Machine Learning for Arterial Blood Pressure Prediction

Jessica Zheng*   JZHENG22@MIT.EDU
MIT, Cambridge, MA

Hanrui Wang*   HANRUI@MIT.EDU
MIT, Cambridge, MA

Anand Chandrasekhar   ANANDC@MIT.EDU
MIT, Cambridge, MA

Aaron D. Aguirre   AAGUIRRE1@MGH.HARVARD.EDU
Center for Systems Biology, Massachusetts General Hospital, Harvard Medical School, Boston, MA
Cardiology Division, Corrigan Minehan Heart Center, Massachusetts General Hospital, Harvard Medical School
Wellman Center for Photomedicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA
Healthcare Transformation Lab, Massachusetts General Hospital, Harvard Medical School, Boston, MA

Song Han   SONGHAN@MIT.EDU
MIT, Cambridge, MA

Hae-Seung Lee   HSLEE@MTL.MIT.EDU
MIT, Cambridge, MA

Charles G. Sodini   SODINI@MTL.MIT.EDU
MIT, Cambridge, MA

Abstract

High blood pressure is a major risk factor for cardiovascular disease, necessitating accurate blood pressure (BP) measurement. Clinicians measure BP with an invasive arterial catheter or via a non-invasive arm or finger cuff. However, the former can cause discomfort to the patient and is unsuitable outside Intensive Care Unit (ICU). While cuff-based devices, despite being non-invasive, fails to provide continuous measurement, and they measure from peripheral blood vessels whose BP waveforms differ significantly from those proximal to the heart. Hence, there is an urgent need to develop a measurement protocol for converting easily measured non-invasive data into accurate BP values. Addressing this gap, we propose a non-invasive approach to predict BP from arterial area and blood flow velocity signals measured from a Philips ultrasound transducer (XL-143) applied to large arteries close to heart. We developed the protocol and collected data from 72 subjects. The shape of BP (relative BP) can be theoretically calculated from these waveforms, however there is no established theory to obtain *absolute* BP values. To tackle this

challenge, we further employ data-driven machine learning models to predict the Mean Arterial Blood Pressure (MAP), from which the absolute BP can be derived. Our study investigates various machine learning algorithms to optimize the prediction accuracy. We find that LSTM, Transformer, and 1D-CNN algorithms using the blood pressure shape and blood flow velocity waveforms as inputs can achieve 8.6, 8.7, and 8.8 mmHg average standard deviation of the prediction error respectively without anthropometric data such as age, sex, heart rate, height, weight. Furthermore, the 1D-CNN model can achieve 7.9mmHg when anthropometric data is added as inputs, improving upon an anthropometric-only model of 9.5mmHg. This machine learning-based approach, capable of converting ultrasound data into MAP values, presents a promising software tool for physicians in clinical decision-making regarding blood pressure management.

Data and Code Availability In this study, we recorded ultrasound data and the gold standard cuff BP measurements from various healthy volunteers within the age group of 18 to 60 years. The study was approved by IRB and was conducted after receiving written, informed consent from all partici-

---

* These authors contributed equally

pants. Individuals with any prior cardiovascular disease other than hypertension were excluded from the study. More detailed information about data collection can be found in Section 3.1. We have made our code available at this link.

Institutional Review Board (IRB) Our research requires IRB approval. All the human study described in the paper was approved by Institutional Review Board (IRB) at MIT (1405006402).

1. Introduction

Hypertension is a major cardiovascular risk factor; hence, accurate blood pressure (BP) measurement is of high clinical interest. In an Intensive Care Unit (ICU), physicians can use an invasive arterial catheter to measure the BP waveform with high resolution. In non-ICU settings, arterial catheters are not used, and clinicians must rely upon isolated spot measurements from a non-invasive arm-cuff device. These devices measure systolic and diastolic pressure values and cannot output the arterial BP (ABP) waveform. As a result, these devices cannot track the *dynamic nature* of the arterial system and measure the beat-to-beat variations of BP, which may reveal information on the underlying pathophysiology. Alternatively, clinicians can use non-invasive devices based on the volume clamping method to acquire the ABP waveform from peripheral blood vessels like the digital arteries in the finger. Since the BP waveform measured at large arteries proximal to the heart (aorta, subclavian, brachial) are clinically more valuable than the distal digital arteries, there is strong interest in developing non-invasive devices to measure the ABP waveform in the larger arteries. Such a device may offer a quantitative method to perform rapid hemodynamic profiling of patients who cannot undergo invasive BP measurements. Clinicians may use such non-invasive devices in an emergency room, step-down clinical ward, outpatient clinic, or even in a home setting to monitor the ABP waveform of the patient for quick diagnosis.

In this work, we present an ML model to convert data from large arteries recorded from an ultrasound transducer to the mean value of the ABP waveform, referred to as the Mean Arterial Pressure (MAP). Using the MAP, we can obtain the absolute ABP waveform by centering the relative ABP waveform (computed directly from the ultrasound device) on the MAP value. Ultrasound devices are best suited for non-invasive MAP monitoring for several reasons.

First, ultrasound devices are low-risk, non-invasive devices that are easy to use with fewer clinical precautions and may be easily utilized outside of an ICU. Second, these devices can measure physiological signals (arterial diameter and blood flow velocity waveform) in their respective absolute units. These signals have a close relationship with the pulse pressure (PP) of the subject; one may easily derive their relationship with the PP using basic fluid-dynamic equations, as discussed elsewhere. Third, clinicians can easily use ultrasound devices to record data from non-superficial arteries, unlike many other non-invasive methods such as photoplethysmogram. While ultrasound devices can produce accurate pulse pressure waveforms, there is no established theory that provides *absolute* blood pressure values. Therefore, estimating the MAP from ultrasound devices generally requires establishing a baseline using a cuff-based device, which is inconvenient, inaccurate (due to long distance to heart) and cannot provide MAP results in real time. In this paper, we propose a novel approach that leverage data-driven ML models to learn the knowledge of absolute blood pressure and predict the MAP from the ultrasound devices. In summary, the contributions of this work are three folded:

• We design the data collection protocol and use Philips transducer device to measure and collect ultrasound data from 72 subjects.

• We present three kinds of machine learning models (LSTM, Transformer and 1D-CNN) to learn the absolute blood pressure that cannot be deducted from existing theory using ultrasound signals.

• We show that with the ML model, the ultrasound-based prediction can achieve 7.9mmHg error std, which is 1.6mmHg lower than the model based on anthropometric data only. We also provide comprehensive discussions of experimental results.

The remaining of the paper is presented as follows: We first discuss previous work on measuring absolute blood pressure, then our data collection method, followed by a detailed description of the algorithms to estimate MAP, and the results.

2. Background and Related Work

2.1. Blood Pressure Measurement

The current standard for measuring BP involves using an arm cuff device that inflates and deflates a physical elastic cuff around the upper left arm (Ward and Langton, 2007). However, this method only provides the systolic and diastolic values and cannot es-

timate a continuous BP waveform; hence, there is a strong interest in developing devices that can monitor absolute BP continuously. A few investigators explored light-based transducers, often referred to as photoplethysmogram (PPG) sensors (Kyriacou and May, 2022), in combination with machine learning algorithms as an alternative for cuff-based BP measurement (Xing and Sun, 2016). The vasomotor tone and skin temperature heavily influence PPG signals from the distal arteries; hence, these measurement locations may not be ideal for long-term BP tracking (Allen, 2007). PPG sensors are not used on the proximal arteries, for instance, the carotid blood vessel, as light-based measurements cannot record data from non-superficial arteries. Ultrasound-based technology can potentially be an ideal sensor for measuring data from blood vessels deeply embedded from the skin's surface (Wang et al., 2021a). Quite a few investigators have used ultrasound-based sensors to measure BP waveform features (Wang et al., 2018), (Seo et al., 2015). Wang et al. (2018) developed a methodology to design ultrasound sensors to record morphology of the BP waveforms measured from the carotid artery. Seo et al. (2015) used data recorded from the ultrasound sensors for monitoring pulse pressure from the carotid artery. There also exists work using transmission line model based method (Harabedian, 2022; John, 2004) to estimate the bp from ultrasound signals but the performance is not satisfactory (error mean -19.6, std 28.8mmHg). Hill et al. (2021) calibrate the non-invasive signals measured from the finger using a photoplethysmogram sensor to obtain absolute BP with ML models. In contrast, we estimate blood pressure from the neck that offers a wide spectrum of clinical utility, ranging from investigating the pressure at which the heart feeds blood to the brain to an opportunity to study the hemodynamic status of patients with Traumatic Brain Injury (TBI). Measuring the pressure at which the heart feeds blood to the brain may help clinicians guide therapy for TBI patients while titrating medications to improve blood circulation to the brain. Moreover, measured data from the peripheral circulation at an arterial location distal from the heart that is often influenced by the temperature of the extremities. We addressed this problem by using data recorded from an arterial site closer to the heart, which is less susceptible to body temperature, offering us an advantage on the signal quality.

2.2. Machine Learning for Time Series Data

There has also been extensive work on using ML models for time series data, particularly with CNN, LSTM and Transformer models. For instance, Zhao et al. (2017), Fauvel et al. (2021) and Wang et al. (2021c) employed 1D-CNN-based models for time series classification. Kim and Kang (2019) applied both 1D-CNN and LSTM algorithms for time series prediction, and Widiputra et al. (2021) and Essien and Giannetti (2019) used CNN-LSTM stacked models for time series forecasting. Recently, attention-based Transformers (Vaswani et al., 2017; Devlin et al., 2018; Radford et al., 2019; Brown et al., 2020; Wang et al., 2022) are becoming main-stream models because of its ability to model both short-term and long-term dependencies. Yang et al. (2021) and Rußwurm and Körner (2020) leveraged transformer models for time series classification, and Grigsby et al. (2021) and Wu et al. (2020) used transformer models for time series forecasting. Radford et al. (2022) leverage transformer (GPT) model for speech recognition and translation. There are also works (Wang et al., 2020b,a, 2021b; Zhang et al., 2020) on improving efficiency of Transformer models for fast time series analysis.

3. Methods

3.1. Data Collection

In this study, we recorded ultrasound data and the gold standard cuff BP measurements from various healthy volunteers within the age group of 18 to 60 years. The study was approved by the committee on the use of Humans as Experimental Subjects at MIT (IRB #1405006402) and was conducted after receiving written, informed consent from all participants. Individuals with any prior cardiovascular disease other than hypertension were excluded from the study.

A total of 72 subjects volunteered for the study. The data from 8 subjects were excluded due to missing information and the anthropomorphic details of the 64 participants are illustrated in Table 1. The data was collected using a well-defined clinical protocol approved by the IRB. The study began with a resting phase. The subject was allowed to rest in a supine posture for a period of 5 minutes to stabilize the heart rate and prevent any BP fluctuations or fast breathing during the experimental interval. Cuff blood pressure was recorded at the end of the

resting phase using a standard omron 7-series blood pressure monitor. These cuff BP measurements were measured from the left upper arm and were recorded twice. The average of two BP measurements was reported as the subject's starting BP. Next, an ultrasound probe (Philips: Transducer XL-143, EPIQ system) was placed over the left neck of the subject after requesting the individual to tilt the neck by 45 degrees towards the right side to expose the left common carotid artery. Sufficient ultrasound gel was used between the transducer and skin to improve the signal-to-noise ratio of the recorded images. The transducer was gently placed over the location of the blood vessel distal from the carotid bifurcation. During this experimental interval, the B-mode ultrasound data were recorded for a period of 30 s. The focus and depth of the ultrasound images were adjusted to get sufficient contrast for the B-mode images. The investigator confirmed that the recorded images were clear and the subject did not swallow saliva during this period to prevent motion artifact on the measured data. Next, the pulsed doppler technique was used to record blood flow velocity at the same arterial location. Note that the B-mode images and blood flow velocity waveforms were recorded with ECG gating for synchronization. The experiment terminated with a BP measurement over the left upper arm which was recorded as the ending blood pressure of the subject. The B-mode images and pulse doppler data were processed to get the arterial diameter and blood flow velocity waveform. In summary, we measure the arterial diameter waveform and blood flow velocity waveform from each subject and we also collect the anthropometic data.

Table 1: Statistics on the distribution of MAP and anthropometric data features in out collected dataset. ρ is the Spearman correlation coefficient of the feature with the MAP value.

	Mean	Stdev	ρ
MAP (mmHg)	93.24	10.94	1.00
Age (yrs)	30.44	12.69	0.52
Heart rate (bpm)	73.14	11.72	0.27
Height (cm)	166.7	9.530	0.13
Weight (kg)	66.84	19.04	0.39

3.2. Dataset Split

Since the dataset is quite small (64 subjects in total), we require the validation set to have a sufficient range of MAP values in order for the model to converge. If the range of target MAP values is not sufficiently large, the model can achieve a low error on the validation set by predicting values near the mean MAP value, and will thus fail to converge on the training and testing sets. Therefore, we split the dataset into $k = 5$ folds as follows:

1. Sort the subjects by MAP value and split them into n/k consecutive buckets of size k, where $n = 64$ is the total number of subjects.

2. For each bucket, randomly assign a different subject from the bucket to each of the k folds.

Using this method, we obtain 5 folds with 12-13 subjects per fold. We then use 3 folds for training, 1 for validation, and 1 for testing. In addition, we repeat this process using 3 different random seeds in order to further reduce the variance of the final evaluated metric.

3.3. Input Features

3.3.1. Blood Pressure Shape Waveform

We scale the raw arterial area waveform A by the compliance in equation 5 to obtain the blood pressure shape waveform bp_shape as seen in equation 1. An example of the full blood pressure shape waveform can be seen in Figure 1.

$$\texttt{bp\_shape} = \frac{A - \text{mean}(A)}{\texttt{AC}} \tag{1}$$

The arterial area waveform A is computed using the formula:

$$A = \pi \frac{D^2}{4} \tag{2}$$

Where D is the arterial diameter that is directly measured from the ultrasound data.

While the compliance value used in the experiments is computed using measurements from the arm cuff, it can also be calculated directly from synchronized arterial area (A) blood flow velocity (V) waveforms using equation 3 (Seo et al., 2015) where ρ is blood density.

$$\texttt{AC} = \frac{4 \cdot \left| \frac{d}{dF} \sqrt{A(t)} \right|^2_{\text{reflection free period}}}{\rho} \tag{3}$$
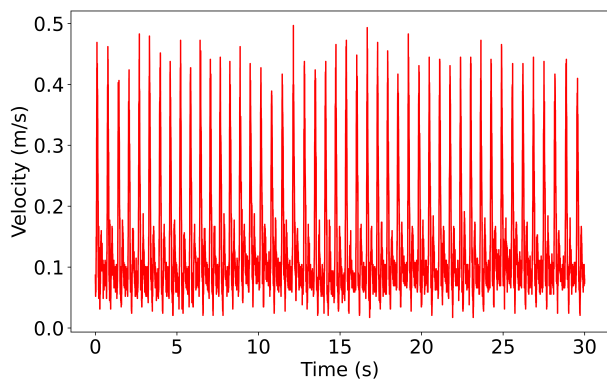
$$F(t) = V(t) \cdot A(t)$$

3.3.2. Blood Flow Velocity Waveform

We also use the raw blood flow velocity waveform measured from the ultrasound device, as shown in Figure 1.
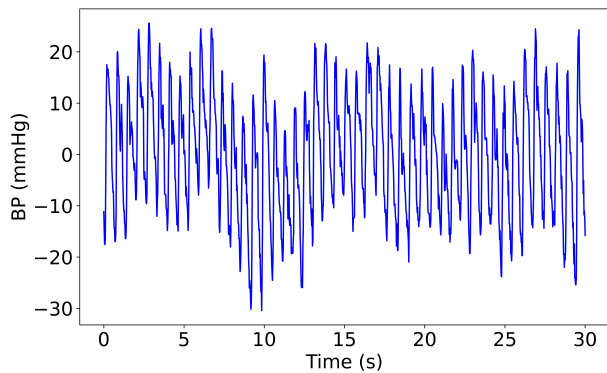
(*a*) The full bp_shape waveform

(*b*) The full (V) waveform

Figure 1: The full 30-second blood pressure shape (bp_shape) and blood flow velocity (V) waveforms

3.3.3. Anthropometric Features

We also include the age, heart rate, height, weight, and sex as inputs to some of the models.

3.4. Prediction Target

3.4.1. MAP

We obtain the ground truth MAP by calibrating the arterial area waveform (A) using the systolic and diastolic blood pressure measurements (SBP and DBP) from the arm cuff according to equation 4 (Chemla et al., 1998). Then our models are trained to predict this value.
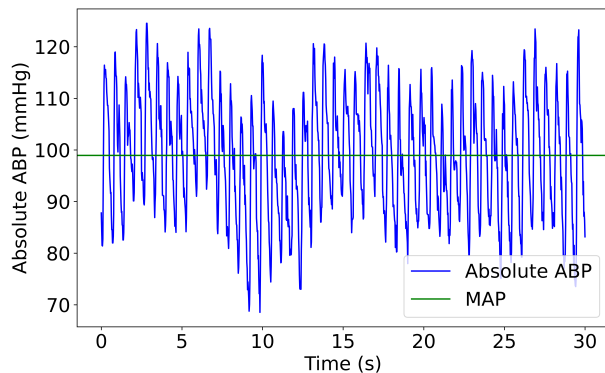

Figure 2: Reconstructed absolute bp shape.

$$\begin{aligned} \text{MAP} &= \text{DBP} + \frac{\text{mean}(A) - \min(A)}{\text{AC}} \\ &= \text{DBP} + \frac{\text{mean}(A) - \min(A)}{\max(A) - \min(A)} \cdot \text{PP} \qquad (4) \\ &= \text{DBP} + \frac{\text{mean}(A) - \min(A)}{\max(A) - \min(A)} \cdot (\text{SBP} - \text{DBP}) \end{aligned}$$

Where AC is the arterial compliance and PP is the pulse pressure, defined by the following equations:

$$\text{AC} = \frac{\max(A) - \min(A)}{\text{PP}} \qquad (5)$$

$$\text{PP} = \text{SBP} - \text{DBP} \qquad (6)$$

The MAP from equation 4 is recorded as the ground truth to be compared with the values obtained from various ML models. Using the MAP value, we can center the relative blood pressure shape waveform bp_shape on the MAP value to obtain the full absolute ABP waveform as shown in figure 2.

3.5. Models

We compare the performance of 5 different models for predicting mean arterial blood pressure:

1. Anthro-only model: A fully-connected model (Haykin, 1994) trained using only anthropometric data: Age, height, weight, sex, and heart rate.

2. MLP: A fully-connected model (Haykin, 1994) trained using the blood pressure shape and blood flow velocity waveforms.

3. LSTM: An LSTM model (Hochreiter and Schmidhuber, 1997) trained using the blood pressure shape and blood flow velocity waveforms.
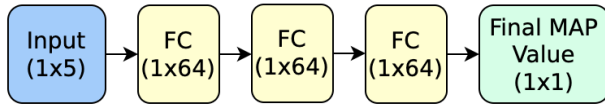
431

Figure 3: Anthro-only fully-connected model architecture

4. Transformer: A transformer encoder model (Vaswani et al., 2017) trained using the blood pressure shape and blood flow velocity waveforms.

5. CNN: A 1d convolutional neural network (Goodfellow et al., 2016) trained using the blood pressure shape and blood flow velocity waveforms.

Each of the models are trained for 200 epochs with Adam optimizer (Kingma and Ba, 2015) and learning rate 0.001.

3.5.1. ANTHRO-ONLY MODEL

In order to establish a baseline for comparison for the MAP prediction models and to determine how much information is contained in the arterial area and blood flow velocity waveforms, we first train a fully-connected model using *only* anthropometric data as inputs.

Inputs As inputs to this model, we use five scalar anthropometric data features: the age, height, weight, sex, and heart rate of the subject.

Model We first project the inputs (1×5) into a new feature map of dimension 1×64 using an initial fully-connected layer. We then feed this into three fully-connected layers of dimension 64 and ReLU activation function (Agarap, 2018), along with a final fully-connected layer to regress the MAP value. The model architecture is shown in Figure 3.

3.5.2. MLP

We first experiment with a simple fully-connected model using the two waveforms as inputs, to determine if more complex models are necessary for this prediction task.
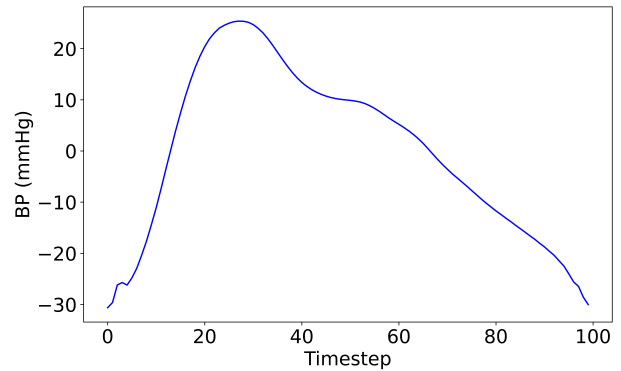
Inputs As inputs to this model, we use the ensembled (1×100) blood pressure shape and blood flow velocity waveforms as inputs.

Model We first project the inputs (100×2) into a new feature map of dimension 100×128 using an initial fully-connected layer. We then feed this into three fully-connected layers of dimension 128 and

ReLU activation function (Agarap, 2018), along with a final fully-connected layer to regress the MAP value.

3.5.3. LSTM

Inputs We use two waveforms as inputs to the model: the blood pressure shape waveform (bp_shape), and the blood flow velocity waveform (V). The original measurements taken from the ultrasound device contain around 30 seconds of data, with 30-50 heartbeats per waveform. However, these measurements are quite noisy, so we average each of the two waveforms into a single ensembled heartbeat and resample them into 1×100 vector features, as shown in Figure 4.

(*a*) The ensembled bp_shape waveform
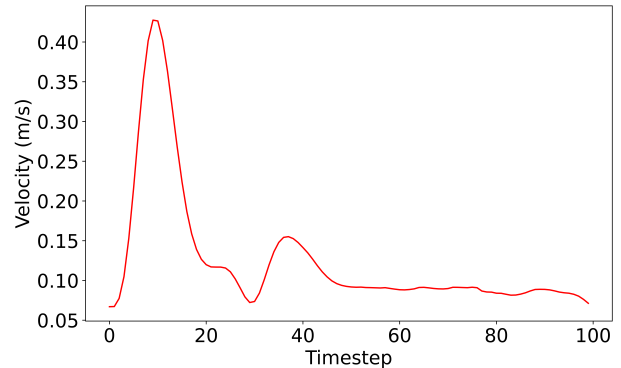
(*b*) The ensembled (V) waveform

Figure 4: The ensembled blood pressure shape (bp_shape) and blood flow velocity (V) waveforms

Model We first project the inputs $(100\times n$, where n is the number of input features) into a new feature map of dimension 100×256 using a fully-connected layer. We then feed this into a unidirectional LSTM with 3 layers, followed by a final fully-connected layer to regress the MAP value, as shown in Figure 5.
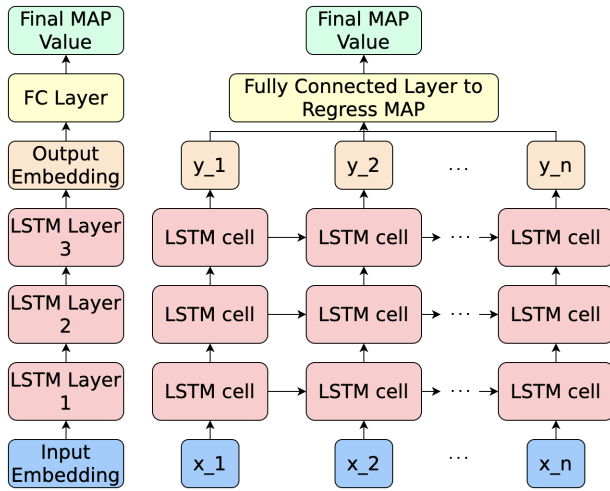
Figure 5: LSTM model architecture

3.5.4. Transformer

Inputs Just as in the LSTM model, this model uses the ensembled (1×100) blood pressure shape and blood flow velocity waveforms as inputs.

Model We first project the inputs $(100\times n$, where n is the number of input features) into a new feature map of dimension 100×256 using a fully-connected layer, and we convert this into input embeddings via a positional encoding layer. We then feed this into a transformer encoder module with 4 layers and 8 attention heads, along with a final fully-connected layer to regress the MAP value. The model architecture is shown in Figure 6.
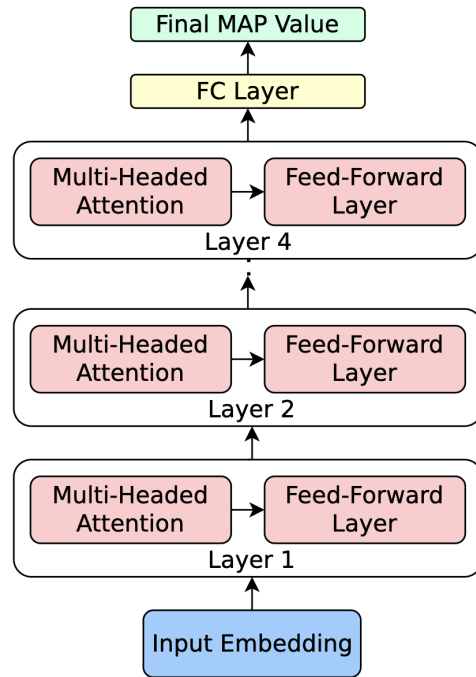
3.5.5. CNN

Inputs Just as in the LSTM model, this model uses the ensembled (1×100) blood pressure shape and blood flow velocity waveforms as inputs.

Model We first project the inputs $(100\text{x}n$, where n is the number of input features) into a new feature map of dimension 100×64 using a fully-connected layer. We then feed this into three layers of 1D convolution with stride 1, kernel size 3, padding 1, and dropout 0.3, along with a final fully-connected layer to regress the MAP value. The model architecture is shown in Figure 7.
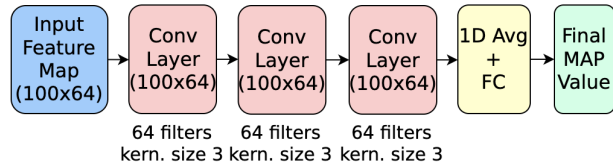
3.5.6. Adding Anthropometric Data as Additional Inputs

We also experiment with adding anthropometric data (age and heart rate) as additional features alongside the two waveform inputs. To do this, we create a fea-

Figure 6: Transformer model architecture

Figure 7: CNN model architecture

ture map of dimension 100x4 using the 4 features (2 scalar, 2 vector) as inputs to the LSTM, Transformer and CNN models.

3.6. Ablation Studies

Without regularization, we find that the Transformer, LSTM, and CNN models severely overfit to the training data. To address this issue, we perform ablation studies on the dropout ratio, L2 norm, and also model size in order to reduce the amount of overfitting. Following the standard ML training protocol, we select the best hyperparameters on the validation set and test on the testing set.

Furthermore, as stated earlier, the arterial compliance (`AC`) value used in the experiments is computed using equation 5 which uses measurements from the arm cuff, as opposed to using equation 3 which is calculated directly from synchronized arterial area (A) and blood flow velocity (V) waveforms. So, in order

to verify that the information contained in the two waveforms improves the accuracy of the estimated MAP value compared to the anthro-only model, we perform an ablation study substituting the arterial area waveform for the `bp_shape` waveform as input to the model, such that the arterial compliance is not used.

3.7. Beat-to-beat Calibration

To increase the amount of input data available for training, we experiment with splitting the original 30-second raw waveforms into individual beats instead of using one single ensembled beat. To determine the MAP value for each beat, we compute a linear equation $\mathtt{MAP}_k = m \cdot \mathrm{mean}(a_k) + c$, where a_k is the arterial area waveform of the k-th beat, and $\mathtt{MAP}_k$ is the calibrated MAP value for the k-th beat. We then solve for the values of m and c by calibrating based on the first beat for each subject as well as the MAP and DBP values from the ensembled beat, so we have the equations:

$$
\begin{aligned}
\mathtt{MAP} &= m \cdot \mathrm{mean}(a_1) + c \\
\mathtt{DBP} &= m \cdot \min(a_1) + c
\end{aligned}
\tag{7}
$$

Solving for m and c and substituting in equation 5, we get the following equation for the calibrated MAP values:

$$
\begin{aligned}
\mathtt{MAP}_k = {} & ((\mathtt{MAP} - \mathtt{DBP}) \cdot \mathrm{mean}(a_k) \\
& + \frac{\mathrm{mean}(a_1) * \mathtt{DBP} - \min(a_1) * \mathtt{MAP}}{\mathrm{mean}(a_1) - \min(a_1)}
\end{aligned}
\tag{8}
$$

Using this method, we produce a single MAP value for each beat for each patient, resulting in a total of 2210 data points overall. We find that the calibrated MAP values can have a large range for a single patient, as shown in Figure 8.

3.8. Evaluation Criteria

To quantitatively compare the performance of different models, we compute the average standard deviation of the prediction error on the test set across all folds as the metric for evaluation. We use this metric instead of the raw prediction error because we can use post-processing methods to correct biases in the MAP predictions.
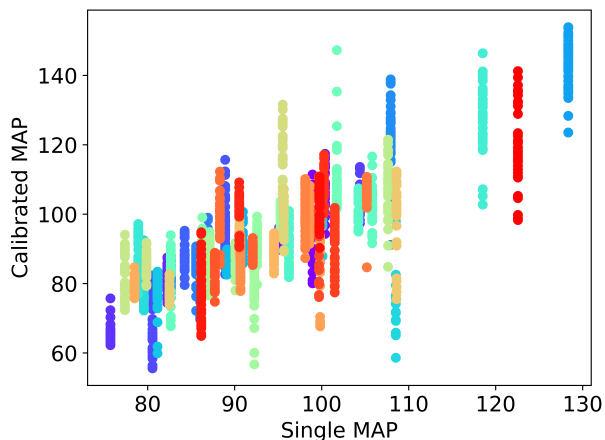

Figure 8: Calibrated MAP values vs. Single (ensembled) MAP value, where each color represents a different subject

4. Evaluation

4.1. Ensembled Beat Results

Using only the two waveforms (blood pressure shape and blood flow velocity), we find that all four models improve upon the anthro-only data model. As shown in figure 9 and table 2, the transformer model achieves an average error standard deviation of 8.7 mmHg, the LSTM model achieves 8.6 mmHg, and the CNN achieves 8.8 mmHg, compared to the anthro-only model which achieves 9.5 mmHg. The MLP model with waveforms as inputs also improves upon the anthro-only model, but to a lesser extent, achieving 9.1 mmHg.

We also experiment with adding anthropometric data (age and heart rate) as features in addition to the two waveforms, to see if this information can further improve the models' performance. With the CNN model, adding these features improves the average error standard deviation from 8.8 mmHg to 7.9 mmHg. The transformer model improves to a lesser degree, decreasing the error standard deviation from 8.7 mmHg to 8.4 mmHg. With the LSTM model, we find that the model does not improve, achieving an average error standard deviation of 8.8 mmHg. The RMSE, mean, and standard deviation of the error are shown in figure 9 and table 2.

Furthermore, to evaluate the extent of overfitting, we show the results of the RMSE, mean error, and the average error standard deviation on the test set compared to the train set in figure 9. We find that
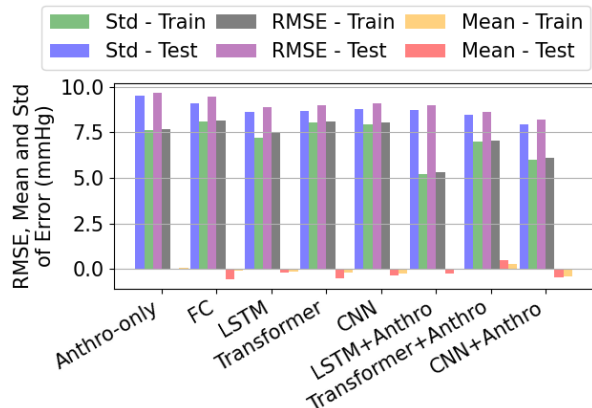
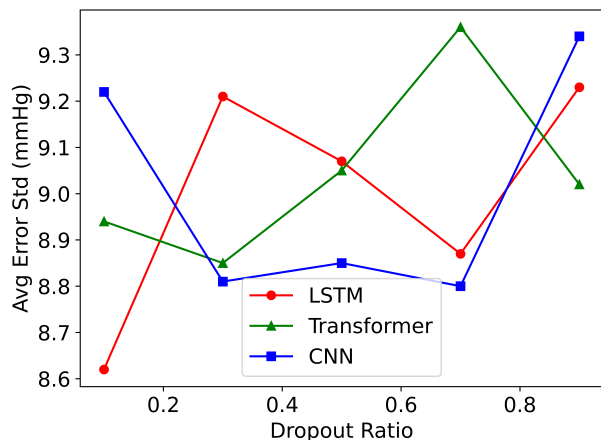Figure 9: RMSE, mean and standard deviation of the error on the test set vs. training set for each model

Figure 11: Average error std (mmHg) vs L2 Norm

Figure 10: Average error std (mmHg) vs dropout ratio
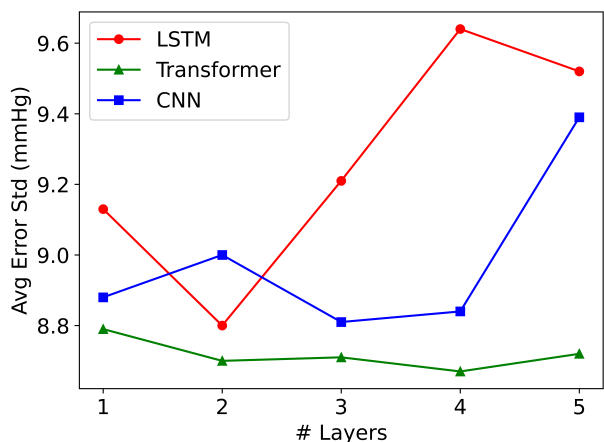
Figure 12: Average error std (mmHg) vs Model Size (# Layers)

while there is overfitting, it is not severe for most models.

4.2. Ablation Studies

4.2.1. Dropout Ratio

We perform an ablation study on the effect of the dropout ratio applied after each layer for the LSTM, Transformer, and CNN models using only waveforms as inputs. We find that the CNN and Transformer models especially tend to benefit from a moderate amount of dropout as shown in Figure 10. Table 3 shows the average error standard deviation for each of the three models with varying dropout ratios.
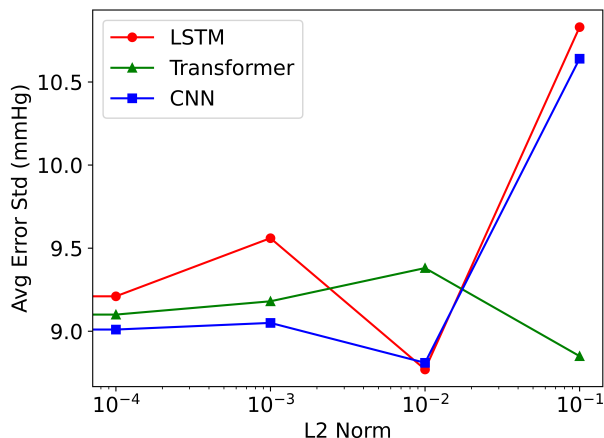
4.2.2. L2 Norm

We also perform an ablation study on the effect of the L2 norm on the three models. We find that the performance improves for all three models with sufficiently large L2 norm as shown in Figure 11. Table 4 shows the average error standard deviation for each of the three models with different L2 norms.

4.2.3. Model Size

We perform an ablation study on the model size for all three models by varying the number of layers in each model. We find that a larger model size does not necessarily improve performance, as shown in Figure 12. Table 5 shows the average error standard deviation and number of parameters for each of the three models with different model sizes.

Table 2: RMSE, Mean and standard deviation of the prediction error for 7 models: Anthro-only, MLP (waveforms only), LSTM (waveforms only), Transformer (waveforms only), CNN (waveforms only), Transformer + anthropometric data, CNN + anthropometric data. The average RMSE, mean error, and std of the error are shown, along with the number of parameters and FLOPs of each model.

Model	Mean Error	Error Std	RMSE	# Params	MFLOPs
Anthro-only	-0.011	**9.51**	9.67	12.9K	0.026
MLP	-0.548	**9.10**	9.45	50.1K	9.98
LSTM	-0.192	**8.62**	8.90	1.58M	317.5
Transformer	-0.481	**8.67**	8.98	3.16M	629.3
CNN	-0.339	**8.81**	9.11	37.3K	7.44
LSTM + anthro	-0.240	**8.75**	8.98	4.73M	317.6
Transformer + anthro	0.478	**8.37**	8.63	3.16M	629.4
CNN + anthro	-0.424	**7.92**	8.20	37.4K	7.48

Table 3: Ablation Study: Dropout Ratio vs. Average Error Std (mmHg)

Dropout	LSTM	Transformer	CNN
0.1	8.62	8.94	9.22
0.3	9.21	8.85	8.81
0.5	9.07	9.05	8.85
0.7	8.87	9.36	8.80
0.9	9.23	9.02	9.34

Table 4: Ablation Study: L2 Norm vs. Average Error Std (mmHg)

L2 Norm	LSTM	Transformer	CNN
0	9.05	9.07	9.13
0.0001	9.21	9.10	9.01
0.001	9.56	9.18	9.05
0.01	8.77	9.38	8.81
0.1	10.83	8.85	10.64

4.2.4. Using Area Waveform As Input

We also try using the arterial area waveform (instead of `bp_shape` waveform), velocity waveform, and anthropometric data (age, heartrate) as inputs. Figure 9 and table 6 shows the average error standard deviation, mean error, and RMSE for each of the 3 models. We find that the CNN and Transformer models still outperform the anthro-only model, achieving 9.1 and 8.9 mmHg std of the error respectively.

4.3. Beat-to-beat Calibration

We analyze the result of using the beat-to-beat calibration method described in section 3.7. To compute the error in a uniform way compared to the ensembled beat setting, we first average out the measurements for each patient to produce a single prediction and target value per patient. We then compute the average standard deviation of the error for these measurements. Table 7 displays the average standard deviation of the error when training and testing on individual calibrated beats. In terms of the raw metrics, we find that the ensembled beat method outperforms the beat-to-beat calibration method. It is likely that this is partially due to the error introduced through the calibration, since it depends heavily on the waveform measurements of the first beat. However, within the predicted values for each patient, the correlation between predicted and ground truth MAP values is quite strong. This is shown in figure 13, which displays the prediction results for a single fold using the LSTM model. This suggests that it may be possible to achieve better results with a more accurate calibration method.

5. Conclusion

We utilize ML algorithms to estimate Mean Arterial Blood Pressure from ultrasound measurements of arterial area and blood flow velocity waveforms. In our work, we use patient data we collected by ourselves. We evaluate the performance of three algorithms: LSTM, Transformer, and CNN, and we compare them to a anthro-only fully-connected model

Table 5: Ablation Study: Model Size (# Params) vs. Average Error Std (mmHg)

# Layers	LSTM		Transformer		CNN	
	#Params	Error Std	#Params	Error Std	#Params	Error Std
1	0.53M	9.13	0.79M	8.79	12.6K	8.88
2	1.05M	8.80	1.58M	8.70	25.0K	9.00
3	1.58M	9.21	2.37M	8.71	37.3K	8.81
4	2.11M	9.64	3.16M	8.67	49.7K	8.84
5	2.63M	9.52	3.94M	8.72	62.0K	9.39

Table 6: Ablation Study: Using area (instead of BP shape) and velocity waveforms + anthropometric data (age, heartrate) as inputs

	Error Std Per Seed			
Model	1	2	3	Avg
LSTM	9.71	9.35	10.29	**9.79**
Transformer	8.98	8.86	8.75	**8.87**
CNN	9.09	9.26	8.90	**9.08**

Table 7: Standard deviation of the error (mmHg) per random seed and on average for each model using calibrated beats and MAP values

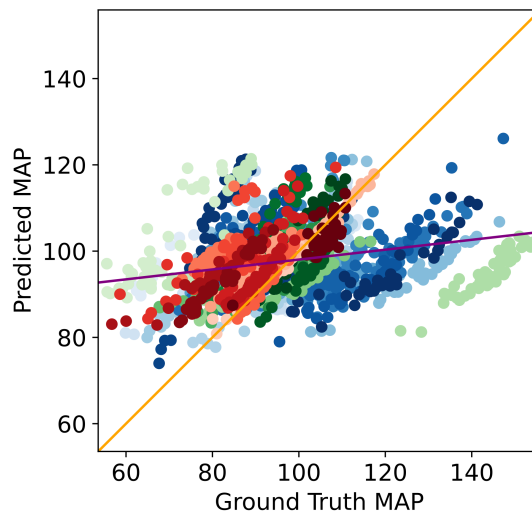|	Error Std Per Seed			
Model	1	2	3	Avg
Anthro-only	13.30	12.81	13.29	**13.13**
LSTM	14.09	13.82	14.02	**13.98**
Transformer	17.70	17.72	17.69	**17.70**
CNN	18.05	17.97	18.13	**18.05**

Figure 13: Prediction results for one fold using the LSTM model with calibrated beats and MAP values. Each color represents a different patient.

trained with only anthropometric data. Compared to the average error standard deviation of 9.5 mmHg with the anthro-only model, the LSTM with only waveform data can achieve an average error standard deviation of 8.6 mmHg, and the CNN algorithm with both waveform and anthropometric data can achieve an average error standard deviation of 7.9 mmHg.

Acknowledgments

References

Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018.

John Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological measurement*, 28(3):R1, 2007.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

Denis Chemla, Jean-Louis Hébert, Catherine Coirault, Karen Zamani, Isabelle Suard, Patrice Colin, and Yves Lecarpentier. Total arterial compliance estimated by stroke volume-to-aortic pulse pressure ratio in humans. *American Journal of Physiology-Heart and Circulatory Physiology*, 274(2):H500–H505, 1998.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Aniekan Essien and Cinzia Giannetti. A deep learning framework for univariate time series prediction using convolutional lstm stacked autoencoders. In *2019 IEEE International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)*, pages 1–6, 2019.

Kevin Fauvel, Tao Lin, Véronique Masson, Élisa Fromont, and Alexandre Termier. Xcm: An explainable convolutional neural network for multivariate time series classification. *Mathematics*, 9 (23), 2021. ISSN 2227-7390.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

Jake Grigsby, Zhe Wang, and Yanjun Qi. Long-range transformers for dynamic spatiotemporal forecasting, 2021.

Jeanne Harabedian. *Modeling the Arterial System to Improve Ultrasound Measurements of Hemodynamic Parameters*. PhD thesis, Massachusetts Institute of Technology, 2022.

Simon Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.

Brian L Hill, Nadav Rakocz, Ákos Rudas, Jeffrey N Chiang, Sidong Wang, Ira Hofer, Maxime Cannesson, and Eran Halperin. Imputation of the continuous arterial line blood pressure waveform from noninvasive measurements using deep learning. *Scientific reports*, 11(1):15755, 2021.

Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8): 1735–1780, 11 1997.

LR John. Forward electrical transmission line model of the human arterial system. *Medical and Biological Engineering and Computing*, 42:312–321, 2004.

Sangyeon Kim and Myungjoo Kang. Financial series prediction using attention lstm, 2019.

Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*. 2015.

Panicos A Kyriacou and James M May. Photoplethysmography: New trends and future directions. In *Photoplethysmography*, pages 469–487. Elsevier, 2022.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision. *arXiv preprint arXiv:2212.04356*, 2022.

Marc Rußwurm and Marco Körner. Self-attention for raw optical satellite time series classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169:421–435, 11 2020.

Joohyun Seo, Sabino J Pietrangelo, Hae-Seung Lee, and Charles G Sodini. Noninvasive arterial blood pressure waveform monitoring using two-element ultrasound system. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 62(4): 776–784, 2015.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, page 5998–6008. 2017.

Chonghe Wang, Xiaoshi Li, Hongjie Hu, Lin Zhang, Zhenlong Huang, Muyang Lin, Zhuorui Zhang, Zhenan Yin, Brady Huang, Hua Gong, et al. Monitoring of the central blood pressure waveform via a conformal ultrasonic device. *Nature biomedical engineering*, 2(9):687–695, 2018.

Chonghe Wang, Baiyan Qi, Muyang Lin, Zhuorui Zhang, Mitsutoshi Makihata, Boyu Liu, Sai Zhou, Yi-hsi Huang, Hongjie Hu, Yue Gu, et al. Continuous monitoring of deep-tissue haemodynamics with stretchable ultrasonic phased arrays. *Nature biomedical engineering*, 5(7):749–758, 2021a.

Hanrui Wang, Zhanghao Wu, Zhijian Liu, Han Cai, Ligeng Zhu, Chuang Gan, and Song Han. Hat: Hardware-aware transformers for efficient natural language processing. In *Annual Conference of the Association for Computational Linguistics*, 2020a.

Hanrui Wang, Zhekai Zhang, and Song Han. Spatten: Efficient sparse attention architecture with cascade token and head pruning. In *2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, pages 97–110. IEEE, 2021b.

Hanrui Wang, Pengyu Liu, Jinglei Cheng, Zhiding Liang, Jiaqi Gu, Zirui Li, Yongshan Ding, Weiwen Jiang, Yiyu Shi, Xuehai Qian, et al. Quest: Graph transformer for quantum circuit reliability estimation. *arXiv preprint arXiv:2210.16724*, 2022.

Hanrui Wang et al. *Efficient algorithms and hardware for natural language processing*. PhD thesis, Massachusetts Institute of Technology, 2020b.

Junlu Wang, Su Li, Wanting Ji, Tian Jiang, and Baoyan Song. A t-cnn time series classification method, 11 2021c.

Matthew Ward and Jeremy A Langton. Blood pressure measurement. *Continuing education in anaesthesia, critical care & pain*, 7(4):122–126, 2007.

Harya Widiputra, Adele Mailangkay, and Elliana Gautama. Multivariate cnn-lstm model for multiple parallel financial time-series prediction. *Complexity*, 2021, 10 2021.

Neo Wu, Bradley Green, Xue Ben, and Shawn O'Banion. Deep transformer models for time series forecasting: The influenza prevalence case, 2020.

Xiaoman Xing and Mingshan Sun. Optical blood pressure estimation with photoplethysmography and fft-based neural networks. *Biomedical optics express*, 7(8):3007–3020, 2016.

Chao-Han Huck Yang, Yun-Yun Tsai, and Pin-Yu Chen. Voice2series: Reprogramming acoustic models for time series classification. 2021.

Zhekai Zhang, Hanrui Wang, Song Han, and William J Dally. Sparch: Efficient architecture for sparse matrix multiplication. In *2020 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, pages 261–274. IEEE, 2020.

Bendong Zhao, Huanzhang Lu, Shangfeng Chen, Junliang Liu, and Dongya Wu. Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics*, 28(1):162–169, 2017.