Title

ECG data-based Blood Pressure Estimation

Authors' and Mentor's Names

1. **Authors**: Rishav Kumar (B19ME066) and Om Solanki (B19CSE061)

2. **Mentor:** Dr Binod Kumar

Abstract

Blood pressure (BP) is the outcome of the mechanical activity of the heart, whereas the Electrocardiograph (ECG) signal shows the electrical activity of the heart. Mechano-Electric Coupling is a word that has been used in earlier research to describe the link between the mechanical and electrical processes of the heart. There is a novel approach for estimating blood pressure that just requires the ECG signal. It is suggested to use machine learning (ML) approaches to estimate blood pressure from electrocardiograph data on devices with limited resources, such as wearables. In order to estimate BP, Adaptive Boosting Regression (AdbootsR) and ANN regression are used in this study. The feature vectors are created using samples of the ECG data over a specific time period. The findings of this investigation support the nonlinear correlation between BP and the ECG signal.

Introduction

Blood pressure (BP) detection is an important aspect of human health evaluation and preventative lifestyle adaptation. The assessment of blood pressure as a routine clinical activity can point to a variety of disorders, including heart failure. An Electrocardiograph (ECG) signal reflects the electrical activity of the heart; it visually portrays the difference in electrical potential across different measurement locations in the body throughout a cardiac cycle. The ECG waveform illustrates and provides data on the electrical activity of the heart, whereas the BP waveform provides information about the mechanical activity of the heart. To understand the activity of the heart and the formation of blood pressure as a result of its function, two independent and interacting viewpoints of electrical and mechanical activity should be investigated. Cardiology researchers use the term Mechano-Electric Coupling (MEC) to describe how the electrical activity of the heart is linked to its mechanical action. This estimation may be accomplished by a series of procedures that include the extraction of significant parameters/features from data samples acquired using diagnostic information such as electrocardiography. ECG data processing is a non-invasive and low-cost method for cardiac monitoring and hence useful for blood pressure estimation using feature extraction and subsequent analysis to search for situations like hypertension (high BP).

With the introduction of wearable ECG monitoring devices, it is vital to equip them with a particular processing power to enable in-situ data analysis for rapid BP calculation. For that, we can use machine learning algorithms. The suggested approach just requires ECG data that can be obtained non-invasively. BP signals are periodic in nature, according to the frequency of heart beat. This signal has two bounds: the upper bound of BP is defined as Systolic Blood Pressure (SBP), and the lower bound is defined as Diastolic Blood Pressure (DBP). Aside from SBP and DBP, the third measure is Mean Arterial Pressure (MAP), which is the average blood pressure during a cardiac cycle.

We explored various strategies for determining blood pressure from ECG signals, evaluating their accuracy and attempting to determine which approach produced the best results. The BP estimate and subsequent categorization into three categories (normal, prehypertension, and hypertension) may be conducted using automated end-to-end analysis of ECG data as shown in the below table.

Category	**SBP(mm Hg)**	**Operator**	**DBP(mm Hg)**
Normal	<=90 >=90 and <120	OR AND	<=60 >=60 and <79
Prehypertension	>=120 and <140	OR	>=80 and <90
Hypertension	>= 140 and <160 >=160 >=140 >=180	OR OR AND OR	>=90 and <100 >=100 <90 >=110

Table 1

Design Problem Formulation

Task 1: Our prime objective in this project is to estimate BP (SBP & DBP) values based on ECG data. We must classify the blood pressure reading into three groups (normal, prehypertension, and hypertension) using the SBP and DBP values shown in the table above.

Task 2: Another team has worked on the classification task, which entails classifying data into normal, prehypertension, and hypertension classes using ML. Assuming this team achieves 100% accuracy, we must expand on this work and develop individual models for the three classes in order to predict blood pressure. The SBP and DBP values of the patient must be passed via the classification model if it has predicted that they fall into a particular class.

Targeted Deliverables:

1. ML-based methodology for BP prediction
2. Documentation and report

Possible Methods to solve the problem

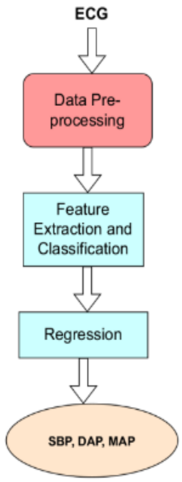
First, we extract features from the ECG data. Based on prior research, a novel data pre-processing strategy is recommended; we will go into more depth about this approach later.

Method 1: The model is trained using an Adaboost regressor and an Artificial Neural Network (ANN)-based regressor on the pre-processed data. Additionally, there are two ways of doing this task:

- **a)** After preprocessing the data and obtaining a dataset with each sample containing four R-peaks, the values are directly used for training.
- **b)** After preprocessing the data and obtaining a dataset with four R-peak values for each sample, we will eliminate any datasets with irregularly spaced consecutive peak values. We have opted for an uncertainty of +- 10%. From there, we obtained a dataset that was used for training and had a consistent distance between R-peak values.

Method 2: A classifier model is used on the pre-processed data to identify which of three categories it belongs to (normal, prehypertension, or hypertension). SBP, DBP, and MAP values are lastly determined using an ANN-based regressor or an Adaboost regressor.

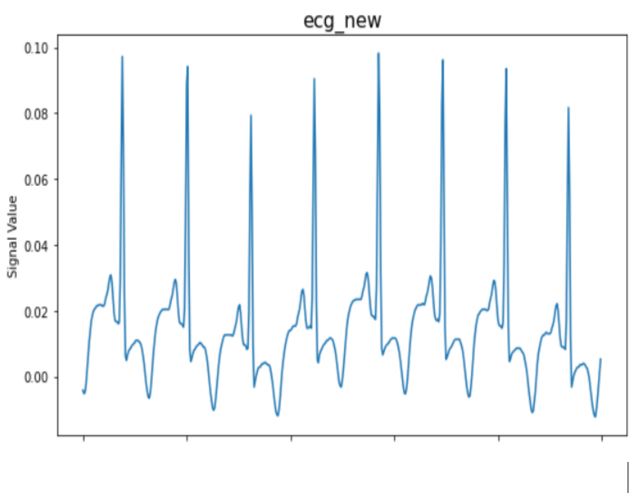
The details of the proposed methods are addressed in the next section.



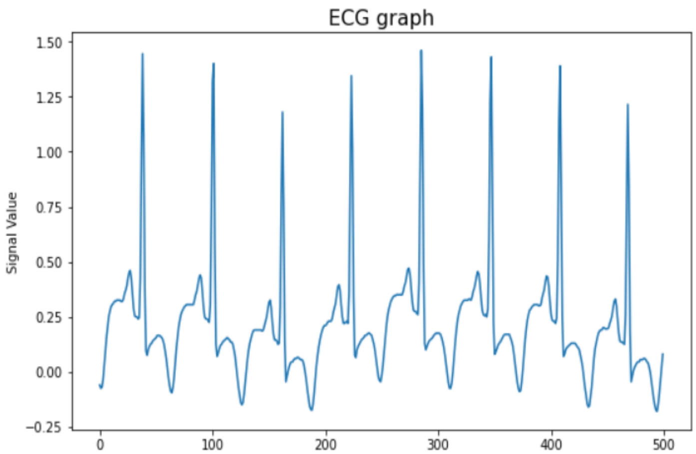
Methodology adopted for the project

Chosen Dataset: The dataset utilized in this work is an open-source version of the Multi-Parameter Intelligent Monitoring in Intensive Care (MIMIC-II) waveform data-set from Physionet. Thousands of signals from 942 ICU patients hospitalized between 2001 and 2008 are included in this edition. These transmissions have a 125 Hz sampling rate. This data set's synchronous Aorta Blood Pressure (ABP), ECG, and Photoplethysmograph (PPG) signals have been smoothed using a basic averaging filter, and any signals with jarring discontinuities or unacceptably high heart rates have been eliminated. Along with the signal blocks that displayed a significant change, those with inappropriate human blood pressure readings were also removed. It is important to note that the data source does not include any extra patient data, such as sex, age, or other demographics.

Data Preprocessing: The signals are then divided into samples of 4 seconds each, resulting in an array of length 500 for each sample. The greatest value is used as the SBP and the minimum value is used as the DBP for the ABP in each segment. A second order band-pass Butterworth filter with corner frequencies of 0.1 Hz and 50 Hz is used to filter the four-second ECG data. Any high frequency noise that may have gotten through owing to the power frequency being lessened as a result. To ensure that the signal does not lose any information, the sampling rate is at least 50 Hz.



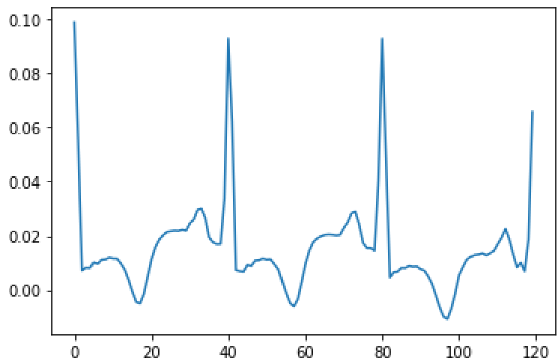
Filtered ECG 4 second signal

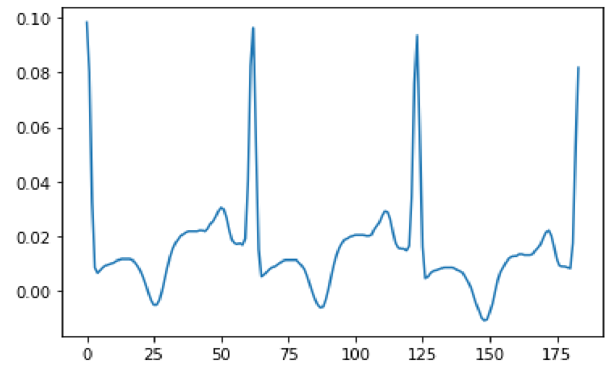


Original ECG 4 second signal

The R-peak value is determined using the built-in library neuronkit2 after the signal has been filtered. The 4-second signals with less than four R-peaks are eliminated, and these R-peak locations are kept. Additionally, the signals with discontinuities are removed.

Due to the clipping of the signal segments, each segment now comprises of the ECG signal in between four consecutive R-peaks. Since the new signal array we got now has a length that can be anywhere between 117 and 450, we must decide on a definite length for the array; in this example, we have chosen the signal length to be 120 samples. We must resample each segment to a length of 120 by obtaining the Fast Fourier Transform (FFT), since some samples of the four R-peak signal are larger than that and others are smaller.





After resampling **Before resampling**

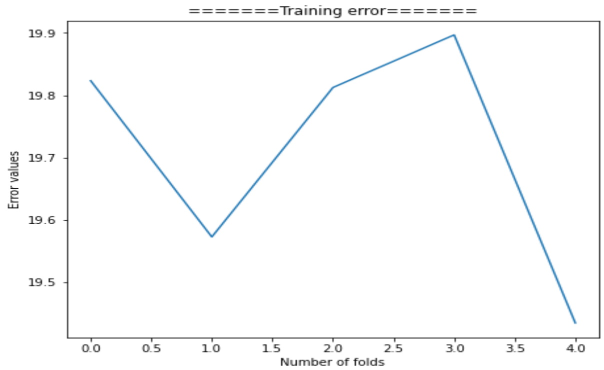
Method 1:

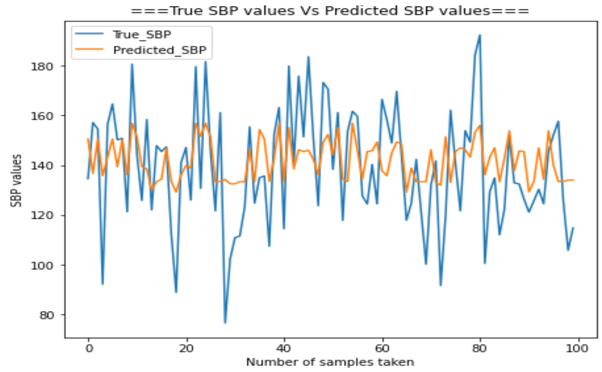
a) Using an Adaboost regressor: In this approach, the data is directtly trained using the adaboost regression algorithm. First, a train-test split with a test size of 0.33 is used to separate the dataset into training and validation sets. For both SBP and DBP, the model is receiving individualised training. Additionally, K fold cross validation is used to identify problems like overfitting or selection bias and to provide information on how the model will generalise to an independent dataset. Here, K=5 has been selected as the value. Furthermore, we determined the Root Mean Squared error (RMSE) at each fold as well as the average RMSE value. The Mean Squared Error (MSE) value has also been determined in addition to the RMSE.

For SBP:

```
Average RMSE over 5 folds: 19.70789982236872
```

```
MSE :15.811903849515 23
```

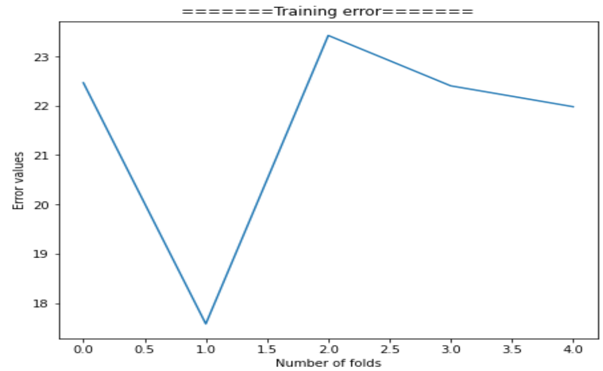


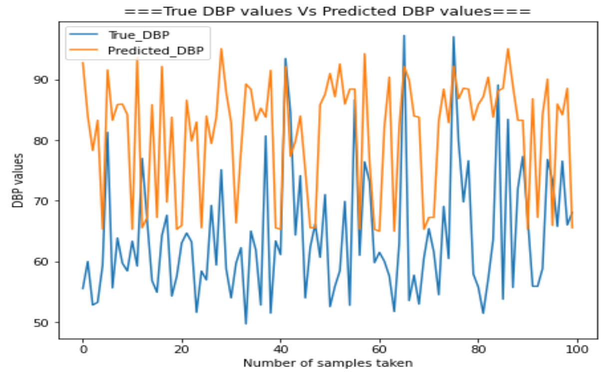


For DBP:

```
Average RMSE over 5 folds: 21.570187087585726
```

```
MSE :19.299382262678787
```

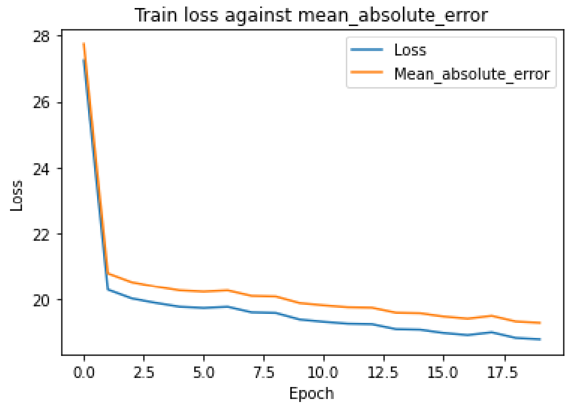


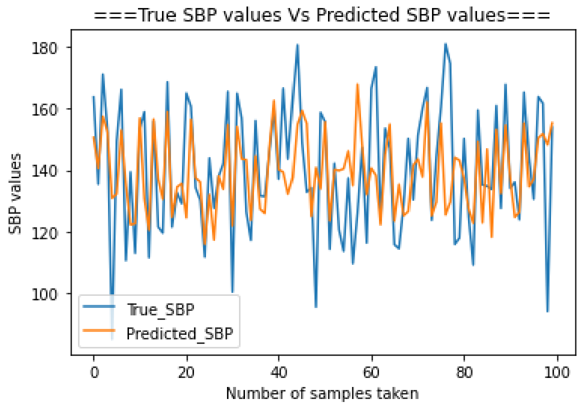


Using an ANN-based regressor: In this approach, an ANN-based regression algorithm is used to directly train the dataset. First, a train-test split with a test size of 0.33 is used to separate the dataset into training and validation sets. For both SBP and DBP, the model is receiving individualised training. The model used is a sequential model with three hidden layers made up of 1024, 512, and 64 neurons, respectively. In order to prevent overfitting, dropout has also been applied at each layer. Each layer uses a different dropout of 0.5, 0.5, and 0.25. With a learning rate of 0.01, we employed an activation function for the hidden layer of the Relu and Adam Optimizer. Since we only predict the SBP and DBP values, the output layer only contains one neuron. 100 epochs are used to train the model. Additionally, we determined the neural net RMSE value and plotted the loss along with the mean absolute loss.

For SBP:

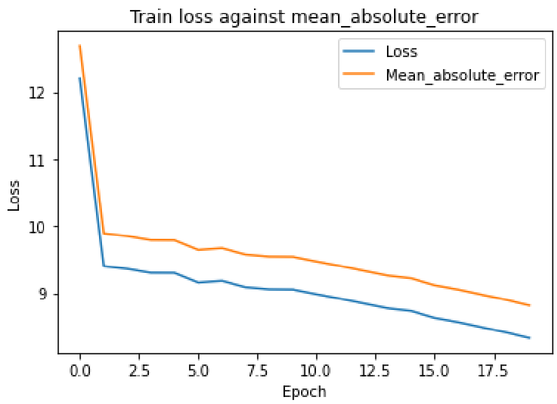
```
Neural Net RMSE: 15.186525424780317
```

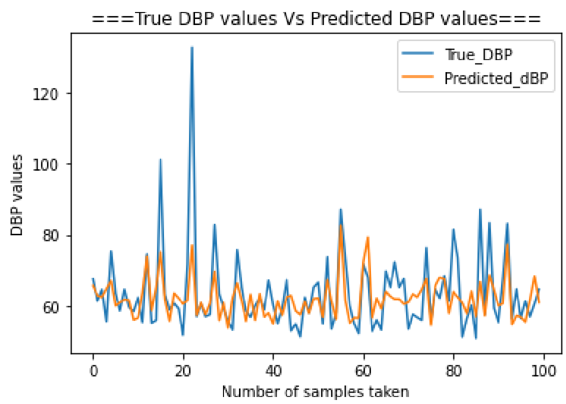




For DBP:

```
Neural Net RMSE: 9.09528037081269
```





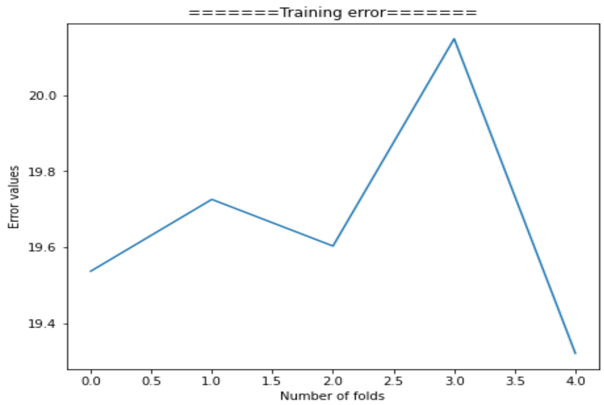
b) In this approach, we also train both the Adaboost and ANN models using the same procedure and parameters, but we also perform additional data preprocessing before training the models. After gathering the dataset with four R-peak values, we additionally examined the separation between adjacent R peak values. Additionally, we estimated the typical separation between two R-peaks and compared their separation. We deleted a data point if the distance (dist) between them is more than +-10% and does not fall within the range (dist - 0.1*dist, dist + 0.1*dist).

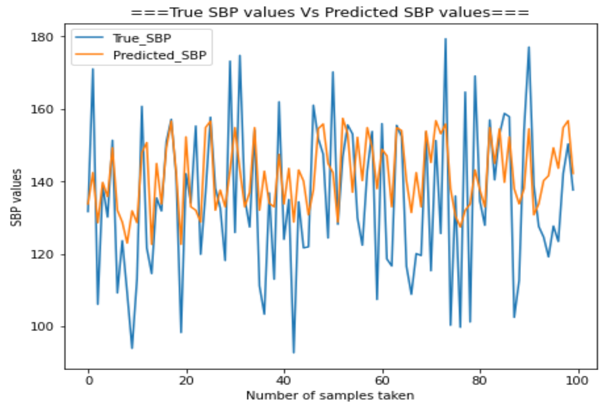
Using an Adaboost regressor:

For SBP:

`Average RMSE over 5 folds: 19.666706219209686`

`Mean Absolute Error: 15.850666406829212`

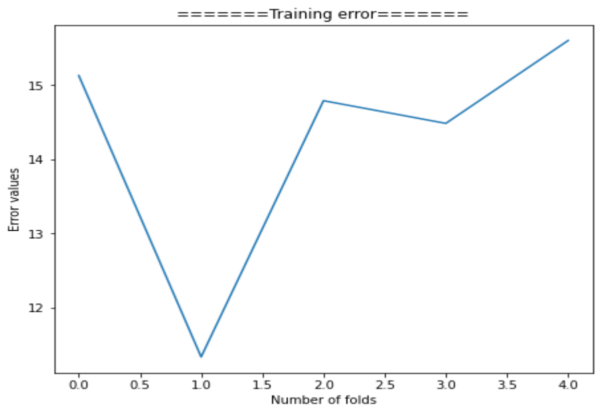


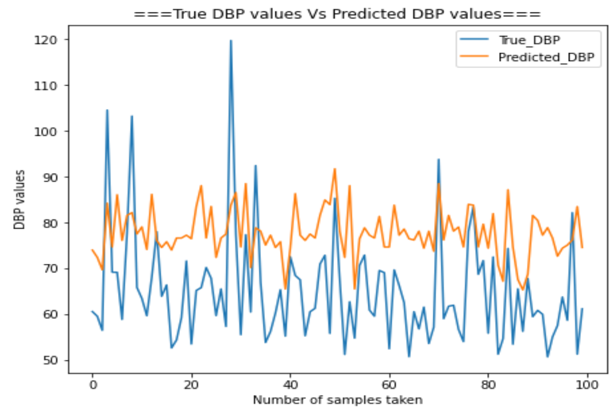


For DBP:

`Average RMSE over 5 folds: 14.26789541173926`

`Mean Absolute Error: 13.70891259352805`

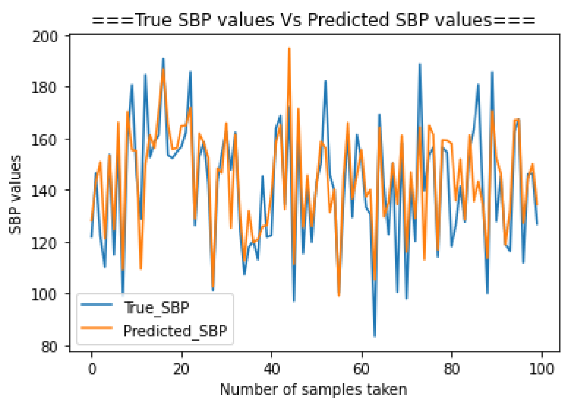


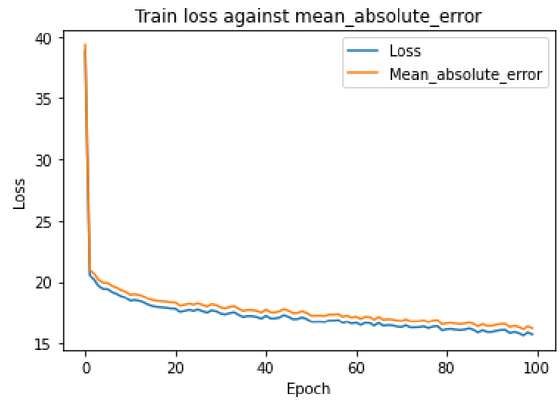


Using an ANN-based regressor:

For SBP:

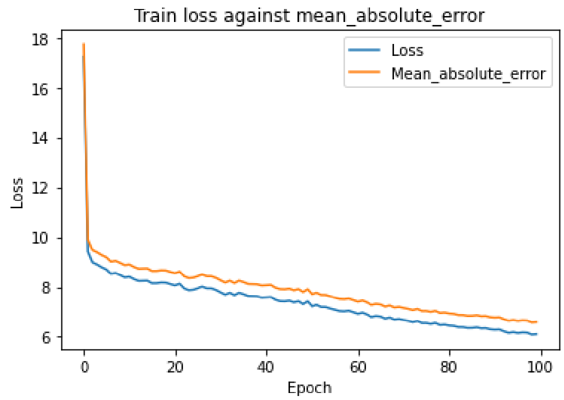
```
Neural Net RMSE: 12.757742688569234
```

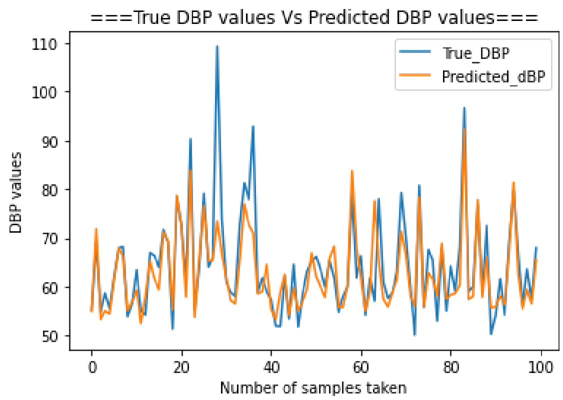




For DBP:

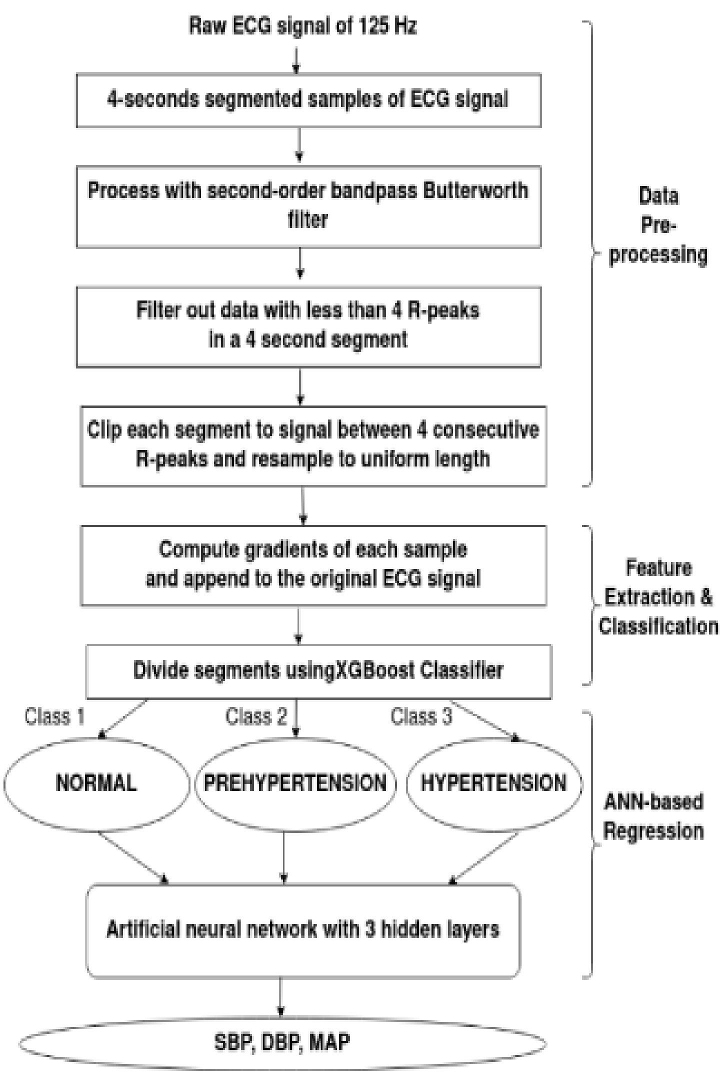
```
Neural Net RMSE: 6.338174902260865
```





Method 2:

With this approach, the model is trained separately for three classes for both SBP and DBP. The classification problem, which involves applying ML to divide data into classifications for normal, prehypertension, and hypertension, has been worked on by another team. We must build on this work and create unique models for the three classes in order to predict blood pressure, assuming this team achieves 100% accuracy. If the classification model predicts that the patient belongs to a specific class, the SBP and DBP readings of the patient must be passed through it. So, using Table 1, we categorise the data after the preprocessing work. Then, we obtained three different datasets. Individually, using the same settings as before, we apply the Adaboost and ANN models to this dataset.

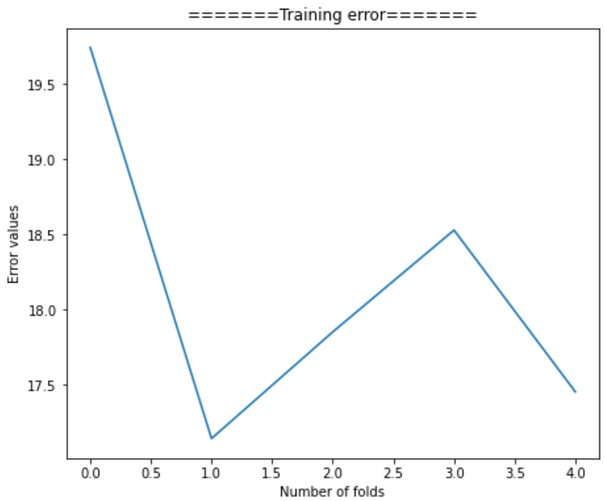


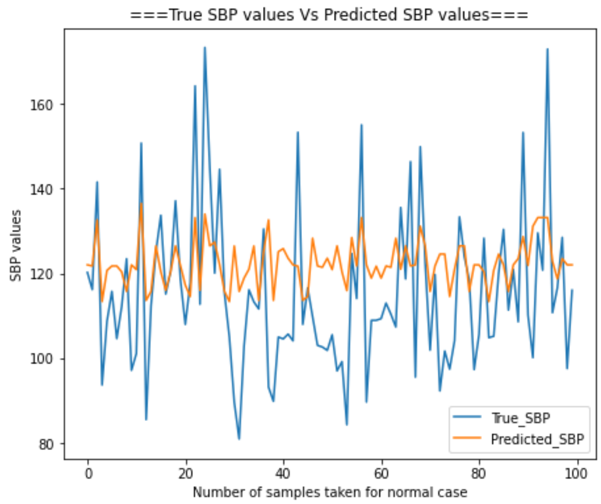
Using an Adaboost regressor:

Class 1 (Normal Category):

For SBP:

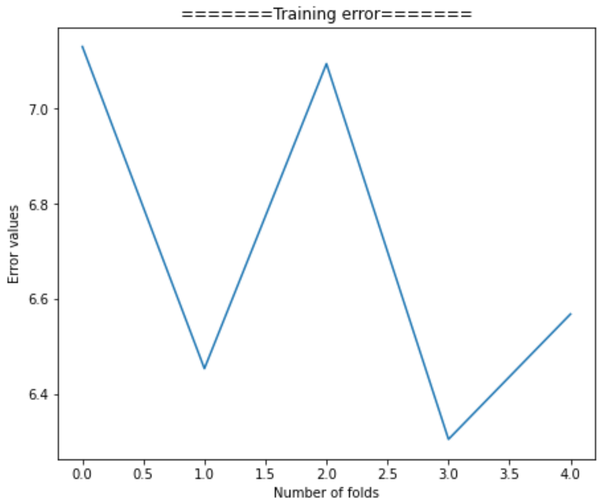
```
Average RMSE over 5 folds: 18.144848787869638
```

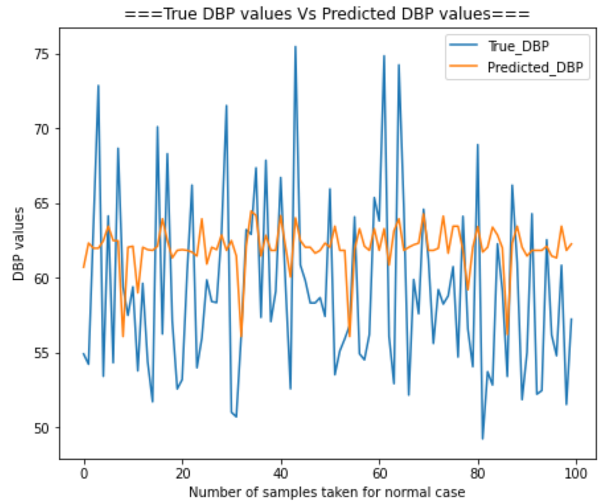




For DBP:

```
Average RMSE over 5 folds: 6.71007001446061
```

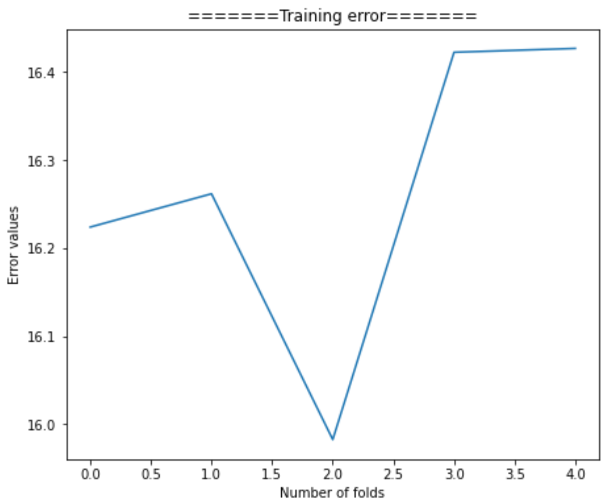


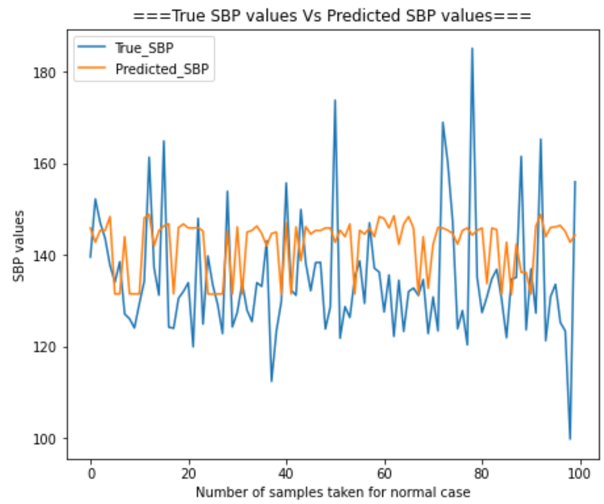


Class 2 (Prehypertension Category):

For SBP:

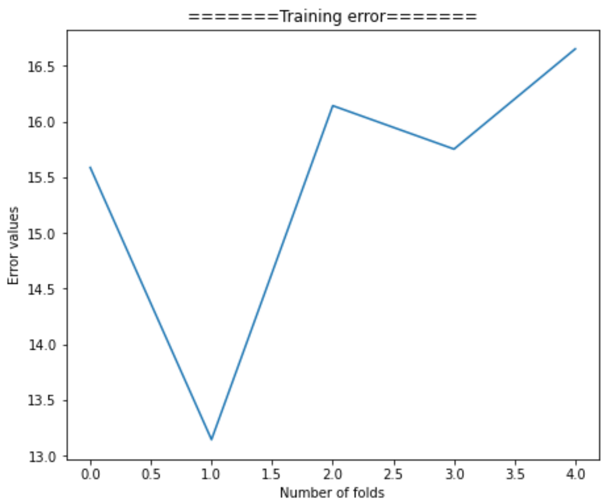
```
Average RMSE over 5 folds: 16.263392699571902
```

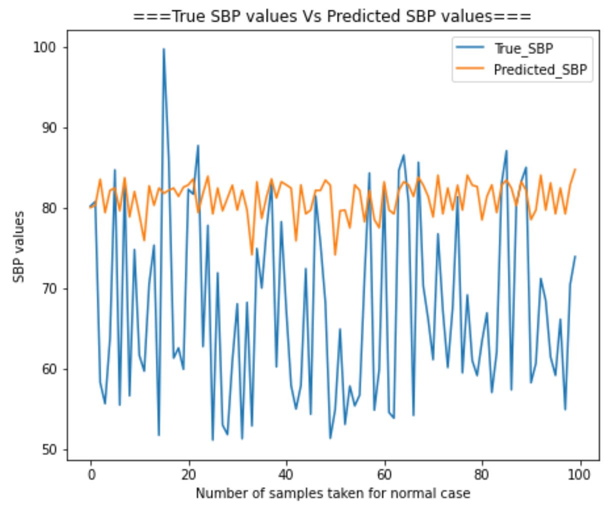




For DBP:

```
Average RMSE over 5 folds: 15.455149841896423
```

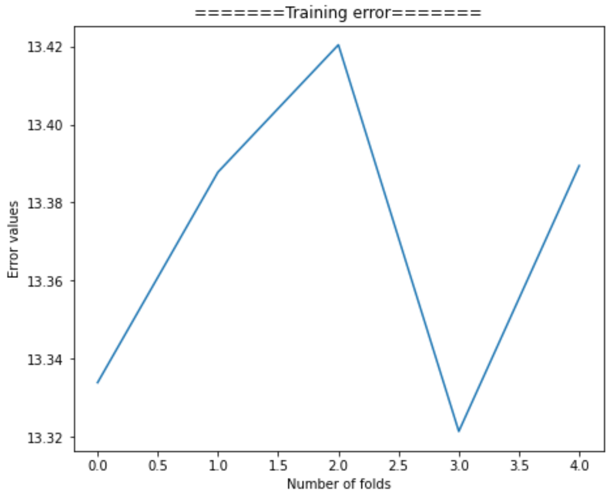


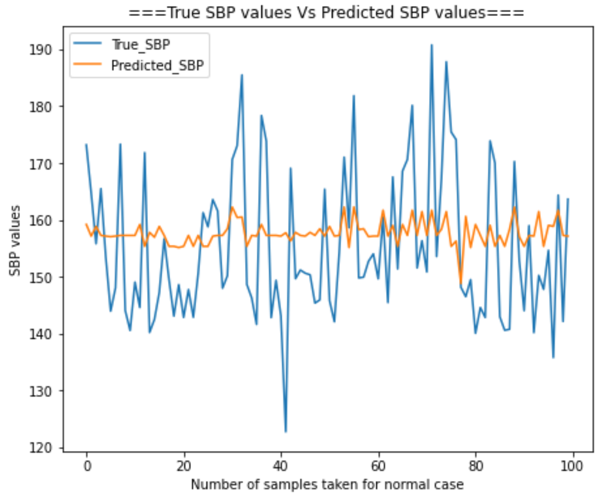


Class-3 (Hypertension Category) :

For SBP:

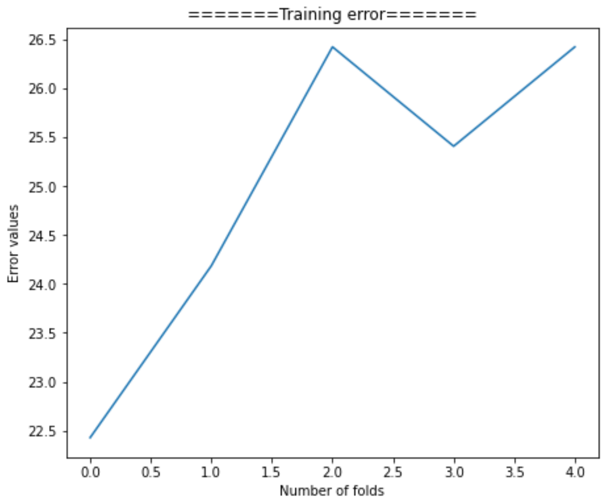
```
Average RMSE over 5 folds: 13.370559769713472
```

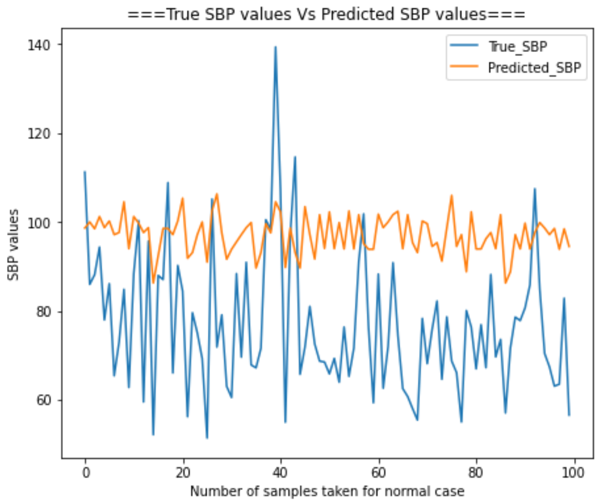




For DBP:

```
Average RMSE over 5 folds: 24.97219942291945
```



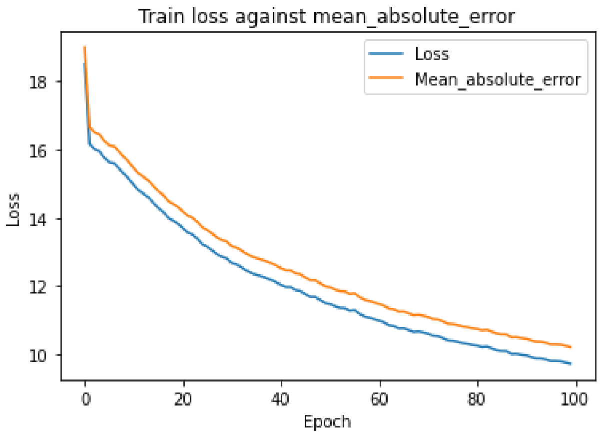


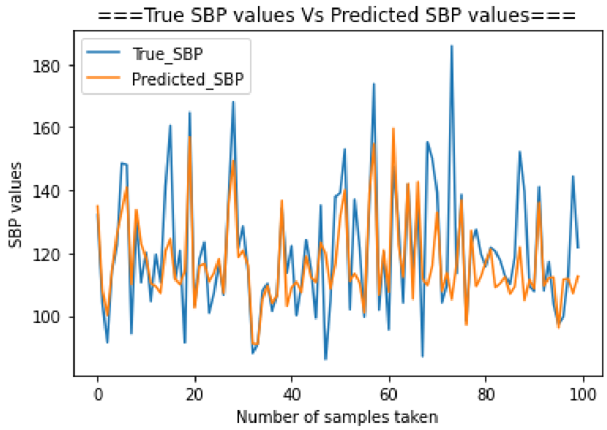
Using an ANN-based regressor:

Class-1 (Normal Category):

For SBP:

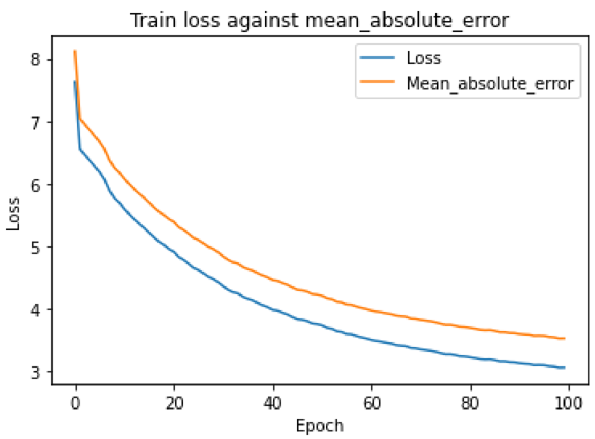
```
Neural Net RMSE: 12.465272120081679
```

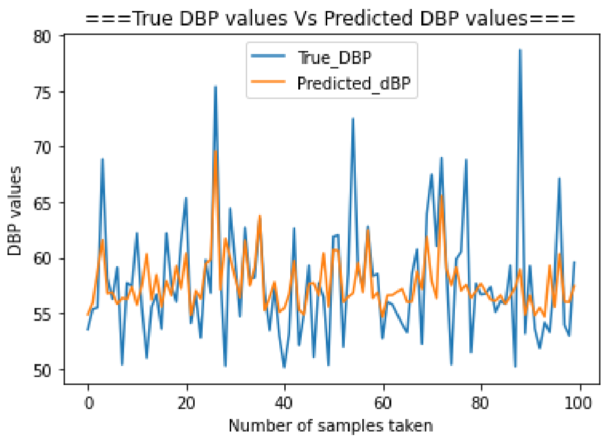




For DBP:

`Neural Net RMSE: 4.676271347181978`

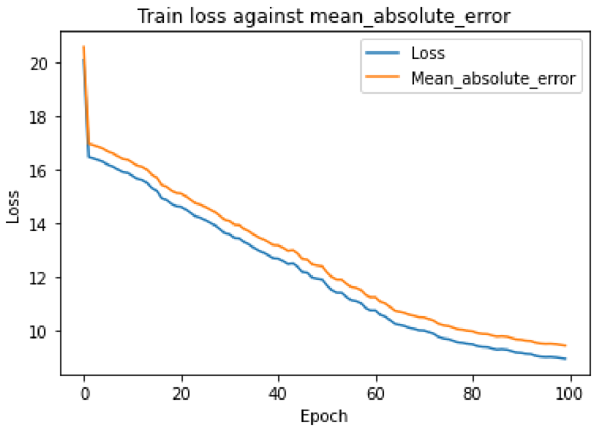


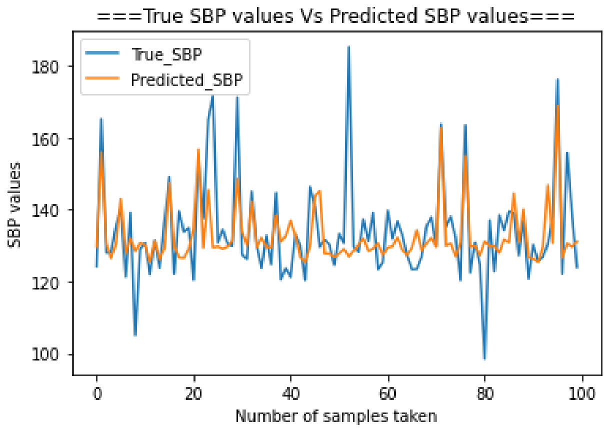


Class-2 (Prehypertension Category):

For SBP:

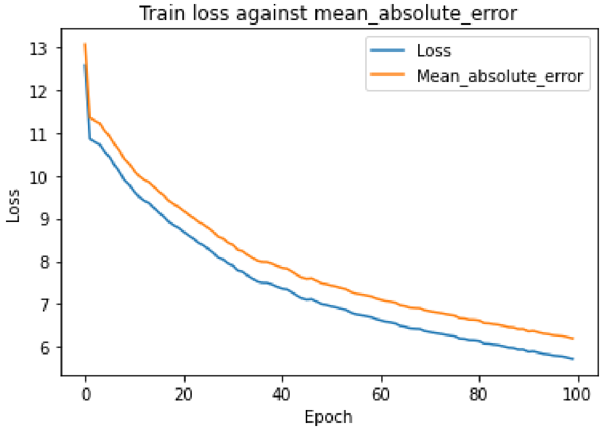
`Neural Net RMSE: 10.745985247730248`

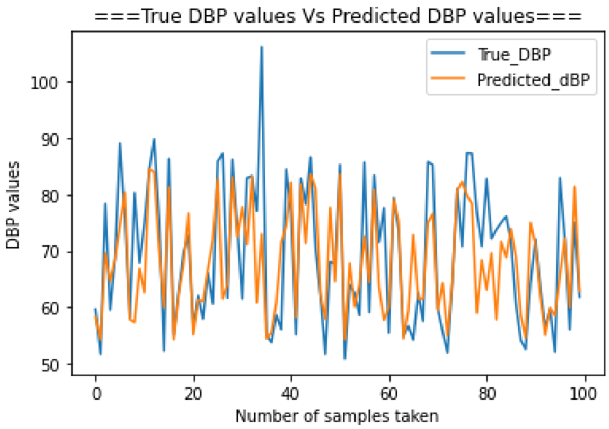




For DBP:

```
Neural Net RMSE: 8.13407420841933
```

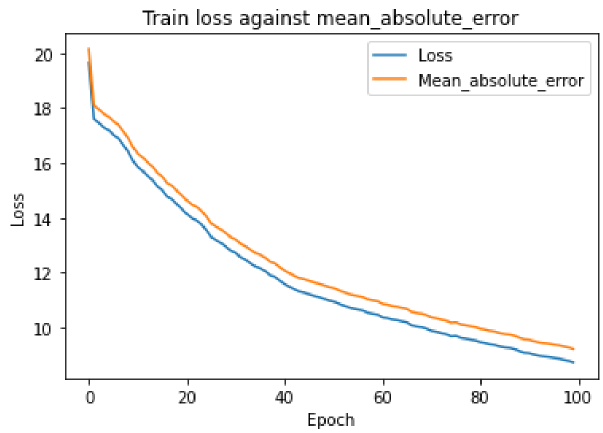


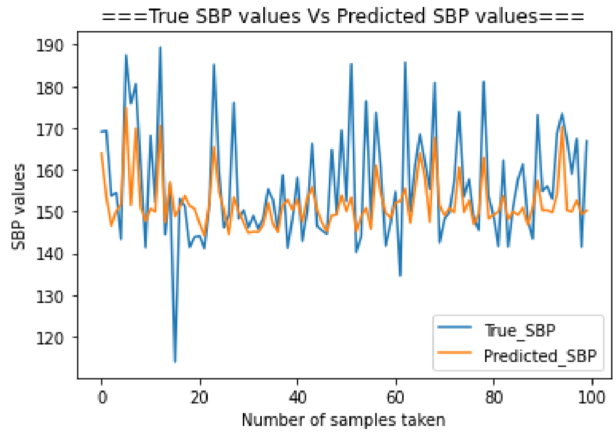


Class-3 (Hypertension Category) :

For SBP:

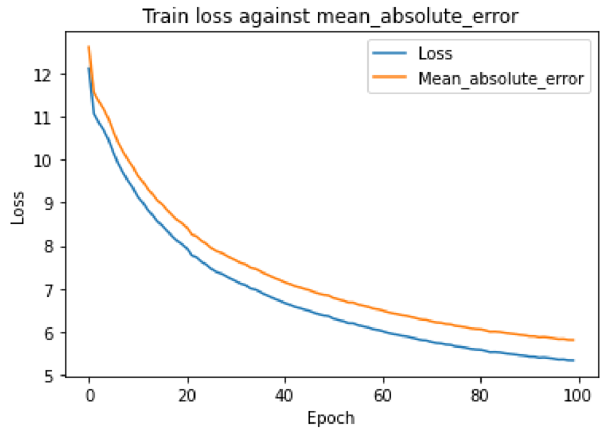
```
Neural Net RMSE: 11.185960777543254
```

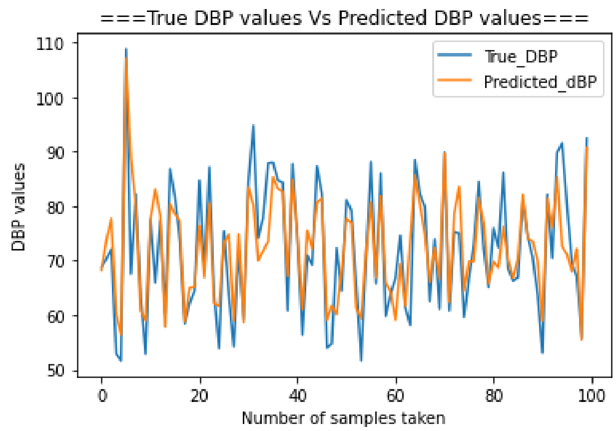




For DBP:

```
Neural Net RMSE: 8.85719609788793
```





Justification for the design choices

AdaboostR technique is one of the top estimator algorithms, according to the findings of prior studies. The AdaboostR technique uses an array of basic estimators to build the full estimator. Despite the fact that each simple estimator yields poor results when compared to the best estimate, it may be used to create weak estimators and assign each one. Most AdaboostR implementations employ the Decision Tree Regression (DTR) technique as the basic estimator. AdaBoostR is fundamentally thought of as a metaestimator that begins by fitting a regressor into the first dataset and then fits more copies of the regressor on the same dataset; the error of the current prediction is used as the criterion for modifying the weights of instances.

ANNs may pick up on really complex structures. It works pretty well on large dataset.Artificial Neural Networks have the ability to learn the complex relationship between the features and target due to the presence of activation function in each layer.The activation function is the function that is responsible for introducing non-linearity in the relationship. In our case, the output layer must contain a linear activation function. Each layer can also have regularizers associated with it. Regularizers are responsible for preventing overfitting. When some or all of the variables are categorical, Adaboost is able to learn a perfect classifier. ANNs CAN NOT. When we have relatively tiny datasets, adaboost are fantastic. operate effectively with the majority of medically related datasets and pictures.

Results and Analysis

Method-1

S.No	Data type	Regressor	RMSE(SBP)	RMSE(DBP)
1	Non-Uniform	Adaboost	19.70789982236872	21.570187087585726
2	Non-Uniform	ANN	15.186525424780317	9.09528037081269
3	Uniform	Adaboost	19.666706219209686	14.26789541173926
4	Uniform	ANN	12.757742688569234	6.338174902260865

Method-2

In this method, we are using non uniform data so that the model can be generalized.

S.No	Regressor	Class	RMSE(SBP)	RMSE(DBP)
1	Adaboost	Normal	18.144848787869638	6.71007001446061
2	Adaboost	Prehypertension	16.263392699571902	15.455149841896423
3	Adaboost	Hypertension	13.370559769713472	24.97219942291945
4	ANN	Normal	12.465272120081679	4.676271347181978
5	ANN	Prehypertension	10.745985247730248	8.13407420841933

6	**ANN**	**Hypertension**	11.185960777543254	8.85719609788793

Conclusions

In this research, a method for BP estimate utilising ML approaches was suggested. The need for calibration due to differences in readings is avoided, which is one of the key advantages of ML-assisted BP estimation. The characteristics are retrieved from the ECG data using a pre-processing method. These features are used by the ANN and adaboost regressor to predict the values of SBP and DBP.These are done by the 2 methods. Based on the results of two method it can be concluded that if we perform classification first and on that we apply regression to predict SBP and DBP it will give less error(method-2).The suggested model may be included into devices for low-cost and remote health supervision since it uses ECG signals to perform additional functions like heart-rate and associated health parameter monitoring.

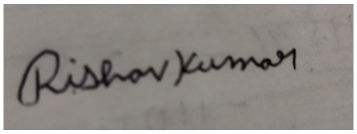
References

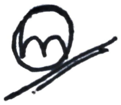
1. https://arxiv.org/ftp/arxiv/papers/2008/2008.10099.pdf
2. https://www.kaggle.com/datasets/mkachuee/BloodPressureDataset/discussion
3. https://www.kaggle.com/code/stephenmugisha/bloodpressure-analysis
4. Blood Pressure Estimation from ECG Data Using XGBoost and ANN for Wearable Devices Sourav Banerjee, Binod Kumar, Member, IEEE and Jai Narayan Tripathi, Senior Member, IEEE
5. ECG-Based Blood Pressure Estimation Using Mechano-Electric Coupling Concept SEYEDEH SOMAYYEH MOUSAVI1 , MOSTAFA CHARMI1 , MOHAMMAD FIROUZMAND2 , MOHAMMAD HEMMATI1 , MARYAM MOGHADAM3 , and YADOLLAH GHORBANI4
6. https://www.analyticsvidhya.com/blog/2021/08/a-walk-through-of-regression-analysis-using-artificial-neural-networks-in-tensorflow/

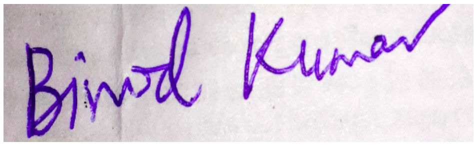
Declaration

We solemnly declare that the project report is based on our own work carried out during the course of our study. We further certify that

1. The work contained in the report is original and has been done by us under the general supervision of our supervisor.
2. Whenever we have used materials from other sources, we have given due credit to them in the text of the report and given their details in the references.







Rishav Kumar Om Solanki

Signature of the Students

Dr Binod Kumar

Signature of the Supervisor