A Sparse Regression based Approach for Cuff-less Blood Pressure Measurement

Monika Jain  
IIIT-Delhi, ECE  
New Delhi, India  
monikaj@iiitd.ac.in

Niranjan Kumar  
IIIT-Delhi, ECE  
New Delhi, India  
niranjank@iiitd.ac.in

Sujay Deb  
IIIT-Delhi, ECE  
New Delhi, India  
sdeb@iiitd.ac.in

Angshul Majumdar  
IIIT-Delhi, ECE  
New Delhi, India  
angshul@iiitd.ac.in

Abstract— This paper proposes a sparse regression based approach for accurate continuous Blood Pressure (BP) monitoring. ECG and Finger PPG signals serve as the input from which 32 parameters are extracted. Not all parameters are indicative of BP; to automatically trim the redundant parameters a sparse regression based approach is proposed. To build the BP predicting model the necessary parameters and their corresponding weights are learned using data from 99 subjects. The learned model is applied on 10 test subjects. The ground truth BP is measured using a clinically proven professional automatic digital BP monitor OMRON HBP1300. The BP prediction results show that the SBP/DBP mean absolute error and error standard deviation, with OMRON monitor as a reference, is 4.43/2.46 and 4.90/3.31 mmHg respectively, which falls under the standard allowable error mentioned by Association for the Advancement of Medical Instrumentation for estimation of BP. We have compared our work with other BP prediction techniques (Linear Regression and Feed Forward Neural Network) and have seen that our proposed method yields considerably better results, especially for diastolic BP.

Keywords— ECG, PPG, BP, Sparse Regression

I. INTRODUCTION

Easy and affordable health monitoring has become major concern to sustain the type of lifestyle we have today. Despite the always improving medical facilities, a large section of the populace remains oblivious to their health. A possible way to avoid this can be a compact, affordable and continuous health monitoring device which could be owned and accessed by all. According to an estimate made by WHO, by the year 2025, about 1.56 billion adults would be living with cardiovascular diseases, which would certainly lead to premature deaths.

Amongst many physiological parameters, Blood Pressure (BP) is an important biomarker for keeping a check on cardiovascular ailments. There is a race between a number of products and processes for measuring the BP with non-invasive and cost effective approaches. However, the accuracy and robustness of these devices is questionable and improving them is an active area of research. Some very popular approaches for BP prediction are based on furnishing its relationship with parameters extracted from various cardiac signals. Few widely used cardiac signals are ECG, PPG, ICG and PCG. Previous research has only extracted a few parameters from the ECG and PPG signals. They tried explaining the BP based on these.

This work proposes to extract a large number of parameters from ECG and PPG signals. Obviously not all the parameters are useful for explaining the BP. Instead of employing domain knowledge to select the parameters, we follow a data driven approach. BP prediction is formulated as a regression problem. In the learning phase, the few explanatory parameters are automatically selected so as to explain the BP measurements. Selection of a few parameters is enforced via sparse regression. During test or operational phase, the selected parameters and their corresponding regression weights are used to predict the BP of new individuals. The results are also compared with linear regression and feed forward neural network (FFNN) using Resilient Back Propagation training (RBP) algorithm [1, 2]. The comparative study showed that our proposed method is more promising than other approaches.

II. LITERATURE REVIEW

A significant amount of work has been done on oscillometric BP estimation. A promising approach for measuring the BP by using the envelope of the oscillometric pulse was proposed in [1]; this method was out-performed by exploring other parameters extracted from the oscillometric pulse [2, 3]. It is has been empirically proven that BP depends greatly on pulse transit time (PTT) [4, 5, 6, 7, 8]. Many researchers have also suggested a frequency dependent model for estimation of SBP using PTT [9]. However, there is a considerable amount of work which have shown that BP can be calculated from other parameters too [2, 7, 10]. Estimation of BP parameters such pulse ejection period (PEP), pulse arrival time (PAT) and pulse wave velocity (PWV) has also been deemed as a promising approach [2, 11].

Maximum Amplitude Algorithm (MAA) is the most commonly mentioned and used technique in the literature [12]. Apart from MAA, many machine learning based techniques such as adaptive neuro-fuzzy inference system (ANFIS), principle component analysis [13], artificial neural network [1] and Gaussian mixture model [14] have been employed over the years for improving BP prediction. Many studies concluded that RBP algorithm is more efficient than any other Neural Network (NN) training algorithm for BP.
Read the ECG and PPG signals, Remove the DC component, normalize the signals and Set a threshold. Find: all peaks lying above the threshold, which gives R-peaks (R). Find: minima lying between R and R_{i-1} which gives S-peaks (S). Find: 2nd maxima lying between S and S_{i-1}, which gives T-peaks (T). Find: 3rd maxima lying between S and S_{i-1}, which gives P-peaks (P). Find: minima lying between P and R_{i-1}, which gives Q-peaks (Q). Find: mid-point lying between S and T, which gives Point-A (A). Find: minima lying between T and P_{i-1}, which gives Point-C (C). Find: mid-point lying between T and C_{i-1}, which gives Point-B (B). Find: minima of PPG pulse in the time interval between t[R_{i}] and t[R_{i-1}], which gives PPG_foot_{i}. Find: maxima of PPG pulse in the time interval between t[R_{i}] and t[R_{i-1}], which gives PPG_peak_{i}. Find: zero slope point between PPG peak_{i} and PPG_foot_{i}, which gives dicrotic notch_{i}. Calculate the parameters as given in Table I. Stop

**TABLE I** PARAMETERS EXTRACTED FROM ECG AND PPG SIGNALS

<table>
<thead>
<tr>
<th>Source</th>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
</table>
| **ECG** and **PPG** | PAT1* | interval between R-peak and PPG-peak^*
| | PAT2# | interval between R-peak and PPG-foot^#
| | PTT1 | PAT1-PEP^*
| | PTT2 # | PAT2-PEP^*
| | PEP | Interval between Q-peak and point B^*
| | Heart rate | No. of R-peaks in 10 sec x 6
| | LVET*# | interval between point A and point B^*
| | EMS* | PEP+LVET^*

**AND**

<table>
<thead>
<tr>
<th>Source</th>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PPG</strong></td>
<td>X^<em>, Y^</em>, Z^</td>
<td>height of diastolic peak, dicrotic notch and systolic peak respectively [15]</td>
</tr>
</tbody>
</table>

**Mean PPG Peak Distance^*#**

\[
\sum_{i=1}^{n} (T_{S_{i+1}} - T_{R_{i}})/D
\]

**Mean PPG Foot Distance^*#**

\[
\sum_{i=1}^{n} (T_{S_{i+1}} - T_{A_{i}})/D
\]

**X^*, Y^*, Z^**

height of diastolic peak, dicrotic notch and systolic peak respectively

In this study, the peak detection algorithm shown in Fig. 1 was used to detect the significant peaks of the ECG and PPG signals. This algorithm detects P, Q, R, S, T peaks and points A, B and C in the ECG signal and Foot, Peaks and Dicrotic notches in the PPG signal, which are here denoted as P_i, Q_i, R_i, S_i, T_i, A_i, B_i, C_i, PPG_peak_{i}, PPG_foot_{i} and dicrotic_notch_{i}. The subscript i = \{1, 2, ..., D\}, where ‘D’ denotes the number or cardiac cycles in a given time period.

In past few decades, a number of BP estimation techniques have been introduced. These studies were able to draw the correlation between BP and a number of these parameters, listed in Table I, were computed using these detected peaks. For calculating these parameters, First and last cardiac cycle recorded was discarded. This is equivalent to discarding R_1, S_1, T_1, A_1, B_1, C_1, PPG_peak_{1}, PPG_foot_{1} and dicrotic_notch_{1} (first cardiac cycle) and P_D, Q_D, R_D (last cardiac cycle). A pictorial view for calculating the parameters (in Table I) can be obtained from Fig. 2. In Table I, one can notice that most of parameters, on which DBP depend, are extracted from PPG signal. This is because arterial stiffness is more related to DBP than SBP, as suggested by [11]. In this study, we will see how these feature leads to a more accurate DBP estimation, which was so far being ignored.

**B. Sparse Regression**

In past few decades, a number of BP estimation techniques have been introduced. These studies were able to draw the correlation between BP and a number of these parameters, listed in Table I, were computed using these detected peaks. For calculating these parameters, First and last cardiac cycle recorded was discarded. This is equivalent to discarding R_1, S_1, T_1, A_1, B_1, C_1, PPG_peak_{1}, PPG_foot_{1} and dicrotic_notch_{1} (first cardiac cycle) and P_D, Q_D, R_D (last cardiac cycle). A pictorial view for calculating the parameters (in Table I) can be obtained from Fig. 2. In Table I, one can notice that most of parameters, on which DBP depend, are extracted from PPG signal. This is because arterial stiffness is more related to DBP than SBP, as suggested by [11]. In this study, we will see how these feature leads to a more accurate DBP estimation, which was so far being ignored.
parameters derived from ECG and PPG signals; these studies either considered too many parameters that were unnecessary for explaining BP or considered too few parameters that did not fully explain the BP. Moreover these studies failed to accurately predict DBP. In this work we propose a simple linear regression based model for BP prediction.

\[y_{\text{N+1}} = A_{\text{N+1}} x_{\text{N+1}} + \eta_{\text{N+1}} \] (1)

where ‘\(y\)’ represents the vector of measured BP, ‘\(A\)’ is the matrix of parameter vectors for each sample, ‘\(x\)’ consists of the corresponding regression and weights and \(\eta\) is the error/noise assumed to be Normally distributed. Here \(N\) is the total number of samples (subjects) and \(m=32\) is the total number of parameters considered.

We know that 32 parameters for predicting BP is an overkill. However, we do not want to manually trim the parameters; we go for a completely data driven approach based on Orthogonal Matching Pursuit (OMP) [18]. This leads to a sparse \(x\). The interpretation of a sparse \(x\) is that the positions in \(x\) having 0’s correspond to parameters that are not useful for predicting BP: only the few non-zero coefficients in \(x\) and the corresponding parameters (columns) in \(A\) are enough to accurately predict the BP.

For a sparse regression vector we would ideally like to solve the \(l_p\)-minimization problem:

\[
\arg \min_{x} \|y - Ax\|_p \quad \text{such that} \quad \|x\|_0 \leq \tau
\] (2)

The \(l_0\)-norm counts the number of non-zero elements in \(x\), minimizing it is NP hard. We follow an approximate greedy solution based on Orthogonal Matching Pursuit (OMP).

IV. EXPERIMENTAL RESULTS

To estimate BP on the basis of the parameters extracted in Section III(A), we used three approaches: 1. Neural Network regression using RBP as training algorithm [1,2]; 2. Linear regression; 3. Orthogonal Matching Pursuit. This section discusses the accuracy of each method in terms of Normalized Mean Square Error (NMSE), Mean Absolute Error (MAE) and Error Standard Deviation (ESD).

A. Data Description

ECG and PPG signals are collected from 109 subjects aged between 20 to 55 years. The entire dataset has been collected from three sources: (1) 67 datasets collected from our data acquisition system- SET A; (2) 20 datasets picked from MIMIC database [19]-SET B; (3) 22 datasets picked from The University of Queensland Vital Sign database [20]-SET C.

SET A was collected in house hence we had full control on the collection procedure. 15 subjects from SET-A were asked to climb 24 set of stairs twice before taking the readings. This was done to take the reading when the subject is undergoing physical exertion. Remaining 52 subjects were asked to relax for 5 minutes before taking the readings. The ground truth BP for the subjects were collected via an Automatic Digital BP OMRON HBP1300, after two minutes of recording the ECG and PPG. The training data consisted of 99 subjects: 57 subjects from SET A and entire SET B and SET C. The remaining 10 subjects were used for testing.

5.2. Results

For a sparse regression vector ‘\(x\)’, the \(l_p\)-minimization problem shown by eq. 2 was solved for \(\tau = \{1, 2,\ldots,32\}\). Every ‘\(x\)’ obtained for different values of \(\tau\) was used to predict the BP 10 test subjects. The NMSE (%), MAE (mmHg) and ESD (mmHg) for each \(\tau\) was calculated against the standard device, OMRON HBP1300 (as shown in Fig. 3).

The result reflected that NMSE, MAE and ESD at \(\tau = 19\) is better than rest of the values of \(\tau\). Error is higher for \(\tau < 19\) while it almost saturates for \(\tau > 19\). This means that BP can be best predicted with only 19 parameters out of 32 that we gave as an input. When the number of parameters are fewer than 19, the error is high owing to the fact that all the parameters required for BP estimation has not been accounted for. For too many parameters (19+) the error increases because they obfuscate the prediction by accounting for irrelevant information. What is surprising it that, these 19 parameters were different for both SBP and DBP, as marked in Table I. We compare our proposed technique with basic linear regression – this has been used profusely for BP prediction. We also compare with a NN based regression model (trained with RBP). In prior studies [1, 2] it was shown that the NN outperforms many other BP prediction techniques.

While using NN based approach (Keeping the test and training data exactly same as that used in OMP), we used two non-linear hidden layers with tangent sigmoid transfer function and an output layer with a linear transfer function. Including hidden layer(s) enables the network to exhibit non-linear behaviour. The input to NN was the 32 parameters which we extracted. During training the network with 99 data sets, 70% of the data was used for training and 15% each was used for validation and testing. The network was then tested over 10 subjects for predicting the unknown BP corresponding to the input parameters. Lowest error using NN based approach was obtained with 40 neurons in the hidden layer. Table II shows the errors obtained using NN based approach, linear regression and OMP.

Two tailed dependent samples t-test [21] was also conducted to verify that there is no statistically significant difference between BP predicted from the used approaches and measured BP. To conduct the t-test, the differences between the predicted and measured BP values should be normally distributed. This condition was found satisfied as the skew and kurtosis levels for every method were lesser than the maximum acceptable range for conducting the t-test (skew < [2.0] and kurtosis < [9.0]), as suggested by [22]. The null hypothesis: “There is no statistically significant difference between the BP predictions made by the used approaches and measured BP, was accepted- for each approach the calculated t-value was lesser than the critical value of ‘t’ and \(p<0.05\) with degree of freedom equal to 9.
Table III shows the 95% confidence interval obtained using t-test. The results in Table II and III can be pronounced as:

1. The SBP/DBP prediction error is high while using NN with RBP as training algorithm. MAE and ESD are greater than the maximum error allowed by the Association for the Advancement of Medical Instrumentation (AAMI) for BP prediction (MAE < 5mmHg and ESD < 8mmHg) [23].
2. The DBP estimation error using linear regression falls off the range allowed by AAMI, while SBP estimation is acceptable. Judging for both SBP and DBP estimation, this method is not acceptable.
3. OMP gives a more accurate estimation for both SBP and DBP. The computed error satisfies the standards set by AAMI [23].

It can be seen that, while using OMP, DBP estimation has better as compared to SBP. The reason being PPG, which is the index of arterial stiffness, is more related to DBP than SBP [11]. We have included a large number of PPG parameters for DBP prediction. So, DBP is being calculated using the parameters it actually depends upon. As a result, the accuracy of DBP estimation has increased compared pre-existing works. This can also be noticed in [26, 27]. Thus we have an approach which works equally well for predicting SBP as well as DBP.

We also compared our final results with few recently published works based on MAE (E2) and ESD (E1), which other researchers obtained at different databases (shown in Table IV). The table shows that our proposed approach is better as compared to others and works equally well for SBP as well as DBP estimation.

![Graph showing error during BP prediction using OMP for each value of τ (Number of parameters used)](image)

**Fig. 3:** Error during BP prediction using OMP for each value of τ (Number of parameters used), with OMRON as standard device

### 1. Conclusion

The paper proposes a new approach for estimating Blood Pressure based on ECG and finger PPG signals. This is not a new problem; many studies have extracted a limited number of parameters from these signals and used variants of regression for learning the predictive model. The choice of the extracted parameters were based on the domain knowledge and the predictive accuracy of such methods were dependent on the choice of the extracted parameters.

In this work we extract a relatively large number of parameters. A data driven approach, based on sparse regression is proposed to select the parameters from the training data. At the same time, the linear predictive model is built. We compare our proposed approach with a Neural Network based regression and linear regression techniques. In this work we show that the previous methods are not suitable for clinical practice since the error rates are outside the allowable limits. Our method outperforms NN on the average and complies with the standard set by Association for the Advancement of Medical Instrumentation.

### Performance Comparison of Proposed Approach with NN Based Approach and Linear Regression

<table>
<thead>
<tr>
<th>Method Used (no. of Parameters used)</th>
<th>Systolic BP</th>
<th>Diastolic BP</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN with RBP (32)</td>
<td>9.14</td>
<td>9.87</td>
</tr>
<tr>
<td>Linear regression (32)</td>
<td>4.47</td>
<td>4.16</td>
</tr>
<tr>
<td>OMP (19)</td>
<td>4.51</td>
<td>4.43</td>
</tr>
</tbody>
</table>

E1: Normalized Mean Square Error (%); E2: Mean Square Error (mmHg); E3: Error Standard Deviation (mmHg)

### 95% Confidence Intervals Obtained Using t-Test

<table>
<thead>
<tr>
<th>Method Used</th>
<th>SBP</th>
<th>DBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN using RBP</td>
<td>(-7.32, -12.03)</td>
<td>(-7.90, -5.08)</td>
</tr>
<tr>
<td>Linear regression</td>
<td>(-6.16, 0.39)</td>
<td>(-6.59, 3.95)</td>
</tr>
<tr>
<td>OMP</td>
<td>(-5.90, 1.22)</td>
<td>(-1.86, 2.89)</td>
</tr>
</tbody>
</table>

### Comparison with Recently Published Works

<table>
<thead>
<tr>
<th>Authors</th>
<th>SBP</th>
<th>DBP</th>
<th>Method Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choudhury et al. 2014 [26]</td>
<td>0.78</td>
<td>13.1</td>
<td>Windkessel model</td>
</tr>
<tr>
<td>Ahmed et al., 2015 [27]</td>
<td>6.00</td>
<td>3.90</td>
<td>Predicted BP using PPG</td>
</tr>
<tr>
<td>Proposed</td>
<td>3.98</td>
<td>4.83</td>
<td>OMP</td>
</tr>
</tbody>
</table>

### II. Acknowledgement

Authors would also like to acknowledge the support provided by ITRA project, funded by DEITY, Government of India, under grant with Ref. No. ITRA/15(57)/Mobile/HumanSense/01.

### References


