Analysis of Machine Learning Algorithm for Sleep Apnea Detection Based on Heart Rate Variability

Muhammad Zakariyah[1], Umar Zaky[2]

[1]Department of Informatics, University Technology of Yogyakarta, Indonesia
[2]Department of Information System, University Technology of Yogyakarta, Indonesia
[1]muhammad.zakariyah@staff.uty.ac.id, [2]umar.zaky@staff.uty.ac.id

Abstract - Sleep apnea is a common problem with health implications ranging from excessive daytime sleepiness to serious cardiovascular disorders. The method for detecting and measuring sleep apnea is through breathing monitoring (polysomnography), which is time consuming and relatively expensive. Cardiovascular which is closely related to heart performance activities allows the use of electrocardiogram (heart rate variability) features to detect sleep apnea. This study aims to compare the results of sleep apnea detection using several machine learning algorithms. A total of 2,445 data were divided into 1,834 data as learning sets and 611 data as test sets. Evaluation of 10-fold cross-validation using all HRV features shows that neural network algorithm has the best performance compared to decision tree algorithm, k-nearest neighbor, and support vector machine with an accuracy rate (82.44% in the learning set, 79.21% in the test set consecutively), precision (85.54% and 82.70%), f-measure (87.70% and 85.67%), and AUC (0.867 and 0.832). Based on the results of performance testing using only selected HRV features (CVRR, HF, SD1/SD2 Ratio, and S-Region), the K-Nearest Neighbors, Support Vector Machine, and Neural Network algorithms experienced a decrease in performance. The use of all HRV features is recommended compared to only using selected HRV features, so it can help detect the presence/absence of sleep apnea much better.

Keywords: Sleep apnea detection; machine learning; heart rate variability; electrocardiogram

I. INTRODUCTION

Sleep apnea is a common medical disorder that occurs when the walls of the throat relax and narrow, causing obstruction of the airways during sleep. This condition inhibits the supply of oxygen that enters the body, which then decreasing oxygens blood levels. Chronic fatigue and daytime sleepiness despite having slept for more than seven hours, are some of the common symptoms in people with sleep apnea [1]. The other most common symptoms of sleep apnea are snoring, stop breathing, and gasping for air during sleep. Many people are not aware that they are experiencing symptoms of sleep apnea, so the diagnosis of this disease is often overlooked or even untreated.

Sleep apnea has implications as a considerable morbidity for cardiovascular disease [2], mental illness, even affect a person's quality of life [3-4]. Polysomnography (PSG) is a tool for examine patients' physiological condition during sleep. Data collection through PSG has two main drawbacks, it takes a long time and quite expensive [5]. To overcome the weakness of PSG, several alternatives have been proposed using physiological signals, such as stomach signals [6], air flow [7], chest signal [8], and oxygen saturation [9].

Sleep apnea can be more severe while untreated. The limitations of sleep laboratories and PSG in each health facility make it difficult to predict the severity of sleep apnea. Therefore, other methods are needed to be able to detect, even predict sleep apnea. Breathing, blood pressure, body temperature, and heart rate are closely related to the autonomic nervous system. Breathing disorders during sleep that interfere with temporal harmony, also cause disturbances to the autonomic nervous system and in the long-term cause complications in the heart, brain, blood circulation and metabolism function.

Autonomic nervous system activity can be monitored through heart rate, namely Heart Rate Variability (HRV) [10]. Many studies on HRV have been carried out for decades. HRV has been used to detect several disorders such as stress [11-12], insomnia and epilepsy [13], impaired cognitive function [14], even used to predict mortality from sepsis [15].

Heart rate variability is one of the analytical methods commonly used to determine changes in heart activity. The HRV method focuses on changes in time interval oscillations and heart rate speed, so that the physiological condition of the body can be known based on its heart rhythm. Heart rate variability is a physiological phenomenon that reflects an indicator of the autonomic nervous system that works subconsciously to control body work/activities [16]. In individuals who have sleep disorders, there is an imbalance of autonomic nervous system function, and this imbalance in autonomic nervous system function can then be measured using HRV.

This study has a hypothesis that sleep apnea can be detected using machine learning algorithm based on heart rate variability features. An objective evaluation method for the analysis of the features/parameters that most influence the early diagnosis of sleep apnea, is very important for prevention, health promotion, and quality of life improvement. With the development of wearable devices to monitor heart rate activity, it is possible to monitor a person's physical condition in daily life. This study aims to find features and compare the performance of several machine learning algorithms that can be used to detect sleep apnea.

This research is important to do considering the lack of awareness about the importance of early detection of sleep apnea, which is often known after complications of other diseases arise due to sleep apnea. The use of HRV as a parameter to detect sleep apnea is also expected to be an alternative to using PSG which is quite expensive and takes a long time to get the results. In addition, HRV can also be generated through a smart watch, which makes it possible to monitor the presence or absence of sleep apnea. Through the results of the discovery of HRV features and the most suitable algorithm for sleep apnea detection, it is hoped that the model can be developed by embedding it in a wearable health device (smart watch).

II. METHOD

A. Dataset

The dataset used in this study is an ECG signal recording data [17], taken with 100 Hz frequency sampling [18]. The initial dataset is data from the ECG recording of 6 subjects while sleeping for 7 hours with a total of 2445 data. The data is divided into 1834 data as learning sets and 611 data as test sets. Both learning set and test set have been evaluated by sleep apnea experts, to provide a label/annotation (presence or absence of sleep apnea) every minute.

B. Signal Pre-processing

Pre-processing data consists of three stages: windowing stage, peak detection, and R-R interval (Fig. 1). The windowing stage is the stage to break down data from each subject (RAW ECG Data), which initially lasted 7 hours into data with a minute duration, so that each subject had ± 420 data.

The peak detection stage is the stage for processing the signal in each window. ECG signal data for each window, processed using the Two Moving Average Algorithm [19-20] to detect the peak of each QRS Complex (R-Peak) pulse. Each R-Peak detected in each window, will be used to calculate the distance between beats (difference between two R-Peaks), which is commonly known as the R-R Interval. The new time series data will be obtained from the results of the R-R Interval process.

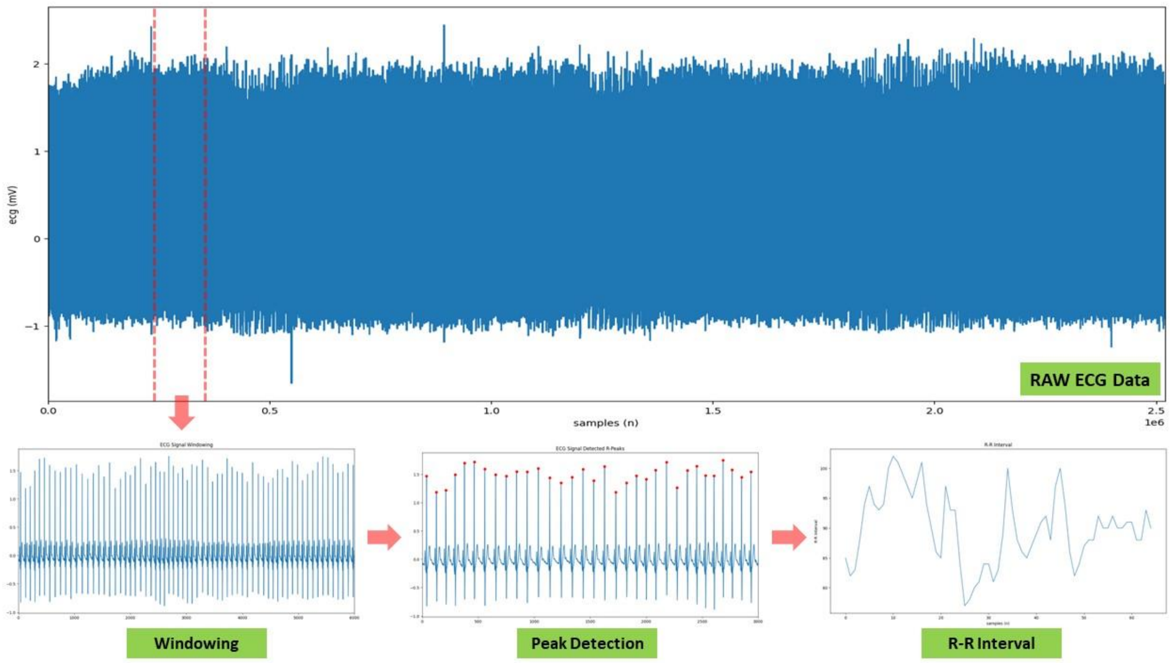

Fig. 1 Pre-processing data stages

C. Heart Rate Variability Features Extraction

Time series data obtained from R-R Interval then being extracted into HRV features. The HRV features consists of linear measurements and non-linear measurements. Linear measurement consists of time domain analysis and frequency domain analysis, while non-linear measurement consists of poincaré plots analysis and sample entropy (sampen). Table I shows HRV features extracted from R-R Interval.

After extracting HRV features, annotating/labeling is carried out on each data. Labeling is done on all data (learning set and test set). Labeled data in the learning set is used to train the model in machine learning, while labeled data in the test set is used to test whether the detection results in each algorithm match the label that should be.

D. Machine Learning Classifiers

Sleep apnea detection in this study was carried out using a classification algorithm in machine learning. Based on previous research, there are several algorithms that can be used to classify. Comparison in this study uses four classification algorithms according to previous studies: Decision Tree [15], K-Nearest Neighbors [8], Support Vector Machine [9], dan Neural Network [6]. The device specifications and HRV features used for classification in all algorithms have been standardized.

E. K-Fold Cross-Validation

Testing is carried out using k-Fold Cross-Validation, which is a re-sampling procedure to evaluate machine learning models on a limited sample of data [21]. The number of groups (k) in this study is 10, so the dataset will be divided into 10 groups (fold) and a total of 10 iterations of cross-validation will be implemented. Fig. 2 shows the steps on 10k-Fold Cross-Validation. For each iteration of the k-Fold Cross-Validation, a machine learning classification algorithm is used to measure the performance. Performance measurement of the classification algorithm is performed using a confusion matrix and Receiver Operating Characteristic curve (ROC curve) [22].

Confusion matrix is a performance measurement for machine learning classification problems, where the output is table with 4 different combinations of predicted values and actual values: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). This table is useful for measuring Recall, Precision, Accuracy, and AUC (Area Under ROC Curve). AUC is a graph that shows the performance of the classification model across all classification thresholds. This curve plots two parameters, specifically True Positive Rate (TPR) and False Positive Rate (FPR).

TABLE I
HEART RATE VARIABILITY FEATURES

Measurement	Feature	Unit	Description
Time Domain Analysis	MeanRR	ms	Average of R-R Interval.
	SDRR	ms	Standard deviation of R-R Interval
	CVRR	ms	Coefficient of Variance of R-R Interval
	RMSSD	ms	The mean squared distance of the R-R Interval consecutively
	pNN50	%	Percentage difference between adjacent R-R intervals that are greater than 50 ms
	pNN20	%	Percentage difference between adjacent R-R intervals that are greater than 20 ms
Frequency Domain Analysis	VLF	ms^2	Absolute power of very low frequency (0.0033-0.04 Hz)
	LF	ms^2	Absolute power of low frequency (0.04-0.15 Hz)
	HF	ms^2	Absolute power of high frequency (0.15-0.4 Hz)
	LF/HF	%	LF to HF Rasio
Poincaré Plot Analysis	SD1	ms	The standard deviation of the poincaré plot is perpendicular to the identity line
	SD2	ms	The standard deviation of the poincaré plot along the identity line
	SD1/SD2	%	SD1 to SD2 Ratio
	S-Region	ms	The area of ellipse that represents total HRV
Sample Entropy	Sampen		A measure of regularity and complexity of a time series

ms = milisecond; ms^2 = milisecond squared

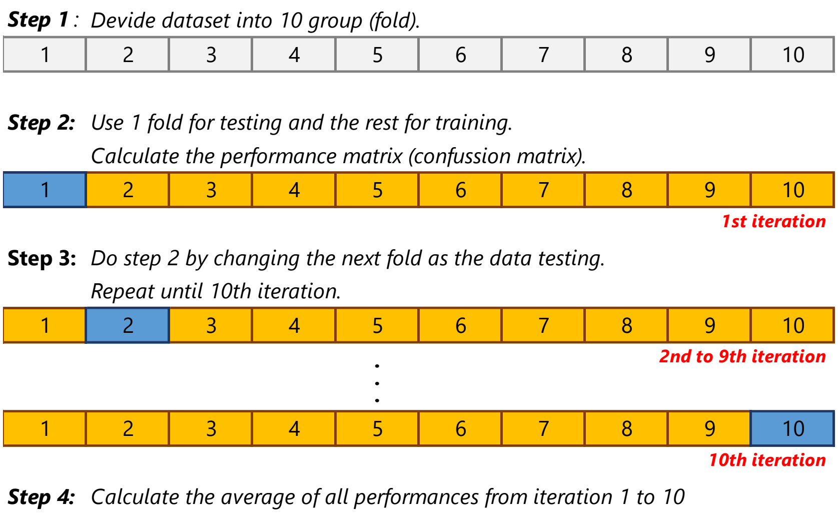

Fig. 2 Steps in k-Fold cross-validation with k=10

III. RESULTS AND DISCUSSION

A. *Classification Parameters on Machine Learning*

The algorithms explored in this study are some of the well-known and frequently used algorithms. Comparison of several classification algorithms is performed, considering that there is no one universal algorithm that can be used to solve all classification problems [23]. Through this comparison, hope that it can provide an overview of the capabilities and performance of each algorithm in detecting sleep apnea. The performance of each algorithm is strongly influenced by the parameters used. Therefore, parameters used in this study are the default parameters in RapidMiner tools. Table II shows the default parameters of each classification algorithm.

B. *Evaluation Using All HRV Features*

Performance evaluation included: Accuracy, Recall, Precision, Area Under ROC Curve (AUC), and F-Measure. Table III shows the test results of each classification algorithm using all HRV features in the learning set. The results of this study indicate that NN has the highest accuracy rate of 82.44%, followed by SVM (81.19%), and KNN (80.04%). The DT algorithm has the worst performance among the machine learning algorithms tested, with accuracy (74.97%), precision (73.84%), F-Measure (84.52), and AUC of 0.603. However, this algorithm has the best recall value, which is 98.82%. Overall, the NN algorithm using default parameters in the learning set has the best results among the four algorithms compared.

Testing is also performed on the test set data, where the data has never been included in the training process (learning). However, the test set data has been given label, so that it can be used to test performance as in the learning set data. Performance results for test set data as shown in Table IV.

TABLE II
DEFAULT PARAMETERS OF EACH CLASSIFICATION ALGORITHM

No.	Classification Algorithm	Parameters	Value
1.	Decision Tree (DT)	Split Criteria	Gain Ratio
		Maximal Depth	10
2.	K-Nearest Neighbors (KNN)	K	5
3.	Support Vector Machine (SVM)	Kernel Type	Dot
		Kernel Cache	200
		C	0,01
		Convergence Epsilon	0,001
		Maximum Iteration	100000
4.	Neural Network (NN)	Training Cycle	200
		Learning Rate	0,01
		Hidden Layer	10
		Momentum	0,9

TABLE III
LEARNING SET PERFORMANCE RESULTS USING ALL HRV FEATURES

Classification Algorithm	Accuracy (%)	Recall (%)	Precision (%)	F-Measure (%)	AUC
DT	74.97 ± 1.85	**98.82 ± 0.85**	73.84 ± 1.42	84.52 ± 1.06	0.603 ± 0.026
KNN	80.04 ± 2.85	88.95 ± 2.50	83.30 ± 1.90	86.03 ± 2.16	0.799 ± 0.046
SVM	81.19 ± 2.42	92.98 ± 1.90	82.16 ± 1.77	87.24 ± 1.83	0.824 ± 0.028
NN	**82.44 ± 3.97**	89.98 ± 6.35	**85.54 ± 3.02**	**87.70 ± 4.09**	**0.867 ± 0.037**

TABLE IV
TEST SET PERFORMANCE RESULTS USING ALL HRV FEATURES

Classification Algorithm	Accuracy (%)	Recall (%)	Precision (%)	F-Measure (%)	AUC
DT	73.16 ± 2.25	97.63 ± 1.93	72.82 ± 1.88	83.42 ± 1.90	0.585 ± 0.021
KNN	74.14 ± 6.27	86.04 ± 6.77	78.70 ± 4.35	82.21 ± 5.30	0.714 ± 0.065
SVM	70.21 ± 1.83	**98.11 ± 2.17**	70.43 ± 1.31	82.00 ± 1.63	0.796 ± 0.053
NN	**79.21 ± 5.01**	88.86 ± 5.00	**82.70 ± 4.80**	**85.67 ± 4.90**	**0.832 ± 0.052**

Performance of NN algorithm when tested on the test set data, still shows the best accuracy, precision, f-measure, and AUC (79.21%, 82.70%, 85.67%, and 0.832 respectively) compared to other algorithms. Different performance is shown in the DT algorithm, which previously showed the worst performance. The test results on the test set data, show that SVM algorithm has the lowest level of accuracy (70.21%), precision (70.43%), and f-measure (82.00%). However, this algorithm also has the highest recall performance.

C. Heart Rate Variability Features Selection

The purpose of HRV features selection is to find the best HRV feature set to build a model in detecting the presence or absence of sleep apnea. Feature selection technique for labeled data is using supervised learning models. The feature selection technique with the supervised learning model consists of: Embedded Method, Wrapper Method, and Filter Method.The features of the HRV extraction are then selected based on the test results of the three methods involving all the data. The test results for each feature selection technique with a supervised learning model can be seen in Table V.

Wrapper Method provides better performance than filter-based or embedded feature selection [24]. Similar results are also seen in Table V which shows that based on three feature selection methods, the wrapper method provides better accuracy, precision, and f-measure values. The wrapper method has two feature selection techniques, namely Forward Selection and Backward Elimination.

Forward selection was chosen because it produced a better score than backward elimination (accuracy, precision, and f-measure are 79.96%, 81.20%, and 86.50%, respectively). This method works iteratively, starting with using one feature to be tested using machine learning algorithms, then calculating its accuracy/performance. Features that have the highest level of accuracy will be retained. In the next iteration, this feature is then added with other HRV features. New feature combination then trained using a learning algorithm, so that new accuracy results are obtained. This is done continuously, until no significant increase in accuracy is found. The combination of HRV features with the highest performance will be selected as the features that will be used for classification in machine learning.

As seen in Table VI, feature selection using forward selection produces a weight which will be used to determine whether the feature will be used to detect sleep apnea or not. Based on these results, features with 1 value will be used as a parameter in the machine learning classification. On the other hand, features with 0 value will not be used in the classification process. According to the results in the Table VI, the HRV features that will be used for the machine learning classification process are CVRR, HF, SD1/SD2 Ratio, and S-Region.

D. Evaluation Using Selected HRV Features

Selected features are tested using same data (learning set and test set) as previuos testing. This is done to see the impact of using only selected HRV features on the machine learning classification algorithm performance. The results of learning set performance using only selected HRV features are shown in the Table VII.

TABLE V
TEST RESULTS ON EACH FEATURE SELECTION TECHNIQUE

No.	Feature Selection Technique	*Accuracy* (%)	*Recall* (%)	*Precision* (%)	*F-Measure* (%)
1.	**Embedded Method**	75.38 ± 2.71	98.58 ± 0.80	74.30 ± 2.34	84.74 ± 1.19
2.	**Wrapper Method**				
	Forward Selection	**79.96 ± 1.09**	92.54 ± 2.48	**81.20 ± 2.10**	**86.50 ± 2.27**
	Backward Elimination	78.29 ± 2.22	94.08 ± 1.72	78.70 ± 1.79	85.71 ± 1.75
3.	**Filter Method**				
	Weight by Information Gain	74.68 ± 1.40	98.34 ± 1.00	73.78 ± 1.16	84.31 ± 1.07
	Weight by Correlation	74.40 ± 1.78	98.11 ± 2.05	73.65 ± 1.60	84.14 ± 1.80
	Weight by Chi-Square	74.84 ± 2.90	98.46 ± 0.75	73.94 ± 2.62	84.46 ± 1.17
	Weight by Relief	74.19 ± 1.87	**98.88 ± 1.41**	73.24 ± 1.91	84.15 ± 1.62

TABLE VI
FORWARD SELECTION RESULTS ON HRV FEATURES

No.	Feature	Weight
1.	MeanRR	0
2.	SDRR	0
3.	**CVRR**	**1**
4.	RMSSD	0
5.	pNN50	0
6.	pNN20	0
7.	VLF	0
8.	LF	0
9.	**HF**	**1**
10.	LF/HF Ratio	0
11.	SD1	0
12.	SD2	0
13.	**SD1/SD2 Ratio**	**1**
14.	**S-Region**	**1**
15.	SampEn	0

Accuracy describes the ratio between the number of correct predictions and the total number of predictions. Among the four classification algorithms tested, DT algorithm has the highest accuracy (80.10%) followed by the NN algorithm (79.77%), KNN (76.23%), and SVM (71.38%). The f-measure was performed to calculate for false positives and false negatives from unbalanced dataset. The F-measure on DT algorithm also has the highest value (86.57%), although this value does not show a significant difference with the results on NN algorithm (86.10%). Performance comparison between using all HRV features and only selected HRV features in the learning set data, can be seen in Fig. 3.

As shown in Fig. 3 using only CVRR, HF, SD1/SD2 Ratio, and S-Region features, made DT algorithm shows a significant change in performance. The accuracy increased from 74.97% to 80.10%. Likewise with precision (81.66%) and f-measure (86.57%). But overall, the NN algorithm still shows good performance results, even though the accuracy rate has decreased, from 82.44% to 79.77%. A significant decrease in accuracy was even shown by KNN algorithm (from 80.04% to 76.23%) and SVM (from 81.19% to 71.38%). Performance of the test set data is also carried out to compare the selected features as shown in Table VIII.

Table VIII shows the NN algorithm still has the highest accuracy (77.42%) when tested using test set data, even though only using the selected HRV feature. The same result is also shown by the NN algorithm on the f-measure (84.79%). To compare the performance of each algorithm between using all HRV features and only using selected HRV features in the test set data can be seen in Fig. 4.

TABLE VII
LEARNING SET PERFORMANCE RESULTS USING SELECTED HRV FEATURES

Classification Algorithm	Accuracy (%)	Recall (%)	Precision (%)	F-Measure (%)	AUC
DT	**80.10 ± 2.90**	92.11 ± 2.88	81.66 ± 3.36	**86.57 ± 3.10**	0.753 ± 0.054
KNN	76.23 ± 2.84	86.74 ± 3.41	80.41 ± 1.84	83.46 ± 2.39	0.765 ± 0.040
SVM	71.38 ± 1.26	**99.84 ± 0.33**	70.76 ± 0.92	82.82 ± 0.49	0.799 ± 0.024
NN	79.77 ± 2.26	90.61 ± 3.34	**82.02 ± 1.43**	86.10 ± 2.00	**0.800 ± 0.041**

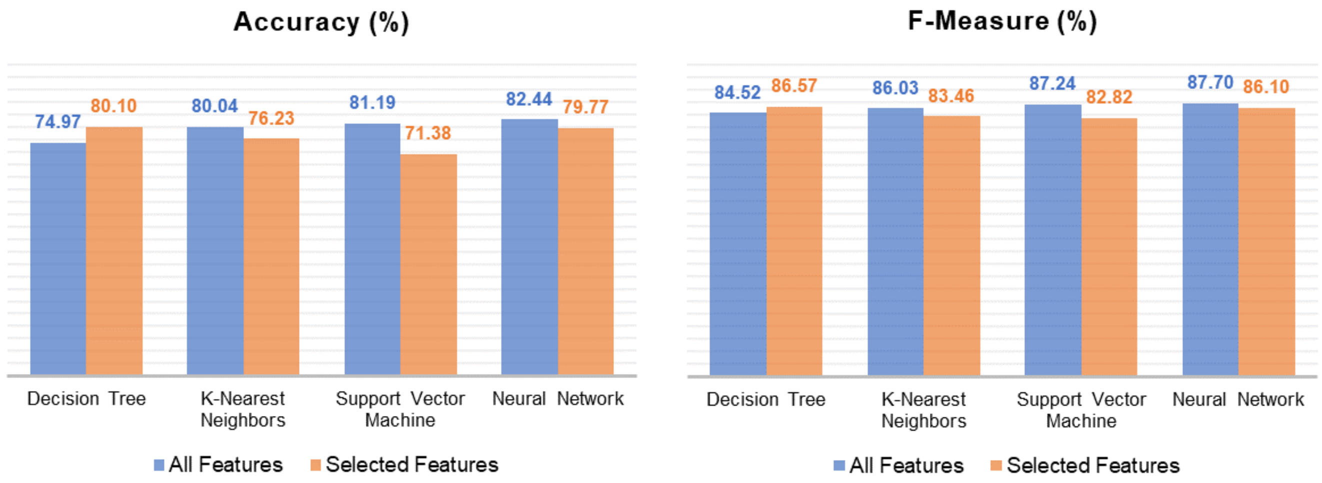

Fig. 3 Performance comparison between all features and only selected features in learning set data

TABLE VIII
TEST SET PERFORMANCE RESULT USING SELECTED HRV FEATURES

Classification Algorithm	Accuracy (%)	Recall (%)	Precision (%)	F-Measure (%)	AUC
DT	74.14 ± 3.61	94.56 ± 3.51	74.94 ± 4.00	83.61 ± 3.74	0.692 ± 0.055
KNN	70.04 ± 5.51	85.31 ± 5.80	74.86 ± 3.73	79.74 ± 4.54	0.666 ± 0.076
SVM	69.07 ± 0.83	**100.00 ± 0.00**	69.07 ± 0.83	81.71 ± 0.00	0.773 ± 0.062
NN	**77.42 ± 3.86**	90.54 ± 4.55	**79.73 ± 3.33**	**84.79 ± 3.85**	**0.777 ± 0.057**

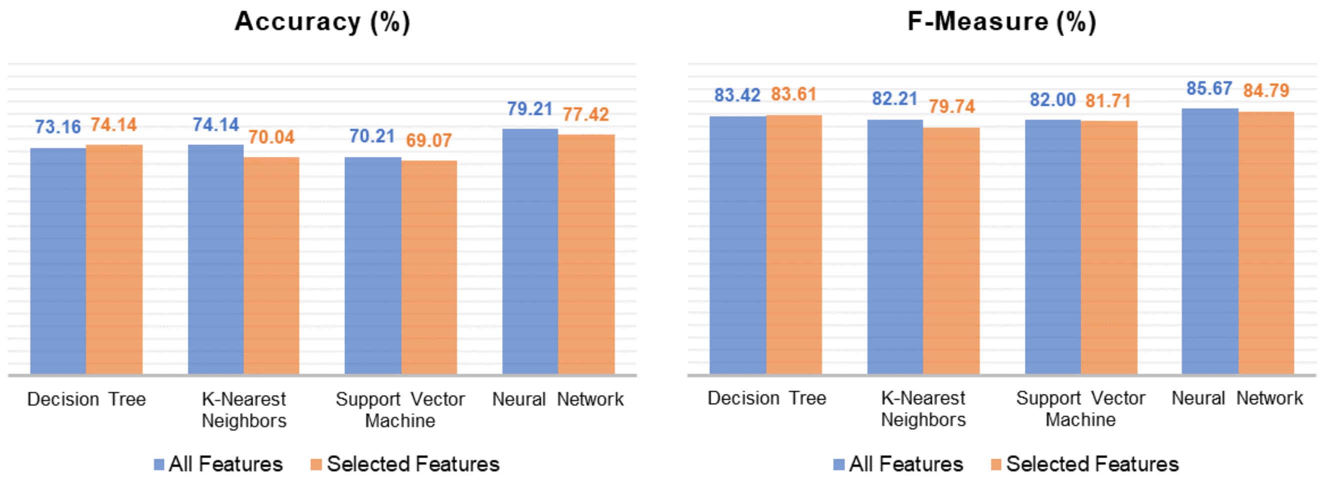

Fig. 4 Performance comparison between all features and only selected features in test set data

Performance results with only selected features on the test set data gives not much different compared to performance on learning set data (Table VII). Although there is an increase in the test set data using only selected features in DT algorithm, but the results do not show significant difference compared to use all HRV features (from 73.16% to 74.14%). The NN algorithm still shows the best performance among other algorithms with accuracy (77.42%), precision (79.73%), f-measure (84.79%), and AUC (0.777).

The results of comparison between using all features, and only using a few features generated from the feature selection process in this study, were also found in other studies on different objects. The KNN [25-26], NN [27] and SVM [26-28] algorithms also experienced a decrease in the level of accuracy, recall, precision, and F1-measure, when tested using only a few selected features.

The decrease in accuracy results in some of these algorithms is inversely proportional to the results shown in the DT algorithm. This study shows that the accuracy of DT algorithm only increases when using a few selected features. Similar results were also found in the study conducted by Wang & Li [29], which showed that DT algorithm when used after feature selection had increased accuracy. However, the increase in accuracy is also accompanied by an increase in the number of tree nodes. Increasing the number of trees will have an impact on making it more difficult to understand tree decisions. Thus, the DT algorithm is only suitable for data that has small sample size, because DT is more easily influenced by irrelevant features.

In this study, forward selection technique in feature selection phase shows the highest level of accuracy. Forward selection works well when the optimal subset has small number of features, but it cannot remove features that become obsolete after the addition of other features [30]. Different feature selection methods produce highly variable results, depending on the essence of feature selection itself. The feature selection method only selects features that have good performance, not relevant features [29]. The performance of a feature is determined by the evaluation of the features, the features that have been selected, the search strategy, and so on. The selection of relevant features according to Thenkabail, Enclona, & Ashton [31] is the features that still have a high frequency even though they are tested by different methods.

IV. CONCLUSION

Based on the results of 10-fold cross-validation test using all HRV features, neural network algorithm has the best performance than decision tree, k-nearest neighbour, and support vector machine algorithms with an accuracy rate (82.44% in the learning set, 79.21% in the test set), precision (85.54% in the learning set, 82.70% in the test set), f-measure (87.70% in the learning set, 85.67% in the test set), and AUC (0.867 in the learning set, 0.832 in the test set). Performance results using only selected HRV feature show that decision tree algorithm shows significant increase in performance, especially in terms of accuracy, which was initially only 74.97% to 80.10%. However, the accuracy of this algorithm does not have a significant value compared to neural network algorithm (79.77%). Performance test using only selected HRV feature, the three algorithms (K-Nearest Neighbors, Support Vector Machine, and Neural Network) experienced a decrease in performance. Thus, the use of all HRV features is recommended compared to only using selected HRV features, so that it can help detect the presence/absence of sleep apnea better. Further research can be done by improving the performance of neural network algorithm by optimizing the default algorithm parameters, so that it can produce much better performance.

ACKNOWLEDGEMENT

This research was supported by Ministry of Education, Culture, Research, and Technology through Penelitian Dosen Pemula Scheme. We thank our colleagues who provided insight and expertise that greatly assisted the research, although they may not agree with all the interpretations/conclusions of this paper.

REFERENCES

[1] N. F. Watson, M. S. Badr, G. Belenky, D. L. Bliwise, M. Buxton, D. Buysse, D. F. Dinges, J. Gangwisch, M. Grandner, C. Kushida, R. K. Malhotra, J. L. Martin, S. Patel, S. F. Quan and E. Tasali, "Recommended Amou of Sleep for a Healthy Adult: A Joint Consensus Stateme of the American Academy of Sleep Medicine and Sle Research Society," *Journal of Clinical Sleep Medicir* vol. 11, no. 6, pp. 591-592, 2015.

[2] J. Tietjens, D. Claman, E. Kezirian, T. De Marco, Mirzayan, B. Sadroonri, A. Goldberg, C. Long, Gerstenfeld and Y. Yeghiazarians, "Obstructive Sle Apnea in Cardiovascular Disease: A Review of t Literature and Proposed Multidisciplinary Clinic Management Strategy," *J Am Heart Assoc*, vol. 8, no. 1, e010440, 2019.

[3] H. Azimi, P. Xi, M. Bouchard, R. Goubran and F. Knoef "Machine Learning-Based Automatic Detection of Centr Sleep Apnea Events From a Pressure Sensitive Mat," *IEE Access*, vol. 8, pp. 173428-173439, 2020.

[4] M. E. Tagluk and N. Sezgin, "Classification of Sle Apnea through Sub-band Energy of Abdominal Effo Signal Using Wavelets + Neural Networks," *J Med Sy* vol. 34, no. 6, pp. 1111-1119, 2010.

[5] B. L. Koley and D. Dey, "Real-Time Adaptive Apnea a Hypopnea Event Detection Methodology for Portab Sleep Apnea Monitoring Devices," *IEEE Transactions Biomedical Engineering*, vol. 60, no. 12, pp. 3354-336 2013.

[6] A. Thommandram, J. M. Eklund and C. McGrego "Detection of apnoea from respiratory time series da using clinically recognizable features and kN classification," in *35th Annual International Conference the IEEE Engineering in Medicine and Biology Socie (EMBC)*, Osaka, Japan, 2013.

[7] B. Ma, Z. Wu, S. Li, R. Benton, D. Li, Y. Huang, M. Kasukurthi, J. Lin, G. M. Borchert, S. Tan, Y. Meihong a J. Huang, "A SVM-Based Algorithm to Diagnose Sle Apnea," in *2019 IEEE International Conference Bioinformatics and Biomedicine (BIBM)*, San Diego, C USA, 2019.

[8] J. Dong, “The role of heart rate variability in spor physiology (Review),” *Experimental and Therapeu Medicine,* pp. 1531-1536, 2016.

[9] E. Tobaldini, L. Nobili, S. Strada, K. Casali, A. Braghiro and N. Montano, “Heart rate variability in normal a pathological sleep,” *Front. Physiol,* vol. 4, no. 294, pp. 12, 2013.

[10] G. Forte, F. Favieri and M. Casagrande, “Heart Ra Variability and Cognitive Function: A Systemat Review,” *Frontiers in Neuroscience,* vol. 13, no. 710, p 1-11, 2019.

[11] P. Amiri, A. Derakhshan, B. Gharib, Y. H. Liu and M Mirzaaghayan, “Identifying Optimal Features from Hea Rate Variability for Early Detection of Sepsis in Pediatr Intensive Care,” in *2019 41st Annual Internation Conference of the IEEE Engineering in Medicine a Biology Society (EMBC),* Berlin, Germany, 2019.

[12] M. Corrales, C. Torres, G. Esquive, G. Salazar and Orellana, “Normal Values of Heart Rate Variability at Re in a Young, Healthy and Active Mexican population *Health,* pp. 377-385, 2012.

[13] T. Penzel, G. Moody, R. Mark, A. Goldberger and J. Pete “The Apnea-ECG Database,” *Computers in Cardiolog* vol. 27, pp. 255-258, 2000.

[14] G. Moody and R. Mark, “Content: Apnea-ECG Database MIT Laboratory for Computational Physiology, February 2000. [Online]. Availabl https://physionet.org/content/apnea-ecg/1.0.0/. [Access 4 October 2021].

[15] M. Elgendi, M. Jonkman and F. D. Boer, “Frequen Bands Effects on QRS Detection,” in *The 3rd Internation Conference on Bio-inspired Systems and Sign Processing (BIOSIGNALS2010),* Valencia, 2010.

[16] M. Zakariyah and A. Sahroni, “Komparasi Algoritm Deteksi Puncak QRS Kompleks Elektrokardiogram (EK Pada Pasien Penderita Stroke Iskemik,” in *Semin Nasional Informatika Medis (SNIMed),* Yogyakar Indonesia, 2019.

[17] G. James, D. Witten, T. Hastie and R. Tibshirani, *A* Introduction to Statistical Learning, Berlin: Springer, 201

[18] S. Uddin, A. Khan, M. E. Hossain and M. A. Mo “Comparing different supervised machine learni algorithms for disease prediction,” *BMC Medic Informatics and Decision Making,* vol. 19, no. 281, pp. 16, 2019.

[19] D. H. Wolpert, “The Lack of A Priori Distinctions Betwe Learning Algorithms,” *Neural Computation,* vol. 8, no. pp. 1341-1390, 1996.

[20] J. Arunadevi and M. J. Nithya, “Comparison of Featu Selection Strategies for Classification using Rapid Mine *International Journal of Innovative Research in Comput and Communication Engineering,* vol. 4, no. 7, pp. 1355 13563, 2016.

[21] B. Hosseinifard, M. H. Moradi and R. Rostan “Classifying depression patients and normal subjects usi machine learning techniques and nonlinear features fro EEG signal,” *Computer Methods and Programs Biomedicine,* vol. 109, no. 3, pp. 339-345, 2013.

[22] R.-C. Chen, C. Dewi, S.-W. Huang and R. E. Carak “Selecting critical features for data classification based machine learning methods,” *Journal of Big Data,* vol. no. 1, pp. 1-26, 2020.

[23] M. A. Salam, A. T. Azar, M. S. Elgendy and K. M. Foua “The Effect of Different Dimensionality Reducti Techniques on Machine Learning Overfitting Problem *International Journal of Advanced Computer Science a Applications (IJACSA),* vol. 12, no. 4, pp. 641-655, 2021

[24] Y. Khourdifi and M. Bahaj, “Heart Disease Prediction a Classification Using Machine Learning Algorith Optimized by Particle Swarm Optimization and A Colony Optimization,” *International Journal of Intellige Engineering and Systems,* vol. 12, no. 1, pp. 242-2 2019.

[25] Y. Y. Wang and J. Li, “Feature-selection ability of t decision-tree algorithm and the impact of featu selection/extraction on decision-tree results based hyperspectral data,” *International Journal of Remo Sensing,* vol. 29, no. 10, pp. 2993-3010, 2008.

[26] S. Khalid, T. Khalil and S. Nasreen, “A survey of featu selection and feature extraction techniques in machi learning,” in *Science and Information Conferen* London, United Kingdom, 2014.

[27] P. Thenkabail, E. Enclona and M. Ashton, “Accura assessment of hyperspectral waveband performance f vegetation analysis application,” *Remote Sensing Environment,* vol. 91, no. 3-4, pp. 354-376, 2004.

[28] A. Carneiro-Barrera, A. Díaz-Román, A. Guillé Riquelme and G. Buela-Casal, “Weight loss and lifesty interventions for obstructive sleep apnoea in adul Systematic review and meta-analysis,” *Obes Rev,* vol. 2 no. 5, pp. 750-762, 2019.

[29] D. Durgan and R. J. Bryan, “Cerebrovascul Consequences of Obstructive Sleep Apnea,” *J Am He Assoc,* vol. 1, no. 4, p. e000091, 2012.

[30] A. Bravi, G. Green, C. Herry, H. E. Wright, A. Longtin, P. Kenny and A. J. E. Seely, “Do physiological a pathological stresses produce different changes in he rate variability?,” *Front Physiol,* vol. 4, no. 196, pp. 1-1 2013.

[31] U. Zaky, A. Anggara, M. Zakariyah and I. Fathulla “Poincaré Plot Method for Physiological Analysis of t Gadget Use Effect on Children Stress Level,” *Jurn Online Informatika,* vol. 7, no. 1, pp. 46-55, 2022.