FEATURE EXTRACTION OF PPG BY CONVOLUTIONAL KERNEL AND ESTIMATION OF ABP BY HIGHER-ORDER REGRESSION

A PREPRINT

Hisashi Ikari
Circanoid Healthcare
Saitama, Japan
hisashi@ikari.io

January 27, 2022

ABSTRACT

It is important to measure the blood pressure resulting from ischemic stroke or myocardial infarction during the long term of daily life to provide awareness of the need to increase healthy life expectancy. However, it is difficult to measure blood pressure for a long period of time using an invasive cartel or a cuff system with compression. Therefore, in this study, we estimate ABP from PPG, which is a non-invasive optical blood volume measurement. We used deep learning to obtain robust features for two groups of variables, including local variation and potential vascular elasticity. In addition, the direct projection from PPG to ABP is difficult and difficult to explain. Therefore, we set the problem of blood pressure estimation by reducing the difficulty to a regression problem with a rounded projection, which is relatively easy to explain. As a result, we obtained a MAE of 3.39 and an STD of 5.88, which are close to those of previous studies. Although more research on individual differences and robustness between groups is needed, the results are promising for future development in medical practice where explanations are needed.

Keywords Machine Learning · Deep Learning · Biometric Sensing · Photoplethysmogram · Arterial Blood Pressure

1 Introduction

1.1 The need for non-invasive vital measurement to reduce preventive care and healthcare costs

In recent years, health care costs have increased due to the aging of society. The increase in healthcare costs is attributed to the gap between healthy life expectancy and average life expectancy. Preventive medicine is one of the measures to narrow the gap between healthy life expectancy and average life expectancy. Preventive medicine is based on the measurement of vital signs in daily life to predict the degree of vitality leading up to disease and to extend healthy life expectancy through measures to mitigate disease before it occurs.

In 2015, 68% of hemorrhagic strokes in Japan were caused by hypertension, and more than 70% of cerebral infarctions, whether lacunar, atherothrombotic, or cardiogenic, have hypertension. In order to alleviate this problem, preventive medicine, such as daily exercise, is effective [Toyoda et al. [2022]].

However, the accurate measurement of blood pressure requires invasive measurements. Invasive measurements involve inserting a cannula directly into the arterial pathways of the radius, thigh, dorsum of the foot and upper arm to measure arterial waves. This places a high physical burden on the subject. In comparison, the cuff method of non-invasive measurement measures blood pressure from the variation in pressure of the pulse wave obtained by compression against the artery. The non-invasive cuff method places less of a burden on the measurement subject than the invasive method. However, due to the need for compression, routine long-term measurements are also difficult to perform.

* Circanoid Healthcare is the trade name of a sole trader and this study is the result of a disinterested individual’s efforts.
Therefore, it is necessary to develop a method that enables long-term blood pressure measurement in daily life with less burden on the subject of measurement, which is expected to contribute to preventive medicine and increase healthy life expectancy.

### 1.2 Advances in data-driven deep learning time series analysis for vital measurements

Vitals are time-continuous and are measured according to a time resolution. In other words, they are time series data. The analysis to obtain various characteristic variations from time series data is called time series analysis, and Fourier transform and mode decomposition are typical examples of methods to obtain characteristic variations. These analyses are heuristic analyses, in which the analyst obtains characteristic variations in a time series from evidence or his own empirical rules in order to satisfy his objective, which is a sequential problem approach.

On the other hand, the effectiveness of pre-learning by deep learning has been shown ([Erhan et al., 2010](#)). This prior learning is a data-driven approach to find the characteristic variability that satisfies the objective from time series data as an inverse problem. The Fourier transform of the heuristic analysis is based on a combination of persistent periodicity and a fixed time unit. Mode decomposition, on the other hand, allows for local variations but assumes a fixed time unit. In contrast to these preconditions, deep learning pre-training dynamically combines the sizes of the convolutional kernel windows and further removes the precondition of fixed time units. As a result, the deep learning prior is able to focus on more fine-grained variations of the time series, regardless of local variations or the start and end time of the measurement, and obtain the characteristic variations that satisfy the objective by solving the inverse problem.

Therefore, the data-driven pre-training of deep learning is suitable for the long-term measurement of vitals in daily life, without being affected by local fluctuations in vitals caused by exercise, or by the fragmented start and end of measurements due to the charging of the measurement device. It is also expected to be able to obtain characteristic variations that satisfy various objectives in a data-driven manner.

### 1.3 Higher-order development of regressions to take account of individual differences in vital measurements

Blood pressure is a continuous quantity that is not categorical. Therefore, for the purpose of blood pressure, the problem of deep learning is not a discrimination problem but a regression problem. Regression can lead to errors in the regression and the data when the explanatory variables follow different distributions. Also, when the posterior distribution of the explanatory variable is a mixture of distributions, such as individual differences in the posterior distribution of the explanatory variable, where the posterior distribution follows the prior distribution of the parameter, a single regression will likewise introduce errors in the regression. In general, the former problem is solved by a generalized linear regression that takes into account the different posterior distributions. The latter problem is solved by a hierarchical Bayesian model that obtains the regression while inferring the prior distribution as a parameter of the posterior distribution by Bayes’ theorem ([Magni et al., 1998](#)).

However, a continuous vital quantity such as blood pressure can vary over a short period of time (seconds), in addition to individual differences that do not change over time, such as BMI. For example, local variations caused by exercise, or unintentional interruptions depending on the measurement device, can be considered. For this reason, it is necessary to estimate higher-order distributions from the data, taking into account both the long and short term.

In recent years, the field of machine learning has developed an ensemble of higher-order decision trees to explain the distribution of the data for problems requiring such higher-order configurations. We hypothesize that this higher-order structure, as well as local variability in data-driven time series, can explain local variability due to exercise in the short term, momentary interruptions due to measurement devices, and individual differences in BMI over time.

### 1.4 Data-driven time series analysis of PPG and estimation of ABP by higher-order regression

An example of non-invasive biometric measurement is photoelectric volumetric pulse wave recording (PPG), which is achieved by shining a light on the human skin with an LED and measuring the intensity change of the reflected light of the blood flow with a photodiode PPG is the most widely used blood flow measurement method in the wellness field.

For the accurate measurement of blood pressure, the arterial blood pressure (ABP) method is used, as described above. This method involves inserting an arterial pressure measuring catheter or elastomeric needle into an artery and using a transducer to directly measure arterial pressure. This method of measurement places a heavy burden on the person being measured. However, in terms of accuracy, it meets the requirements for medical use.

Both PPG and ABP are blood flow measurements and their waveforms are similar. Therefore, many attempts have been made to estimate ABP from PPG; the cohort datasets for PPG and ABP are available on Physionet as the MIMICII dataset. Therefore, applying the data-driven time series analysis and higher-order regression described in this paper to
MIMICII, including PPG and ABP, and estimating ABP from the most commonly used PPG with high accuracy may be useful to satisfy non-invasive measurements.

2 Materials and Methods

2.1 Assessment of blood pressure estimation

The British Hypertension Society (BHS) and The Association for the Advancement of Medical Instrumentation (AAMI) have established a standard for the physiology and biomedical engineering of blood pressure. The estimation will be evaluated according to these indices.

The BHS Standard ([E et al. (1993)]) is a measure of performance in blood pressure estimation, with grades of A, B, C and D based on the percentage of cumulative error in the estimated blood pressure. The error here is the Mean Absolute Error (MAE). The error of the estimated blood pressure is ranked A with 60% of the cumulative percentage of estimates below 5 mmHg, 85% of the cumulative percentage of estimates below 10 mmHg and 95% of the cumulative percentage of estimates below 15 mmHg; rank B, 50%, 75% and 90% respectively; rank C, 40%, 65% and 85% respectively. The BHS Standard is listed in (Table 1).

The index of performance for the estimation of AAMI blood pressure is called the AAMI Standard (Association for the Advancement of Medical Instrumentation et al. [2003]) and is evaluated by the mean and standard deviation of the error of the estimated blood pressure. The error here is the Mean Error (ME). The indices are a mean of less than 5 mmHg, a standard deviation of less than 8 mmHg, and more than 85 subjects. The AAMI Standard is listed in (Table 2).

| Grade | ≤5mmHg  | ≤10mmHg | ≤15mmHg |
|-------|---------|---------|---------|
| A     | 60%     | 85%     | 95%     |
| B     | 50%     | 75%     | 90%     |
| C     | 40%     | 65%     | 85%     |

Table 2: The Association for the Advancement of Medical Instrumentation (AAMI) Standard

| Blood Pressure | ME (mmHg) | STD (mmHg) | Number of subjects |
|----------------|-----------|------------|--------------------|
| DBP            | ≤5        | ≤8         | ≥85                |
| MAP            | ≤5        | ≤8         | ≥85                |
| SBP            | ≤5        | ≤8         | ≥85                |

2.2 Dataset

Beth Israel Deaconess Medical Center, Laboratory for Computational Physiology (LCP), Massachusetts Institute of Technology (MIT) and Philips Healthcare et al. collaborated to provide vitals, diagnoses, procedures, and medications for anonymous patients at Beth Israel Deaconess Medical Center between 2001 and 2012. Patient information includes gender, age, place of residence, occupation, and education, but does not include personally identifiable demographic information. This dataset, called Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II) (Saeed et al. [2002]), contains simultaneous measurements of PPG and ABP for 942 patients at 125 Hz 8-bit. (Kachuee et al. [2015]) conducted a study to remove measurements of PPG and ABP from MIMIC II that did not fall within the upper and lower limits of physiologically abnormal vitals. They also used the Daubechies 8 (db8) wavelet to remove abnormal vital waveforms. As a result, the maximum systolic arterial blood pressure (SBP), the minimum diastolic blood pressure (DBP) and the average blood pressure (ABP) were narrowed down to only those blood pressures where 60 mmHg ≤ DBP ≤ 130 mmHg and 80 mmHg ≤ SBP ≤ 180 mmHg. The datasets with these removals are available separately in the UCL Machine Learning Repository and the Kaggle Dataset. The basic statistics fall within the ranges shown in (Table 3).

https://archive.ics.uci.edu/ml/datasets/Cuff-Less+Blood+Pressure+Estimation

For ease of access and pre-processing, we use the dataset provided by (Kachuee et al. [2015]), in their Kaggle Dataset. The sampling rate of this dataset is 125Hz and we have defined the epoch as 10 seconds. This means that the unit of data points is 1250 data points. Fractional data points that did not fall within this range were excluded without zero
Estimation of ABP by higher-order regression problem using PPG with prior learning

2.3 Proposed Methodology

2.3.1 Preprocessing

We used a dataset provided by (Kachuee et al. [2015]) on Kaggle. As mentioned above, this dataset has a defined upper and lower limit of blood pressure and is denoised using a wavelet transform. The wavelet transform uses Daubechies 8 (db8) as the mother wavelet for the convolution and removes the low frequency components between 0Hz and 0.25Hz and the frequency components between 250Hz and 500Hz. For thresholding, we use soft thresholding to reconstruct the decomposed signal by the wavelet transform, narrowing it down to only those signals that satisfy the threshold. The noise is removed by this process. The PPG signals are normalized for easy discrimination by machine learning.

Our pre-processing, as described above, first divides the data points into units of 1250 data points. We do not use zero padding and we exclude fractional data points. As our blood pressure estimation is a regression problem, we obtain for these 1250 data points the mean, the maximum and the minimum of the ABP over the same time range. This mean, maximum and maximum ABP are the objective variables in the regression problem. The explanatory variable is the local variability of the PPG, which is obtained in a data-driven way using deep learning. An example of the training data obtained for the regression problem is shown in (Figure 1).

Figure 1: Example of a training data set of 10 sets of 10 seconds with a sampling rate of 125 Hz. Here, PPG is the original signal as the explanatory variable from which the features are derived, and ABP is the objective variable to be estimated in 10 seconds.

2.3.2 Mini Random Convolutional Kernel Transform

The Random Convolutional Kernel Transform (Rocket) (Dempster et al. [2020a]) is a method in the field of time series analysis that adapts the convolutional kernel of deep learning to further improve the accuracy of algorithms for extracting readable characteristic units of variation in time series, called shapelet (Yang et al. [2016]). MiniRocket...
Further reduces the computational complexity of the algorithm by restricting the convolutional kernel to only those parameters that can be characterised by a variety of time series data and by limiting the variability to only positive variability.

A shapelet represents a unit of characteristic variability. The algorithm that extracts them creates a tree structure of all the sub-sequences of the time series data, called a shapelet tree, in order to filter and evaluate the candidate sub-sequences. For each of these sub-sequences, the distance to the time series data is calculated to obtain the units of variation characteristic of the time series. The convolution kernel is a convolution of an arbitrary number of kernels with an arbitrary width, and the ratio between the maximum value and the positive value is obtained. The ratio of positive values obtained here is called PPV, and if we can divide this PPV statistic into subsets, it will represent a pattern, a positive contribution, characterising the time series data. The PPV obtained here can be regarded as a shapelet, a shapelet tree can be constructed, the distance between the shapelet tree and the time series data can be calculated, and the characteristic unit of variation can be obtained. In this paper, MiniRocket is used to extract the characteristic variation. An example is shown in (Figure 2).

Figure 2: Example of the positive contribution of a set of 10 characteristic variations over 10 seconds obtained with MiniRocket. This represents the PPV of the convolution kernel of the PPG for 10 s obtained with MiniRocket: for MiniRocket it is PPV only and only positive contributions are obtained.

2.3.3 Light Gradient Boosting Machine

In this case, the regression problem is specified in order to satisfy the desired blood pressure. We also assume a higher-order regression problem. The explanatory variables are obtained from the PPG in a data-driven way and the characteristic variability that satisfies the regression of the blood pressure is obtained by a deep learning convolution kernel. In the convolutional kernel, the shapelets are obtained in a tree structure. Therefore, to solve higher-order regression problems, we use a machine learning mechanism based on decision trees.

Examples of decision tree-based machine learning schemes for solving regression and discrimination problems are eXtreme Gradient Boosting (XGBoost) ([Chen and Guestrin 2016]) and Light Gradient Boosting Machine (LightGBM) ([Ke et al. 2017]). Both of them perform ensemble learning of weak hypotheses, and there is no significant difference in their performance. However, we use LightGBM in this paper because LightGBM has a more complex structure due to Leaf-Wise and requires less computer resources for training. A simple example is shown in (Figure 3).

Figure 3: An example of the hierarchy of higher-order regressions we will attempt. In these hierarchical contributions, we find the combinations of explanatory variables that satisfy the regression of the target variable, the blood pressure.

3 Experiments

We use the Python 3.7 programming language and runtime environment to estimate the maximum ABP (DBP), minimum ABP (SBP), and mean ABP (MAP) of a 10-second period of ABP from PPG. The computing resource is Google Colaboratory with Google Compute Engine of Google Cloud Platform as the backend. The performance of this back-end computing resource is 13Gb memory, 200Gb disk space, and NVIDIA Tesla P100 GPU. Due to the previous generation of the system, the K80 may only be able to achieve a tenth of the performance of the P100 in LightGBM training. You can reproduce this case on the Google Colaboratory from the following link.
3.1 Parameters of Convolutional Kernel Transform

In this paper, we use MiniRocket to obtain characteristic variations from PPG by tsai (Oguiza [2020]). The parameters given to MiniRocket are the size of the convolutional kernel, the number of convolutional kernels, and the dilations as the range of extension of the size of the convolutional kernel, which are 9, 84, and 32, respectively (Table 5). The maximum characteristic variation obtained is at most 10,000.

Table 5: Parameters of Mini Random Convolutional Kernel Transform

| kernel size | kernel number | kernel dilations |
|-------------|---------------|------------------|
| 9           | 84            | 32               |

3.2 Parameters of Light Gradient Boosting Machine

In this paper, we use the LightGBM to solve the blood pressure regression problem with a decision tree from the characteristic variation. The parameters given to the LightGBM are gbdt for boosting, mae for metrics, 0.1 for the learning coefficient, 31 for the number of leaves in the tree structure, 20 for the minimum number of data per leaf, and other parameters related to learning (Table 6). The other parameters related to learning were set to the default values specified by lightgbm.

Table 6: Parameters Light Gradient Boosting Machine

| boosting type | metrics | learning coefficient | leaves in tree | minimum data per leaf |
|---------------|---------|----------------------|----------------|-----------------------|
| gbdt          | mae     | 0.1                  | 31             | 20                    |

3.3 K-Fold Cross Validation

The evaluation of the estimation of blood pressure is assessed by a 10-crossover test. If at first sight the accuracy of the estimation appears to be high, but the dataset is biased in a way that is not representative of the whole population, it is the accuracy of an unusual dataset. To avoid this, we double-check the accuracy of the estimation on subsets of several different datasets and evaluate the estimation in a generalized way. The value of the evaluation is the MAE, which is the average of all the 10 cross-validation tests.

4 Results

In this case, we use MiniRocket to obtain the characteristic variation for 10 seconds of an arbitrary period. The characteristic variability is not affected by the peaks of the PPG waveform due to the data-driven convolution kernel, and it does not have the assumption of permanent periodicity, so we assume that the local variability contributes to the regression. From the characteristic variability obtained here, we make a further assumption that we can find explanatory variables to solve regression problems involving individual differences in higher-order decision trees. Thus, the hypothesis is robust to periods, to constant periodicity, to noise and to individual differences.

The MAE obtained for the estimation results is 3.39±5.88 overall, with a minimum ABP (DBP) of 2.46±3.47, a mean ABP (MAP) of 2.65±3.37 and a maximum ABP (SBP) of 5.05±6.47 (Figure 4). In the previous study, we set up a projection problem for PPG and ABP and showed that the projection was accurate, including the phase difference. However, in this study, we simplified the problem by setting up a regression problem for the estimation of DBP, MAP, and SBP from PPG, and assumed that the accuracy could be further improved by limiting the value of the estimation to the target variable only. As a result, a slight improvement in accuracy is observed, and the readability of the basis of the regression is thought to be further improved by using DBP, MAP and SBP as the objective variables of the regression problem, and the convolution kernel as the explanatory variable. We believe that it is meaningful to improve the accuracy and the readability of the regression problem by estimating it in a way similar to deep learning.

4.1 Assessment according to British Hypertension Society Standards (BHS Standards)

The grade of the blood pressure estimate was determined by comparing the grade with the criteria set out in the BHS. In the BHS, all criteria must be met to achieve a given grade. In this study, DBP was 88.40% (≤5mmHg),
Estimation of ABP by higher-order regression problem using PPG with prior learning

Figure 4: Scatter plot of validation data for all 10 cross-validation tests obtained. The overall ABP was 3.39±5.88, the minimum ABP (DBP) was 2.46±3.47, the mean ABP (MAP) was 2.65±3.37 and the maximum ABP (SBP) was 5.05±6.47. The error in the minimum blood pressure (DBP) was small, showing a tendency for error to occur with each increase in blood pressure value.

96.73% (≤10mmHg), and 98.75% (≤15mmHg), all of which were Grade A. MAP was 85.91% (≤5mmHg), 96.25% (≤10mmHg), and 98.65% (≤15mmHg), all of which were Grade A. SBP was 68.86% (≤5mmHg), 87.78% (≤10mmHg), and 98.65% (≤15mmHg), all of which were Grade A. Overall, 81.06% for ≤5mmHg, 93.59% for ≤10mmHg, and 96.89% for ≤15mmHg were classified as Grade A (Table 7)(Figure 5). Previous studies using the same dataset used in this study have shown that the error in the maximum blood pressure (SBP) is high. This study also showed a high SBP error.

Table 7: Assessment according to British Hypertension Society Standards (BHS Standards)

| Grade | ≤5mmHg | ≤10mmHg | ≤15mmHg |
|-------|--------|---------|---------|
| DBP   | A      | 88.40%  | 96.73%  | 98.75%  |
| MAP   | A      | 85.91%  | 96.25%  | 98.65%  |
| SBP   | B      | 68.86%  | 87.78%  | 93.28%  |
| Overall | A | 81.06% | 93.59% | 96.89% |

Figure 5: Distribution of errors in the targets assessed by British Hypertension Society Standards (BHS Standards). Overall, 81.06%, 93.59%, and 96.89% correspond to grade A.

4.2 Assessment according to The Association for the Advancement of Medical Instrumentation Standard (AAMI Standard)

The results of the blood pressure estimation were compared to the criteria set out in the AAMI Standard. This assessment has also been used as an indicator in many publications. The AAMI standard requires that the ME should be less than 5 mmHg, the STD less than 8 mmHg and that there should be at least 85 subjects. The results of this study showed that DBP was 0.00 for ME and 4.26 for STD, MAP was -0.02 for ME and 4.29 for STD, and SBP was -0.02 for ME and 8.20 for STD (Table 8)(Figure 6).
Table 8: Assessment according to Association for the Advancement of Medical Instrumentation Standard (AAMI Standard)

|       | ME (mmHg) | STD (mmHg) | Number of subjects |
|-------|-----------|------------|-------------------|
| DBP   | 0.00      | 4.26       | 984               |
| MAP   | -0.02     | 4.29       | 984               |
| SBP   | -0.02     | 8.02       | 984               |
| Overall | -0.01   | 5.88       | 984               |

Figure 6: Mean and standard deviation according to the assessment of the Association for the Advancement of Medical Instrumentation Standard (AAMI Standard). Overall, ME is -0.01 and STD is 5.88. The error in the estimation of blood pressure increases with increasing blood pressure and the SBP does not meet the standard.

4.3 Possible increase in SBP error due to mixing of groups of hypertensive subjects

So far, there is a large error in the SBP. This may be due to the inclusion of a hypertensive group of subjects and the contribution of latent vascular elasticity. Therefore, we assume that the relationship between PPG values and blood pressure values is dichotomous and that two significant groups will appear if we cluster the training data characterized by local variations in the time series. Here, we used kmeans clustering to cluster the data into two groups (Figure 7).

As a result, we obtained two groups, one with a large blood pressure relationship and the other with a small PPG value. When the elasticity of the blood vessels is high, the blood pressure increases, even though the blood volume is close. Therefore, without estimating the potential elasticity of the vessels, the relationship between PPG values and blood pressure values can be used to discriminate between them, and to some extent the regression can be satisfied. However, the regression is more stringent, which may have contributed to the larger error. In addition, since one group cannot be estimated from the other during training, it is necessary to include the data of both groups in the training data (Figure 8).

Figure 7: The kmean clustering diagram of MiniRocket features in the training data. There is an association between the values of the standardized PPG peaks and high blood pressure, based on which the training data is broadly dichotomized. This means that we need to include both in the training data. Without the hypertension training data, the estimation of hypertension is not possible.

4.4 Comparison with Existing Methods

The results of this study are compared with the results of the previous study. There is a difference in the dataset and its preprocessing between the previous study and this study. Therefore, it is difficult to make a direct comparison. However, it is possible to evaluate the results of the previous studies as a rough guide, because they all report the distribution of the error of the estimation results and the mean and standard deviation of the error. In this section, the results of each previous study are applied to the BHS Standard and the AAMI Standard and compared with the results of this study.
Estimation of ABP by higher-order regression problem using PPG with prior learning

Figure 8: Diagram of the comparison between PPG and ABP of the two clusters. The right figure is cluster 1 and the left figure is cluster 0. The group on the right has a higher blood pressure and is hypertensive.

All previous studies have used MIMIC II, but some have narrowed the sample. The BHS Standard has progressed from heuristic analysis of PPG waveforms to estimation using deep learning pre-training, as described in the table. The BHS Standard, as shown in the table, has progressed from heuristic analysis of PPG waveforms to estimation using deep learning pre-training, and its grade has increased. The AAMI Standard has been met for DBP and MAP in all studies since Kachuee et al. In this case, the SBP has been evaluated slightly below the standard, and there is a possibility that the SBP standard can be met by tuning the parameters of the decision tree (Table 9).

Table 9: Comparison with Existing Methods (Targeted at DBP, MAP and SBP in that order.)

| Study               | MIMIC II | Input       | BHS Standard | AAMI Standard |
|---------------------|----------|-------------|--------------|---------------|
| Kachuee et al. [2015] | 942 subs | PPG and ECG | B, C, D      | 6.34, 7.52, 12.38 |
| Kachuee et al. [2017] | 942 subs | PPG and ECG | B, C, D      | 5.35, 5.92, 11.17 |
| Mousavi et al. [2019]| 441 subs | PPG         | A, B, C      | No rating value |
| Slapnij et al. [2019] | 510 subs | PPG         | No rating value | 9.43, unknown, 6.88 |
| Ibtehaz and Rahman [2020] | 942 subs | PPG         | A, A, B      | 3.45, 2.31, 5.73 |
| Our Results         | 942 subs | PPG         | A, A, B      | 2.46, 2.65, 5.05 |

5 Conclusions

In today’s ageing society, it is important to increase healthy life expectancy in order to reduce health care costs. One of the causes of death is ischemic stroke and myocardial infarction. One of the factors of these diseases is hypertension, and it is thought that measuring blood pressure in daily life and presenting its objective value are effective as preventive medicine until the disease occurs. However, in the measurement of blood pressure, if it is medical grade accuracy, the direct measurement of blood pressure using an invasive cartel, and the cuff type measurement that presses the measurement part of the body even if it is a non-invasive measurement are mainly used. These methods place a heavy burden on the body and are difficult to measure on a daily basis for a long time.

In this paper, we propose a method to measure the maximum (SBP), mean (MAP), and minimum (DBP) of ABP during a given period of time from PPG to MIMIC II, which provides the waveform of PPG and ABP measured simultaneously. By using the data-driven nature of recent deep learning, we set up a high-order regression problem that automatically finds local variations and is easy to explain, and took measures to reduce the difficulty of the estimation problem by changing it from a mapping problem to a regression problem.

As a result, DBP=A, MAP=A, SBP=B in BHS Standard, and DBP, MAP, and SBP of ME in AAMI met the standard, and DBP and MAP met the standard in STD, and SBP was 8.02, which was slightly less than the standard. In addition, from the features obtained from the pre-training of deep learning, we obtained two groups, one is hypertensive group and the other is healthy group, and regardless of the mixture of these two groups, we can solve the regression problem even when latent variables such as the elasticity of blood vessels can be included in the mixture of the two groups, on the assumption that these groups are included in the training data. The results show that it is possible to solve the regression problem even when latent variables such as vascular elasticity can be included in a mixture of two groups.

Furthermore, the robustness of individual differences and between groups needs to be explored, and evaluation and analysis using cross-validation tests that take into account individual differences between hypertensive and healthy groups is necessary. However, the estimation of ABP from PPG by a regression problem that satisfies the criteria of the
Estimation of ABP by higher-order regression problem using PPG with prior learning

BHS Standard and the AAMI Standard and is easy to explain may be a meaningful development in the medical field, where explanations are always necessary.

References

Kazunori Toyoda, Sohei Yoshimura, Michikazu Nakai, Masatoshi Koga, Yusuke Sasahara, Kazutaka Sonoda, Kenji Kamiyama, Yukako Yazawa, Sanami Kawada, Masahiro Sasaki, Tadashi Terasaki, Kaori Miwa, Junpei Koge, Akiko Ishigami, Shinichi Wada, Yoshitaka Iwanaga, Yoshihiro Miyamoto, Kazuo Minematsu, Shotai Kobayashi, and Japan Stroke Data Bank Investigators. Twenty-Year Change in Severity and Outcome of Ischemic and Hemorrhagic Strokes. *JAMA Neurology*, 79(1):61–69, 01 2022. ISSN 2168-6149. doi:10.1001/jamaneurol.2021.4346. URL https://doi.org/10.1001/jamaneurol.2021.4346

Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, and Samy Bengio. Why does unsupervised pre-training help deep learning? *Journal of Machine Learning Research*, 11(19):625–660, 2010. URL http://jmlr.org/papers/v11/erhan10a.html.

P. Magni, R. Bellazzi, and G. De Nicolao. Bayesian function learning using mcmc methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1319–1331, 1998. doi:10.1109/34.735805

O’Brien E, Petrie J, Littler W, de Swiet M, Padfield PL, Altman DG, Bland M, Coats A, and Atkins N. An outline of the revised british hypertension society protocol for the evaluation of blood pressure measuring devices. In *Journal of Hypertension*, volume 11, pages 677–679, 1993. doi:10.1097/00004872-199306000-00013.

Association for the Advancement of Medical Instrumentation et al. *American national standard. manual, electronic or automated sphygmomanometersansi*. Technical report. AAMI SP10-2002, 2003.

M. Saeed, C. Lieu, G. Raber, and R.G. Mark. Mimic ii: a massive temporal icu patient database to support research in intelligent patient monitoring. In *Computers in Cardiology*, pages 641–644, 2002. doi:10.1109/CIC.2002.1166854.

Mohamad Kachuee, Mohammad Mahdi Kiani, Hoda Mohammadzade, and Mahdi Shabany. Cuff-less calibration-free blood pressure estimation using pulse transit time. In *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1006–1009, 2015. doi:10.1109/ISCAS.2015.7168806.

Angus Dempster, Francois Petitjean, and Geoffrey I. Webb. Rocket: exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery*, 34:1454–1495, 2020a.

Yi Yang, Qilin Deng, Furao Shen, Jinchao Zhao, and Chaomin Luo. A shapelet learning method for time series classification. In *2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 423–430, 2016. doi:10.1109/ICTAI.2016.0071.

Angus Dempster, Daniel F Schmidt, and Geoffrey I Webb. MINIROCKET: A very fast (almost) deterministic transform for time series classification. *arXiv*:2012.08791, 2020b.

Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’16, pages 785–794, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4322-2. doi:10.1145/2939672.2939785. URL http://doi.acm.org/10.1145/2939672.2939785

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30:3146–3154, 2017.

Ignacio Oguiza. tsai - a state-of-the-art deep learning library for time series and sequential data. Github, 2020. URL https://github.com/timeseriesAI/tsai.

Mohamad Kachuee, Mohammad Mahdi Kiani, Hoda Mohammadzade, and Mahdi Shabany. Cuffless blood pressure estimation algorithms for continuous health-care monitoring. *IEEE Transactions on Biomedical Engineering*, 64(4):859–869, 2017. doi:10.1109/TBME.2016.2580904.

Seyedeh Somayyeh Mousavi, Mohammad Firouzmand, Mostafa Charmi, Mohammad Hemmati, Maryam Moghadam, and Yadollah Ghorbani. Blood pressure estimation from appropriate and inappropriate ppg signals using a whole-based method. *Biomedical Signal Processing and Control*, 47:196–206, 2019. ISSN 1746-8094. doi:https://doi.org/10.1016/j.bspc.2018.08.022. URL https://www.sciencedirect.com/science/article/pii/S1746809418302209.

Gaper Slapnjar, Nejc Mlakar, and Mitja Lutrek. Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network. *Sensors (Basel, Switzerland)*, 19, 2019.

N. Ibtehaz and Mohammad Sohel Rahman. Ppg2abp: Translating photoplethysmogram (ppg) signals to arterial blood pressure (abp) waveforms using fully convolutional neural networks. *ArXiv*, abs/2005.01669, 2020.