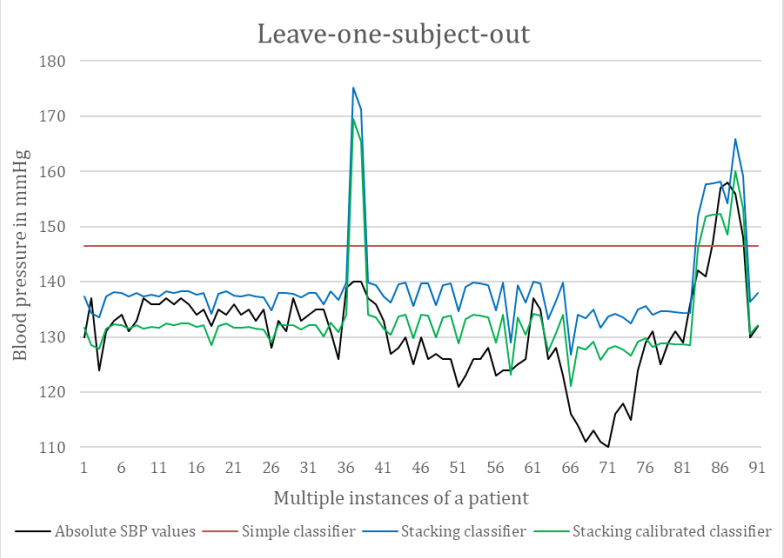


**Fig. 2.** Leave-one-subject-out for patient X.

### 3.3 Feature analysis

Observing the experimental results, it can be noticed that the models are able to approximate the actual absolute BP values. In order to provide more insights

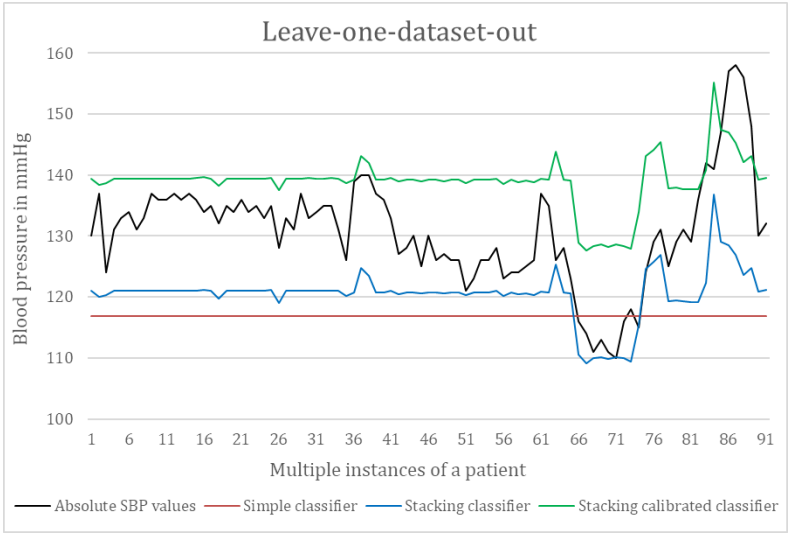


**Fig. 3.** Leave-one-dataset-out for patient X.

into the usefulness of the complexity features, we provided additional analysis comparing each feature value with respect to the real BP values. Considering a sequence of actual BP values within 6 hours period, in Figure 4 we depicted the features values (y-axis) depending on the BP values sorted ascendingly. It can be seen that the increase in the BP values, influences the absolute value and the variability of the features.

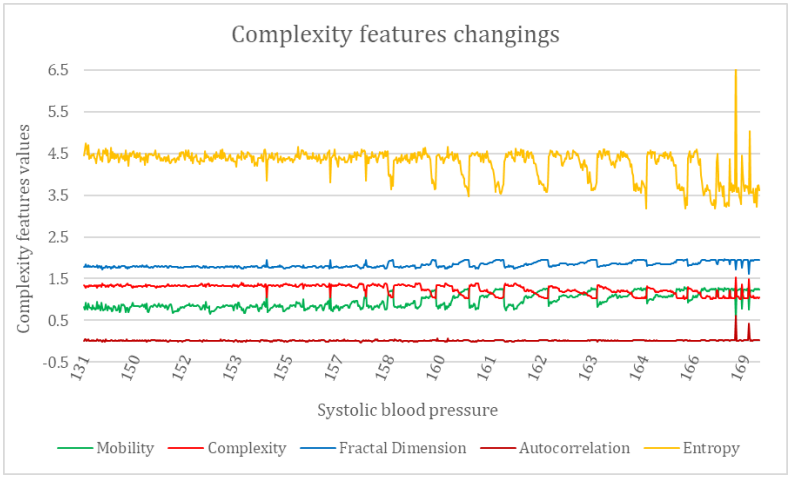


**Fig. 4.** BP changings tracking for patient X.

Next, in Figure 5 we present box-whisker plots to visualize the shape of the distributions, the mean value, and the variability of each complexity feature with respect to the three BP classes. It can be seen that for some of the features, e.g., Mobility, Complexity and Entropy, just the mean value itself has a discriminatory power for the three classes. In addition to the mean value, the variability of the feature values has some additional information. Even though, in some cases, the variability of the feature values may indicate noise in the data.

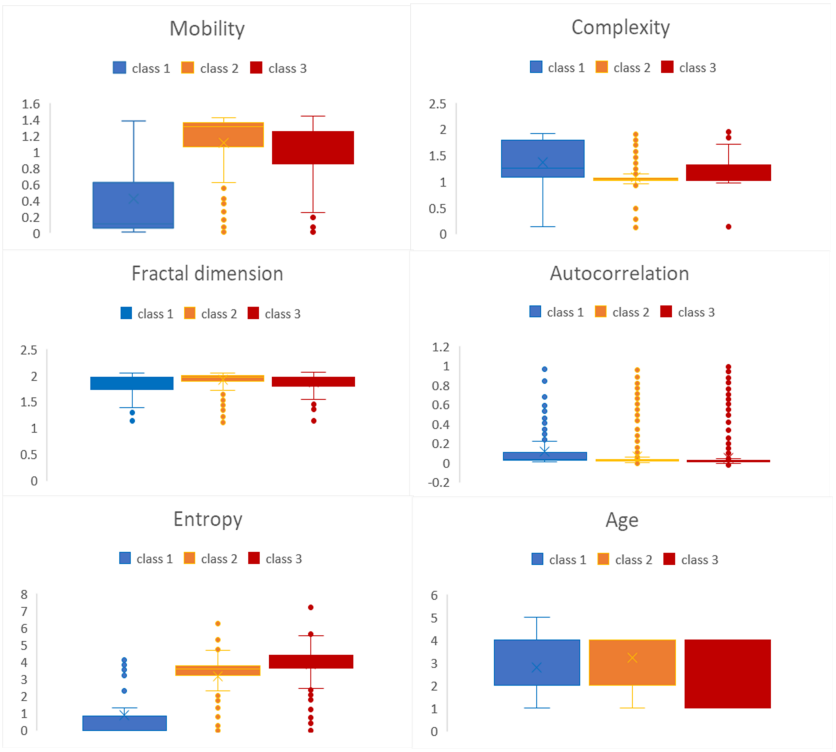


**Fig. 5.** Box-Whiskey plots per class for the complexity features.

### 3.4 Devices evaluation

Even though all participants groups are measured with different devices, we provide a comparison between the performance of the used wearable bio-sensors and the clinical monitors regularly used in medical practice, regarding our results. The comparison is made as a ratio between the prediction errors obtained from the experiments. Considering the gained measurements, the groups of participants (healthy, with brain injury and unhealthy), and the types of devices (bio-sensors and clinical monitors), we have compared the results for two main

**Table 7.** Concession difference.

| BP | Group | Prediction (P) | | | Calibration (C) | | | Ratio P | Ratio C |
|---|---|---|---|---|---|---|---|---|---|
| | | MAE (M) | RMSE (R) | R - M | MAE (M) | RMSE (R) | R - M | | |
| **SBP** | WS | 8.01 | 8.87 | **0.87** | 6.41 | 7.14 | **0.74** | **2.6** | **3.4** |
| | CM | 8.07 | 10.31 | **2.24** | 7.63 | 10.11 | **2.49** | | |
| **DBP** | WS | 8.04 | 9.59 | **1.55** | 7.02 | 8.63 | **1.61** | **1.0** | **1.1** |
| | CM | 11.82 | 13.37 | **1.55** | 7.88 | 9.68 | **1.80** | | |

groups. The first group encompasses the healthy and unhealthy participants measured with wearable sensors technology (WS), and the second group comprise the participants with brain injuries measured with the regularly used clinical monitors (CM) in the hospitals. In Table 7 we have used the differences between the RMSE and MAE (R-M) to calculate the differences within the calibration and the prediction models in order to compute the ratios of errors between WS and CM. The differences between the RMSE and MAE were used to show the disparity of the actual SBP and DBP outputs and the predicted/calibrated outcome in the participants readings. The calculations show that the errors are approximately the same for all the participants. Having computed the difference R - M for both the prediction and the calibration models, we calculated the ratios for both cases. Ratio P for SBP shows that the CM group suffers 2.6 times worse errors than the WS group; whereas in the DBP case the errors are the same. Ratio C is in regard of the condition after the calibration - the CM group suffers 3.4 times worse errors (even more than in the prediction case); whereas for DBP the errors are only 1.1 time worse.

## 4 Discussion

Our BP estimation system based on ECG sensor inputs enabled the reliable monitoring of various BP parameters on data obtained from 51 different subjects and 4 different ECG sensors. In the traditional train-validation-test evaluation of our method [24], we achieved MAE (non-calibrated/calibrated) measured in mmHg of 8.6/7.7 for SBP, 18.2/9.4 for DBP, and 13.5/8.13 for MAP. Performing a completely different approach to evaluate the methodology as presented in this paper, it turned out that the performance can be significantly improved. The error on an unseen dataset, using another sensor, is 13 for the SBP and 12.3 for the DBP prediction (see no calibration LODO results in Table 5). If sensor-specific labelled dataset is provided, the MAE decreases to 8.2 for SBP and 8.7 for DBP prediction (see no calibration LOSO inter-dataset results in Table 3). Moreover, if person-specific labelled data is provided, the MAE decreases to 6.7 for SBP and 6.6 for DBP (see calibration results for LODO in Table 6). These results are close to a certified medical device for BP estimation ($\pm 5$ mmHg, and the SD within 8 mmHg according to BHS and AAMI standards [16]). Considering the time performance of the method, once the prediction model is built, the predictions can be considered real-time calculations.

## 5 Conclusions

Our method estimates SBP, DBP and the MAP from ECG sensor data. The method was tested on 51 different subjects and 4 different ECG sensors - part of which we obtained from online database and the rest from the database that we created. By performing leave-one-subject out evaluation, the method achieved results close to a certified medical device, especially when sensor-specific and person-specific labelled data is provided.

The proposed solution has promising real-world applications in civilian and military environments, however it should be tested with a dataset containing hundreds of diverse participants in variety of medical conditions. In the future work, the method could be enriched by an activity recognition module [7], or context-based BP estimation [8].

## References

1. Blood pressure databases., http://www.webcitation.org/6ulZxAGP8
2. Ahonen, L., Cowley, B., Torniainen, J., Ukkonen, A., Vihavainen, A., Puolamäki, K.: Cognitive collaboration found in cardiac physiology: Study in classroom environment. PloS one **11**(7), e0159178 (2016)
3. BEREKSI-REGUIG, M.A., Bereksi-Reguig, F., Ali, A.N.: A new system for measurement of the pulse transit time, the pulse wave velocity and its analysis. Journal of Mechanics in Medicine and Biology **17**(01), 1750010 (2017)
4. Biosignals., B.: Emotion faros. 2016., http://www.megaemg.com/products/faros/
5. Cliff, D.P., McNeill, J., Vella, S., Howard, S.J., Kelly, M.A., Angus, D.J., Wright, I.M., Santos, R., Batterham, M., Melhuish, E., et al.: The preschool activity, technology, health, adiposity, behaviour and cognition (path-abc) cohort study: rationale and design. BMC pediatrics **17**(1), 95 (2017)
6. Ding, H., Sarela, A., Helmer, R., Mestrovic, M., Karunanithi, M.: Evaluation of ambulatory ecg sensors for a clinical trial on outpatient cardiac rehabilitation. In: Complex Medical Engineering (CME), 2010 IEEE/ICME International Conference on. pp. 240–243. IEEE (2010)
7. Gjoreski, M., Gjoreski, H., Luštrek, M., Gams, M.: How accurately can your wrist device recognize daily activities and detect falls? Sensors **16**(6), 800 (2016)
8. Gjoreski, M., Luštrek, M., Gams, M., Gjoreski, H.: Monitoring stress with a wrist device using context. Journal of biomedical informatics **73**, 159–170 (2017)
9. Goldberger, A.L., Amaral, L.A., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E.: Physiobank, physiotoolkit, and physionet. Circulation **101**(23), e215–e220 (2000)
10. Hacks., C.: 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
11. Hailstone, J., Kilding, A.E.: Reliability and validity of the zephyr bioharness to measure respiratory responses to exercise. Measurement in Physical Education and Exercise Science **15**(4), 293–300 (2011)
12. Hsiu, H., Hsu, C.L., Wu, T.L.: A preliminary study on the correlation of frequency components between finger ppg and radial arterial bp waveforms. In: Biomedical and Pharmaceutical Engineering, 2009. ICBPE'09. International Conference on. pp. 1–4. IEEE (2009)

13. Ilango, S., Sridhar, P.: A non-invasive blood pressure measurement using android smart phones. IOSR Journal of Dental and Medical Sciences **13**(1), 28–31 (2014)
14. Johnstone, J.A., Ford, P.A., Hughes, G., Watson, T., Garrett, A.T.: Bioharness multivariable monitoring device: part. i: validity. Journal of sports science & medicine **11**(3), 400 (2012)
15. Johnstone, J.A., Ford, P.A., Hughes, G., Watson, T., Mitchell, A.C., Garrett, A.T.: Field based reliability and validity of the bioharness multivariable monitoring device. Journal of sports science & medicine **11**(4), 643 (2012)
16. Jones, D.W., Hall, J.E.: The national high blood pressure education program (2002)
17. Kim, N., Krasner, A., Kosinski, C., Wininger, M., Qadri, M., Kappus, Z., Danish, S., Craelius, W.: Trending autoregulatory indices during treatment for traumatic brain injury. Journal of clinical monitoring and computing **30**(6), 821–831 (2016)
18. Miettinen, T., Myllymaa, K., Westeren-Punnonen, S., Ahlberg, J., Hukkanen, T., Töyräs, J., Lappalainen, R., Mervaala, E., Sipilä, K., Myllymaa, S.: Success rate and technical quality of home polysomnography with self-applicable electrode set in subjects with possible sleep bruxism. IEEE journal of biomedical and health informatics (2017)
19. Mitchell, G.F.: Arterial stiffness and hypertension. Hypertension **64**(1), 13–18 (2014)
20. Morales, J., Díaz-Piedra, C., Di Stasi, L.L., Romero, S., et al.: Low-cost remote monitoring of biomedical signals. In: International Work-Conference on the Interplay Between Natural and Artificial Computation. pp. 288–295. Springer (2015)
21. Nitzan, M.: Automatic noninvasive measurement of arterial blood pressure. IEEE instrumentation & measurement magazine **14**(1) (2011)
22. Rosendorff, C., Lackland, D.T., Allison, M., Aronow, W.S., Black, H.R., Blumenthal, R.S., Cannon, C.P., De Lemos, J.A., Elliott, W.J., Findeiss, L., et al.: Treatment of hypertension in patients with coronary artery disease. Hypertension **65**(6), 1372–1407 (2015)
23. Sahoo, A., Manimegalai, P., Thanushkodi, K.: 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. pp. 285–289. IEEE (2011)
24. Simjanoska, M., Gjoreski, M., Gams, M., Madevska Bogdanova, A.: Non-invasive blood pressure estimation from ecg using machine learning techniques. Sensors **18**(4), 1160 (2018)
25. Technology., Z.: Zephyr bioharness 3.0 user manual. 2017, https://www.zephyranywhere.com/media/download/bioharness3-user-manual.pdf
26. Thomas, S.S., Nathan, V., Zong, C., Soundarapandian, K., Shi, X., Jafari, R.: Biowatch: A noninvasive wrist-based blood pressure monitor that incorporates training techniques for posture and subject variability. IEEE journal of biomedical and health informatics **20**(5), 1291–1300 (2016)
27. Winderbank-Scott, P., Barnaghi, P.: A non-invasive wireless monitoring device for children and infants in pre-hospital and acute hospital environments (2017)