APS360 PROJECT FINAL REPORT
PREDICTING BLOOD PRESSURE WITH RNNS

Haroon Akram
Student# 1007001968
haroon.akram@mail.utoronto.ca

Alex Liu
Student# 1007030008
alexx.liu@mail.utoronto.ca

Elizabeth Tang
Student# 1007711843
elizabeth.tang@mail.utoronto.ca

Sophia Tao
Student# 1008048569
sophiaa.tao@mail.utoronto.ca

ABSTRACT

Cuff-based blood pressure (BP) devices, which are the standard method of BP measurement, are limited by their inability to capture the continuous changes in BP over time. Due to this, they are unable to utilize continuous fluctuation for early detection of cardiovascular conditions. This project attempts to design a neural network using Recurrent Neural Networks (RNNs) to predict blood pressure using data points from photoplethysmographs (PPGs) and electrocardiograms (ECGs) from 84 subjects. The final network implements 3 stacked Long Short-Term Memory (LSTM) layers, has 2 features in the input, and has 128 features in the hidden state. Overall, the model is able to learn the patterns between the biological signals and make reasonable predictions for BP but tends to overfit. In addition, there are several ethical considerations such as limitations in the training data including data bias from demographic factors. Furthermore, due to the complex nature of this problem and the significant impact of the results, the product should not be used as a standalone diagnostic tool. —-Total Pages: 9

1 INTRODUCTION

Blood pressure, which is the measure of the force of blood on the blood vessels, is an important aspect of health as it allows blood to flow and deliver nutrients throughout the body [1]. Conditions relating to blood pressure can often have serious or life-threatening effects. As blood pressure changes over time, this fluctuation is not captured in current methods of blood pressure measurement such as cuffs [1]. Thus, current methods are unable to utilize the blood pressure variation, which may be useful in early detection and prevention. This project intends to use neural networks to cufflessly predict continuous blood pressure using PPGs and ECGs. The machine learning approach suits the chosen data since it can handle the large-scale datasets and capture the intricate patterns of PPGs and ECGs [2]. However, there were still many challenges and limitations in throughout this project which will be discussed along with the results in this report.

2 ILLUSTRATION/FIGURE

Figure 1 shows a poster designed by the team that introduces the idea of the project.

3 BACKGROUND/RELATED WORK

Since continuous and cuffless blood pressure measurement methods can transform cardiovascular disease detection, considerable research in the field has been done since the early 2000s. These methods are mainly based on PPG and ECG data.

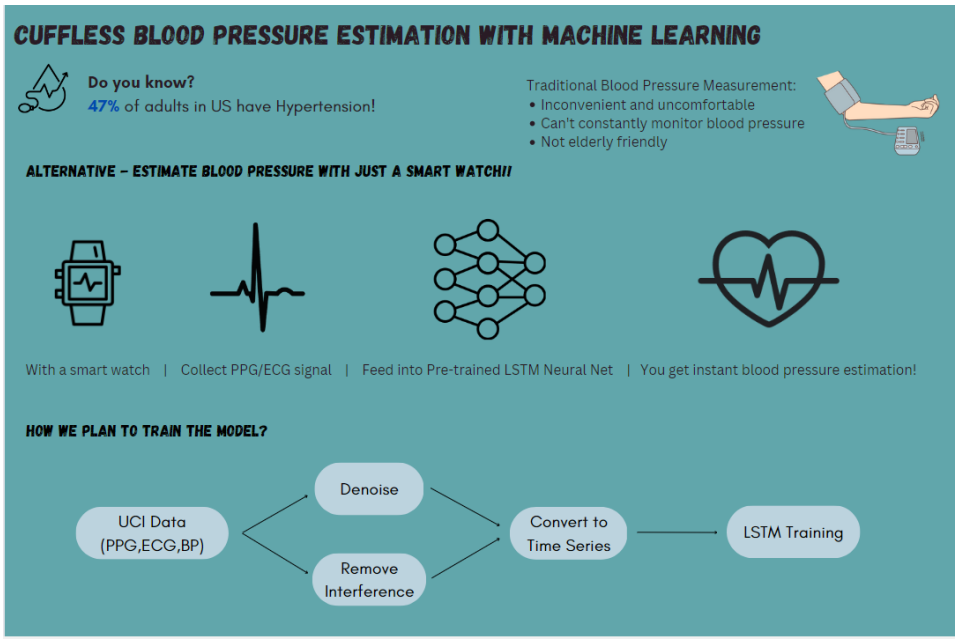

Figure 1: An illustration figure that summarizes the idea of the project.

Blood pressure estimation using Pulse Wave Velocity (PWV) is a pioneering method in cuffless blood pressure measurement. PWV, denoting the speed of the pressure wave propagation in the blood vessels, has demonstrated a correlation to blood pressure. The computation of PWV uses the Pulse Transit Time (the time taken for the pulse to travel between two sites on the body) and the artery length. These contributions include work from Chen et al. [3]., Nitzan et al. [4], etc. However, as this method requires signals from two distant sites, the different signals can introduce synchronicity issues.

With developments in machine learning, a more practical approach is Pulse Wave Analysis, which focuses on characteristic feature extraction and signal processing from a PPG waveform. Using four features extracted from the waveform, Teng and Zhang [5] were able to build a linear regression model (Figure 2). However, since the relationship between BP and PPG signals is non-linear, the predicted results lack accuracy. Kurylyak et al. [6] then created a NN model based on 21 temporal features which outperformed the linear regression model (Figure 3).

Given that blood pressure fluctuates with time, a major shortcoming of the two machine learning models is the failure to consider previous time steps, which causes the accuracy of predictions to deteriorate over time. To consider time step changes, Tanveer and Hasan [7] used two stacked long short term memory (LSTM) layers, which specialize in time domain tasks, in their network. Further improvements include using automated feature extraction by the NN rather than traditional feature engineering techniques. Lo et al. [8] also utilized LSTM to predict blood pressure using PPG and ECG, which improved the accuracy and had low prediction errors.

4 Data Processing

The data used by the team for training and validation is from the UCI machine learning repository "Cuff-Less Blood Pressure Estimation" dataset [9]. The dataset provides synchronized and preprocessed PPG, ECG, and Arterial Blood Pressure (ABP) data. Figure 4 shows the visualization of the raw data.

Our team tried processing the raw data with a wavelet-based denoising technique. However, no significant improvement in signal quality was seen since the data is already preprocessed. In order for the LSTM to learn the time series relationship of the data, the data was converted into a time

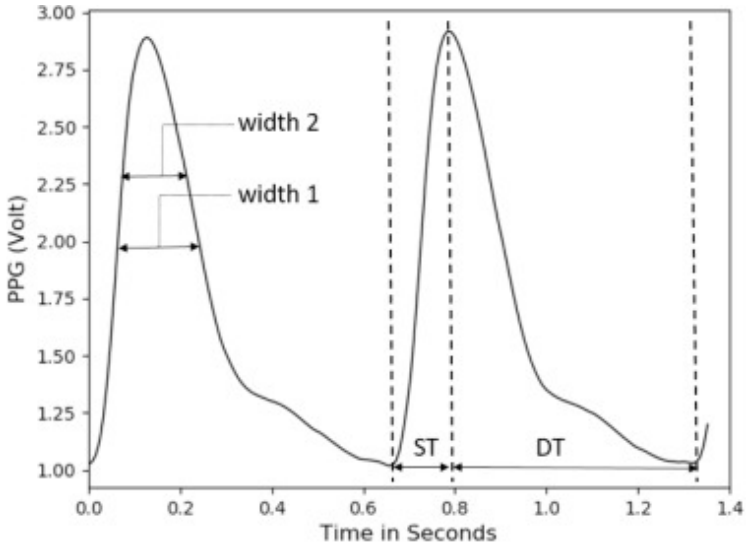

Figure 2: The four PPG features extracted from the waveform by Teng and Zhang [3].

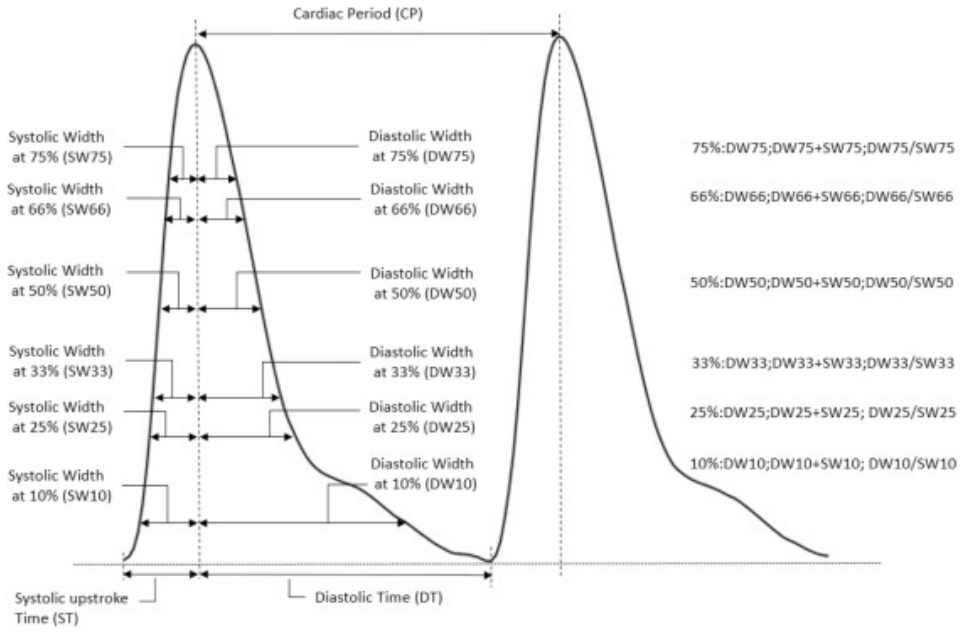

Figure 3: 21 temporal features used by Kurylyak et al. [8].

series structure. The time series structured data is in the form of a 2D data frame. The first row of the data frame is the first 32 data points (32 is our chosen time window hyperparameter) in the 1D array. The second row is from the second data point to the 33rd data point, and the third row is from the third data point to the 34th data point, and so on.

The entire database contains data for over 3000 different subjects. Our team randomly selected 84 subjects from the database and did a train-validation split base on the subjects with a ratio of 4:1. The PPG, ECG, and ABP data were normalized separately since each of them has a different unit.

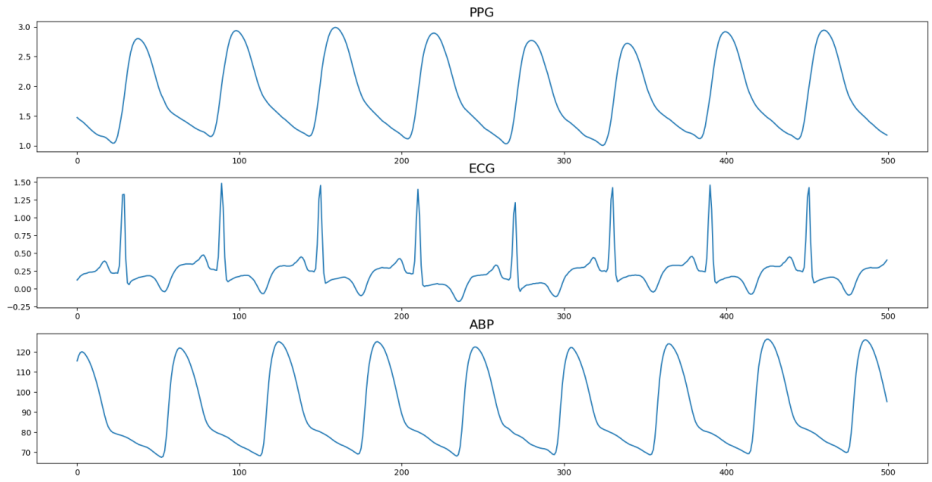

Figure 4: A snapshot of the PPG, ECG, and ABP raw data used for training.

5 ARCHITECTURE

Due to the time-series nature of our data, the final neural network model our team has settled on is an LSTM. An LSTM is a type of Recurrent Neural Network (RNN) with the ability to learn long-term dependencies [10]. This is done by using a cell-state that helps decide whether to keep or lose dependencies [11].

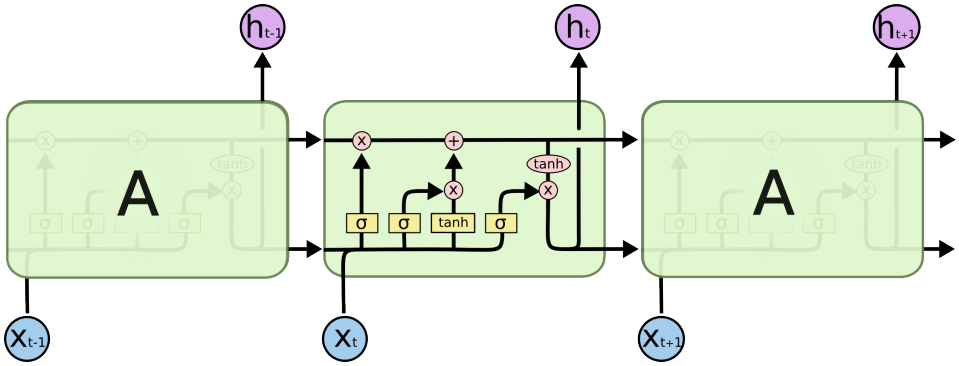

Figure 5: A diagram of an LSTM's architecture [11].

As shown in Figure 5, each repeating module is labelled with "A". Each of these modules contains four layers with one hyperbolic tangent activation layer and three sigmoid activation layers. Furthermore, each module has pointwise operations like addition, multiplication, and a hyperbolic tangent. A gate is the combination of a sigmoid layer and a pointwise multiplication. The gates control the passage of information to and from its cell state and other modules. This process of controlling information in the cell-state gives an LSTM its ability to learn long-term dependencies.

Our final model uses PyTorch's torch.nn.LSTM class to implement a basic LSTM as described above [12]. We then stacked three of these LSTMs as hidden layers for our stacked LSTM model. In a stacked LSTM, each hidden LSTM layer feeds the next LSTM layer. Then, we apply a linear output layer to improve the interpretability of our results [13]. At the end, we apply a Savitzky-Golay low-pass filter to help "smooth" out our noisy predictions [14] to better follow how actual blood pressure looks. Table 1 shows the parameters and layers behind our LSTM setup. Figure 6 shows the architecture of our LSTM model.

Table 1: Basic LSTM parameters, their descriptions and the values we set them to for our final model.

Parameter	Description	Value
input_size	The number of features expected in our input.	2
hidden_size	The number of features in the model's hidden state.	128
num_layers	The number of LSTMs we have stacked on each other.	3
batch_first	Whether the input and output tensors are provided by us.	True

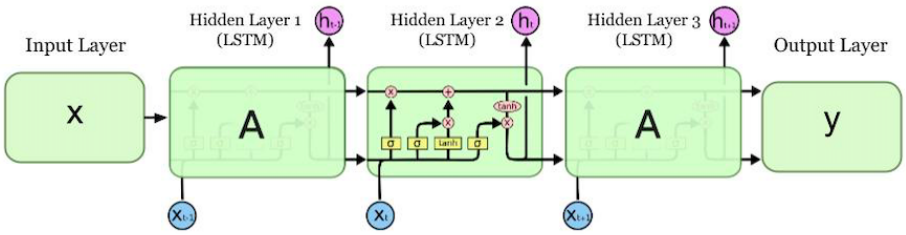

Figure 6: A diagram of our LSTM model architecture.

6 Baseline Model

The team chose to use a simple autoregressive model called AR(100) as the baseline model [15]. From the time series-based data, AR is one of the simplest baseline models we can use to judge our LSTM model. The statsmodels Python module provides the AutoReg() function, which takes two inputs: the training data and the appropriate lag value, to perform our baseline analysis [16]. The lag value used is 100, meaning the current value is based on the values in the previous 100 data points. Then, the function fit() is called to train the baseline model on the dataset and use the model to predict future values.

The baseline model is trained on the first 1000 data points of blood pressure to predict the blood pressure for the next 2000 data points. Figure 7 shows the performance of the model. Predictions are in red while ground truth is in blue.

Measuring blood pressure depends on many different factors within a complex biological system so a simple model such as AR is not sufficient in predicting blood pressure. It is unable to predict based blood pressure based on the ECG and PPG, which the LSTM primary model can. Since the AR model is only able to predict using previous values of blood pressure and not ECG and PPG, it would not be able to adapt to changes.

7 Quantitative Results

The team utilized the accuracy of the model at peaks and troughs of blood pressure and measured the mean-squared error (MSE) loss for the training and validation predictions to measure accuracy. The peaks and troughs of blood pressure are important when it comes to accuracy since they are primarily used to diagnose a patient with blood pressure issues such as high blood pressure and low blood pressure [9]. As a result, if the peak and trough predictions are accurate, this is a good indicator of whether the model performs well for the purposes of this project. Furthermore, we used MSE loss to measure how well our model is performing. An example of this can be seen in Figure 8.

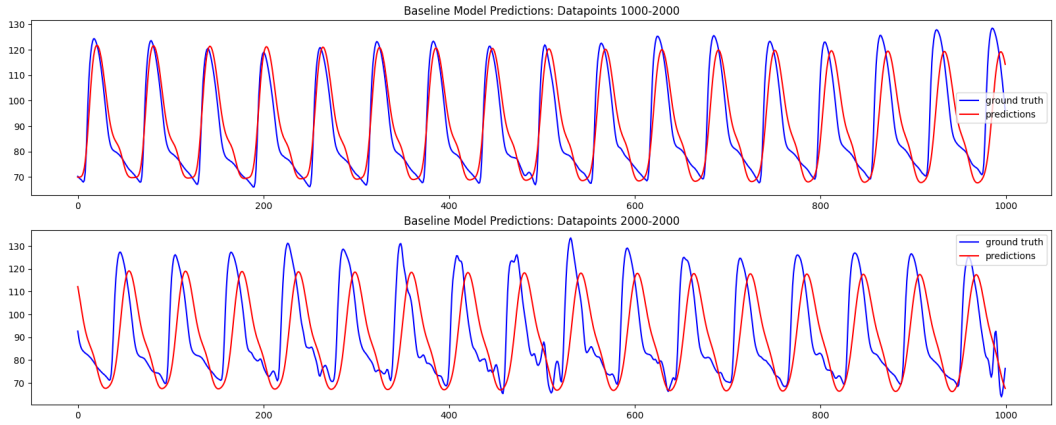

Figure 7: The baseline model predictions for blood pressure compared to the ground truth.

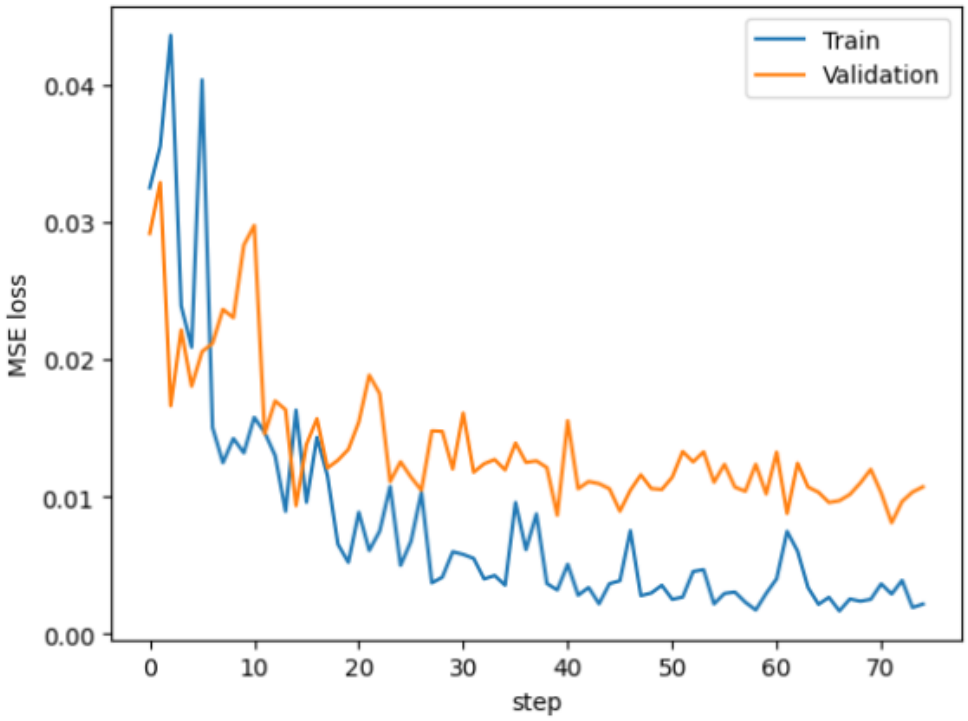

Figure 8: Primary model MSE loss for validation and train.

8 Qualitative Results

The quantitative measurements must be understood in the context of our problem because of how blood pressure is measured. Blood pressure is measured by the two pressures when the heart contracts (systole) and relaxes (diastole) during its pumping cycle [17]. In the graph, the diastolic pressure is the lows/dips in the curve. The systolic pressure is at the peaks/maximum pressure in the graph. Therefore, it is more important to look at the distance between the ground truth low and the predicted low for diastolic, and the ground truth high and the predicted high for the systolic. Figure 9 shows an example of the training and validation results modelled against the ground truth.

The training results are more accurate and have less fluctuation than the validation data which is expected. Although the predicted line is much less smooth than the ground truth, this is not important, rather it is important to see that the model is able to reach the proper highs and lows. Our model performs well from the hyperparameter tuning and smoothing function modifications we made.

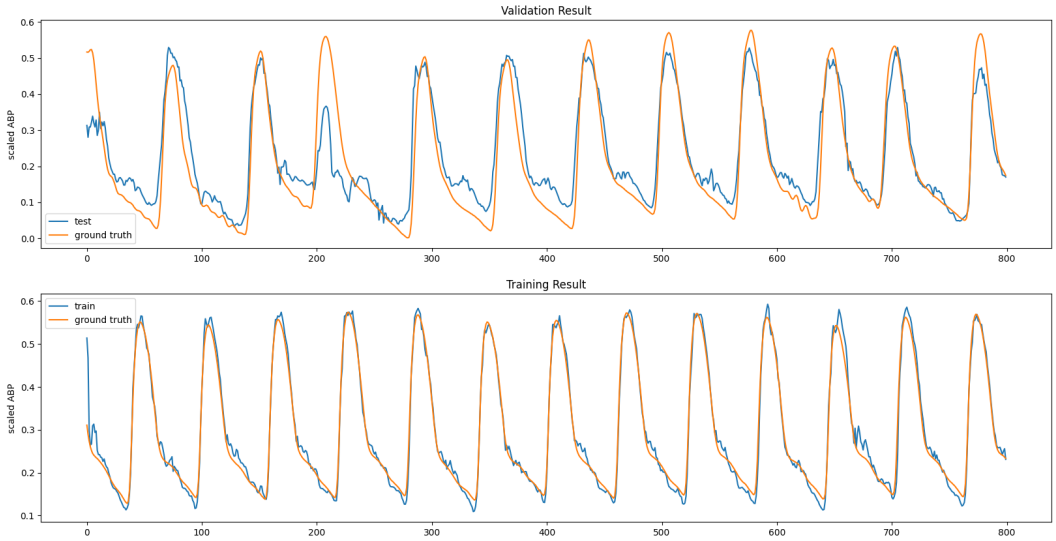

Figure 9: Primary model training and validation result.

Although our model is trained on a dataset of 68 subjects, the model may still struggle to model outliers. For example, a patient may have a medical condition that causes their blood pressure to change but not their PPG or EKG, so that change might not be reflected in the results from our model. It is difficult to predict blood pressure from such a diverse population of people and these predictions depend on the limitations of our training dataset.

9 Model Evaluation on New Data

To accurately evaluate the model's performance, the model is finally tested on a completely different database that is never used before during the training process. The MIMIC-IV Waveform Database has a large collection of PPG, ECG, and ABP signals and measurements from patients in intensive care units [18]. Since MIMIC-IV is a huge and representative database that is widely used in research, the team randomly selected a few subjects in the database as the test dataset. The PPG and ECG data are processed the same way so that they can be fed into the trained model. A snapshot of the prediction of the test data is shown in Figure 10.

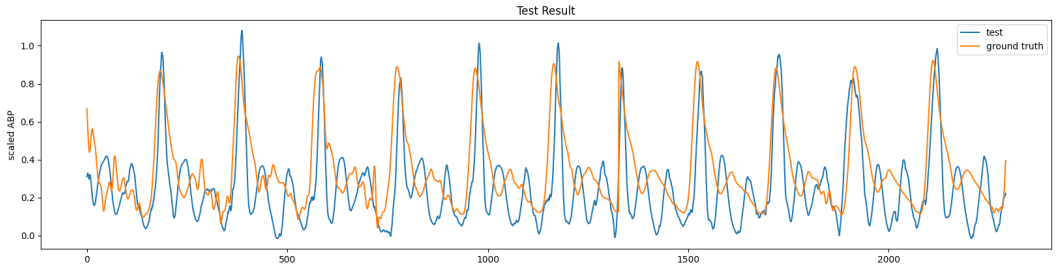

Figure 10: Model's prediction on the test data compared to the ground truth.

Since the source of the test data is completely different from the training data, there are many variables that cannot be controlled, e.g. machine that records the data, preprocessing technique, etc.

Surprisingly, the model still performs very well on the test data despite the interferences of those factors. The model is able to capture the general blood pressure pattern and the predicted waveform is in phase with the ground truth. More importantly, although the predicted waveform does not align with the ground truth perfectly, the model is capable of predicting systolic and diastolic blood pressure accurately. The MSE of the test data prediction is 0.021.

10 Discussion

Overall, the model is able to learn the relationship between the biological signals and make reasonable predictions on the blood pressure. If the variety of training data is limited, the training loss can be reduced to close to zero, as shown in Figure 8. This proves the model's capability to make predictions, but on the other hand, also shows that the model is prone to overfitting. However, since biological signals can vary largely between individuals when a large amount of data is introduced during training, the model can struggle to understand the underlying pattern of blood pressure. Therefore, it is important to control the diversity of the dataset as a hyperparameter.

From Figure 9, we can see that the validation predictions are not as good as the training results, but still performs well. It is able to capture the general increasing or decreasing trend of blood pressure. However, it is observed that at around the 200th datapoint in Figure 9, the model is unable to predict the blood pressure peak. This may be because there is a sudden increase in blood pressure, potentially related to a sudden change in PPG or ECG, and the model is unable to understand this abnormal behaviour.

In Figure 10, it is observed that the test accuracy is not as good as the validation results, and the predicted blood pressure waveform shape is less similar to the ground truth as well. This is mainly due to the difference between the test database and the train/validation database. Note that the ground truth blood pressure shape in the test dataset is also different from that in the train/validation dataset. This can be due to either individual differences or database differences, including data recording process and preprocessing technique, etc. As a result, it is understandable that the predictions on test data are less stable.

11 Ethical Considerations

Since we are working with medical information, the training data has several limitations such as inadequate representation, data bias from demographic factors, and poor data quality from outliers and missing information. Further, our model has several limitations such as generalization for a diverse population, uncertainty in results, and external biological factors that the model cannot account for. Due to these limitations, ethical issues can arise in several use cases such as if it is used as a standalone diagnostic tool. If the system is used to diagnose patients without a qualified healthcare professional to verify results, there could be misdiagnosis and many risks to the patient's safety. Both patients and professionals must be informed of the risks and understand the system [19].

12 Project Difficulty/Quality

This project has several difficulties. One of the difficulties is in data collection. For instance, to measure ABP at an accuracy level required for this project ourselves, we would require an invasive procedure called cannulation of a peripheral artery [20]. This is when a cannula is inserted into an artery which allows you to measure ABP at a high level of accuracy. Any other method is not reliable or accurate enough for this project. As a result, our team relied on existing datasets. Furthermore, the project itself is not a simple task. Research papers that have attempted the same project have had difficulties in getting accurate results. Some of these attempts include using Pulse Wave Velocity (PWV) to calculate ABP [21], using a linear regression model to predict ABP [5], and a feedforward neural network model to predict ABP [6]. However, all of these attempts still had issues with generating inaccurate results due to the complex nature of this problem.

13 IMPORTANT LINKS

Our project is being done on GitHub. Here is the link to our project's repository.

GitHub Repository Link: https://github.com/elizabethtang/APS360Project

14 REFERENCES

[1] InformedHealth.org, “What is blood pressure and how is it measured?,” 23-May-2019. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK279251/. [Accessed: 6-Aug-2023]

[2] IBM, “What is deep learning?” [Online]. Available: https://www.ibm.com/topics/deep-learning. [Accessed: 6-Aug-2023]

[3] Y. Chen, C. Wen, G. Tao, M. Bi, G. Li, Continuous and noninvasive blood pressure measurement: a novel modeling methodology of the relationship between blood pressure and pulse wave velocity, Ann. Biomed. Eng. 37 (11) (2009) 2222–2233.

[4] M. Nitzan, B. Khanokh, Y. Slovik, The difference in pulse transit time to the toe and finger measured by photoplethysmography, Physiol. Meas. 23 (1) (2001) 85.

[5]X.F. Teng, Y.T. Zhang, Continuous and noninvasive estimation of arterial blood pressure using a photoplethysmographic approach, September, IEEE, Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439) Vol. 4 (2003) 3153–3156.

[6]Y. Kurylyak, F. Lamonaca, D. Grimaldi, A neural network-based method for continuous blood pressure estimation from a PPG signal, May, in: Instrumentation and Measurement Technology Conference (I2MTC), 2013 IEEE International, IEEE, 2013, pp. 280–283.

[7] M. Tanveer, M. Hasan, Cuffless Blood Pressure Estimation from Electrocardiogram and Photoplethysmogram Using Waveform Based ANN-LSTM Network, 2018, arXiv Preprint arXiv: 1811.02214.

[8] F. P. . -W. Lo, C. X. . -T. Li, J. Wang, J. Cheng and M. Q. . -H. Meng, ”Continuous systolic and diastolic blood pressure estimation utilizing long short-term memory network,” 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Korea (South), 2017, pp. 1853-1856, doi: 10.1109/EMBC.2017.8037207.

[9] NHS, ”What is blood pressure?” [Online]. Available: https://www.nhs.uk/common-health-questions/lifestyle/what-is-blood-pressure/. [Accessed: 6-Aug-2023]

[10] Intellipaat, “What is LSTM? Introduction to Long Short Term Memory,” 13-Jun-2023. [Online]. Available: https://intellipaat.com/blog/what-is-lstm/?US. [Accessed: 6-Aug-2023]

[11] C. Olah, “Understanding LSTM Networks.” [Online]. Available: https://colah.github.io/posts/2015-08-Understanding-LSTMs/. [Accessed: 6-Aug-2023]

[12] “LSTM” [Online]. Available: https://pytorch.org/docs/stable/generated/torch.nn.LSTM.html [Accessed: 6-Aug-2023]

[13] “Neural networks regression: Scaling the outputs or using a linear layer?” 7-Jul-2023. [Online]. Available: https://saturncloud.io/blog/neural-networks-regression-scaling-the-outputs-or-using-a-linear-layer [Accessed: 6-Aug-2023]

[14] ”Savitzky-Golay filtering - MATLAB”, https://www.mathworks.com/help/signal/ref/sgolayfilt.html. [Accessed: 6-Aug-2023]

[15] J. Fernando, “What Are Autoregressive Models? How They Work and Example,” 06-Apr-2022. [Online]. Available: https://www.investopedia.com/terms/a/autoregressive.asp. [Accessed: 14-Jun-2023]

[16] “statsmodels.tsa.ar model.AutoReg.” [Online]. Available: https://www.statsmodels.org/dev/generated/statsmodels.tsa.ar model.AutoReg.html. [Accessed: 14-Jun-2023]

[17] “Understanding blood pressure readings,” www.heart.org, https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings (accessed Aug. 10, 2023).

[18] B. Moody, S. Hao, B. Gow, T. Pollard, W. Zong, and R. Mark, ”MIMIC-IV Waveform Database (version 0.1.0),” PhysioNet, 2022. [Online]. Available: https://doi.org/10.13026/9mw-f949. [Accessed: 6-Aug-2023]

[19] D. D. Farhud and S. Zokaei, "Ethical Issues of Artificial Intelligence in Medicine and Healthcare," Iranian Journal of Public Health, vol. 50, no. 11, pp. i-v, 2021.

[20] Y. Nguyen, V. Bora, "Arterial Pressure Monitoring," National Library of Medicine. Available: https://www.ncbi.nlm.nih.gov/books/NBK556127/. [Accessed: 6-Aug-2023]

[21] C. El-Hajj and P. A. Kyriacou, "A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure," vol. 58, p. 101870, 2020, doi: 10.1016/j.bspc.2020.101870. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809420300264. [Accessed: 6-Aug-2023]