MEKATRONIKA
e-ISSN: 2637-0883
VOL. 4, ISSUE 1, 56 – 69
DOI: https://doi.org/10.15282/mekatronika.v4i1.7758

ORIGINAL ARTICLE

Multiclass Classification of Systolic and Diastolic Peaks on Open Benchmark PPG Signals: Performance Evaluation

Noor Iiza Simon[1], Asrul Adam[1], Low Chen Yik[1], Zuwairie Ibrahim[1] and Mohd Ibrahim Shapiai[2]

[1]College of Engineering, Universiti Malaysia Pahang, 26600 Pahang, Malaysia.
[2]Malaysia-Japan International Institute of Technology (MJIIT), University Technology Malaysia, Jalan Sultan Yahya Petra, 54100 K.Lumpur, Malaysia.

ABSTRACT – Photoplethysmogram (PPG) signals contain valuable health information that is in the relation between the volumetric variations of blood circulation and the cardiovascular and respiratory systems. This study introduces the performance evaluation on open clinical benchmark PPG signals with a multiclass neural network with random weights (NNRW) classification method for systolic peak and diastolic point detection. The best performance of the peak and point detection is crucial to be achieved at the early stage for extracting further valuable information in addition to future predictions of cardiovascular-related illness. Various open clinical datasets of PPG signals have been introduced, however, there is a lack of information on peak annotations. Due to the lack of peak annotation information, it is time-consuming to be prepared. One suitable clinical benchmark dataset with peak annotation information for peak detection has been previously evaluated, however, it cannot be generalized and rely upon only one dataset. Therefore, for generalization, there is a new open clinical benchmark dataset that is found in the year 2018 and our own collected data from normal participants is utilized in this study. The findings exhibit more convincing overall accuracy and *Gmean* of testing results with 94.86 and 94.74 percent, respectively. The findings of the comparison with previous work indicate that the proposed methodology to predict PPG-based multi-class systolic and diastolic points is more generalizable.

ARTICLE HISTORY
Received: 29th May 2022
Revised: 13th June 2022
Accepted: 24th June 2022

KEYWORDS
Computational
Artificial intelligence
NNRW classifier
Photoplethysmography
Systolic
Diastolic

INTRODUCTION

Photoplethysmogram (PPG) is a non-invasive periodic but very nearly irregular and random signal. The PPG signals are obtained through an optical method by a pulse oximeter. Pulse oximetry device is a small clip-like device, lightweight, non-invasive, and painless test. It is standard practice by medical practitioners to determine the blood's oxygen saturation level. It is convenient and quickly measures even little changes in blood's oxygen that are being carried from the heart to parts of the human, such as a finger, toes, and earlobe. The device can be used to monitor the health of people suffering from any condition that can affect blood oxygen levels including various lung diseases, heart failure, and congenital heart defects due to non-awareness to monitor the possibility to risk of cardiovascular disease.

PPG signals contain a wealth of information that can provide significant information about an individual's health and risk of cardiovascular disease. In this study, features are extracted related to systolic peak and diastolic point information from the PPG waveform in the time domain. In literature, researchers have categorized PPG features based on different methods, for example, time-domain [1]–[14] , frequency domain [2], [12], [15]–[17], time-frequency domain [12], [18], and others [12], [17], [18]. Example of time-domain PPG-based features including systolic peak and diastolic point amplitude, peak to peak interval amplitude and time, pulse width, rising time, crest time, pulse area, rising slope, width phase, pulse peak ratio, pulse area ratio, inflection point area ratio, augmentation index, large artery stiffness index, and reflection index. Elgendi et al. and Chowdury et al., have also introduced some useful frequency-domain PPG features from the derivation of first and second derivative methods related to systolic and diastolic peak information [2], [12]. In addition, the time-frequency domain PPG features which are *mNPV*, percentile, mean absolute deviation (MAD), inter quartile range (IQR), skewness, kurtosis, Shannon's entropy and spectral entropy were also used in estimating SBP and DBP [12], [18].

The systolic peak and diastolic point can be predicted to occur during the peripheral pulse wave on PPG signals. The systolic peak can be described as the outcome of the direct wave of pressure that moves from the left ventricle to the part located away from the body core. It also acts as the local maxima of the signals to be related to a diastolic point in estimating the vasoconstriction of the subject. Vasoconstriction is a condition where the difference of amplitude between systolic and diastolic points. Whereas the diastolic point can be explained to be the outcome of the wave of pressure

*CORRESPONDING AUTHOR | N. Simon | ✉ lizasimon2011@gmail.com.my

reflected by arteries located at the lower region of the body [4]. These diastolic points act as local minima of the signals and can be used to calculate the inter-beat interval (IBI) data.

Clinical databases are the collections of observational data on people who meet certain criteria for a patient's or normal subject's health. The applications of these databases differ depending on whether the data comes from a single institution, many clinical facilities, or a population. Clinical datasets consist of more sensitive information about subjects. In literature, different open clinical datasets can be used for benchmarking for example MIMIC [17], MIMIC-I [1], [8], MIMIC-II [4], [16], [18], MIMIC-III [11], Aboy et al. [3], [6], [13], D.Liu et al. [7], [9], and the latest one is obtained from Liang et al. [12]. All of these datasets can be utilized for investigating systolic peak and diastolic point prediction. Some clinical datasets contain abnormal health information of disease, for example, diabetes, stroke, hypertension, vertebrobasilar insufficiency, and also have known cardiovascular problems [19], [20]. Normal datasets from healthy persons without healthcare issues such as hypertension, irregular cardiac rhythm, heart failure, or coronary heart disease are also available [10], [14]. Furthermore, there is a type of dataset that does not have a history of cardiovascular disease, neurological or respiratory disease [2], [5]. It can be concluded that every individual dataset has its specific intention.

The features of PPG signals have been extracted using various approaches. For example, L. Wang et al. [21], in the literature, have used 72 subjects from the MIMIC database and then proposed 22 features using a feed-forward artificial neural network (FFANN) classifier. Kurylyak et al. [1] and S. Shukla [8] have proposed 21 features and 10 features, respectively from database MIMIC-I on 100 subjects. They have implemented a multi-back-propagation artificial neural network (MBPANN) and Multi-task Gaussian Processes (MTGPs) classifier. In addition, X. Xing et al. [16], Kachuee et al. [4], and Hasanzadeh et al. [18] have implemented database MIMIC II on 1000 subjects as their case study and implemented artificial neural network (ANN) and Adaptive Boosting (AdaBoost) algorithm method. Chen et al. [11] also have retrieved a clinical database from MIMIC III on 510 subjects, using a support vector machine (SVM) as a classifier.

Furthermore, different datasets of PPG signals have been utilized such as D. Liu et al. [22] have utilized the vital signs dataset on 32 subjects with healthy data-driven from 7 volunteers. S. Khalid et al. have proposed a method using Random Forest Regression (RFR), Zhang and Wang [7] have proposed 13 features using geometry algorithm-back propagation neural network (GA-BPNN) on this clinical dataset. In addition, Kuntamalla et al. [3], Tee et al. [6] and T.Hann et al. [13] have retrieved clinical dataset from Aboy et al. [19]. They have used the neural network with random weights (NNRW) method to detect systolic peaks and diastolic points using a clinical database on 47 subjects. Otherwise, Chowdhury et al. [12] have extracted 107 features on 126 subjects that from dataset Liang et al. [20] has utilized Gaussian Process Regression (GPR) method.

Next, Gao et al. [5] have proposed healthy data-driven PPG signals based on the camera lens and pulse oximeter with an SVM classifier. A novel algorithm proposed by Elgendi [2] can detect accurate systolic peak detection under challenging condition subjects where PPG signal is measured after exercise activities. Xu et al. [14] have proposed 15 features and Cho et al. [10] have extracted 16 features on the subject that from a healthy adult male. They have implemented an ANN as a classifier. Other researchers, Zhang and Wang [7] and Chen et al. [11] have used the mean impact value (MIV) to determine the impact of features on the classifier to eliminate feature redundancy.

The major goal of this research is to assess the performance of the NNRW-based approach for systolic-diastolic peak prediction from PPG signals. Ten PPG features of systolic and diastolic peaks are generated using an open clinical China database by Liang et al. [20]. Furthermore, the NNRW classifier has the potential to be used to classify systolic-diastolic points using the China database. The performance of this dataset using NNRW is still opened for investigation.

The overall performance evaluation of this study is then compared with the results from T. Hann et al. [13] so that can guarantee the generalization of method performance. In the next section of the paper, we will explain the material and methodology, data collection, validation experiments, and the results and discussions.

MATERIAL AND METHODOLOGY

The signal pre-processing strategy, the peaks candidates identification and manual annotation processes, the features extraction method, and the NNRW multi-class classification method to predict systolic-diastolic points are all discussed in detail in Section 2.

Figure 1 shows several important steps of an overall multi-class systolic-diastolic classification process. In the beginning, PPG signals from the database that is obtained by Liang et al. [20] were first assessed to check signal quality. Due to the signals consists of noises, it suggests going through the filtering process to smooth the signals. Next, true peak candidate identification is done automatically by our algorithm and manually annotates the true systolic-diastolic peaks and points. The data was then distributed into two sets at random for training and testing. The features are extracted based on the time domain which is contained a total of 20 features. A new dataset based on the 20 features was generated, and the dataset is randomly divided into two sets with 75 and 25 percent of the ratio. 75 % of the dataset is used to evaluate the training and validation classification algorithm's accuracy and the remaining 25 % is unseen dataset being used to test the algorithm's performance. Both training and testing datasets go through the classification stage using the NNRW classifier.

Dataset

The dataset applied for this work is produced by Liang et al. [20] and can be retrieved at https://doi.org/10.6084/m9.figshare.5459299 [23]. The PPG dataset of 219 subjects were containing 657 data segments is recorded over a three-minute duration each. The participants range in age from 20 to 89 years old, and the records

include information on cardiovascular diseases. Data for subjects are gathered from the Guilin People's Hospital in Guilin, China, and is based on complete clinical data. Furthermore, this dataset can be utilized to conduct research and identify PPG signal quality, to discover the relationship between the PPG waveform and cardiovascular disorders such as hypertension or diabetes, as well as evaluating latent feature information included in PPG signals.

This dataset was originally collected at 1000 Hz and had three segments per participant. Every signal has a length of 2.1 seconds and recorded 2100 data points. It was measured using the Omron HEM-7201 (Omron Company, Kyoto, Japan) and the PPG signal was taken at the fingertip of the left index finger while the arterial blood pressure was acquired from the right forearm. In addition, the subject's gender, age, height, and weight, as well as heart rate, systolic and diastolic pressures, were all also measured during the data collection session.

Another dataset was collected that contains normal PPG dataset among Universiti Malaysia Pahang (UMP) undergraduate students volunteers between range age 20 and 30 years and healthy background. A total of 45 participants will be conducted to measuring the PPG signals, blood pressure, pulse oximetry (SpO_2), and heart rate using two devices which are NONIN AVANT 2120 and OEM III. The PPG signal had to be taken in three segments, every segment recorded 300 data points with a signal duration of 5 seconds per participant. Table I shows the details of the dataset.

The data produced by Mateo Aboy et. al [19] was used in a previous paper T. Hann et al. [13] have provided their validation dataset is publicly available to retrieve. The dataset was containing PPG signal of intracranial pressure (ICP), arterial blood pressure (ABP), and pulse oximetry (SpO_2) signals of randomly 210 subjects who had histories admitted for traumatic brain injury (60 subjects), sepsis (60 subjects), and cardiac conditions (90 subjects) in Pediatric Intensive Care Unit (PICU) at Doernbecher Children's Hospital, Oregon Health & Science University. A total of 43539 PPG waveforms are recorded, then only 60 minutes duration are randomly selected for each subject.

The dataset consists of two PPG signals with expert annotated detections generated by two different experts (DT and JM) was sampled at 125 Hz with a resolution of 8-bits. The first algorithm's performance assessed using the expert manual annotations (DT) on 15866 PPG waveform corresponding to two 60 minute records. While the second performance of the algorithm on 2649 PPG waveform randomly selected was assessed against the two expert manual annotations (DT and JM).

Overview of Algorithm

The raw PPG signals were filtered before this dataset signals were sent for identification of true peak and point. The pre-processing signal methodology via the identification of true peaks and points is highlighted in the earlier stage of the flowchart in Figure 1, which is the completed phase after manual annotation of systolic-diastolic points for the raw signals. To find all variable peak and point candidates, a three-point sliding window approach is used [24]. The annotation procedure is then performed manually by an expert to distinguish between true and false peaks in the signals. Our prior works can be used to learn more about this methodology. The steps are followed by feature extraction and feature classification techniques [13].

Before the classification stage, there are two ways for the training and testing sets. In both training and testing, the K-fold cross validation approach, with K set to 4, is employed. The NNRW classifier is trained using training sets, which are then intended at the end to produce the classification's performance. The testing sets are then used with the trained classifier to produce testing results.

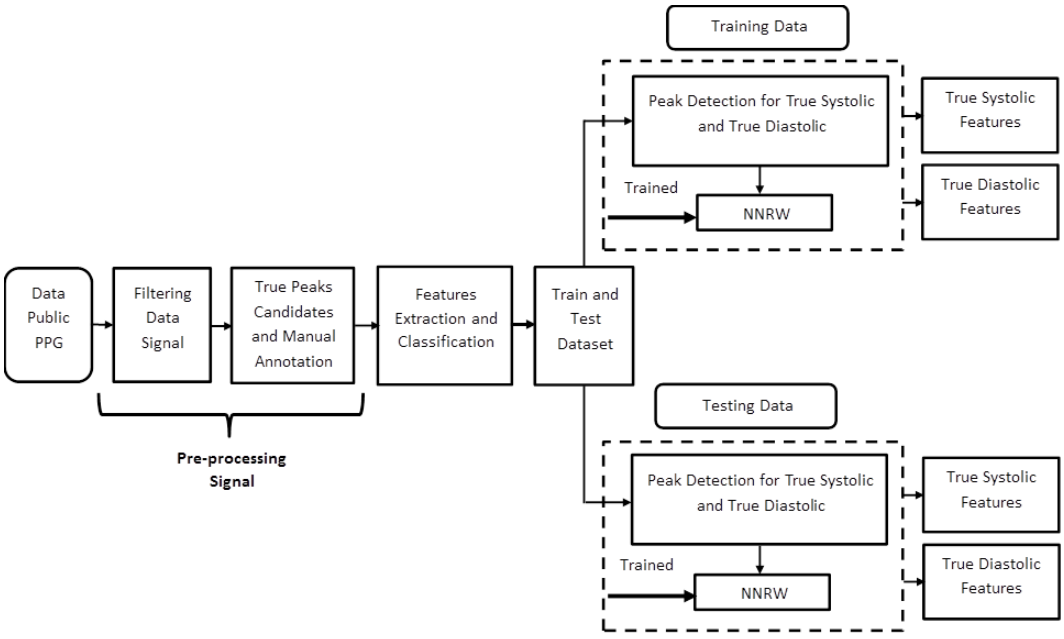

Figure 1. The proposed algorithm flowchart for multi-class systolic-diastolic prediction using the NNRW classifier.

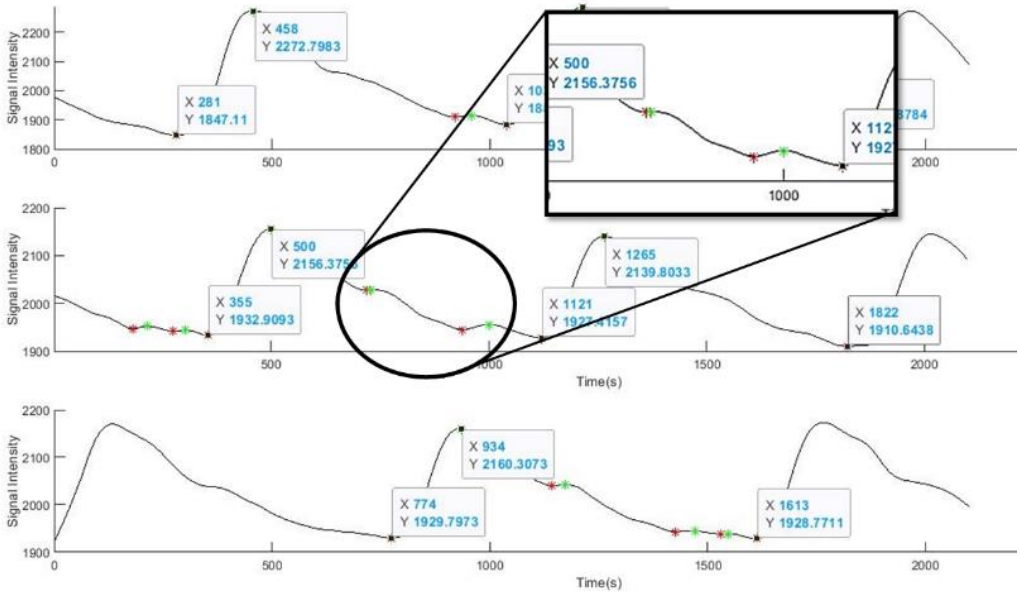

Figure 2. An example of illustration for manual annotation of raw PPG signals' systolic-diastolic points.

Table I. Details of Dataset

	Male	Female	Age Group (Years)
Clinical (219)	105	114	20 to 89
Normal (45)	30	15	20 to 30
Total (263)	135	129	

Pre-processing Signals

Before feature extraction, the raw PPG signals were filtered and then go through several stages of pre-processing, which are summarised in two parts; 1) peak candidates identification and 2) annotation of systolic-diastolic points.

Peak candidates identification

This section discusses the process of identifying systolic-diastolic peaks candidates that are used in the time domain of PPG signals. The PPG signals have to be filtered to eliminate noise and unwanted peaks from the signals before proceeding to the peak candidates identification stage. Then, using the concept of a peak, peak candidates will be determined, with the systolic peak being the local maxima in between two local minima in a waveform and the valley being the local minima of the PPG waveform. The procedure is explained in detail in [13]. Then the steps are followed by manual annotation for both systolic-diastolic peak points. The manual annotation can let peak detection reduce the probability of selecting the false peak.

Manual Annotation of Systolic-Diastolic Points

After the systolic-diastolic point candidates are identified, the process proceeds with manual annotation technique in which means the identification of the actual location of systolic-diastolic points. In this study, for Liang et. al.'s dataset, there are 654 samples of PPG signals from 218 subjects were utilized while for 45 healthy volunteer datasets, there are recorded 135 samples of PPG signals. The manual annotation of actual systolic peaks and diastolic point processes are performed by an expert who is a knowledgeable person related to PPG signals.

The raw PPG signals or line plotted shown in Figure 2 is an illustration of the process annotation. The point for the systolic candidate is indicated by green color, while for the diastolic candidate is highlighted by red color. Referring to Figure 2 there are many systolic-diastolic point candidates have been identified. The annotated systolic-diastolic point candidates that are chosen by an expert can be defined as true peaks and points. The remaining systolic-diastolic point candidates that are not chosen can be defined as false peaks and points. For Liang et. al.'s dataset, the possible actual systolic peaks that have been annotated are 1311 from 3144 candidates. The possible actual diastolic points that have been annotated are 1809 from 3144 candidates. Furthermore, for the healthy volunteer dataset, 642 possible actual systolic

peaks have been annotated from 1000 peak candidates and 774 possible actual diastolic peaks from 1000 points candidates.

Feature Extraction

Twenty features in total are extracted from the signals. For systolic peaks, there are 10 extracted features. For diastolic points, there are also 10 extracted features. All 20 time-domain features are categorized into three groups as described in detail in our previous paper [13]. The three groups are amplitudes, widths, and slopes. Based on seven parameters of peak candidate which is emphasized in [6], it is used to calculate all the 20 features. A detailed explanation to find a peak candidate and the 20 features can be referred to in [13].

K-Fold Cross Validation

In the total of the PPG signal, only 6288 good qualities signal features and reference BPs were used 4-fold cross validation method. The cross validation approach is used to test the algorithm by dividing the original data into many partitions. After that, the data have been divided into two sets, such as training and testing sets using NNRW algorithms.

The PPG data is divided into four partitions for 4-fold, each of which contains training and testing sets. for training and validation, a total of 75 percent of data are used, and the remaining 25 percent of the data being used to evaluate the model's test accuracy. Furthermore, the training set is split up into validation and training sets at random, resulting in two datasets with the same class ratio distribution. The testing set consists of unseen PPG data.

Three folds were utilized to train an algorithm for each iteration, while the remaining fold was used to test its performance. The procedure was repeated until four iterations had been accomplished. Eventually, to obtain the highest accuracy for evaluation performance, the process of cross validation method is conducted 50 times.

The features from the PPG signals dataset were trained with NNRW algorithms using a 4-fold cross validation approach. Due to its fast learning capability proposed by Schmidt et al., the NNRW classifier is recognized as a generalized classification method [25]. The training for NNRW was used to predict the systolic peak and diastolic points. The NNRW model is set up to 3 layered neural networks including input, hidden, and output layers. The layer consists of 10 nodes for the input layer, 500 nodes for the hidden layer, and only one node for the output layer. The number of nodes for the hidden layer is found based on the experiment which is discussed in Section III.

The NNRW classification method indicates as a variety of linear system that can be modeled as $H\beta = T$ mathematically, where H is the output matrix of the hidden layer, β is the output weights with $L \times m$, as the output weights matrix, and T is target yields with $N \times m$ matrix. The equation for β and T matrixes are shown in Equation (1) and (2), respectively [10], [11]. The m defines the number of neurons at the output layer.

$$\beta = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix}_{Lxm} \tag{1}$$

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{Nxm} \tag{2}$$

The NNRW classifier's output function for a given unknown sample, x, is defined as follows: (3). Aside from that, the output matrix at the hidden layer, H, is shown in equation form in (4) and (5).

$$fc(x) = hc(x)\beta \tag{3}$$

$$H = \beta = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} \tag{4}$$

$$H = \begin{bmatrix} g(\sum_{i=1}^{d} a_{i1}x_{1i} + b_1) & \cdots & g(\sum_{i=1}^{d} a_{iL}x_{i1} + b_L) \\ \vdots & \ddots & \vdots \\ g(\sum_{i=1}^{d} a_{i1}x_{Ni} + b_1) & \cdots & g(\sum_{i=1}^{d} a_{iL}x_{Ni} + b_L) \end{bmatrix}_{NxL} \tag{5}$$

From equation form in (5), g denoted as the activation function for hidden neurons, a denoted for random input weights with $d \times L$ matrix, x denoted as input matrix with $N \times L$ matrix, b denoted as the biases that are randomly generated in the hidden layer with $1 \times L$ matrix. The matrixes described as N for a random different sample, L is the total of hidden neurons, d is the total of neurons in the inputs layer, and I is described as the output for the ith hidden neuron for $(x_1 \ldots x_d)$.

The sigmoidal function is chosen in this study as an activation function that is utilized for normalization purposes. Following the normalization step, a linear function will be assigned to the output layer's neuron. The equation form in (6)

demonstrates the minimum-norm least-squares approach, in which β is the linear system of $H\beta = T$. While equation form in (7) shows the smallest norm least-squares approach, where H^+ denote as Moore-Penrose pseudo-inverse of H.

$$\|H(a_1, \cdots, a_L, b_1, \cdots, b_L)\beta - T\| = \min_{\beta}\|H(a_1, \cdots, a_L, b_1, \cdots, b_L)\beta - T\| \quad (6)$$

$$\beta = (H^T H)^{-1} H^T T = H^+ T \quad (7)$$

The training of the NNRW classifier can be summarized that contains three important stages in this process. The first stage has to generate the random input weights, a_i, and the random biases in the hidden neurons, b_i. Follow the next stages by the calculation of the output matrix, *H*. Finally, the weights are then calculated with $\beta = H^+ T$.

Performance Evaluations

Evaluation of Peaks Detection Algorithm

Table II recalls the categorized criteria of the performance evaluation proposed for the systolic-diastolic points detection algorithm. To evaluate the performance of the algorithm, four criteria were used in the standard metric for evaluation which are the average of value (*Mean*), the maximum of value (*Max*), the minimum of value (*Min*), and standard deviation (*STD*) as shown in Table II. In that table, all equations used X_i as the actual value of peak and *n* as the number of samples. Along with that table emphasize five other evaluation aspects including accuracy (*Acc*), precision (*+P*), sensitivity (*Se*), specificity (*Sp*), and *F-score* along with equations form and descriptions based on guideline by the European Society of Hypertension (ESH) and European Society of Cardiology (ESC) [26]–[28].

The accuracy (*Acc*) is the number of correct predictions made by the model over all kinds of predictions made. In addition, precision (*+P*) measures the proportion of a peak candidate, actually had a correctly true peak. Recall or sensitivity (*Se*) is the measured proportion of a peak candidate by the algorithm as the correctly true peak. While specificity (*Sp*) represents the proportion of occurrences the true non-peak of a peak candidate is accurately detected. *Sp* is the exact opposite of *Se*. The *F-score* depends on the relationship between *+P* and *Se*. If one number is really small between *+P* and *Se*, then the *F-score* kind of raises a flag and also is more closer to the smaller number.

To evaluate overall results, two measures are used: overall accuracy and *Gmean*. The equation forms (8) and (9) are showed below. Noted that *TPR* is also denoted for sensitivity (Se_i) or True Peak Rate (TPR_i), and *TNR* is denoted for specificity (Sp_i) or True Non-Peak Rate (TNR_i).

Evaluation by ESH and ESC

Another guideline, which is provided jointly by the ESH and ESC, categorizes the measurement of hypertension into seven classes of a confusion matrix. Based on this standardization, we assigned the output into four classes which are True-Systolic (TS_a) as Class 0, False-Systolic (FS_b) as Class 1, True-Diastolic (TD_c) as Class 2, and False-Diastolic (FD_d) as Class 3, which are Figure 3 shows that confusion matrix. The goal of utilizing this confusion matrix is to figure out what each class's score represents.

TS_{aa} is defined as the true peak that correctly detected the apex point of a systolic peak candidate. While FS_{bb} is denoted as the true peak that correctly detected the apex point of a false systolic peak candidate. The prediction values of (E_{ba}, E_{ca}, E_{da}) and (E_{ab}, E_{cb}, E_{db}) predict the false peak that is incorrectly designated non-peak point of a systolic peak candidate and false systolic peak candidate, respectively. While the prediction of (E_{ab}, E_{ac}, E_{ad}) and (E_{ba}, E_{bc}, E_{bd}) estimated for the false non-peak that is incorrectly detected true peak point of systolic peak candidate and false systolic peak candidate, respectively.

Furthermore, for diastolic point detection, TD_{cc} descript as the true peak that correctly detected the apex point of a diastolic point candidate. While the true peak that correctly detected the apex point of a false diastolic point candidate denote as FD_{dd}. The prediction values of (E_{ac}, E_{bc}, E_{dc}) and (E_{ad}, E_{bd}, E_{cd}) predict the false peak that is incorrectly designated non-peak of a diastolic point candidate and false diastolic point candidate, respectively. While the prediction values of (E_{ca}, E_{cb}, E_{cd}) and (E_{da}, E_{db}, E_{dc}) estimated for the false non-peak that is incorrectly detected true peak of diastolic point candidate and false diastolic point candidate, respectively. The result for the confusion matrix along with the five evaluation metrics for every individual class is given in Table II.

		Predicted			
		TS_a	FS_b	TD_c	FD_d
Actual	TS_a	TS_{aa}	E_{ab}	E_{ac}	E_{ad}
	FS_b	E_{ba}	FS_{bb}	E_{bc}	E_{bd}
	TD_c	E_{ca}	E_{cb}	TD_{cc}	E_{cd}
	FD_d	E_{da}	E_{db}	E_{dc}	FD_{dd}

Figure 3. Four Classes Classification of Confusion Matrix.

Table II. Four Classes Classification Results for Clinical Participants.

Performance Evaluation	Classes Classification	Training Results				Testing Results			
		Mean (%)	*Maximum (%)*	*Minimum (%)*	*Standard deviation*	*Mean (%)*	*Maximum (%)*	*Minimum (%)*	*Standard deviation*
Accuracy	TS	99.03	99.08	98.96	0.03	98.45	98.60	98.32	0.10
	FS	98.98	99.03	98.91	0.05	98.53	98.66	98.30	0.11
	TD	97.10	97.26	96.98	0.08	96.18	96.32	95.93	0.13
	FD	97.08	97.20	96.92	0.08	96.20	96.48	95.92	0.14
Precision	TS	97.83	97.96	97.67	0.11	97.12	97.34	96.76	0.22
	FS	98.50	98.83	98.04	0.25	97.44	98.06	96.97	0.33
	TD	95.78	96.38	95.37	0.30	94.16	94.71	93.29	0.56
	FD	94.32	94.79	93.33	0.46	92.29	92.87	91.53	0.46
Sensitivity	TS	98.95	99.09	98.80	0.09	97.77	98.15	97.49	0.23
	FS	96.73	97.10	96.43	0.19	95.74	96.18	95.11	0.34
	TD	94.10	94.57	93.85	0.21	92.63	93.06	92.00	0.35
	FD	94.32	94.79	93.33	0.46	92.48	93.08	91.67	0.46
Specificity	TS	99.10	99.15	99.03	0.05	98.81	98.90	98.65	0.09
	FS	99.61	99.70	99.49	0.07	99.34	99.50	99.22	0.09
	TD	97.85	98.02	97.76	0.08	97.69	97.93	97.33	0.24
	FD	97.87	97.94	97.78	0.06	97.28	97.47	97.07	0.13
F-score	TS	98.39	98.47	98.28	0.06	97.44	97.69	97.23	0.16
	FS	97.61	97.72	97.45	0.11	96.58	96.87	96.03	0.25
	TD	94.94	95.19	94.74	0.13	93.38	93.61	92.98	0.21
	FD	94.32	94.79	93.33	0.46	92.38	92.98	91.60	0.45

RESULT AND DISCUSSION

The results of the peak detection algorithm, as well as the NNRW classifier that used in this study, are summarised in this section. It can be seen that the final results of systolic-diastolic points using NNRW are conducted into two experiments. The first experiment was conducted to measure the performance using the various number of neurons in the hidden layer. This experiment also evaluates the algorithm's performance using the 10 systolic and 10 diastolic features. After that, after comparing the results of overall accuracy and *Gmean* accuracy of training and testing results, the number of hidden neurons was chosen as the most optimal. As a result, 500 hidden neurons outperformed compared with all the cases.

Furthermore, the performance of the second investigation was assessed using five assessment metrics: accuracy (*Acc*), precision (*+P*), sensitivity (*Se*), specificity (*Sp)*, and *F-score*. The performance of the NNRW algorithm was also estimated using the convincing overall accuracy and *Gmean* of testing results with 94.86 and 94.74 percent, respectively.

Experiment 1: Performance Evaluation based on Different Hidden Neuron Number from 10 to 1000

Clinical Data Participants

This experiment is carried out with two distinct datasets of systolic and diastolic features, as well as their annotated peaks and points, respectively. The experiment is repeated ten times and the outcomes for the training and testing processes are evaluated using two measurement metrics: overall accuracy and *Gmean* accuracy. The results also contain the percentage of the value for evaluation standard metrics prediction such as the average of value (*Mean*), maximum (*Max*), and minimum (*Min*) value, also standard deviation (*STD*) for comparison purposes. The numbers of hidden

neurons are suitable and needed for the NNRW classifier to get greater generalization ability by setting from 10 to 1000 numbers of hidden neurons.

Referring to Figure 4, for Liang et. al.'s data, the *Gmean* accuracy of training and testing sets for 218 clinical PPG data participants between the values of a mean percentage to the increment of hidden neuron values. For training results, it can be observed that there is an increment in mean percentage achieved above 96 percent with the setting of 500 hidden neurons. For the testing results, upon reaching 100 neurons, it can be observed that there is a decrement in mean percentage lower than 94 percent. Next, between the ranges of 100 neurons to 500 neurons, the mean percentage stayed between the ranges of 94.46 percent to 96.07 percent, below 96.07 percent with 1.61 percent of the difference between the lowest to highest the mean value.

When the 500 to 600 neurons are setting up, the mean percentage show some increment between the ranges of 96.07 percent to 96.28 percent, with 0.21 percent of the difference from lowest to highest the mean value. When the 600 to 700 neurons are set up, there is 0.14 percent of the difference, the mean percentage shows increment between the ranges of 96.28 percent to 96.42 percent. Hidden neurons value between the ranges 700 to 800 neurons show some increment in between 96.42 and 96.66 percent with 0.24 percent, the highest difference range compared to other hidden neurons range. The ranges between 800 to 900 hidden neurons show 0.17 percent of the difference between 96.66 to 96.83 percent, and the highest mean is 97.05 percent between the ranges 900 to 1000 of hidden neurons. The result from the *Gmean* accuracy of testing sets with mean percentage began to decrease when neurons value in between 500 to1000. The mean percentage shows the ranges between 94.74 percent decreased to 94.01 percent.

Although, the mean *Gmean* accuracy of training sets can further increase with neurons value up to 1000. The mean *Gmean* accuracy of testing sets began to reduce if more than 500 neurons are set up. Based on this experiment, the evaluation from above 500 hidden neurons can contribute becomes a relatively longer computational time. The overall evaluation results for 500 hidden neurons for clinical participants are tabulated in Table III.

Normal Data Participants

Figure 5 shows the *Gmean*'s accuracy of training and testing results from in-between values of a mean percentage to a different hidden neuron value. It can be observed from the training results show some increment in mean percentage up above 94.08 percent with the setting of 500 hidden neurons, while for the testing results, it shows the accuracy is above 90.15 percent. The performance starts to decline when the number of hidden neurons is increased. It can be seen that once 100 neurons are set up, the increase in mean percentage becomes lower than 91.31 percent for training results and less than 89.91 for the testing result. Next, between the ranges of 100 to 500, the mean percentage remained the ranges from 91.31 to 94.08 percent and below 94.08 percent with 2.77 percent of the difference between the lowest to highest the mean value.

It can be concluded that when the neuron is setting up to 500, the *Gmean* for training results are obtained to 94.08 percent whereas testing results have reached their maximum, which is 90.15 percent. The training performance continues increasing up to 1000 neurons. However, for the testing set, it began to decline. Referring to Fig. 5 and Table IV, 500 hidden neurons for normal participants have been chosen as the optimal hidden neurons value for classification.

Experiment 2: Performance Evaluation using 500 Neurons.

Clinical Data Participants

This experiment purposely evaluates the algorithm's performance based on the chosen optimal number of neurons in the hidden layer. To identify various classes of peaks and points, the first step of annotation for the systolic peak and the diastolic point has been carried out at the pre-processing signal process. After that, the peak and point candidates will be compared to annotated peaks and points as four classes which are True Systolic (TS), False Systolic (FS), True Diastolic (TD), and False Diastolic (FD).

Table II. recalls the result of five important evaluation metrics for TS and FS classes and the comparison is tabulated in Table V. Table V. shows the four evaluation metrics for mean training results of TS score above 98.39 percent. Only precision (*+P*) metric with score 97.83 percent, lowest than others metrics. This could be due to the high number of predicted results that have been incorrectly classified as TS. For the FS score, four metrics score above 97.61 percent, while the remaining score with 96.73 percent, for sensitivity (*Se*) metrics.

Table V. also arranges the comparison of mean training results from five significant evaluation metrics for TD and FD classes (see results in Table II.). The results appeared as only one metric evaluation from both classes scored above 97.85 percent, which is from specificity (*Sp*) metrics. While remaining four others metrics (*Acc*, *+P*, *Se*, and *F-score*) showed scores between 94.10 percent to 97.10 percent from both classes, respectively. *Sp* in both classes was the highest score with almost 97.90 percent. Although, the same relationship between TD and FD where a high occurrence of falsely predicted value as TD can be linked to the high occurrence of falsely predicted value as non-FD, or either. Therefore, the *F-score* result can be affected depending on the relationship between *+P* and *Se* result.

Referring to the results in Table V., it can conclude that the accuracy of the four classes performs well, especially as the classification of TS and FS. The accuracy for TD and FD both almost at 97 percent makes a high accuracy performance result to our objective in identifying diastolic point. Without excluding the accuracy of TS and FS, both classes performed with a near-perfect result almost in 99 percent.

Normal Data Participants

Referring to experiment 1 for normal data participants, the 500 neurons are chosen as the optimal number of hidden neurons to put along with the NNRW classifier. Table VI. illustrates the result of five significant evaluation metrics for TS and FS classes of classification for training and testing results while Table VII. draws a comparison for mean training results between the classes. That table estimates the four evaluation metrics for mean training results of TS score very well above 97.19 percent. Only sensitivity (*Se*) metric with score 94.56 percent, lowest than others metrics. For the FS score, four metrics score above 98.47 percent, while the remaining score with 97.01 percent, for precision (+*P*) metrics. The results show that there is a relationship between TS and FS.

Table VII. also arranges the evaluation of mean training results from five significant performance metrics prediction for TD and FD classes (see results in Table VII.). The results also appeared as the four evaluation metrics for mean training results of TD score very well above 90.72 percent. Only sensitivity (*Se*) metric with score 83.24 percent, lowest than others metrics. For the FD score, four metrics score above 96.70 percent, while the remaining score with 95.18 percent, for specificity (*Sp*) metrics.

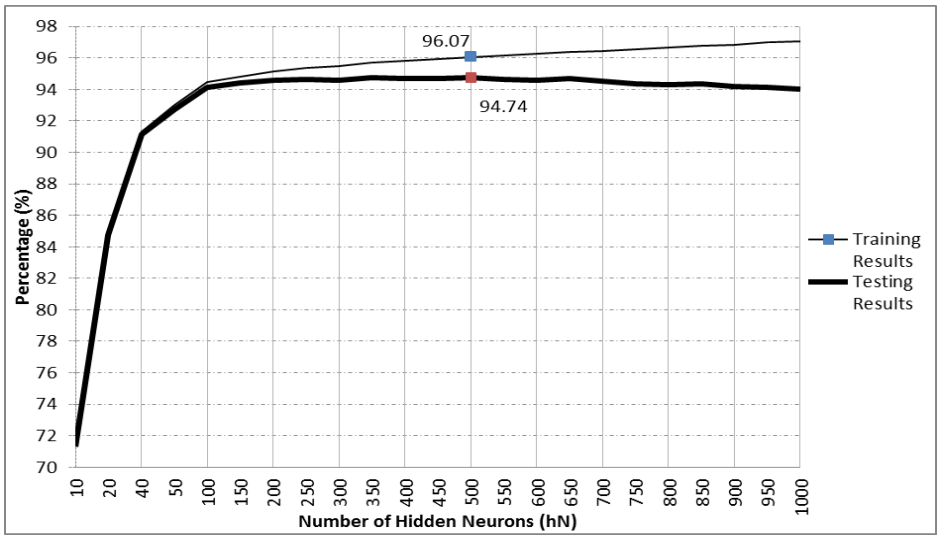

Figure 4. The comparison for performance number of hidden neurons using a mean percentage of clinical PPG data participants from *Gmean* accuracy of training and testing result.

Table III. Overall Evaluation Results Using 500 Neurons (Clinical Participants).

	Training Results				Testing Results			
	Mean (%)	**Max (%)**	**Min (%)**	**Std. Dev (%)**	**Mean (%)**	**Max (%)**	**Min (%)**	**Std. Dev (%)**
Overall Accuracy	96.17	96.29	96.10	0.06	94.86	95.09	94.66	0.14
Gmean	96.07	96.18	95.97	0.07	94.74	94.96	94.53	0.14

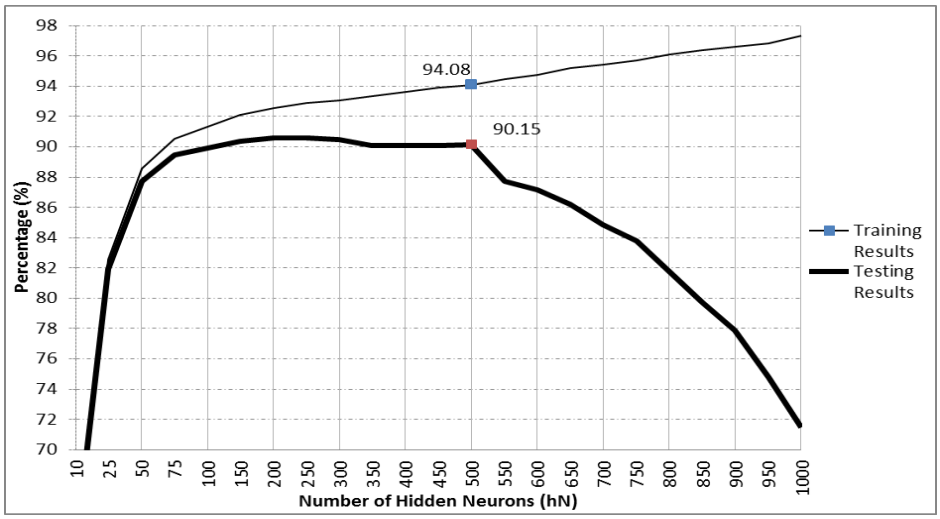

Figure 5. Performance evaluation on different neurons value using a mean percentage of normal PPG data participants from *Gmean* accuracy of training and testing results.

Table IV. Overall Evaluation Results Using 500 Neurons (Normal Participants).

	Training Results				Testing Results			
	Mean (%)	Max (%)	Min (%)	Std. Dev (%)	Mean (%)	Max (%)	Min (%)	Std. Dev (%)
Overall Accuracy	95.74	95.94	95.54	0.13	92.68	92.99	90.37	0.43
Gmean	94.08	94.41	93.79	0.19	90.15	90.51	88.02	0.43

Comparison with the Previous Study

Table VIII. compares the results for detection systolic peak and diastolic point with the previously published methods [13] that use dataset from Mateo Aboy et al. [19]. The same classification we have used along with the NNRW classifier and the features extraction process. As is seen in Table VIII., our proposed method has resulted in a smaller prediction percentage for two performance evaluation metrics, using overall accuracy and *Gmean* result compared to the method proposed in [13].

The mean percentage result from both papers estimated at 95.60 percent and 94.72 percent for overall accuracy and *Gmean* of testing results, respectively. We believe that this is due to consists of the detection system of expert annotation (JM and DT) from the dataset [19], as well as using a more accurate pre-processing process for detecting key points in the signals. A dataset in [19] contains 4603 sample signals randomly select from 47 subjects. While, in this work, all 2100 data per signal from 218 clinical subjects and 45 volunteers as such, many signal parts belong to the same individuals were shuffled and then partitioned into training and testing sets. This means that in this work, some samples of the test subjects were in the training sets, which makes the prediction task significantly complex compared to work in [13] where no sample from the test subjects are in the training set.

Table IX., recalls the comparison result of five significant evaluation metrics for TS and FS classes between this work and paper in [13]. This table noticed that from both research works, four evaluation metrics of TS score very well above 98.39 percent. After compared and analyzed that only precision (*+P*) metric with the mean percentage 98.41 percent from both papers, lowest than others metrics result. This could be due to the highest number of cases in both comparisons that were incorrectly predicted as TS.

For the FS score comparison result, four metrics score above 99.46 percent, while the remaining score with 99.02 percent, for sensitivity (*Se*) metrics in previous work, followed as well as result in this work. The lower result of 98.58 percent from both *Se* could be due to the highest number of results incorrectly predicted as non-FS. From the result, we can identify both existing works on the same track to the close correlation of TS and FS.

Table IX. also arranges the comparison of results between this study and the previous work [13], from five significant evaluation metrics for TD and FD classes. The results in previous work seem like only one metric from both classes was the highest score with almost 99.98 percent, which is from specificity (*Sp*) metrics. While remaining four others metrics (*Acc*, *+P*, *Se*, and *F-score*) showed scores between 99.86 percent to 99.95 percent from both classes, respectively. The mean percentage from both papers, which is from *Sp* metrics show scored above 97.67 percent, highest than others metrics score result. Although, the same relationship between TD and FD in paper [13], followed as well as the result in this study.

By summarizing Table IX., we can conclude that the overall performance evaluation of this study compared with the results from the previous paper, T. Hann et al. [13] are performed well, too might guarantee the quality of method performance using NNRW-based method for multi-class classification systolic and diastolic peak detection of PPG signals. Later, in the future, we can investigate the seven different labels of hypertension to each of the ground truth and estimated BP values.

$$Overall\ accuracy = \frac{Sum\ of\ Correct\ Choices}{Sum\ of\ All\ Choices} \times 100 = \frac{\sum_{i=1}^{I}(TP_i + TN_i)}{All\ Predictions\ Made} \times 100 \tag{8}$$

$$G_{mean} = (Sensitivity_a \times Sensitivity_b \times Senstivity_c \times Senstivity_d)^{\frac{1}{4}} \times 100 = \sqrt[4]{TPR_a \times TPR_b \times TPR_c \times TPR_d} \times 100 \tag{9}$$

Table V. Comparison of Mean Training Result For Four Classes Classification (Clinical Participants)

Class	Classes Classification	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F-score (%)
0	True-Systolic peak (TS)	99.03	97.83	98.95	99.10	98.39
1	False-Systolic peak (FS)	98.98	98.50	96.73	99.61	97.61
2	True-Diastolic point (TD)	97.10	95.78	94.10	97.85	94.94
3	False-Diastolic point (FD)	97.08	94.32	94.32	97.87	94.32

Table VI. Four Classes Classification Results for Normal Participants.

Performance Evaluation	Classes Classification	Training Results				Testing Results			
		Mean (%)	*Maximum (%)*	*Minimum (%)*	*Standard deviation*	*Mean (%)*	*Maximum (%)*	*Minimum (%)*	*Standard deviation*
Accuracy	TS	99.02	99.07	98.91	0.05	96.88	97.52	96.24	0.39
	FS	99.02	99.07	98.91	0.05	97.46	97.69	97.20	0.18
	TD	96.70	96.82	96.61	0.06	93.65	94.48	93.04	0.42
	FD	96.70	96.82	96.61	0.06	93.60	94.48	92.95	0.56
Precision	TS	99.98	100.00	99.90	0.04	96.62	97.52	95.76	0.61
	FS	97.01	97.16	96.66	0.14	95.01	95.70	93.84	0.51
	TD	99.70	100.00	99.11	0.29	89.74	92.89	86.44	1.75
	FD	99.85	100.00	99.56	0.14	94.90	96.57	93.28	0.91
Sensitivity	TS	94.56	94.84	93.92	0.26	86.05	88.66	83.16	1.71
	FS	99.98	100.00	99.94	0.02	97.32	97.80	96.23	0.53
	TD	83.24	83.70	82.71	0.31	76.81	78.16	75.43	0.91
	FD	99.85	100.00	99.56	0.14	96.44	97.69	95.53	0.80
Specificity	TS	99.99	100.00	99.98	0.01	99.36	99.53	99.19	0.12
	FS	98.65	98.72	98.49	0.06	97.76	98.08	97.19	0.24
	TD	95.18	95.31	95.03	0.09	97.87	98.58	97.16	0.38
	FD	95.18	95.31	95.03	0.09	92.45	93.22	91.73	0.50
F-score	TS	97.19	97.35	96.86	0.14	90.98	92.86	89.02	1.16
	FS	98.47	98.55	98.30	0.07	96.14	96.52	95.78	0.28
	TD	90.72	91.08	90.49	0.19	82.68	84.74	81.20	1.07
	FD	99.85	100.00	99.56	0.14	95.66	96.99	94.46	0.80

Table VII. Comparison of Mean Training Result For Four Classes Classification (Normal Participants).

Class	Classes Classification	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F-score (%)
0	True-Systolic peak (TS)	99.02	99.98	94.56	99.99	97.19
1	False-Systolic peak (FS)	99.02	97.01	99.98	98.65	98.47
2	True-Diastolic point (TD)	96.70	99.70	83.24	95.18	90.72
3	False-Diastolic point (FD)	96.70	99.85	99.85	95.18	99.85

Table VIII. Comparison of Overall Performance with existing work.

Overall Performance Evaluation	Participants Category	Work / Publication Year	Hidden Neurons	Training Results				Testing Results			
				Mean (%)	Max (%)	Min (%)	Std. Dev. (%)	Mean (%)	Max (%)	Min (%)	Std. Dev. (%)
Overall Accuracy	Clinical	This work	500	96.17	96.29	96.10	0.06	94.86	95.09	94.66	0.14
		[13] / 2019	700	99.67	99.70	99.64	0.01	99.26	99.39	99.19	0.04
	Normal	This work	500	95.74	95.94	95.54	0.13	92.68	92.99	90.37	0.43
	Mean (%)			97.19	97.31	97.09	0.07	95.60	95.82	94.74	0.20
Gmean	Clinical	This work	500	96.07	96.18	95.97	0.07	94.74	94.96	94.53	0.14
		[13] / 2019	700	99.68	99.71	99.64	0.01	99.27	99.40	99.20	0.04
	Normal	This work	500	94.08	94.41	93.79	0.19	90.15	90.51	88.02	0.43
	Mean (%)			96.61	96.77	96.47	0.09	94.72	94.96	93.92	0.20

Table IX. Comparison of four classes classification with other published work.

Classes Classification	Participants Category	Work / Publication Year	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F-score (%)
True-Systolic peak (*TS*)	Clinical	This work	99.03	97.83	98.95	99.10	98.39
		[13] / 2019	99.72	98.98	99.90	99.67	99.44
	Mean (%)		99.38	98.41	99.43	99.39	98.92
	Normal	This work	99.02	99.98	94.56	99.99	97.19
	Mean (%)		99.26	98.93	97.80	99.59	98.34
False-Systolic peak (*FS*)	Clinical	This work	98.98	98.50	96.73	99.61	97.61
		[13] / 2019	99.72	99.90	99.02	99.97	99.46
	Mean (%)		99.35	99.20	97.88	99.79	98.54
	Normal	This work	99.02	97.01	99.98	98.65	98.47
	Mean (%)		99.24	98.47	98.58	99.41	98.51
True-Diastolic point (*TD*)	Clinical	This work	97.10	95.78	94.10	97.85	94.94
		[13] / 2019	99.95	99.86	99.93	99.98	99.89
	Mean (%)		98.53	97.82	97.02	98.92	97.42
	Normal	This work	96.70	99.70	83.24	95.18	90.72
	Mean (%)		97.92	98.45	92.42	97.67	95.18
False-Diastolic point (*FD*)	Clinical	This work	97.08	94.32	94.32	97.87	94.32
		[13] / 2019	99.95	99.86	99.86	99.98	99.86
	Mean (%)		98.52	97.09	97.09	98.93	97.09
	Normal	This work	96.70	99.85	99.85	95.18	99.85
	Mean (%)		97.91	98.01	98.01	97.68	98.01

CONCLUSION

In conclusion, this study has accomplished the performance evaluation for a multi-class NNRW-based approach for systolic-diastolic peak prediction from PPG signals on two different open benchmark datasets. The first dataset is an open clinical China database and the second dataset is data collected from undergraduate students, Universiti Malaysia Pahang (UMP). It is evaluated using two experimental setups, then compared the results with previous research so that can guarantee the generalization of method performance.

The earlier module of the assessing work deal with pre-processing signals to eliminate the noise from the PPG signal. It then was sent for predicting true peaks and points after the manual annotation stage of systolic-diastolic points for the raw signals. This is followed by feature extraction and feature classification approached by 4-fold cross validation, and finally generated the performance results of predicted systolic-diastolic peaks during training and testing using the NNRW classifier. The best performance examined that the research achieved good results for prediction of True Systolic peak and True Diastolic peak with 99.03 percent and 97.10 percent for clinical databased respectively while for normal databased almost achieved with 99.02 and 96.70 percent, respectively. The performance testing results to the overall accuracy and *Gmean* for clinical dataset also achieved the result with 94.86 percent and 94.74 percent, respectively while for normal dataset the results obtained 92.68 percent and 90.15 percent, respectively.

The finding results from the two experiments showed the ability of the algorithm to predict systolic and diastolic peaks and points of PPG Signals, respectively. For future work, this algorithm can be employed to predict and estimate any medical health applications related to oxygen's blood including hypertension, diabetes, lung diseases, heart failure, and congenital heart defects.

ACKNOWLEDGMENT

The authors wish to acknowledge Liang et al. [20] for sharing their PPG database. The Malaysian Ministry of Education is funding this research through the Fundamental Research Grant Scheme (FRGS), RDU190144, (Ref: FRGS/1/2018/ ICT02/UMP/02/16).

REFERENCES

[1] Y. Kurylyak, F. Lamonaca, and D. Grimaldi, "A Neural Network-based method for continuous blood pressure estimation from a PPG signal," *Conference Record - IEEE Instrumentation and Measurement Technology Conference*, pp. 280–283, 2013, doi: 10.1109/I2MTC.2013.6555424.

[2] D. Elgendi, M., Norton, I., Brearley, M., Abbott, D., & Schuurmans, "Systolic Peak Detection in Acceleration Photoplethysmograms Measured from Emergency Responders in Tropical Conditions," *PLoS ONE*, vol. Vol. 8, 2013.

[3] S. Kuntamalla and L. Ram Gopal Reddy, "An Efficient and Automatic Systolic Peak Detection Algorithm for Photoplethysmographic Signals," *International Journal of Computer Applications*, vol. 97, no. 19, pp. 18–23, 2014, doi: 10.5120/17115-7686.

[4] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuffless Blood Pressure Estimation Algorithms for Continuous Health-Care Monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 4, pp. 859–869, 2017, doi: 10.1109/TBME.2016.2580904.

[5] S. C. Gao, P. Wittek, L. Zhao, and W. J. Jiang, "Data-driven estimation of blood pressure using photoplethysmographic signals," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2016-Octob, pp. 766–769, 2016, doi: 10.1109/EMBC.2016.7590814.

[6] H. Tee, P. D. I. C. A., Rao, P. D. P., Adam, A., Potti, C., & Rajagopal, "Systolic Blood Pressure Peak Classification of PPG Signals using Time-domain Features," *Journal of Medical Informatics and Computational Methods in Engineering and Health Sciences, Japan.*, 2016.

[7] Y. Zhang and Z. Wang, "A hybrid model for blood pressure prediction from a PPG signal based on MIV and GA-BP neural network," *ICNC-FSKD 2017 - 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, pp. 1989–1993, 2018, doi: 10.1109/FSKD.2017.8393073.

[8] S. N. Shukla, "Estimation of blood pressure from non-invasive data," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 1, no. 3, pp. 1772–1775, 2017, doi: 10.1109/EMBC.2017.8037187.

[9] S. G. Khalid, J. Zhang, F. Chen, and D. Zheng, "Blood Pressure Estimation Using Photoplethysmography Only: Comparison between Different Machine Learning Approaches," *Journal of Healthcare Engineering*, vol. 2018, 2018, doi: 10.1155/2018/1548647.

[10] S. il Cho, T. Negishi, M. Tsuchiya, M. Yasuda, and M. Yokoyama, "Estimation system of blood pressure variation with photoplethysmography signals using multiple regression analysis and neural network," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 18, no. 4, pp. 229–236, 2018, doi: 10.5391/IJFIS.2018.18.4.229.

[11] S. Chen, Z. Ji, H. Wu, and Y. Xu, "A Non-Invasive Continuous Blood Pressure Estimation Approach Based on Machine Learning," *Sensors*, vol. 19, no. 11, p. 2585, 2019, doi: 10.3390/s19112585.

[12] M. H. Chowdhury *et al.*, "Estimating Blood Pressure from the Photoplethysmogram Signal and Demographic Features Using Machine Learning Techniques," *Sensors*, 2020.

[13] T. J. Hann, N. Liza Simon, and A. Adam, "Systolic and Diastolic Multiclass Classification of PPG Signals using Neural Network with Random Weight," 2019. doi: 10.1109/SSCI44817.2019.9002876.

[14] Z. Xu, J. Liu, X. Chen, Y. Wang, and Z. Zhao, "Continuous blood pressure estimation based on multiple parameters from eletrocardiogram and photoplethysmogram by Back-propagation neural network," *Computers in Industry*, vol. 89, pp. 50–59, 2017, doi: 10.1016/j.compind.2017.04.003.

[15] Z. Z. Zhihong Xu, Jiexin Liu, Xianxiang Chen, Yilong Wang, "Continuous blood pressure estimation based on multiple parameters from eletrocardiogram and photoplethysmogram by Back-propagation neural network," *Computers in Industry*, vol. Volume 89, no. 0166–3615, pp. 50–59, 2017, doi: https://doi.org/10.1016/j.compind.2017.04.003.

[16] X. Xing and M. Sun, "Optical blood pressure estimation with photoplethysmography and FFT-based neural networks," *Biomedical Optics Express*, vol. 7, no. 8, p. 3007, 2016, doi: 10.1364/boe.7.003007.

[17] X. Wang, L., Zhou, W., Xing, Y., & Zhou, "A Novel Neural Network Model for Blood Pressure Estimation Using Photoplethesmography without Electrocardiogram.," *J Healthc Eng*, vol. 7804243., 2018.

[18] N. Hasanzadeh, M. M. Ahmadi, and H. Mohammadzade, *Blood Pressure Estimation Using Photoplethysmogram Signal and Its Morphological Features*, vol. 20, no. 8. 2020, pp. 4300–4310. doi: 10.1109/JSEN.2019.2961411.

[19] B. Aboy, M., McNames, J., Tran, T., Tsunami, D., Ellenby, M. S., & Goldstein, "An automatic beat detection algorithm for pressure signals.," *IEEE Transactions on Biomedical Engineering*, vol. 52(10), 16, 2005.

[20] Y. Liang, Z. Chen, G. Liu, and M. Elgendi, "A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China," *Scientific Data*, vol. 5, pp. 1–7, 2018, doi: 10.1038/sdata.2018.20.

[21] L. Wang, W. Zhou, Y. Xing, and X. Zhou, "A novel neural network model for blood pressure estimation using photoplethesmography without electrocardiogram," *Journal of Healthcare Engineering*, vol. 2018, 2018, doi: 10.1155/2018/7804243.

[22] M. G. and S. A. J. Liu, D., "University of Queensland vital signs dataset: development of an accessible repository of anesthesia patient monitoring data for research.," *Anesthesia & Analgesia*, vol. 114(3), pp. 584–9, 2012.

[23] Z. & E. Liang, Y., Liu, G, Chen, "M. Figshare," 2017, doi: https://doi.org/10.6084/m9.figshare.5459299.

[24] I. Adam, A., Ibrahim, Z., Mokhtar, N., Shapiai, M. I., Mubin, M., & Saad, "Feature selection using angle modulated simulated Kalman filter for peak classification of EEG signals.," *Springerplus, 5(1), 1580.*, 2016.

[25] W.F. Schmidt; M.A. Kraaijveld; R.P.W. Duin, "Feedforward neural networks with random weights," *Paper presented at the Proceedings., 11th IAPR International Conference on Pattern Recognition. Vol.II. Conference B: Pattern Recognition Methodology and Systems, 1992.*, 1992.

[26] M. G. O'brien, E.; Waeber, B.; Parati, G.; Staessen, J.; Myers, "Blood pressure measuring devices: Recommendations of the European Society of Hypertension.," *BMJ*, vol. 322, pp. 531–536., 2001.

[27] G. Mancia et al., "2013 ESH/ESC guidelines for the management of arterial hypertension: the Task Force for the Management of Arterial Hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC),"," *Blood Press.*, vol. 22, no. 4, pp. 193–278, 2013.

[28] B. Stergiou, G.S.; Alpert, B.; Mieke, S.; Asmar, R.; Atkins, N.; Eckert, S.; Frick, G.; Friedman, "A universal standard for the validation of blood pressure measuring devices: Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization (AAMI/ESH/ISO) Collaboration Statement.," *Hypertension 2018*, vol. 71, pp. 368–374., 2018.