Fast Cuffless Blood Pressure Classification with ECG and PPG signals using CNN-LSTM Models in Emergency Medicine

Ivan Kuzmanov, Ana Madevska Bogdanova, Magdalena Kostoska, Nevena Ackovska
Ss. Cyril and Methodius University,
Faculty of Computer Science and Engineering
Skopje, North Macedonia
ivan.kuzmanov@students.finki.ukim.mk,
{ana.madevska.bogdanova, magdalena.kostoska, nevena.ackovska}@finki.ukim.mk

Abstract—Cuffless blood pressure (BP) measurement is gaining a lot of attention as a promising new technology that can be embedded in a patch-like biosensor device. Electrocardiogram (ECG) and photoplethysmogram (PPG) waveforms are non-invasive by their nature - they can be recorded without sending any electrical impulses to the human body. These signals present different aspects of the cardiovascular system, thus using both of the signals for blood pressure classification seems like a viable strategy. Quick estimation of the blood pressure during the triage process in cases of natural disasters with many injured subjects, is an essential measure for following the hemostability of the injured. The main goal of this study is to develop a two-class classification model (Hypotension and Nothypotension) for fast prediction of the blood pressure category by utilizing ECG and PPG signals, in order to detect a BP sudden drop. The developed deep learning models are based on the LSTM architecture and its variants, CNN-LSTM. We also conducted three class classification model. The models were trained and tested using the data from the UCI Machine Learning Repository Cuff-Less Blood Pressure Estimation dataset with 12000 instances. The best result in the two-class model is AUROC = 0.74.

Keywords—blood pressure (BP) estimation, triage, electrocardiogram (ECG), photoplethysmogram (PPG), long short term memory (LSTM), CNN-LSTM, artificial neural network, deep learning

I. INTRODUCTION

A multisensor patch-like device that can be attached to an injured’s chest in mass casualties events could improve the START triage process. This device is capable of collecting and analysing information on vital parameters such as respiration rate (RR), heartbeat (heart rate – HR), blood oxygen saturation level (SpO2), blood pressure and body temperature [1]. Our interest in this research is to utilize the embedded electrocardiogram (ECG) and photoplethysmogram (PPG) sensors in the patch-like biosensor in order to follow the injured’s hemostability (if there is an internal bleeding).

Blood pressure is the force that the blood exerts on the walls of the blood vessels, as the heart pumps blood. The heart on average beats 60 - 100 times a minute [2]. During each beat the heart performs a cardiac cycle that consists of two phases: the systole, when the heart contracts pumping the blood into the arteries and the diastole, when the heart relaxes after a contraction. With every beat of the heart, the blood pressure changes from the maximum, systolic (SBP), to the minimum, diastolic (DBP) [3]. This is a metric that is often measured by health care professionals during a medical checkup. ECG represents the electrical activity of the heart, while PPG shows the changes of blood volume in the microvascular tissue. The codependency among BP, ECG and PPG has been explored in many different studies [4] [5] [6] [7].

During the overview of the literature we have realized that the BP hypotension category (SBP<90, DBP<60) is not in the focus of most researches, while it is important for the project presented in [1]. Typical symptoms of hypotension are dizziness and fainting. Severely low blood pressure can deprive the brain and other vital organs of oxygen and nutrients, leading to a life-threatening condition - shock. Since our interest is to monitor the injured patient’s hemostability in emergency situations, it’s important to note that a sudden drop in the BP might be caused by an internal bleeding, leading to a change in the triage priority of giving a medical treatment (to triage label ‘immediate’). As a main interest in this paper, we explore the BP categorisation as a two-class classification problem (classes Hypotension and Nohypotension).

As a proposed model for solving the categorization of the BP is a type of recurrent neural network - Long short term memory, LSTM. They are capable of processing sequential information, e.g. time series data (ECG, PPG signals), and are specifically built to be able to follow long term dependencies. This study focuses on CNN-LSTM models. The developed model architecture supports the idea of fast categorisation - the algorithm is simple and able to function in real time.

The proposed methodology is consisted of the following steps:

1) Cleaning the data, by excluding signals that have missing values or the difference between the SBP and DBP is less than 20 or greater than 100, or the value for DBP is greater than 130;
2) The data segments in the longer signals are transformed into smaller ones with uniform length - a format of the input vectors passed to the network models;
3) The ECG and PPG signals are transformed first by standardization and then by filtration;
4) BP categorization - two BP categories (First experiment) and three BP categories (Second Experiment);
5) Building the classification CNN-LSTM models;
6) Evaluation of the models.

The rest of the paper is organized as follows. Section 2 presents an overview on similar papers or related researches. The methodology is described in Section 3. The results of the experiments and the discussion are presented in Section 4, including the introduction of the dataset, the preprocessing part and the used methods. Section 5 presents the conclusion.

II. RELATED WORK

In this section, several researches on the subject of interest are explored - solutions for noninvasive continuous blood pressure estimation utilizing ECG and/or PPG signals. Some of these solutions attempt to manually extract relevant features from the signals, like pulse arrival (PAT) time that has been proven to have a correlation to blood pressure. Other authors rely on deep learning neural networks to automatically find and extract features.

Pulse transit time (PTT) measures the time taken for the arterial pulse pressure wave to travel from the aortic valve to a peripheral site. PAT is defined as the difference between the R-peak of an ECG and the systolic peak of a PPG. While the features are similar, they are not interchangeable [8]. They are both proven markers of BP with the latter gaining prominence in the recent publications [6] [9] [7]. Those features are not the only morphological characteristics extracted from the signals. Many papers suggest their own features as they believe contribute to more accurate estimations, i.e. Womersley number [10] that reflects the flow properties of blood. Other studies consider the main reason for the inaccuracy of the estimators in blood pressure is the assumption that artery diameter is a constant. The pulse intensity ratio (PIR) is suggested as an indicator for the evaluation of changes to the artery diameter and its correlation to blood pressure is experimentally proven [11]. In a seven-day experiment on 30 patients it has been shown that patients that suffer from hypertension have both higher average and higher variance PIR [12]. Another study relies on the hypothesis that a healthy biomedical system is highly complex, and that when abnormalities occur, the complexity of the system decreases. This paper utilizes a complexity analysis for feature extraction, achieves MAE of 8.64 mmHg for SBP, 18.20 mmHg for DBP, and 13.52 mmHg for MAP. These results improve further when models are calibrated: MAE of 7.72 mmHg for SBP, 9.45 mmHg for DBP and 8.13 mmHg for MAP [13]. Heart rate is another feature often taken into account since its easy to calculate from the ECG and an elevated heart rate usually indicates elevated blood pressure. Sometimes they are combined with other information about the patients like demographic characteristics. A study working with features extracted from a PPG signal and demographic characteristics achieves with RMSE of 6.74 and 3.59, for SBP and DBP respectively [14].

Other studies develop deep learning models. One proposed solution uses a waveform based hierarchical Artificial Neural Network–LSTM (ANN-LSTM). The lower hierarchy level ANN extracts necessary morphological features from ECG and PPG waveforms and the upper hierarchy level LSTM layers accounts for the time domain variation of the features extracted by lower hierarchy level. The proposed ANN-LSTM network achieves MAE of 1.10 for SBP and 0.58 DBP, and RMSE of 1.56 mmHg for SBP and 0.85 mmHg for DBP [15]. Another study attempt something similar by using a CNN-LSTM where the input into the neural network is the difference of the ECG and PPG signal [16]. The CNN layers extract morphological features, while the LSTM extract temporal features. A hybrid CNN-LSTM model developed to utilize both raw PPG and ECG data and physical features, achieves a MAE of 4.43 ± 6.09 mmHg for SBP and being 3.23 ± 4.75 mmHg for DBP [17].

Most of the referenced papers are using some subset of the MIMIC (Medical Information Mart for Intensive Care) databases. They have a varying degree of success and different approaches. It’s important to note that some models are built with the assumption that the ECG and PPG signals are synchronized [18]. If this condition is not met, the features can’t be extracted either manually or automatically from the combination of ECG and PPG signals.

III. MATERIALS AND METHODS

In this section we describe the dataset, data preprocessing procedures, the developed CNN-LSTM model and the used evaluation metrics.

A. Dataset

The dataset used in this study is UCI Machine Learning Repository Cuff-Less Blood Pressure Estimation Dataset [19], [20]. It contains 12000 instances. Each instance contains only ABP, PPG and ECG signals sampled at 125 Hz.

1) PPG: Photoplethysmograph (PPG), is an optically gained plethysmograph that detects the changes in the volume of blood in the microvascular tissue. PPG gives valuable information for the cardiovascular system and it’s a simple, portable, and a low-cost technology [5]. An example of a PPG signal, utilizing the Nurokit2 package is given in Figure 1.

2) ABP: Arterial blood pressure (ABP) is defined as the force that is exerted by the blood on the arterial wall. Blood pressure is connected with the cardiovascular cycle that has two phases: systole, when the heart contracts and
Fig. 1: Visualization of a PPG waveform using the Neurokit2 package

pumps the blood, and the diastole, when the heart relaxes after contraction [3].

3) ECG: Electrocardiogram (ECG) represents a graph of the electrical activity of the heart through time. The impulses coordinate the contractions of different parts of the heart allowing the blood to flow. An example of an ECG signal, utilizing the Neurokit2 package is given in Figure 2.

Fig. 2: Visualization of an ECG waveform using the Neurokit2 package

B. Preprocessing

1) Data Cleaning: The instances with missing data in any of the three signals are removed. The normal value of the ABP in a healthy person is between 40 and 60. An indication of a serious condition, but also a possible invalid signal if ABP value is less than 20, therefore these instances are excluded. ABP value above 100 is indication of many cardiovascular diseases and this big difference between the SBP and DBP is possible. Nevertheless, the instances where this abnormal condition occurs are not taken into consideration.

2) Segmentation: The average length of a signal in this dataset is 27807.5 data points, within the range [1000, 74000]. In this study, segments with 1000 data points at 125 Hz frequency are considered, or 8 seconds in time domain. This is done in order to avoid the exclusion of the instances with signal length of 8 sec. This approach increases the number of samples, but decreases their information gain. To further increase the number of training samples, multiple segments are taken from the longer signals, with no overlap. Each segment contains unique data. Different studies utilize signals with different length. More research needs to be done to find the optimal signal length.

3) Standardization: Standardization is the process of transforming different feature values to be on the same scale. Each signal is transformed to have a mean value 0 and variance value 1.

4) Filtering: Signals can be distorted while recording or transmitting, because of the presence of a noise. To eliminate both types of noise (low and high frequencies), the signals are filtered.

The ABP is not filtered in order to avoid a distortion in the SBP and DBP values. For filtering the ECG and PPG signals, both notch and bandwidth are typically used [21]. In this study the bandwidth filter of order 5 for ECG and the bandwidth filter of order 4 for PPG is used.

5) BP categorization: In order to prepare the labeling of the data, the values for SBP and DBP are calculated using the ABP values by finding the local maxima for SBP and local minima for DBP and averaging them [22]. An illustration of SBP and DBP extraction from ABP waveform is presented in Figure 3.

Fig. 3: Illustration of SBP and DBP extraction from ABP waveform

In this study, two different BP classification problems are explored - data is divided in two BP categories (First Experiment) and three BP categories (Second Experiment).

First experiment

The main goal of this research, as elaborated in the Introduction, is to develop a model to follow the hemostability of an injured patient by detecting the possible inner bleeding. This condition can be detected by a low BP - hypotension. Therefore, the data is divided in two BP categories: Hypotension and Nothypotension, presented in Table I.
The built model is an multivariate CNN-LSTM, an ensemble of two neural networks, one for each variable (PPG and ECG), with the same structure that link to the same output layer:

- First experiment: 2 neurons representing the blood pressure categories (0: Hypotension, 1: Nothypotension).
- Second experiment: 3 neurons representing the blood pressure categories (0: Normal, 1: Prehypertension, 2: Hypertension).

D. Evaluation Metrics

In this section, the used metrics in order to evaluate the models are elaborated. Since we developed classification models, the following evaluation metrics are used: accuracy, precision, recall, f1 score, AUROC values. Also, Support, as well as Macro and Weighted averaging schemas are presented. The schemas average the metrics into a single value to evaluate the performance of the model:

1) Support: The number of samples in the testing dataset labeled as a class.
2) Macro average: Macro average is a schema used in multiclass classification, i.e. calculating corresponding metrics independently and averages the results.
3) Weighted average: Weighted average is schema that averages metrics like precision and recall for all classes in a dataset by assigning weights to each class depending on their importance.

IV. RESULTS AND DISCUSSION

In this section, the results obtained using the aforementioned methodology and the chosen LSTM-CNN model are presented.

A. Model Training and Evaluation

In the scope of this research, many different model structures have been explored. The model presented below was the most successful in the research so far. The network has a batch size of 128 and runs for 20 epochs.

In the presented experiments 85% of the dataset is used to train the model, while the remaining 15% is used for testing, i.e. evaluating the model’s results. A validation dataset containing 10% from the training dataset is used to prevent overfitting. An ADAM optimizer with a learning rate of 0.03 was used to train the models. The learning rate is selected utilizing a naïve approach, by experimenting with different values - it is concluded that this value gives the best loss.

The First Experiment

The overall accuracy of the best model is 0.76. The model evaluation is given in Table III with the class labels: ‘0’ as Hypotension and ‘1’ as Nothypotension. The AUROC value is 0.74. The precision for the Hypotension class is small-scale, partly due to the fact that this is the minority class. However, reasonably good recall and f1 scores rates are achieved. The model takes 147.14s to train and uses 148 MB of RAM.
The BP classification problem. The achieved reasonably good performance in classifying hypertension is presented as a simplified version of the classic machine learning setup. Real-life situations require the classification process to be delivered in real time, as the model needs to function with no regard of the person’s age and gender. The only information that can be used are the morphological characteristics. This study attempts to build a generalized model, a model that’s expected to perform satisfactorily on any patient.

The model is limited by the fact that it works with fixed segments instead of dynamic ones. Blood pressure varies during the span of the day. The changes to BP can happen gradually. Other works have proven that calibrating a model to a patient produces far better results than using a fixed segments. Blood pressure varies during the span of the day. The changes to BP can happen gradually. Other works have proven that calibrating a model to a patient produces far better results than using a fixed segments instead of dynamic ones. Blood pressure varies during the span of the day.

Using an extensive dataset from different sources will help to overcome the pointed obstacles.

V. Conclusion

The obtained results give an excellent starting point to develop a classification model for BP estimation that can be embedded on a compact patch-like biosensor that contains ECG and PPG sensors. Once the model is built, the classification process is delivered in real time, as the real-life situations require.

The First experiment with the two classes is an idea that presents itself as a simplified version of the classic BP classification problem. The achieved reasonably good recall rates (0.71 and 0.78 for the two classes) and the f1 score rates, lead us to believe that further research in this direction would be productive. We also tried a four-class classification (Hypotension, Normal, Prehypertension and Hypertension), but the obtained results were not satisfactory.

There are several reasons that impacted the results in the First and Second experiment - the value of the segment size (8 sec.) might be too short of an interval to accurately deduce the blood pressure. Further study focusing on this aspect is required. Still, the model can distinguish between the different categories and training the same model on a larger dataset can improve the results.

The model is limited by the fact that it works with fixed segments instead of dynamic ones. Blood pressure varies during the span of the day. The changes to BP can happen gradually. Other works have proven that calibrating a model to a patient produces far better results because it allows the model to learn the patient specific morphological characteristics. This study attempts to build a generalized model, a model that’s expected to perform reasonably well on any patient.

Using an extensive dataset from different sources will help to overcome the pointed obstacles.

B. Discussion

The obtained results give an excellent starting point to develop a classification model for BP estimation that can be embedded on a compact patch-like biosensor that contains ECG and PPG sensors. Once the model is built, the classification process is delivered in real time, as the real-life situations require.

The First experiment with the two classes is an idea that presents itself as a simplified version of the classic BP classification problem. The achieved reasonably good recall rates (0.71 and 0.78 for the two classes) and the f1 score rates, lead us to believe that further research in this direction would be productive. We also tried a four-class classification (Hypotension, Normal, Prehypertension and Hypertension), but the obtained results were not satisfactory.

There are several reasons that impacted the results in the First and Second experiment - the value of the segment size (8 sec.) might be too short of an interval to accurately deduce the blood pressure. Further study focusing on this aspect is required. Still, the model can distinguish between the different categories and training the same model on a larger dataset can improve the results.

The model is limited by the fact that it works with fixed segments instead of dynamic ones. Blood pressure varies during the span of the day. The changes to BP can happen gradually. Other works have proven that calibrating a model to a patient produces far better results because it allows the model to learn the patient specific morphological characteristics. This study attempts to build a generalized model, a model that’s expected to perform reasonably well on any patient.

Using an extensive dataset from different sources will help to overcome the pointed obstacles.

### Table III: Results - First experiment

<table>
<thead>
<tr>
<th></th>
<th>precision</th>
<th>recall</th>
<th>f1-score</th>
<th>support</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.Hypotenion</td>
<td>0.86</td>
<td>0.71</td>
<td>0.80</td>
<td>1355</td>
</tr>
<tr>
<td>1.Normal</td>
<td>0.91</td>
<td>0.78</td>
<td>0.84</td>
<td>5240</td>
</tr>
<tr>
<td>accuracy</td>
<td></td>
<td></td>
<td>0.76</td>
<td>6635</td>
</tr>
<tr>
<td>macro avg</td>
<td>0.69</td>
<td>0.75</td>
<td>0.70</td>
<td>6635</td>
</tr>
<tr>
<td>weighted avg</td>
<td>0.82</td>
<td>0.76</td>
<td>0.78</td>
<td>6635</td>
</tr>
</tbody>
</table>

In Table IV the confusion matrix from the First experiment is presented.

### Table IV: Confusion matrix - First experiment

<table>
<thead>
<tr>
<th></th>
<th>Hypotension</th>
<th>Nothypotension</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>996</td>
<td>1165</td>
</tr>
<tr>
<td>Predicted labels</td>
<td>399</td>
<td>4075</td>
</tr>
</tbody>
</table>

### Table V: Results - Second experiment

<table>
<thead>
<tr>
<th></th>
<th>precision</th>
<th>recall</th>
<th>f1-score</th>
<th>support</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.Normal</td>
<td>0.78</td>
<td>0.70</td>
<td>0.74</td>
<td>2461</td>
</tr>
<tr>
<td>1.Prehypertenion</td>
<td>0.60</td>
<td>0.68</td>
<td>0.64</td>
<td>2605</td>
</tr>
<tr>
<td>2.Hypertension</td>
<td>0.61</td>
<td>0.57</td>
<td>0.59</td>
<td>1521</td>
</tr>
<tr>
<td>accuracy</td>
<td></td>
<td></td>
<td>0.66</td>
<td>6635</td>
</tr>
<tr>
<td>macro avg</td>
<td>0.66</td>
<td>0.65</td>
<td>0.65</td>
<td>6635</td>
</tr>
<tr>
<td>weighted avg</td>
<td>0.67</td>
<td>0.66</td>
<td>0.66</td>
<td>6635</td>
</tr>
</tbody>
</table>

In Table VI the multiclass confusion matrix from this experiment is presented.

### Table VI: Confusion matrix - Second experiment

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Prehypertenion</th>
<th>Hypertension</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>1717</td>
<td>674</td>
<td>99</td>
</tr>
<tr>
<td>1.Prehypertenion</td>
<td>393</td>
<td>1806</td>
<td>454</td>
</tr>
<tr>
<td>2.Hypertension</td>
<td>90</td>
<td>569</td>
<td>862</td>
</tr>
</tbody>
</table>
(f1=0.74; precision=0.78; recall 0.70), but the recognition of the other two classes was not successful.

The summarised results from the first experiment indicate that the two-class classification including the hypertension category should be considered as a solution for the fast BP categorization during the triage process. We believe that this simplified BP classification is more suitable for the patch-like device, since it can help in the detection of the sudden BP drop as the first sign of possible internal bleeding.

Future work includes adjusting the segment size making sure that segments from the same signal are all either in the train or in the test set. Also, we intend refining the results by developing regression models for BP estimation with deep learning approach, using Big Data.

ACKNOWLEDGMENT

This paper has been written thanks to the support of the “Smart Patch for Life Support Systems” - NATO project G5825 SP4LIFE and by the National project IBS4LIFE “Smart Patch for Life Support Systems” - NATO project results by developing regression models for BP estimation.

REFERENCES


