Estimation of Systolic and Diastolic Blood Pressure for Hypertension Identification from Photoplethysmography Signals

Hygo Sousa De Oliveira, Rafael Albuquerque Pinto, Eduardo James Pereira Souto * and Rafael Giusti

Abstract: Continuous monitoring plays a crucial role in diagnosing hypertension, characterized by the increase in Arterial Blood Pressure (ABP). The gold-standard method for obtaining ABP involves the uncomfortable and invasive technique of cannulation. Conversely, ABP can be acquired non-invasively by using Photoplethysmography (PPG). This non-invasive approach offers the advantage of continuous BP monitoring outside a hospital setting and can be implemented in cost-effective wearable devices. PPG and ABP signals differ in scale values, which creates a non-linear relationship, opening avenues for the utilization of algorithms capable of detecting non-linear associations. In this study, we introduce Neural Model of Blood Pressure (NeuBP), which estimates systolic and diastolic values from PPG signals. The problem is treated as a binary classification task, distinguishing between Normotensive and Hypertensive categories. Furthermore, our research investigates NeuBP’s performance in classifying different BP categories, including Normotensive, Prehypertensive, Grade 1 Hypertensive, and Grade 2 Hypertensive cases. We evaluate our proposed method by using data from the publicly available MIMIC-III database. The experimental results demonstrate that NeuBP achieves results comparable to more complex models with fewer parameters. The mean absolute errors for systolic and diastolic values are 5.02 mmHg and 3.11 mmHg, respectively.

Keywords: photoplethysmograph; arterial blood pressure; dilated convolution; hypertension; continuous monitoring

1. Introduction

Cardiovascular diseases, particularly hypertension, pose a significant threat to overall well-being and quality of life, often leading to severe health complications, such as strokes and heart attacks [1,2]. The diagnosis of hypertension usually relies on specific criteria, with Systolic Blood Pressure (SBP) \( \geq 140 \) mmHg and Diastolic Blood Pressure (DBP) \( \geq 90 \) mmHg serving as indicative thresholds [3].

Detecting hypertension early is challenging due to its often asymptomatic nature, and its symptoms, such as headaches, chest pains, and shortness of breath [4], tend to be intermittent. Hypertension is often referred to as a “silent killer” by medical professionals due to its potentially fatal nature [5,6]. For this reason, the continuous monitoring of blood pressure is recognized as the most effective means of addressing the morbidity associated with hypertension [7].

Blood pressure monitoring methods include invasive and non-invasive techniques. While invasive methods are highly precise, they are medically complex and present risks such as infection, in addition to causing discomfort to the patient during cuff inflation and deflation [8,9]. Non-invasive methods make use of equipment like oscillometric and auscultatory devices, such as sphygmomanometers [10], as well as signals collected by devices such as fiber optic sensors [11] and force-sensitive electromechanical film sensors [12].

For blood pressure obtained by activity cardiovascular monitoring, there are two primary methods to detect changes in heart electrical activity and microvascular blood
flow: Electrocardiography (ECG) and Photoplethysmography (PPG). ECG signals are captured through devices employing either disposable electrodes placed on the torso or textile electrodes, similar to those used in conventional heart rate monitors [13]. This technique plays a pivotal role in diagnosing a wide range of cardiac arrhythmias [14].

In comparison, PPG signals are obtained non-invasively at a low cost, aiming to measure changes in blood flow through light absorption and reflection on the skin [15]. In clinical settings, PPG signals are found in pulse oximeters (SpO2) used to measure oxygen saturation. Data provided by these devices enable the extraction of essential physiological parameters, including heart rate, respiratory rate, and tissue/vascular perfusion, enabling the identification of heart conditions, all without the need for invasive blood pressure measurements [16].

Recent technological advancements have paved the way for the development of wearable devices, with a prominent focus on the utilization of PPG sensors capable of capturing and processing human biosignals. These sensors can be seamlessly integrated into a range of wearable accessories, including wristbands, watches, eyewear, headphones, belts, and smart jewelry [17]. In addition, there is a promising avenue that utilizes PPG signals for the estimation of Arterial Blood Pressure (ABP), a parameter traditionally associated with invasive techniques. PPG signals inherently exhibit a close relationship with ABP due to the fluctuations in blood flow. This intrinsic connection allows for the estimation of ABP through the inherent characteristics of PPG signals, achieved through the application of statistical techniques or machine learning algorithms [18,19].

ABP estimation through PPG signals adheres to criteria set by the Association for the Advancement of Medical Instrumentation. Methods include the calculation of Pulse Transit Time (PTT), representing the interval of time taken by blood to reach venous terminations during a heartbeat [20], and of Pulse Wave Velocity (PWV), associated with blood speed in vessels [21].

The related literature highlights the existence of various proposals employing modules for estimating blood pressure through PPG signals [22–24]. However, most of these approaches rely on manual feature extraction processes. In this study, we introduce a deep neural network architecture, named Neural Model of Blood Pressure (NeuBP), meticulously designed to automatically extract crucial features from PPG signals and estimate the ABP signal. The absence of manual feature extraction steps represents a significant advantage, simplifying and expediting the inference process while reducing the potential for human error.

NeuBP incorporates dilated convolutional blocks, allowing for the effective and comprehensive capture of the relationships among these features, increasing the precision level in blood pressure estimation.

In contrast to previous approaches, the evaluation of the proposed model addresses not only the binary question of the presence or absence of hypertension but also ventures into the multiclass classification of different blood pressure categories, including normotension, prehypertension, and hypertension of Grades 1 and 2. This expansion of the scope of analysis reflects the commitment to providing a robust and comprehensive tool for cardiovascular state assessment.

The rigorous evaluation of NeuBP, conducted using data from the renowned MIMIC-III database, validates its effectiveness and reliability. Our experimental results demonstrate not only NeuBP’s ability to rival more complex models, like the one proposed by [19], but also its potential to stand out as a viable and accessible solution for continuous blood pressure monitoring.

Key contributions of this study encompass the following:

- Automatic feature extraction: Unlike existing approaches that rely on manual feature extraction processes, Neural Model of Blood Pressure (NeuBP), introduced in this work, is a deep neural network architecture meticulously designed to automatically extract crucial features from PPG signals and estimate the ABP signal. The absence of
manual feature extraction steps significantly simplifies and expedites the inference process while reducing the potential for human error.

- Alignment of ABP and PPG signals through noise removal: We implement advanced signal processing techniques to effectively eliminate noise and ensure precise alignment between ABP and PPG signals. This synchronization of PPG-ABP segments, paired with the identification of blood pressure labels within the median of systolic and diastolic ABP segments, enhances the accuracy of our model.

- Utilization of dilated convolutional blocks: NeuBP incorporates dilated convolutional blocks, allowing for the effective capture of relationships among features extracted from convolutional layers. This enhances the model’s ability to discern complex patterns in the data, leading to improved precision in blood pressure estimation.

- Comprehensive evaluation: Our experimental results not only demonstrate NeuBP’s ability to rival more complex models but also its potential as a viable and accessible solution for continuous blood pressure monitoring. Additionally, our evaluation not only addresses the binary question of the presence or absence of hypertension but also ventures into multiclass classification of different blood pressure categories, reflecting our commitment to providing a robust tool for cardiovascular state assessment.

The remainder of this article has the following structure: Section 2 explores related works. Next, we present our proposed method in Section 3. Section 4 details the experimentation, covering the training method, evaluation metrics, results, and discussions. Finally, Section 5 provides a comprehensive conclusion, summarizing key findings and outlining potential avenues for future research.

2. Related Works

In the literature, distinct research methodologies for processing and classifying hypertension via PPG have emerged. These approaches can be broadly categorized into two prevalent paradigms for predicting systolic and diastolic values. The first paradigm relies on the application of classical models that utilize statistical data extracted from both public and private datasets. In contrast, the second approach is characterized by studies employing deep learning techniques for the automatic extraction of temporal features from PPG time series.

Zhang and Feng [25] introduced a support vector machine (SVM) approach to estimate ABP using PPG signals and the UQVS database. They partitioned the dataset into training and test subsets, maintaining a non-personalized separation by avoiding data overlap from the same patient in both sets. The evaluation of blood pressure relied on systolic and diastolic points, with metrics including absolute error and relative error. The reported results for systole and diastole were $11.641 \pm 8.202$ mmHg and $7.617 \pm 6.783$ mmHg, respectively.

Xie et al. [26] proposed the evaluation of models for estimating ABP in real time based on linear regression, neural networks, decision trees, bagging, and random forests. The experiments used the University of Queensland Vital Signs (UQVS) database [27], consisting of 11,492 PPG signal records. A random forest model attained errors of $7.59$ mmHg ± $5.39$ mmHg in DBP and $4.21$ mmHg ± $3.24$ mmHg in SBP (absolute mean ± standard deviation).

Kurylyak et al. [28] presented a study on continuous blood pressure estimation using PPG signals. They applied a neural network algorithm to predict systolic and diastolic values, utilizing data from the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) database [29]. Validation involved a holdout approach with data being split as 70:15:15 into training, test, and validation sets, respectively. The neural model achieved mean absolute error (MAE) and standard deviation values for systole and diastole of $3.80 \pm 3.46$ mmHg and $2.21 \pm 2.09$ mmHg, respectively. It is important to note the utilization of a small undisclosed subset of the MIMIC database, comprising over 150,000 cycles of PPG signals with corresponding blood pressure defined by the ABP signal.

Liang et al. [30] proposed the use of deep learning methods to classify blood pressure levels from PPG signals, aiming to detect cardiovascular diseases associated with hyper-
tension in an early stage. They employed data from 121 patients in the publicly available MIMIC III database [29] and transformed temporal data into scalograms using Continuous Wavelet Transform (CWT). During model training, the authors applied transfer learning from the GoogLeNet model to classify ABP levels into Normotensive (NT), Prehypertensive (PHT), and Hypertensive (HT) categories, evaluated using the F1-Score metric. The results demonstrated accuracy of 80.52% for NT versus PHT, 92.55% for NT versus HT, and 82.95% for NT + PHT versus HT.

Slapničar et al. [31] devised a method for estimating blood pressure from PPG signals using machine learning algorithms to select relevant features from statistical data extracted in the time and frequency domains of the PPG signal. The authors employed two configurations for validating their method: k-fold cross-validation (with five subsets) and Leave One Subject Out (LOSO). In the 5-fold validation, the random forest model yielded the best results, with a mean error of 4.90 ± 6.59 mmHg in SBP and 2.21 mmHg ± 3.70 mmHg in DBP. In the LOSO validation, the values were 4.47 ± 5.85 mmHg for systole and 2.02 mmHg ± 2.94 mmHg for diastole.

Slapničar et al. [19] achieved outstanding results by using the explored Spectro-Temporal method for blood pressure estimation on MIMIC-III data [29]. They employed a deep neural network model with three input channels (pure PPG signal, first derivative, and second derivative) in 5 s windows and 786.875 parameters. LOSO validation (with and without personalization) was applied, resulting in the customized ResNet model, which achieved the lowest mean absolute error, with values of 9.43 mmHg for systole and 6.88 mmHg for diastole.

Athaya et al. [2] developed a U-net deep learning architecture that utilizes fingertip PPG signals for non-invasive estimation of the Arterial Blood Pressure (ABP) waveform. The authors collected data about 100 subjects from MIMIC-III database. Using their deep learning model, the tool converted the PPG signal from milliVolts (mV) to ABP measured in mmHg. The mean absolute error metrics achieved for SBP were 3.68 ± 4.42 mmHg, and for DBP were 1.97 ± 2.92 mmHg.

Harfiya et al. [32] proposed a deep learning model to learn how to perform signal-to-signal translation from PPG to Arterial Blood Pressure (ABP). The authors obtained a translated ABP signal, which extracted the SBP and DBP values accordingly to ease the comparative evaluation. With this approach, they achieved an average absolute error below 5 mmHg, with 70% confidence for SBP and 95% confidence for DBP without complex feature engineering.

Qin et al. [33] developed a convolution-based deep autoencoder (DAE) model with the application of multi-domain adversarial training to learn cross-domain features. This approach enabled the authors to achieve a mean absolute error (MAE) of 7.945 and 4.114 for uncalibrated DAE in systolic BP (SBP) and diastolic BP (DBP), respectively. After using 80 s of data for calibration, the MAE of DAE was reduced to 5.424 and 3.144 accordingly.

Nabil et al. [34] created PPG2ABP for non-invasively estimating the continuous Arterial Blood Pressure (ABP) waveform by utilizing Photoplethysmography (PPG) signals. The authors used a two-stage cascaded deep learning-based method to estimate the continuous ABP. To validate the developed tool, the authors applied the mean absolute error metric and obtained values for SBP and DBP of 5.727 ± 9.162 mmHg and 3.449 ± 6.147 mmHg, respectively.

In contrast to the aforementioned approaches, the aim of this study is to develop a neural network-based learning model that takes only the SpO2 signal as input. Furthermore, it is not guaranteed that the training subset is free from leakage of patient data contained in the test and validation subsets. For the proposed method, the dataset used in the experimental protocol was completely partitioned based on patient data, isolating the training, test, and validation subsets.
3. Proposed Method

Our approach aims to discern the behavioral correlation between ABP and PPG sensor data, facilitating the estimation of both systolic and diastolic phases. This process centers around Neural Model of Blood Pressure (NeuBP), depicted in Figure 1. The proposed method encompasses four fundamental stages: (I) data acquisition; (II) pre-processing; (III) construction of a deep neural model featuring dilation between convolutional layers [35]; and (IV) classification, where we ascertain whether the blood pressure classes are aligned with the predicted values.

![Figure 1](image)

**Figure 1.** Overview of the NeuBP Method Generation Process, which is divided into four steps: (I) data acquisition, (II) preprocessing, (III) application of neural model, and (IV) classification of ABP after prediction of systole and diastole labels.

3.1. Database Description

This study utilized patient data sourced from the publicly available MIMIC-III (Medical Information Mart for Intensive Care) database [29], including both PPG and ABP signals. MIMIC-III is a comprehensive database containing diverse information on patients in intensive care units, comprising vital signs, medications, laboratory measurements, clinical notes, fluid balance, procedural and diagnostic codes, imaging reports, hospital stay durations, and survival data. It encompasses records from 53,423 patient admissions, representing data from 38,597 distinct patients, with a signal frequency set at 125 Hz.

Data extraction was performed using two clinical devices: the Philips CareVue Clinical Information System (models M2331A and M1215A; Philips Healthcare, Andover, MA, USA) and iMDsoft MetaVision ICU (iMDsoft, Needham, MA, USA). These devices facilitated continuous collection of physiological measurements such as heart rate, Arterial Blood Pressure, respiratory rate, and SpO2.

Due to the extensive size of the database and its computational demands, we selected a subset of 1629 users for analysis. Within this subset, sample windows of ABP data containing systolic and diastolic values within the intervals of 75 to 165 mmHg and 40 to 80 mmHg were excluded to mitigate outlier effects on the signals and facilitate correlation with hypertension classes.
3.2. Data Pre-Processing

Biological signals often harbor motion-induced artifacts, causing distortions in the signal. To mitigate signal distortion in the PPG data, a Butterworth filter technique for waveform smoothing, using the zero-phase filter FiltFilt, was employed. This filtering technique ensures the preservation of the signal trend behavior by facilitating low-pass and high-pass frequency cutoffs. Specifically, a fourth-order Butterworth filter was employed, with low-pass and high-pass cutoffs set at 0.5 Hz and 8 Hz, respectively. Conversely, due to the ABP signal’s invasive and sensitive waveform nature, the raw data were retained to facilitate systolic and diastolic point extraction.

Following signal filtering, the PPG signals underwent zero-mean and standard deviation scaling normalization (Equation (1)). This normalization strategy aims to minimize potential signal distortions within the segments by adjusting their scale and highlighting possible deviations from patterns observed in normal blood pressure, particularly in cases of hypertension. In Equation (1), \( X_i \) represents the current segment, \( \bar{x}_i \) stands for the mean value, \( n \) denotes the number of samples, \( \sigma \) is the standard deviation, and \( (X_{n(i)}) \) represents the adjusted segment. This pre-processing step ensures that the input signals are appropriately scaled and free from motion-induced artifacts, enhancing the robustness and effectiveness of the subsequent stages in NeuBP.

\[
\sigma = \sqrt{\frac{\sum (x_i - \bar{x}_i)^2}{n}}
\]

\[
X_{n(i)} = \frac{X_i - \bar{x}_i}{\sigma}
\]  

(1)

An important task to undertake when utilizing machine learning algorithms is to determine the input dimension for the intended model. Biosignal data commonly possess vector dimensions, and the length of the vector determines the number of data that feed into the model. Therefore, since datasets often contain long samples that can impede model learning, it is common to use subsamples of these data derived with sliding window techniques. These techniques may incorporate an overlapping element applied to preserve information between windows and increase the number of observations, as depicted in Figure 2.

Figure 2. Segmentation with sliding window and 50% overlap of two time series. The windows are the non-grayed areas.

Both the ABP and PPG signals undergo the segmentation process illustrated in Figure 2. However, the systole (peaks from a part of a time series) and diastole (valley from
a part of a time series) extraction step is applied exclusively to the ABP signal. During each segmentation of the ABP, tuples containing values extracted from the systole and diastole of the segmented time series are stored. This procedure facilitates the computation of mean systole and diastole values, establishing a connection with the corresponding PPG segment. Consequently, this process is crucial to associating a PPG segment with the target values of systole and diastole.

### 3.3. Deep Learning Architecture

To generate systolic and diastolic estimates correlated with the PPG signal, this study explores a deep neural network approach, utilizing feature extraction through vector dimension convolution techniques [36]. Convolutional layers in neural networks mimic the functionality of the Fourier transform, enabling the automatic extraction of frequency characteristics in multiple dimensions. This allows each dimension to capture the most representative values of the highlighted data.

As seen in Equation (2), convolution is a type of correlation between two functions, $f$ and $g$, as one function “traverses” over the other shifted around the $y$-axis. Convolution is used in machine learning as a means to apply custom filters, which are automatically derived from the training data and function as feature extractors [37].

$$f(x) * g(x) = \int_{-\infty}^{\infty} f(\tau)g(x-\tau)d\tau$$  \hspace{1cm} (2)

Within this approach, the dilation technique was incorporated into the model, thus giving it the ability to store characteristics over time in spatial subsamples related to convolutional layers. This dilation mechanism proves valuable by expanding the receptive field without significantly increasing computational costs [38]. In dilated convolution, filters are applied to areas larger than their lengths, skipping values according to their dilation step. This behavior enables a dilated convolutional network to operate on larger scales compared with a regular convolutional network, achieving a similar effect through pooling or strides (the steps that the filter takes to traverse the receptive field).

This causes dimensionality reduction as data go from layer to layer in the neural network. Figure 3 illustrates the behavior of this network, highlighting the impact of dilations $D_1$, $D_2$, $D_4$, and $D_8$. These dilations lead to the generation of an output (S), which is intricately connected to a block featuring respective convolutional layers.

The dilation process unfolds in the following manner: Within a convolutional block, layers are structured with neuron spacings determined by the assigned dilation level. For example, dilation $D_1$ exhibits a spacing of 1 between its neurons. Subsequently, the data extracted by $D_1$ (level 0) are also processed by $D_2$ (level 1), introducing sparser information extraction with a neuron step of 2. The subsequent level, $D_4$ (level 2), carries a dilation of 4, resulting in a spacing of 4 between neurons. This progression continues until reaching the S output layer. The exponential behavior of dilation is repeated for the other dilation levels, fostering the extraction of diverse features within the block's output for the regression of systolic and diastolic points.

The neural model employed in the proposed method is depicted in Figure 3 and consists of following layers: 4 blocks with dilations mechanisms of 8, 6, 4, and 4, accompanied by corresponding receptive filters of 32, 32, 64, and 64; maximum pooling over vector kernels of length 2; normalization; spatial dropout of 50%; global max pooling; and dense layers with 1 neuron in the output.

The dilation block, prior to the global max pooling layer, is connected to the composite feature map extraction layers presented at the top of the image. The output of the neural model consists of two dense neurons with linear output for prediction purposes. NeuBP involves a total of 246.690 final parameters.
Figure 3. Processing core of NeuBP. There are the following steps. (1) Input data; (2) blocks of residual convs with dilatation technique; (3) feature maps with max pooling, normalization, and spatial dropout.

4. Experiments and Results

To optimize the training and evaluation of the model, three distinct configurations were employed:

1. Generalist model (GM): This model is designed to predict systolic and diastolic values using data not encountered during the training phase.
2. Specialist model with retraining (SMR): Built upon the generalist model, this approach involves retraining on a subset of data from a specific user for evaluation during the test phase.
3. Specialist model without retraining (SM): This model starts from a random point in its weights, without leveraging pre-trained weights.

For the GM approach, data partitioning consisted of 1,000,000 samples divided into 7 segments extracted from 977 individuals for training, from 326 individuals for validation, and from another 326 individuals for testing.

In the specialist model training scenario, holdout validation was employed, with 80% of data from an individual being used in the training phase, 10% for validation, and the remaining 10% for testing. This model type is justified by the unique frequency of the biological signal emitted by each individual, making it challenging for a model to distinguish among hypertension categories in different individuals.

By using the same testing environments, the proposed model, NeuBP, was compared to the Spectro-Temporal model as proposed by Slapničar et al. [19].

4.1. Training Method

The models were developed and trained using Python 3.9.12 and the TensorFlow framework (version 2.4.1). In the neural network training phase, weights were initialized in two ways: randomly without pre-training and with pre-training as per the specified scenario. The training process extended over 60 epochs, employing a batch size of 32, the Adam optimizer, and a learning rate set to 0.003.

To ensure optimal model preservation, ModelCheckpoint callbacks consistently saved the best model during training. Additionally, an early stopping callback with a patience of
10 epochs prevented training from exceeding the maximum number of epochs if weights did not exhibit improved adjustment within 10 epochs. The evaluation metrics for early stopping included mean absolute error and mean square error.

4.2. Evaluation Metrics

The assessment of systolic and diastolic value predictions relies on two key metrics: mean absolute error (MAE) and mean square error (MSE).

The MAE gauges the linear disparity between desired and predicted values, as calculated by Equation (3):

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |x_i - y_i|
\]  

(3)

In contrast, the MSE introduces a squared penalty for deviations between target and predicted values, as defined by Equation (4):

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2
\]  

(4)

4.3. Results and Discussion

4.3.1. Generalist Model (GM)

The primary objective of the generalist evaluation was to develop a model based on data from a diverse set of individuals, encompassing the training, test, and validation sets. Figure 4 visually presents the distribution of data about predicted values (Pred) against the target values (True) for the systole (Sis) and diastole (Dias) classes, within the range of 40 mmHg to 200 mmHg.

![Figure 4](image)

Dispersive analysis of values predicted by the GM experiment in relation to actual values in relation about NeuBP and Spectro-Temporal [31] models.

The distribution of true data for the systole category showed an interquartile range (IQR) in which the first quartile (Q1) was \(\approx 104\) mmHg, the second quartile (Q2) or median was very close to 120 mmHg, and the third quartile (Q3) was \(\approx 139\) mmHg. The interquartile range (IQR) \(\approx 35\) mmHg, and the minimum and maximum values corresponded to approximately 58 mmHg and 183 mmHg, respectively. Additionally, there were outliers both below the lower limit and above the upper limit.

Next, the distribution of true data for the diastole category also showed symmetry in the data, with Q1 \(\approx 57\) mmHg; Q2 = 60 mmHg, Q3 \(\approx 63\) mmHg; IQR \(\approx 6\) mmHg; and minimum and maximum values corresponding to 40 mmHg and 84 mmHg, respectively. Moreover, there were outliers above the upper limit of this distribution.

Once the distribution of true data had been observed, the data predicted by NeuBP and the Spectro-Temporal [19] model helped us understand the ability of each model to reproduce the distributions of systole and diastole data in each of the presented experiments.
When observing the predictions made by our model, NeuBP, for the systole category, it is evident that the model maintained the symmetry of true data while presenting lower dispersion, resulting in quartiles $Q_1 = 118$ mmHg, $Q_2 \approx 123$ mmHg, and $Q_3 \approx 130$, showing very close matching, resulting in an IQR interval of $\approx 12$ mmHg. As the minimum and maximum values also followed the flattening of dispersion, many outliers could be noticed, but still within the interval present in the true data. Next, when evaluating the same systole category predicted by the Spectro-Temporal model, the box plot showed positive symmetry, the same $Q_2$ value concerning the true median, and an IQR dispersion of $\approx 20$ mmHg; however, there were many outliers above the upper limit.

When observing the predicted values of the diastole category in relation to the true values, we saw that both models also presented less dispersion: NeuBP with an IQR value of $\approx 7$ mmHg and the Spectro-Temporal model with an IQR of $\approx 8$ mmHg. Both pointed to a $Q_2$ value of 60 mmHg, and the Spectro-Temporal model presented more outliers than NeuBP.

Table 1 presents key metrics, including mean, standard deviation, and mean absolute error (MAE), offering a comparative assessment of NeuBP and the Spectro-Temporal model against the test dataset’s standard deviation and mean. Notably, NeuBP achieved an MAE of 17.69 mmHg for systole and 8.90 mmHg for diastole on the test set. The Table comprehensively displays statistics for systolic and diastolic points from the test dataset, alongside corresponding predictions made by the models in this study.

Table 1. Arithmetic mean, standard deviation and absolute mean error in the GM experiment.

<table>
<thead>
<tr>
<th>Target</th>
<th>NeuBP</th>
<th>Spectro-Temporal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Systole</td>
<td>Diastole</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>23.16</td>
<td>11.68</td>
</tr>
<tr>
<td>Mean</td>
<td>119.0</td>
<td>59.0</td>
</tr>
<tr>
<td>MAE [mmHg]</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

To evaluate the alignment of predicted values with hypertension classes, two classification steps were implemented, conditioning the values to blood pressure intervals defined by the International Hypertension Association [3]. These steps involved (1) classification into four classes (Normotensive, Prehypertensive, Hypertensive Stage 1, and Hypertensive Stage 2) and (2) binary classification (Normotensive and Hypertensive). Figure 5 presents the multiclass confusion matrix and its associated metrics. Both models exhibited a bias towards the Normotensive category, aligning with the median described in the block diagrams.

According to the International Society of Hypertension, blood pressure is considered hypertensive when systolic and diastolic values are in the ranges of $\geq 149$ and $\geq 90$ mmHg, respectively. For binary classification assessing whether the predicted values indicated a clinical picture of hypertension, the Normotensive and Hypertensive classes in Figure 6 were binarized, revealing the confusion matrix and relevant metrics.

Despite binarization, both models achieved the same F1-Score metric for the Hypertensive and Normotensive classes. NeuBP, with fewer parameters, demonstrated superior results in precision and sensitivity metrics. However, generalist models with holdout validation per person exhibit poor performance due to the signal specificity of PPG and ABP for each individual, akin to human fingerprints, making it a challenging problem to solve with this approach alone.
Figure 5. Confusion matrix of hypertension classes achieved with values predicted by NeuBP and Spectro-Temporal model. The highlight values represents the best values reached between models.

Figure 6. Confusion matrix of the values predicted in the GM experiment in relation to the true values in the binary classes. The highlight values represents the best values reached between models.

A more effective strategy for handling signal specificity involves specialist models. These models, with data split based on a single individual, adjust their weights relative to that individual. The subsequent evaluations provide results for specialist model scenarios.
4.3.2. Specialist Model with Retraining (SMR)

For this evaluation scenario, the data split was 80:10:10 for training, validation, and testing, respectively. Initially, 50 users from the MIMIC-III database were selected; however, only 46 users were included in the experiment, meeting the minimum requirement of 9 samples or 63 s for testing and validation. It is essential to note that specialist models were generated based on the pre-trained model from the previous scenario, which constituted the generalist model.

Regarding both NeuBp and the Spectral Temporal model, the dispersions about true and predicted outcomes are shown in Figure 7. The models closely matched the test data distribution pattern, as depicted in Figure 7. Given that the models run on data with consistent patterns within training and test subsets, all of which belong to a single user, this alignment was expected.

![Figure 7](image)

**Figure 7.** Sys—systolic range; Dias—diastolic range; True—target values; Pred—predicted values. Dispersion analysis of the values predicted in the SMR experiment in relation to the true values.

The true dataset for the systole class showed a slight, positive asymmetry, where the median was closer to Q1. It exhibited Q1 = 102 mmHg; Q2 ≈ 116 mmHg; Q3 ≈ 139 mmHg; IQR ≈ 37 mmHg; and lower and upper limits of ≈ 169 mmHg and ≈ 191 mmHg, respectively, along with some outliers in its upper limit. For this category, NeuBP presented identical values for Q1 and Q2, a smaller Q3 of ≈ 137 mmHg, and an IQR of 37 mmHg. The lower and upper limits practically had the same value, showing a predicted data dispersion very close to true data. On the other hand, the Spectro-Temporal model for the systole category also showed predicted values close to the true data, with only a slightly smaller lower tail and more pronounced outliers compared with the upper tail of the box plot.

The same analysis results were observed for the diastole category. Our model, NeuBP, showed a dispersion very close to the true data, with a predicted Q1 of ≈ 65 mmHg, compared with the true Q1 of ≈ 63 mmHg. Both predicted and true values of Q2 and Q3 overlapped, where Q2 was ≈ 61 mmHg and Q3 was ≈ 70 mmHg, resulting in an IQR of ≈ 7 mmHg. When evaluating the Spectro-Temporal model, this model presented a predicted median (Q2) of approximately 1 mmHg above the true values, as well as a slightly higher Q3, with an IQR of ≈ 8 mmHg. The model also showed more pronounced outliers compared with the upper tail of the box plot.

The mean, standard deviation, and MAE values in Table 2 indicate a notable reduction in error values for both models, with NeuBP exhibiting a superior MAE due to its smaller value. Concerning the predicted values for systole, Spectro-Temporal closely approximated NeuBP. Conversely, Spectro-Temporal demonstrated a closer approximation to diastolic values. Examining the means showed that both models presented identical integer values.
Table 2. Arithmetic mean, standard deviation and absolute mean error in the SMR experiment.

<table>
<thead>
<tr>
<th></th>
<th>Target Spectro-Temporal</th>
<th>NeuBP Spectro-Temporal</th>
<th>Standard deviation</th>
<th>Mean</th>
<th>MAE [mmHg]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Systole</td>
<td>Diastole</td>
<td>Systole</td>
<td>Diastole</td>
<td>Systole</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>23.18</td>
<td>12.99</td>
<td>21.94</td>
<td>12.28</td>
<td>22.07</td>
</tr>
<tr>
<td>Mean</td>
<td>116.0</td>
<td>62.0</td>
<td>116.38</td>
<td>62.32</td>
<td>116.68</td>
</tr>
<tr>
<td>MAE [mmHg]</td>
<td>-</td>
<td>-</td>
<td>5.02</td>
<td>3.11</td>
<td>5.75</td>
</tr>
</tbody>
</table>

As depicted in Figure 8, the Normotensive category showed the highest number of data, causing an imbalance in the test dataset. Nevertheless, the models successfully distinguished signals from this class, facilitating accurate binary classification between Normotensive and Hypertensive.

As shown in Figure 9, both models achieved much more optimistic performance, obtaining equality in the F1-Score metric and similarities in the others.

It is important to note that the experiments with the specialist models with NeuBP and Spectro-Temporal started from a general pre-training. Therefore, there is a need to evaluate the performance of a specialist model without any pre-training, in order to determine the best approach for applying a methodology with a specialist model.
4.3.3. Specialist Model (SM)

In the SM experiment, neural models were created without pre-training and adjusted to a specific individual. In practice, models with random weights were created without any adjustments, subjected to data from a single individual in the training, validation, and testing stages. Figure 10 illustrates the behavior of the models with the predicted values in relation to the systolic and diastolic distribution labels. It can be observed that both models faced difficulties, with many scattered values in their predictions, hindering the correlation of labels with blood pressure categories.

Regarding the previous experiment, the data presented in Figure 10 that vary correspond only to the predicted values, since the models were not pre-trained.

In this context, for the predicted data of the systole category, our model exhibited the following dispersion: $Q_1 \approx 103$ mmHg, $Q_2 \approx 113$ mmHg, $Q_3 \approx 136$ mmHg, and IQR $\approx 33$ mmHg. It is also noticeable that there were outliers beyond both the upper limit and the lower limit. The Spectro-Temporal model maintained the following dispersion of data: $Q_1 \approx 100$ mmHg, $Q_2 \approx 112$ mmHg, and $Q_3 \approx 136$ mmHg, along with an IQR $\approx 36$ mmHg. It can be observed that the data predicted by the Spectro-Temporal model had many outliers at both ends of the box plots. The same can be observed for the diastole category, where both models presented many outliers, and the Spectro-Temporal model...
exhibited a larger dispersion concerning the true data, performing worse than NeuBP in this aspect.

Even though the models exhibited low values for the evaluation metrics, NeuBP managed to achieve a standard deviation and mean more correlated with the labels, and it also presented a lower MAE, as shown in Table 3.

Table 3. Arithmetic mean, standard deviation and absolute mean error in the SM experiment.

<table>
<thead>
<tr>
<th></th>
<th>Target</th>
<th>NeuBP</th>
<th>Spectro-Temporal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Systole</td>
<td>Diastole</td>
<td>Systole</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>23.18</td>
<td>12.99</td>
<td>19.35</td>
</tr>
<tr>
<td>Mean</td>
<td>116.0</td>
<td>62.0</td>
<td>115.36</td>
</tr>
<tr>
<td>MAE [mmHg]</td>
<td>-</td>
<td>-</td>
<td>11.03</td>
</tr>
</tbody>
</table>

Although the focus of this experiment was to generate specialized models, there is a notable gap between this experiment and the SMR experiment. Training a specialist model without pre-training may require more samples to adjust the model to individual data points. However, NeuBP demonstrated better performance when applying the F1-Score metric to the Normotensive and Hypertensive Stage 1 classes, as shown in Figure 11.

Figure 11. Confusion matrix of hypertension classes achieved with values predicted by NeuBP and Spectro-Temporal model. The best values among the models is highlighted.

Finally, when visualizing the distribution of the predicted labels from a binary perspective, as shown in Figure 12, the NeuBP model achieved better performance, particularly in the F1-Score metric, failing to outperform Spectro-Temporal only in precision for the Hypertensive class.
Figure 12. Confusion matrix of the values predicted in the SM experiment in relation to the true values in the binary classes, highlighting the best achieved among the analyzed models.

4.4. Discussion

During the execution and revision of this work, several other studies were developed, presenting diverse results. We cannot help but mention them, even though we have not validated such results through experiments. Therefore, in Table 4, we present the main results achieved in each approach by various authors, highlighting the SBP and DBP values. Among the mentioned works, it is noticeable that in more recent studies, there has been a significant exploration of methods involving deep autoencoders, as discussed by Athaya et al. [2], Harfiya et al. [32], Qin et al. [33], and Mehrabadi et al. [34]. Although these works aim to convert the PPG signal into ABP, there is a clear problem regarding the precision of these models when deployed due to the difficulty in eliminating or attenuating motion artifacts. This is because the number of output parameters that model needs for prediction is much larger than just two parameters, as addressed in this study.

Table 4. A comparison among different approaches was conducted, considering well-established related works. Emphasis was placed on the publication year; dataset utilized; input types (raw signal or features extracted); and the outcomes, specifically the errors in mmHg generated for SBP and DBP.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Dataset</th>
<th>Input</th>
<th>SBP (mmHg)</th>
<th>DBP (mmHg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xie et al.</td>
<td>2018</td>
<td>UQVS</td>
<td>PPG’s Feat.</td>
<td><strong>4.21 ± 3.24</strong></td>
<td><strong>7.59 ± 5.39</strong></td>
</tr>
<tr>
<td>Athaya et al.</td>
<td>2021</td>
<td>MIMIC-III</td>
<td>PPG</td>
<td><strong>3.68</strong></td>
<td><strong>2.17</strong></td>
</tr>
<tr>
<td>Harfiya et al.</td>
<td>2021</td>
<td>MIMIC-III</td>
<td>PPG</td>
<td><strong>4.05</strong></td>
<td><strong>2.41</strong></td>
</tr>
<tr>
<td>Qin et al.</td>
<td>2021</td>
<td>MIMIC-III</td>
<td>PPG</td>
<td><strong>4.11</strong></td>
<td><strong>7.95</strong></td>
</tr>
<tr>
<td>Mehrabadi et al.</td>
<td>2022</td>
<td>MIMIC-III</td>
<td>PPG</td>
<td><strong>2.29</strong></td>
<td><strong>1.93</strong></td>
</tr>
<tr>
<td>NeuBP</td>
<td>2024</td>
<td>MIMIC-III</td>
<td>PPG</td>
<td><strong>5.02</strong></td>
<td><strong>3.11</strong></td>
</tr>
</tbody>
</table>

5. Conclusions

In this article, we propose NeuBP, a blood pressure estimation method based on PPG, employing a 1D convolutional architecture with dilation mechanisms.
The results obtained demonstrate the potential of the NeuBP method in the regression task for blood pressure estimation. The application of filtering methods on the PPG signal, outlier attenuation, and extraction of labels from the ABP signal were crucial to achieving competitive results compared with the literature. Thus, predictions were evaluated using regression-related error metrics and classification-linked metrics.

The strategy of employing a neural model with a core composed of dilation blocks with 1D convolutional layers allows for an increase in the scale of features extracted by the model. This approach utilizes a less complex model capable of identifying changes in physiological signals captured by PPG to determine if the individual is affected by arterial hypertension through their variations. As a result, NeuBP demonstrates promise for the blood pressure regression task, indicating greater feasibility for its use in real-world scenarios due to its less complex nature.

In the future, we intend to investigate the generalization capabilities of blood pressure estimation models across diverse datasets obtained from different populations, healthcare settings, or sensor modalities to assess their robustness and applicability in varied contexts. Additionally, we will assess the interpretability and explainability of blood pressure estimation models by employing techniques such as attention mechanisms or feature importance analysis. This approach establishes the groundwork for incorporating our solution into devices such as smartphones and wearables, with the goal of bolstering user confidence and streamlining clinical decision making in the context of hypertension detection.

Author Contributions: Conceptualization, H.S.D.O. and E.J.P.S.; methodology, E.J.P.S. and R.A.P.; validation, R.G.; investigation H.S.D.O. and R.A.P.; resources, H.S.D.O.; data curation, H.S.D.O.; writing—original draft preparation, H.S.D.O. and R.A.P.; writing—review and editing, H.S.D.O., E.J.P.S. and R.A.P. and supervision E.J.P.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by Samsung Electronics of Amazonia Ltda, under the terms of Federal Law No. 8.387/1991, through agreement No. 003/2019, in accordance with Article 48 of Decree No. 6.008/2006. The study also received financial support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES)-Finance Code 001, and the Amazonas State Research Support Foundation—FAPEAM—through the POSGRAD project.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References


**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.