

CARDIAC ARRHYTHMIA PREDICTION USING ENSEMBLE MODEL BASED ON HEART RATE VARIABILITY FEATURES

DICKY LIEGAR AND SANI MUHAMAD ISA*

Computer Science Department, BINUS Graduate Program – Master of Computer Science
Bina Nusantara University

JL. K. H. Syahdan No. 9, Kemanggisian, Palmerah, Jakarta 11480, Indonesia
dicky.liegar@binus.ac.id; *Corresponding author: sani.m.isa@binus.ac.id

Received November 2020; accepted February 2021

ABSTRACT. *Cardiac arrhythmia is an irregular heartbeat condition that may lead to sudden cardiac arrest (SCA) that usually causes death if it is not treated immediately. This study aims to predict cardiac arrhythmia five minutes before its occurrence with high accuracy which could give earlier medical treatment. An ensemble model is proposed to predict the upcoming condition within a five minutes interval by using five minutes duration of heart rate variability (HRV) features from time domain, frequency domain, and nonlinear domain analysis. The result shows that an ensemble model consisting of a neural network, support vector machine, and probabilistic neural network can improve the predictive performance and achieve an accuracy of 97%.*

Keywords: Cardiac arrhythmia, Prediction, Ensemble model, Heart rate variability

1. **Introduction.** Sudden cardiac arrest (SCA) is a sudden and unexpected case of heart failure which causes the heart to not pump blood anymore [1]. If it is not treated within minutes, it usually causes death. Such case of SCA is called sudden cardiac death (SCD). In the United States, SCD still exists and the total number of cases reaches around 325,000 each year [2]. This implies that SCD remains an important cause of mortality.

A disease called cardiac arrhythmia often precedes SCD [3]. This disease occurs when the heart beats either too fast, known as tachycardia, or too slowly, known as bradycardia [4]. Ventricular fibrillation (VF), as a type of tachycardia, is a rapid irregular heartbeat that is the most frequent cause of SCD [5]. VF itself may be developed from ventricular tachycardia (VT), the same rapid heartbeat but with the regular pattern. An extreme form of VT known as ventricular flutter (VFL) is considered as transition phase from VT to VF [6].

Previous researchers have tried to predict cardiac arrhythmia using heart rate variability (HRV) features extracted from electrocardiogram (ECG) [7-9]. All of them have similarities, and they used a single model to predict: Joo et al. [7] and Lee et al. [9] used a neural network while Murukesan et al. [8] used a support vector machine and a probabilistic neural network separately. In a similar topic, Amin et al. [10] used an ensemble model that combined Naïve Bayes and logistic regression for heart disease prediction achieving an accuracy of 87.41%, outperforming its single model accuracy of 84.81% and 85.86%. Therefore, this study attempts to develop ensemble models based on the models that have been used by previous researchers [7-9], aiming for better and more stable predictive performance. These ensemble models will be compared and decided which one is the best, as the proposed model. However, these previous works used different datasets and prediction intervals which cannot be compared directly. Thus, the same experimental procedures will be conducted in this study to compare the results.

The rest of this work is organized as follows: Section 2 describes the literature review of theories and related work for developing the proposed model; Section 3 elaborates the steps of the experiment; Section 4 discusses results of the study; Section 5 closes the study with conclusion and suggestion for future work.

2. Literature Review.

2.1. Neural network. A neural network (NN) is an algorithm inspired by how the human brain works [11]. NN consists of input layer, hidden layer, and output layer, in which each layer can have neurons. The input layer will hold the input data, the output layer will give the output solution of the problem, and the hidden layer is between the input layer and the output layer. A NN that has more than one hidden layer is called a multi-layer perceptron (MLP). The information will be transferred from the input layer to the next layer until the output layer.

2.2. Support vector machine. According to Pradhan [12], a support vector machine (SVM) is a model that can be used for classification, recognition, and prediction. The main idea of SVM is to develop a hyperplane that can separate the data to its corresponding class. For example, in a 2-dimensional space, the hyperplane will be a line that separates the data. The optimal line separator is the one that can maximize the distance between the line to the data. The kernel function may be applied to modifying the dimensionality of non-linear separable data to become linearly separable data.

2.3. Probabilistic neural network. Probabilistic neural network (PNN) is another type of neural network developed to improve model training computation cost which used the backpropagation method before [13]. PNN architecture was introduced with 4 layers which consist of input layer, pattern layer, summation layer, and decision layer [14]. The input layer is just like the normal input layer of the neural network which represents the input features. The number of pattern layer neurons will be the same as the number of training samples. The pattern layer will apply 2 calculations that calculate feature distance between the input data and the training data and then apply radial basis function based on its distance. The number of summation layer neurons will be the same as the number of classes. Each summation layer neuron will sum each pattern layer neurons from the same class. Finally, the decision layer will choose the maximum output from among summation layer neurons, which will represent the predicted class.

2.4. Ensemble learning. According to Sagi and Rokach [15], ensemble learning is a method that combines multiple models to solve some tasks, such as classification and regression. The model itself can be any type of machine learning algorithm. This concept can be illustrated through the story about hundreds of people participating in a weight guessing contest. Among the participants, no one could guess the correct answer. The interesting part comes when the average of all the answers calculated which came close to the correct answer. This means that combining multiple predictions may lead to better predictions.

There are several further explanations of why the ensemble method can improve the performance. Firstly, ensemble methods may avoid overfitting which tends to happen when the model is trained with a small amount of data. Secondly, the multiple models can avoid getting stuck into local optima. Lastly, combining the models will make the search space extended and hence can give better performance.

2.5. Cardiac arrhythmia prediction. There are several approaches to predicting cardiac arrhythmia, such as by analyzing T wave alternans, QT dispersion, and heart rate variability (HRV) [16]. HRV shows variation or fluctuation between heartbeats in some

time interval, compared to heart rate which only shows heartbeats per minute [17]. There are three common HRV measurements or analyses which are from time domain, frequency domain, and nonlinear domain. Analyzing heart conditions using HRV is a robust approach because it is measured from RR intervals that are less affected by ECG noise [18]. This approach had been used by several studies to classify cardiac arrhythmia [19,20]. Then, in a matter of prediction, several studies also used this approach to predict cardiac arrhythmia [7-9].

Joo et al. [7] had done the research to predict VT which is usually used as SCD predictor. They used Spontaneous Ventricular Tachyarrhythmia Database as the dataset and NN with 25 hidden neurons in the first hidden layer, and 30 hidden neurons in the second hidden layer achieved an accuracy of 82.9% (ten seconds before occurrence). Murukesan et al. [8] also tried to predict VF using HRV features. They used MIT/BIH SCD Holter Database for SCA and Physiobank Normal Sinus Rhythm (NSR) for normal control data. They used two different methods, SVM using radial basis function (RBF) kernel with scaling factor of 1 and PNN with spreading value of 0.4, which achieve accuracy of 96.63% and 93.64% (two minutes before occurrence), respectively. In more recent research, Lee et al. [9] had also done the research to predict VT. They used data from monitors of Asan Medical Center patients. As for the model, NN with 5 hidden neurons achieved an accuracy of 73.5% (one hour before occurrence).

Based on [7-9], there are differences in HRV features set used for the prediction, even though all of them used 5 minutes duration of HRV features, as described in Table 1. The time intervals of prediction vary from 10 seconds to 1 hour and the datasets used for the experiment are also various. This means that these works cannot be compared to each other and decide which work achieves the best performance. Thus, experiments using the same datasets will be conducted in this study to compare the performances.

TABLE 1. Heart rate variability features

Domain	Feature	Description	Joo et al. [7]	Murukesan et al. [8]	Lee et al. [9]
Time	MeanNN	Mean of NN interval	✓		✓
	SDNN	Standard deviation of NN interval	✓		✓
	RMSSD	Square root of the mean squared differences of successive NN intervals	✓		✓
	pNN50	Proportion of interval differences of successive NN interval greater than 50 ms	✓		✓
	Outlier	Number of ectopic beats		✓	
	sdHR	Standard deviation of heart rate		✓	
Frequency	VLF	Power in very low frequency range (0-0.04 Hz)	✓		✓
	LF	Power in low frequency range (0.04-0.15 Hz)	✓		✓
	HF	Power in high frequency range (0.15-0.4 Hz)	✓		✓
	LF/HF	Ratio of LF over HF	✓		✓
	aTotal	Total power of the bands		✓	
	pVLF	Percentage of VLF Power over total power		✓	
	pLF	Percentage of LF Power over total power		✓	
Nonlinear	SD1	Standard deviation of points perpendicular to the axis of line-of-identity	✓	✓	✓
	SD2	Standard deviation of points along the axis of line-of-identity	✓		✓
	SD1/SD2	Ratio of SD1 over SD2	✓		✓
	Alpha	Component of detrended fluctuation analysis		✓	

3. **Methodology.** Figure 1 illustrates the steps of the experiment that will be conducted. The experiment starts from data collection, data preprocessing, HRV feature extraction, training and evaluating single models, making ensemble combinations, and evaluating the ensemble models.

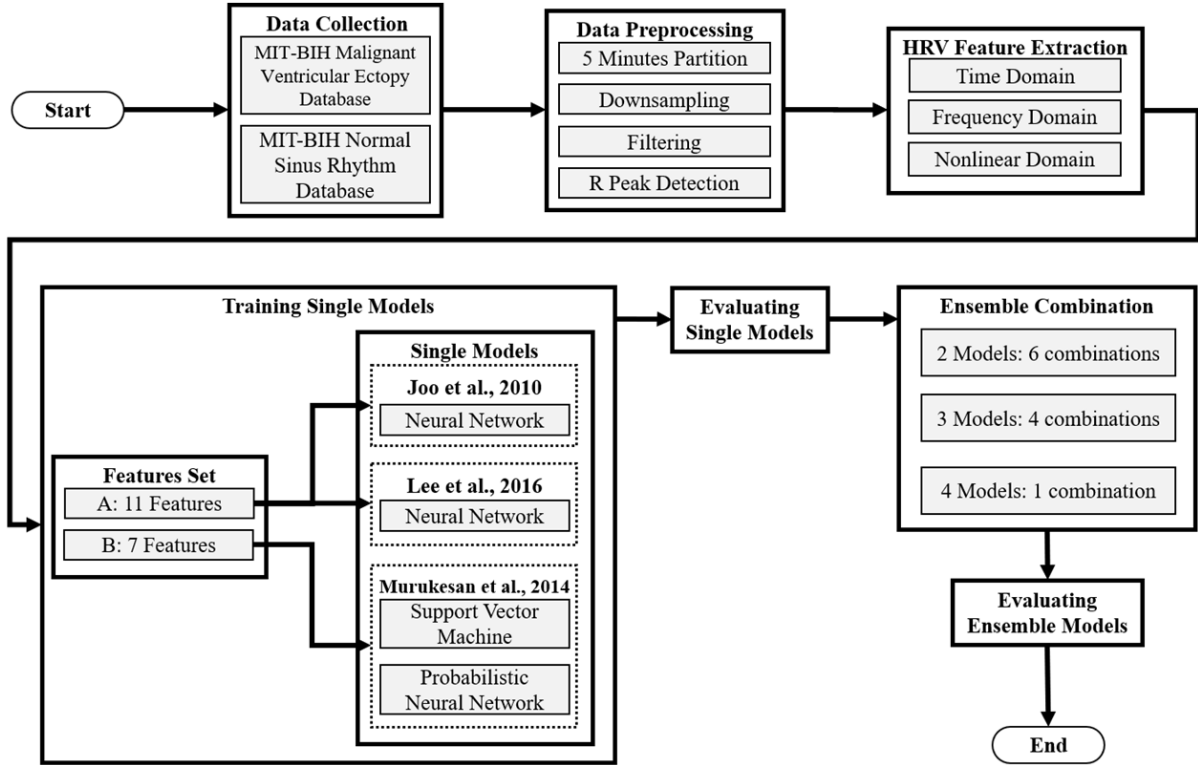


FIGURE 1. Steps of experiment

3.1. **Data collection.** There were two datasets used in this study, collected from PhysioNet repository [21]. The first dataset that will be used is from MIT-BIH Malignant Ventricular Ectopy Database (vfdb). This database consists of 22 half-hour electrocardiogram (ECG) recordings with frequency sampling of 250 Hz which is based on research by Greenwald [22]. The ECG record itself is provided with annotation or label to show what happens with the patient in the corresponding time, such as VT, VFL, and VF. However, the location of R peak is not provided in this database which will be detected in the next process. This database will be used as positive samples.

The second dataset that will be used is from MIT-BIH Normal Sinus Rhythm Database (nsrdb). This database consists of 18 long-term ECG recordings with frequency sampling of 128 Hz. The recordings were gathered from 5 men (26-45 years old) and 13 women (20-50 years old) at Boston's Beth Israel Deaconess Medical Center. The location of R peak is also not provided in this dataset. This database will be used as negative samples.

3.2. **Data preprocessing.** After collecting the ECG records, each ECG record will be sliced with five minutes duration. Each partition of the positive samples will have a five minutes interval from the corresponding events. The number of negative samples will be equal to the number of positive samples to balance the ratio. In this study, a total of 300 samples data will be used for experiment. The positive samples with frequency sampling of 250 Hz will be downsampled to match negative samples with frequency sampling of 128 Hz.

There may exist unwanted signals or noise which can be merged with the true ECG signal during recording. This noise needs to be removed, or else it may cause analyzing

problems which may lead to false diagnosis. The noise itself has its frequency characteristic, then removing or passing a certain range of frequency will produce the true ECG signal. A signal processing called bandpass filter will only allow a certain range of signal frequency that may reduce the noise of the ECG signal.

HRV features which will be extracted in the next process will need the location of R peak in the signal. As stated before, the collected data will have no information about location of R peak. The location of R peak will be detected based on the filtered ECG signal. R peak detection can be done using modified Engzee approach by Lourenço et al. [23], which offers more adaptive detection rather than the original approach by Engelse and Zeelenberg [24] that may not recognize R peak with various ranges of amplitude because of its fixed threshold.

3.3. Feature extraction. After getting the location of the R peak in the ECG signal, the interval between adjacent R peaks (RR interval) will be calculated in millisecond. After that, HRV features will be analyzed based on RR intervals. Then, the value of features will be scaled to normalize the range of value between features. A simple feature scaling called min-max scaling will be applied to the features. The calculation will be based on the minimum and the maximum value of the feature, which will convert the range from 0 to 1.

$$\text{scaled feature} = \frac{\text{feature} - \min(\text{feature})}{\max \text{feature} - \min \text{feature}} \quad (1)$$

3.4. Training and evaluation. The scaled features dataset will be split into multiple partitions for training, validation, and testing set with $K * L$ -Fold Split (nested cross validation) with K value of 3 and L value of 5. The K value represents how many outer partitions or folds will be created from the dataset. An outer fold will be used as a testing set while the other $K - 1$ outer fold will be used as training sets and this process will be repeated K times until all fold is used as testing set. The L value represents how many inner partitions or folds will be created from the $K - 2$ outer fold. An inner fold will be used as a validation set while the other $L - 1$ outer fold will be used as training sets and this process will be repeated for L times until all fold is used as validation set.

Then, single model training will be conducted with each single model taking the corresponding set of features. After finishing single models training, these models will predict the event by using the testing set and will be evaluated by 4 performance metrics: accuracy, F1 score, precision, and recall.

Finally, the trained single models will be combined to make ensemble models as mentioned before. Voting is a simple ensemble method that combines predictions from multiple models to predict the output. In another topic, Mahabub et al. [25] proposed an intelligence detection system based on ensemble voting method for message filtering that outperforms the performance of the single models. In this study, the ensemble model prediction will be based on soft voting, which takes the probability outputs of the models and makes a prediction based on the average of the probability outputs [26], as shown in Equation (2) below. However, not all models produce prediction probability, and models like support vector machines will only produce nominal output. In this case, the Platt scaling method [27] can be used to transform the nominal output to the probability output. The ensemble models will be evaluated the same as the single models.

$$\text{Soft Voting Prediction} = \frac{1}{n} \sum_{i=1}^n \text{Prediction Probability}_i \quad (2)$$

where n = Number of models.

TABLE 2. Single model evaluation results

Single models	Accuracy	F1 score	Precision	Recall
NN [7]	95.0% ($\pm 3.61\%$)	94.88% ($\pm 3.88\%$)	95.77% ($\pm 5.7\%$)	94.67% ($\pm 9.24\%$)
SVM [8]	95.67% ($\pm 4.16\%$)	95.38% ($\pm 4.67\%$)	98.04% ($\pm 1.96\%$)	93.33% ($\pm 9.87\%$)
PNN [8]	95.67% ($\pm 4.04\%$)	95.63% ($\pm 4.0\%$)	96.79% ($\pm 5.55\%$)	94.67% ($\pm 5.03\%$)
NN [9]	94.0% ($\pm 5.0\%$)	93.42% ($\pm 5.69\%$)	99.35% ($\pm 1.13\%$)	88.67% ($\pm 11.02\%$)
Overall metric score	95.08% ($\pm 3.68\%$)	94.83% ($\pm 4.04\%$)	97.49% ($\pm 3.8\%$)	92.83% ($\pm 8.16\%$)

TABLE 3. Ensemble model evaluation results

Ensemble models	Accuracy	F1 score	Precision	Recall
NN [7] + SVM [8]	96.33% ($\pm 3.51\%$)	96.21% ($\pm 3.76\%$)	97.53% ($\pm 4.28\%$)	95.33% ($\pm 8.08\%$)
NN [7] + PNN [8]	96.67% ($\pm 4.04\%$)	96.44% ($\pm 4.45\%$)	98.69% ($\pm 1.13\%$)	94.67% ($\pm 9.24\%$)
NN [7] + NN [9]	96.0% ($\pm 3.61\%$)	95.8% ($\pm 4.01\%$)	97.46% ($\pm 2.87\%$)	94.67% ($\pm 9.24\%$)
SVM [8] + PNN [8]	95.67% ($\pm 3.06\%$)	95.56% ($\pm 3.12\%$)	97.31% ($\pm 3.13\%$)	94.0% ($\pm 5.29\%$)
SVM [8] + NN [9]	96.0% ($\pm 3.61\%$)	95.94% ($\pm 3.62\%$)	97.39% ($\pm 4.53\%$)	94.67% ($\pm 5.03\%$)
PNN [8] + NN [9]	96.67% ($\pm 3.06\%$)	96.51% ($\pm 3.23\%$)	99.31% ($\pm 1.2\%$)	94.0% ($\pm 6.0\%$)
NN [7] + SVM [8] + PNN [8]	97.0% ($\pm 3.0\%$)	96.9% ($\pm 3.2\%$)	98.11% ($\pm 3.27\%$)	96.0% ($\pm 6.93\%$)
NN [7] + SVM [8] + NN [9]	96.0% ($\pm 4.0\%$)	95.82% ($\pm 4.36\%$)	97.53% ($\pm 4.28\%$)	94.67% ($\pm 9.24\%$)
NN [7] + PNN [8] + NN [9]	96.67% ($\pm 4.04\%$)	96.44% ($\pm 4.45\%$)	98.69% ($\pm 1.13\%$)	94.67% ($\pm 9.24\%$)
SVM [8] + PNN [8] + NN [9]	95.67% ($\pm 3.79\%$)	95.52% ($\pm 3.9\%$)	97.96% ($\pm 3.53\%$)	93.33% ($\pm 6.11\%$)
NN [7] + SVM [8] + PNN [8] + NN [9]	96.67% ($\pm 3.51\%$)	96.52% ($\pm 3.8\%$)	98.11% ($\pm 3.27\%$)	95.33% ($\pm 8.0\%$)
Overall metric score	96.3% ($\pm 3.01\%$)	96.15% ($\pm 3.21\%$)	98.01% ($\pm 2.73\%$)	94.67% ($\pm 6.4\%$)

4. **Results and Discussion.** Initially, single model evaluation results are represented in Table 2, showing the average and standard deviation of the models performance metrics. Through this experimental result, the overall accuracy and F1 score are quite similar, with a difference around 0.25%. This shows that these models can predict both events quite well. This may be achieved because the distribution of both events is equal. Among these models, PNN [8] model achieved the highest performance among the single models with an accuracy of 95.67% and F1 score of 95.63%.

Then, combinations of these single models as ensemble models are evaluated with the same evaluation procedure as the single model. The evaluation results of ensemble models are summarized in Table 3. Compared to the single models, the overall ensemble models metric performance achieved higher average and lower standard deviation. This shows that the ensemble models have better predictive performance, more accurate and more

stable. The same pattern as the single models result, the overall accuracy and F1 score metric are quite similar, with a lower difference around 0.15%. Among these ensemble models, a combination of different model algorithms achieved the best result, consisting of NN [7] with SVM [8] and PNN [8]. It achieved the highest accuracy of 97% with the lowest standard deviation of 3% and the highest F1 score of 96.9% with almost the lowest standard deviation of 3.2%. Compared to the best single model, the accuracy is increased by 1.33% and F1 score is increased by 1.27%. Thus, this best ensemble model is chosen as the proposed model.

To get a better representation of performance improvement, Figure 2 shows performance comparison among the single models, NN [7], SVM [8], PNN [8], and NN [9], respectively with the proposed model. The boxplot may represent the minimum and maximum metric score and also how stable is the model, smaller box means a more stable model. As visualized below, the proposed model achieved better and more stable performance compared to the other single models.

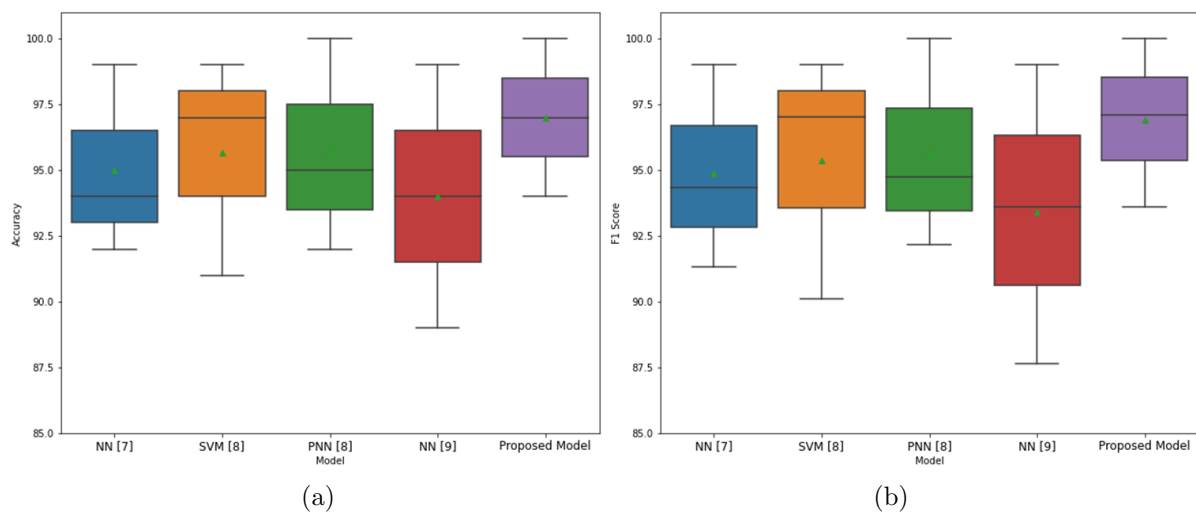


FIGURE 2. (a) Accuracy performance comparison; (b) F1 score performance comparison

5. Conclusion and Future Work. This study illustrates how the ensemble model can be used to improve model performance for cardiac arrhythmia prediction, more accurate, and more stable compared to the single model. We proposed an ensemble model consisting of a neural network, support vector machine, and probabilistic neural network. This model achieves an average accuracy of 97% in predicting cardiac arrhythmia within a five minutes interval with five minutes duration of HRV features from time domain, frequency domain, and nonlinear domain analysis. This study raised several works to be explored. There are different ensemble approaches yet to be implemented for the prediction. Furthermore, a similar experiment can be conducted with a longer prediction interval and larger dataset, as there are limitations to the dataset used in this study.

REFERENCES

- [1] National Heart, Lung, and Blood Institute, *Sudden Cardiac Arrest*, <https://www.nhlbi.nih.gov/health-topics/sudden-cardiac-arrest>, Accessed in 2020.
- [2] Cleveland Clinic, *Sudden Cardiac Death (SCD): Symptoms, Causes*, <https://my.clevelandclinic.org/health/diseases/17522-sudden-cardiac-death-sudden-cardiac-arrest>, 2019.
- [3] D. P. Zipes and H. J. Wellens, Sudden cardiac death, *Circulation*, vol.98, no.21, pp.2334-2351, 1998.
- [4] MedlinePlus, *Arrhythmia*, <https://medlineplus.gov/arrhythmia.html>, 2020.
- [5] Mayo Clinic, *Ventricular Fibrillation*, <https://www.mayoclinic.org/diseases-conditions/ventricular-fibrillation/symptoms-causes/syc-20364523>, 2017.

- [6] Radcliffe Cardiology, *Ventricular Flutter*, <https://www.radcliffecardiology.com/electrophysiology-arrhythmia/ventricular-arrhythmias/ventricular-flutter>, Accessed in 2020.
- [7] S. Joo, K. J. Choi and S. J. Huh, Prediction of ventricular tachycardia by a neural network using parameters of heart rate variability, *2010 Computing in Cardiology*, pp.585-588, 2010.
- [8] L. Murukesan, M. Murugappan, M. Iqbal and K. Saravanan, Machine learning approach for sudden cardiac arrest prediction based on optimal heart rate variability features, *Journal of Medical Imaging and Health Informatics*, vol.4, no.4, pp.521-532, 2014.
- [9] H. Lee, S.-Y. Shin, M. Seo, G.-B. Nam and S. Joo, Prediction of ventricular tachycardia one hour before occurrence using artificial neural networks, *Scientific Reports*, vol.6, 2016.
- [10] M. S. Amin, Y. K. Chiam and K. D. Varathan, Identification of significant features and data mining techniques in predicting heart disease, *Telematics and Informatics*, vol.36, pp.82-93, 2019.
- [11] M. A. Nielsen, *Neural Networks and Deep Learning*, Determination Press, San Francisco, 2015.
- [12] A. Pradhan, Support vector machine – A survey, *International Journal of Emerging Technology and Advanced Engineering*, vol.2, no.8, pp.82-85, 2012.
- [13] D. F. Specht, Probabilistic neural networks, *Neural Networks*, vol.3, no.1, pp.109-118, 1990.
- [14] B. Mohebali, A. Tahmassebi, A. Meyer-Baese and A. H. Gandomi, Probabilistic neural networks: A brief overview of theory, implementation, and application, in *Handbook of Probabilistic Models*, P. Samui, D. T. Bui, S. Chakraborty and R. C. Deo (eds.), Butterworth-Heinemann, 2020.
- [15] O. Sagi and L. Rokach, Ensemble learning: A survey, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol.8, no.4, 2018.
- [16] R. E. Lane, Prediction and prevention of sudden cardiac death in heart failure, *Heart*, vol.91, no.5, pp.674-680, 2005.
- [17] F. Shaffer and J. P. Ginsberg, An overview of heart rate variability metrics and norms, *Frontiers in Public Health*, vol.5, DOI: 10.3389/fpubh.2017.00258, 2017.
- [18] M. Tavassoli, M. M. Ebadzadeh and H. Malek, Classification of cardiac arrhythmia with respect to ECG and HRV signal by genetic programming, *Canadian Journal on Artificial Intelligence, Machine Learning and Pattern Recognition*, vol.3, no.1, pp.1-8, 2012.
- [19] M. G. Tsipouras and D. I. Fotiadis, Automatic arrhythmia detection based on time and time-frequency analysis of heart rate variability, *Computer Methods and Programs in Biomedicine*, vol.74, no.2, pp.95-108, 2004.
- [20] B. M. Asl, S. K. Setarehdan and M. Mohebabi, Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal, *Artif. Intell. Med.*, vol.44, no.1, pp.51-64, 2008.
- [21] A. L. Goldberger et al., PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals, *Circulation*, vol.101, no.23, pp.e215-e220, 2000.
- [22] S. D. Greenwald, *The Development and Analysis of a Ventricular Fibrillation Detector*, Master Thesis, Massachusetts Institute of Technology, 1986.
- [23] A. Lourenço, H. Silva, P. Leite, R. Lourenço and A. L. N. Fred, Real time electrocardiogram segmentation for finger based ECG biometrics, *Biosignals*, pp.49-54, 2012.
- [24] W. A. H. Engelse and C. Zeelenberg, A single scan algorithm for QRS-detection and feature extraction, *Computers in Cardiology*, vol.6, no.1979, pp.37-42, 1979.
- [25] A. Mahabub, M. I. Mahmud and M. F. Hossain, A robust system for message filtering using an ensemble machine learning supervised approach, *ICIC Express Letters, Part B: Applications*, vol.10, no.9, pp.805-812, 2019.
- [26] Z. H. Zhou, *Ensemble Methods: Foundations and Algorithms*, Chapman and Hall/CRC, New York, 2012.
- [27] J. Platt, Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, *Advances in Large Margin Classifiers*, vol.10, no.3, pp.61-74, 1999.