

Comparative Study of Machine Learning Algorithms in Classifying HRV for the Driver's Physiological Condition

Siti Fatimah Abdul Razak ^{1*}, S. N. M. Sayed Ismail ¹, Sumendra Yogarayan ¹,
Mohd Fikri Azli Abdullah ¹, Noor Hisham Kamis ¹, Azlan Abdul Aziz ²

¹ Faculty of Information Science and Technology, Multimedia University, Jln Ayer Keroh Lama, Melaka, 75450 Malaysia.

² Faculty of Engineering and Technology, Multimedia University, Jln Ayer Keroh Lama, Melaka, 75450 Malaysia.

Received 06 March 2023; Revised 06 August 2023; Accepted 11 August 2023; Published 01 September 2023

Abstract

Heart Rate Variability (HRV) may be used as a psychological marker to assess drivers' states from physiological signals such as an electrocardiogram (ECG), electroencephalogram (EEG), and photoplethysmography (PPG). This paper reviews HRV acquisition methods from drivers and machine learning approaches for driver cardiac health based on HRV classification. The study examines four publicly available ECG datasets and analyzes their HRV features, including time domain, frequency domain, short-term measures, and a combination of time and frequency domains. Eight machine learning classifiers, namely K-Nearest Neighbor, Decision Tree, Naive Bayes, Linear Discriminant Analysis, Support Vector Machine, Random Forest, Gradient Boost, and Adaboost, were used to determine whether the driver's state is normal or abnormal. The results show that K-Nearest Neighbor and Decision Tree classifiers had the highest accuracy at 92.86%. The study concludes by assessing the performance of machine learning algorithms in classifying HRV for the driver's physiological condition using the Man-Whitney U test in terms of accuracy and F1 score. We have statistical evidence to support that the prediction quality is different when HRV analysis applies these three sets: (i) time domain measures or frequency domain measures; (ii) frequency domain measures or short-term measures; and (iii) combining time and frequency domains or only frequency domains.

Keywords: Driver Monitoring System; ADAS; Machine Learning; Vehicle Safety.

1. Introduction

The fluctuation in the intervals between adjacent heartbeats is referred to as heart rate (HR), whereas heart rate variability (HRV) is the number of heart beats per minute used to characterize the time series of the interval variation between subsequent heartbeats [1, 2]. HRV is an excellent predictor of human health and is thought to be capable of predicting existing or future health issues such as heart disease like myocardial infarction (MI), atrial fibrillation (AF), ventricular fibrillation (VF), congestive heart failure (CHF), cardiac arrhythmia, etc. [3, 4], and mental health disorders such as anxiety and depression. Moreover, since the European Society of Cardiology and the North American Society of Electrophysiology [5] established recommendations for the use of HRV, this approach has become more widely used and is not only confined to neurology, surgery, exercise physiology, physical activity, and anesthesia [6, 7]. HRV analysis in the time, frequency, and nonlinear domains has also been shown to be conceivable for the detection and identification of driver states [8, 9] in driver status monitoring systems.

* Corresponding author: fatimah.razak@mmu.edu.my

 <http://dx.doi.org/10.28991/CEJ-2023-09-09-013>



© 2023 by the authors. Licensee C.E.J, Tehran, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Since HRV and the autonomic nervous system (ANS) are closely associated, HRV is a popular measurement that may be derived from non-intrusive and wearable physiological measurements [3], such as surface electrocardiograms (ECG), electroencephalograms (EEG), and photoplethysmography (PPG) [10].

The electrocardiogram (ECG) is the gold standard for HRV analysis [11], where the HRV signal from the ECG is derived from the R-R interval [8]. The ECG shows the voltage and time graph of the driver's heart's electrical activity. To establish a single lead channel for recording the ECG signal, at least three surface electrodes must be applied to the skin. Before beginning the recording, the ECG must be set up many times [8]. The heart rate, heart rate variability (HRV), and breathing rate can then be determined using the recorded ECG data and offer important insight into the driver's internal states [10]. Though the results might contain errors because of delay, biological and electromagnetic interference, and the complex ECG signal structure [8]. Suitable QRS detection algorithms must be used to derive the HRV signal. These algorithms must be used to identify acceptable interpolation and resampling, detect the peaks and their R wave, acquire the interval of RR, and provide a consistently sampled tachogram. ECG data that has been transformed into HRV for analysis where the QRS complex is dominant reveals the autonomic nervous system's reaction and enables the prediction of the driver's degree of stress [9].

Besides ECG, an electroencephalogram (EEG) is feasible to extract HRV as a matter of course during EEG data processing by installing a single EEG sensor on the chest [7]. During EEG tests, HRV indicators serve as supplementary assessments of the subject's physiological status. Information from EEG data is mostly employed in studies of brain activity. Mean frequency, energy contents, and bands are examples of frequency domain properties that can be used to determine a driver's state, such as driver fatigueness, which is revealed by the fronto-medial activity θ power [12], whereas standard deviation and average value are time-domain measures that provide information about driver alertness [13]. In a previous study, the EEG-Beat algorithm was proposed to perform automated analysis of HRV from the EEG. The algorithm implements a top-down divide-and-conquer technique to detect the signal peak instead of using QRS complex recognition. However, the algorithm was tested on health patients and found to be inappropriate for online and clinical applications [7]. In driving situations, placement of the EEG electrodes on the driver's head may impair their ability to focus on the road and include noise. Hence, a filter is always used to pre-process the recorded data [13].

Alternatively, PPG, an electro-optical approach, has been proposed to derive and monitor the HRV signal due to its low-cost optical sensors, ease of sensor insertion, non-invasive nature, and inexpensive cost. With each cardiac cycle, PPG signals depict blood volume fluctuations in a superficial body location and pulsatile blood pressure changes [14]. The computation of the HRV signal from the PPG signal uses the inter-beat interval (IBI) or pulse interval (PPI), utilizing software algorithms [8]. The position of a peak in a PPG signal reflects the point in time when a heartbeat occurs. However, PPG signals from wearable devices are sensitive to any small error in identifying the correct location of peaks. Hence, a Bayesian learning system was proposed to improve HRV estimates when the PPG is affected by artifacts [15]. Thus, reliable detection of the position of peaks in the PPG signal is required for HRV computation, which leads to the exact computation of time intervals between successive heartbeats [15].

Researchers frequently use extracted features of HRV from at least 5-minute ECG data [11] to evaluate the health of control subjects with hypertension, stress, and cardiovascular diseases in driver status monitoring systems. Moreover, the driver monitoring system is a complex application where safety is crucial [12]. Hence, the deployment of machine learning systems in this type of system enables optimization of the monitoring while simultaneously providing interpretability of their reasoning. Machine learning is a process that allows computers to self-learn without the use of specialized code and may be used to build systems for data analysis, decision-making, and data preparation in real-world circumstances. This allows researchers to evaluate whether such reasoning is correct. For example, machine learning can demonstrate HRV classification and enhance driver state detection models while allowing interpretability of the learned model. Features derived from physiological signals, combined with machine learning, may offer highly accurate detection and recognition of driver states, which can promote safe driving [3, 13]. Supervised learning and unsupervised learning are machine learning approaches that have been employed to evaluate the reasoning of driver-state monitoring systems. Figure 1 shows the physiological-based driver monitoring system (DMS) components.



Figure 1. Physiological-based DMS components

Open source and private technologies, as well as equipment with higher mobility, cheaper cost, easier operation, and wider accessibility, have lately emerged as new instruments for monitoring individual cardiac health, including HRV recording and analysis. As HRV analysis has advanced, several technologies have been designed and tested, with portable transmission systems proving to be trustworthy. Mobile devices like chest heart rate monitors and smart phones

that have been synced with the transmitters are usually utilized for signal collection. The HRV analysis is done using external software [6]. Moreover, Google Fit, a Google smartphone program produced by Google, allows for the measurement of the number of heartbeats and respiration. Since smartphones have been part of our lives, Google may choose to profit from the opportunity for individuals to monitor their health in several settings, including while driving [14, 15]. Moreover, wearable and non-wearable devices allow drivers to monitor health factors using physiological sensors conveniently.

This study employs eight supervised machine learning classifiers with distinct properties in order to perform comparative analysis on four publicly available drivers' ECG datasets in order to offer the best prediction model for driver state monitoring systems. The study on different combinations of HRV parameters and the performance of the machine learning classifiers was also conducted. Section 2 presents related work on HRV analysis using machine learning approaches. Material and methods used for this study are explained in Section 3, and results are presented in the following Section 4. Lastly, the study is concluded in Section 5.

2. Related Work

Machine learning approaches or models have been implemented for HRV analysis. Most researchers train and test a range of machine learning algorithms before selecting the algorithm that best tackles the problem once all performance data has been collected. In general, HRV analysis employs feature selection, supervised and unsupervised machine learning, as well as deep learning, to classify different driver states based on HRV characteristics.

2.1. Feature Selection

Researchers have used shorter windows to extract information related to physiological function in real-time, in addition to automated diagnosis and categorization [16]. For instance, Castaldo et al. [5] proposed using ultra-short HRV to compensate for short HRV. Six ultra-short HRV features, including MeanNN, StdNN, MeanHR, StdHR, HF, and SD2, demonstrated consistency across all excerpt lengths, which ranged from 1 to 5 minutes when used in IBK to detect drivers' mental health [17]. Feature selection algorithms like Principal Component Analysis (PCA) and Genetic Algorithm (GA) can simplify a classifier, improve classification accuracy, and shorten classification times [18–21] by selecting a subset of the most representative features from ECG data [19] for different types of drivers' state monitoring systems.

Persson et al. [18] implemented Sequential Forward Floating Selection (SFFS) to select the most influential features to predict driver drowsiness from HRV data. The SFFS was equipped with a binary decision tree classifier, five-fold cross-validation, twenty cross-validation trials, and an optimization score that balanced sensitivity and specificity. The SFFS technique was carried out 20 times on various feature selection set divisions since SFFS frequently produces low-dimensional, non-redundant, but noise-sensitive feature sets. The final feature set was composed of the characteristics chosen in 20% of the repeats.

Benchekroun et al. [9] proposed filtering and iterative HRV data imputation using a Gaussian distribution to improve classification accuracy in cases where the signals have a high percentage of missing data. The results show a stable F1 score of 61% compared to other imputation methods, i.e., HRV distribution, variability, and characteristics (DVC), linear, shape-preserving piecewise cubic Hermite (pchip), and spline interpolation using the Python Toolbox HRV.

Hasan et al. [14] applied two methods for selecting features for drowsiness detection, i.e., the ANOVA-F Test and the correlation-based feature selection algorithm. A final list of ensemble characteristics based on the stability selection strategy was created by combining the findings from the two methods. This approach selects an acceptable threshold value for each feature from various measurements based on a predetermined threshold. Then, the characteristics that have values over the threshold are given one point, while the other features are given zero points. As a result, the most crucial features are chosen.

2.2. Supervised Learning

Since 2010, the most popular technique for categorizing different heart diseases and symptoms associated with HRV has been supervised learning. Models like Decision Tree (DT), Linear Discriminant Analysis (LDA), Random Forest (RF), Support Vector Machine (SVM), and k-nearest neighbors (kNN) may predict labels based on related characteristics thanks to supervised models that learn the data and predict labels through learned mapping [16]. Table 1 lists a summary of previous work on HRV classification based on drivers.

For instance, in a driver drowsiness or sleepiness detection system, Persson et al. [18] examined the performance of four distinct binary classifiers: kNN, SVM, AdaBoost, and RF to detect drowsy drivers from binary and multiclass classifications of drowsy drivers. The random forest classifier produced the best overall results for binary classification, which implies that HRV classification requires the use of customized algorithms. In addition, hybrid combinations of

physiological sensors are strongly advised to improve the sensitivity and specificity of the driver drowsiness system. According to Hasan et al. [14], when 13 characteristics from the EEG, EOG, and ECG were used, Artificial Neural Network (ANN) classifiers produced the best overall results compared to KNN, SVM, and RF classifiers.

Moreover, Vincente et al. [8] employed LDA to classify truck drivers ECG data based on driver drowsy detectors and sleep deprivation detectors. Although the detection was very accurate, the findings show that there are several inaccuracies in wireless ECG identification when a vehicle is moving and that some areas of the signal were blank when the truck was moving. This was addressed by Bencheikroun et al. [9], who investigated the effect of imputation on classifiers when the collected signals have many missing data points. However, instead of LDA, RF was combined with the imputation method to determine subjects' arousal state, either relaxed or stressed, in the lab environment.

In another type of driver state monitoring system like driver stress detection and monitoring, Boussejot [11] discovered that SVM-RBF was able to classify HRV parameters, i.e., time, frequency, nonlinear, and time-frequency domains, with 83% accuracy in determining drivers' stress. However, there was no validation or testing conducted in that investigation. Iqbal et al. [22], on the other hand, investigated the performance of logistic regression, Gaussian Naïve Bayes, DT, RF, AdaBoost, and KNN on stress recognition in the Automobile Driver Dataset and SWELL-KW dataset. Results show that DT has the highest classification accuracy, followed by AdaBoost. An automated approach to assessing automobile drivers' stress was proposed using SVM. Findings reveal that manual driving is more demanding compared to autonomous driving [23].

Moreover, when different HRV characteristics from short and ultra-short HRV were used, the IBK classifier performed the best compared to Multilayer Perceptron (MLP), SVM, MLP, DT, and LDA. Six of the 23 ultra-short HRV features (MeanNN, StdNN, MeanHR, StdHR, HF, and SD2) demonstrated consistency across all excerpt lengths (i.e., from 5 to 1 min) when used in a well-dimensioned automatic classifier to detect drivers' mental health [17]. In another work, the synthetic minority over-sampling technique (SMOTE) was applied to reduce the effect of imbalanced data on classification accuracy. The study reveals that RF performed better compared to SVM and MLP in determining drivers' normal and surprise states [24].

In contrast to supervised learning models, unsupervised machine learning models do not require class labels to classify drivers' states. The models apply clustering techniques such as Affinity Propagation, Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH), K-mean, Mini-Batch K-mean, Mean Shift, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), and Ordering Points to Identify the Clustering Structure (OPTICS) for monitoring stress based on acquired data from wearable devices [22]. In a previous study, Wang and Guo employed the supervised ensemble classifier in conjunction with an unsupervised learning classifier to detect stress in drivers' foot galvanic skin response (GSR) data. Their suggested model detected stress with an accuracy of 90.1% [25]. Moreover, the combination of autoencoders and unsupervised deep learning to categorize mental stress related to HRV is a new approach that is expected to gain traction in 2019. Self-organizing map (SOM), a dimensional reduction technique developed through unsupervised learning, may identify the most useful elements needed to accurately characterize stress. To cluster and classify raw HRV data gathered from firefighters, we offer an unsupervised technique that combines autoencoders, i.e., convolutional autoencoder (CAE) and LSTM autoencoder (LAE), and density-based clustering with previous information [26].

Recently, deep learning has been used more often in HRV research to improve real-time automatic categorization. They can detect hidden patterns in input by using hidden layers and iteratively minimizing data mistakes prior to classification. As a result, the algorithm is better at gathering relevant data about the topic under investigation, classification accuracy improves, and fewer characteristics are required for real-time classification [16]. Convolutional neural network (CNN) architectures like LeNet, AlexNet, VGGNet, ResNet, Inception, DenseNet, and EfficientNet, recurrent neural networks (RNN), long-short-term memory (LSTM), and gated recurrent units are examples of popular deep supervised learning methods (GRU) [27–30]. For instance, a new drowsiness detection approach was presented in [31]. The approach uses raw R-R Interval (RRI) time series as inputs and trains the drowsiness detection model with LSTM and an autoencoder. The results revealed that RRI characteristics were superior to HRV features. Additionally, Oskooei et al. [32] presented a CNN-LSTM framework based on the integration of extracted HRV time domain parameters from raw ECG, vehicle data, and contextual data in drivers' stress level detection models in a variety of driving situations. The integration improves the classification accuracy of the classifier based on 27 drivers. In another study, Huang et al. [17] collected 15 drivers' biological data, including EEG (2 channels), ECG, EDA, RSP, and HR, in a simulated driving experiment. The driver's mental workload was classified using CNN, LSTM, and a combination of both, known as CNN-LSTM, after selecting features using XGBoost. The findings show that CNN_LSTM has the highest accuracy of 97.8% when utilizing 3-second samples and outperforms the CNN model in all circumstances.

Table 1. Summary of previous work on HRV classification based on drivers

Application	Machine Learning Approach	Signals
Driver Drowsiness	Random Forest; Support vector machine; Decision Tree; Naive Bayes; 1D CNN2 [22]	ECG
	Linear Discriminant Analysis (LDA) [8]	ECG
	kNN; SVM; Random Forest; Artificial Neural Network [21]	ECG; EOG; PPG
	XGBoost; CNN; ConvLSTM; CNN+LSTM [30]	ECG
Driver Stress	SVM; MLP; RF [24]	ECG
	SVM; KNN; Ensemble [32]	ECG
Auto detection of myocardial infraction (MI)	Random Forest; Support vector machine; Decision Tree; Naive Bayes; 1D CNN2 [31]	ECG
Mental Workload	Linear Discriminant Analysis (LDA) [8]	ECG
Surprise Driving	kNN; SVM; Random Forest; Artificial Neural Network [21]	ECG; EEG; GSR; RESP; RR
Drowsiness; visual inattention; fatigue; cognitive inattention	XGBoost; CNN; ConvLSTM; CNN+LSTM [30]	ECG
	SVM; MLP; RF [24]	ECG

3. Method

In this study, we followed the basic machine learning process as our research framework, which consists of five steps, i.e., acquire, pre-process, extract, classify, and evaluate, as shown in Figure 2. Based on our objectives for the study, we used publicly available datasets that have been used in previous research related to drivers’ physiological state monitoring systems, namely DriveDB, HCILAB, Drozy, and DMD, as described in Table 2. The PTB, PTB-XL, and MIT Affective Road Datasets were not included for further investigation since the datasets do not fulfill the objective of the study. The PTB Dataset and PTB-XL are acquired from clinical records and not specific to drivers, whereas the MIT Affective Road Dataset does not include ECG recordings. The ECG signals acquired then need to be pre-processed to remove noise from the individuals’ movement, respiration, and electrical muscle activity. Besides that, environmental noise or technological aberrations caused by analog and digital signal processing might affect heart rate variability measurements [33]. After that, the analysis of HRV may be performed on long-term (a 24-hour record), short-term (a 5-minute record), or ultra-short-term (less than 5-minute records). Moreover, for short-term recordings, the accuracy of HRV values is dependent on robust digital infinite impulse response (IIR) filters, such as analog models, which can provide NN interval series adequate to reflect physiological signals [34, 35]. Hence, 50Hz notch filtering and a bandpass filter of 0.75 Hz to 35 Hz were applied.



Figure 2. Research framework

Table 2. Datasets used by previous research for driver physiological measures

Dataset	Description
Stress Recognition in Automobile Driver database (DRIVEDB) [36, 37]	The dataset is made up of ECG, EMG (right trapezius), GSR (galvanic skin resistance) measurements taken on the hand and foot, and breathing recordings taken by 13 healthy volunteers while driving on a predetermined route that included city streets, highways, and Z-zones in and around Boston, Massachusetts. The data collection was done to test the feasibility of automated stress detection.
A Dataset of Real World Driving to Assess Driver Workload (HCILAB) [38]	The dataset comprises around 2,500,000 ECG, skin conductance response (SCR), and body temperature samples, as well as a post-hoc video evaluation session of 10 drivers in a real-world motorway, highway, regular streets (50 km/h), and 30 km/h zone. The dataset was used to investigate the changes in drivers’ mental workloads based on road conditions.
PTB Dataset [36, 39]	549 high-resolution 15-lead ECGs with clinical summaries (12 standard leads plus Frank XYZ leads). Each of the 294 participants has one to five ECG records, and they include both healthy volunteers and patients with a variety of cardiac diseases such as myocardial infarction and heart failure.
PTB-XL [40]	The dataset comprises 21,837 clinical 12-lead ECG records of 10 seconds length from 18,885 patients classified based on the SCP-ECG standard including normal, hypertension, myocardial infarction, conduction disturbance and hypertrophy.
MIT Affective Road Dataset [8]	This dataset features 13 drives conducted by 10 drivers in total. The physiological signs recorded were EDA, HR, BR, and skin temperature. Furthermore, GPS data, films (filming inside and outside automobile scenes), and in-car temperature, humidity, and sound level were recorded. A stress measure from low (0) to high (1) was created in real-time based on observation during the driving experiment.
ULg Multimodality Drowsiness Database (DROZY) [41]	Recordings of 14 subjects’ Psychomotor Vigilance Test (PVT) to measure drowsiness which includes Karolinska Sleepiness Scale (KSS), Kinect v2 Sensor, face landmarks and Polysomnography (PSG) i.e., 5 electroencephalogram (EEG) channels, 2 electrooculogram (EOG) channels, electrocardiogram (ECG), and electromyogram (EMG).
Warwick-JLR Driver Monitoring Dataset (DMD) [42]	Includes recordings of EDA and ECG from 13 subjects as well as data from vehicle CAN-bus during driving experiment to assess driver mental workload.

Once the pre-processing is finished, HRV feature analysis can be separated into time, frequency, and non-linear domains. We extract the HRV features using the Python Heart Rate Analysis Toolkit, which includes eight features from time domain measures and ten from frequency domain measures [43]. Besides exploring the time and frequency domains separately, we considered eight short-term measures, combining three features from time domain measures and five from frequency domain measures proposed in Zontone et al. [44] work as listed in Table 3.

Table 3. HRV parameters

Time domain	Frequency domain	Short-term measures
BPM: heart rate in beats per minute	VLF: Absolute power of the very-low-frequency band (0.0033–0.04 Hz) in ms^2 .	IBI: interbeat interval in millisecond.
IBI: interbeat interval in millisecond	LF: Absolute power of the low-frequency band (0.04–0.15 Hz) in ms^2 d.	SDNN: standard deviation of intervals between adjacent beats in millisecond
SDNN: standard deviation of intervals between adjacent beats in millisecond	HF: Absolute power of the high-frequency band (0.15–0.4 Hz) in ms^2 .	RMSSD: root mean square of successive differences between adjacent R-R intervals in millisecond.
SDSD: standard deviation of successive differences between adjacent R-R intervals	LF/HF: Ratio of LF-to-HF power p_total: Total power.	LF: Absolute power of the low-frequency band (0.04–0.15 Hz) in ms^2 .
pNN20: proportion of differences between R-R intervals greater than 20 ms	VLF_perc: Relative powers of VLF band.	HF: Absolute power of the high-frequency band (0.15–0.4 Hz) in ms^2 .
pNN50: proportion of differences between R-R intervals greater than 50 ms	LF_perc: Relative powers of LF band. HF_perc: Relative powers of HF band.	LF/HF: Ratio of LF-to-HF power.
RMSSD: root mean square of successive differences between adjacent R-R intervals in millisecond	LF_nu: Power of LF band in normalized unit.	LF_nu: Power of LF band in normalized unit.
MAD: mean absolute deviation	HF_nu: Power of HF band in normalized unit.	HF_nu: Power of HF band in normalized unit.

In this study, short-term HRV from ECG data was utilized to assess the performance of eight distinct machine learning classifiers, i.e., Adaboost (AB), Gradient Boost (GB), Random Forest (RF), K-nearest neighbors (KNN), Naïve Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), and Linear Discriminant Analysis (LDA), to detect drivers' physiological states, either normal or not. The NB classifier is a straightforward probabilistic classifier based on Bayes' theorem and strong independence assumptions between variables. Prior knowledge is used by the classifier to determine the probabilities of sample data. SVM uses kernels to solve the computational challenge of forecasting. The values of each variable are displayed as dots in a dimensional space with precise coordinates. The RF model is ensemble learning and tree-based, and it is used to build predictive models. The classifier generates a forest out of trees; more trees equals a more resilient forest. RF takes the data samples to form decision trees, calculates each tree, and uses the voting procedure to determine the best result. AdaBoost, or meta-learning, is widely regarded as one of the most effective boosting algorithms. It employs the iterative idea to investigate the flaws of weak algorithms and transform them into robust ones.

The drivers' states in each dataset were classified as normal (0) or abnormal (1) and the performance of all classifiers was assessed based on 70-30 splitting and 10-fold cross validation on accuracy and F1 score when different HRV parameters were included.

4. Results and Discussion

In general, the performance of machine learning classifiers is evaluated based on accuracy, precision, recall, F1 Score and error rates. Even so, accuracy and F1 Score are two measures which are frequently used to assess the quality of classifiers in machine learning. Accuracy represents the ratio between correctly predicted values and all results. Classifiers with a higher accuracy rate are better than those with lower accuracy rate. However, accuracy is only valid if the data we are dealing with is balanced. On the contrary, besides precision and recall, F1 score is used to validate the performance of classifiers on imbalanced datasets. The F1 score indicates how accurate the model is by indicating how many correct classifications are produced. It does not miss positive outcomes and predicts negative ones as well. Hence, in this study accuracy and F1 score metrics were used in evaluating the performance of the machine learning classifiers.

Figures 3 to 10 show the accuracy and F1-score of the classifiers which are evaluated in this study. The four dataset names were suffixed with the extracted features from the ECG i.e., *t* (time domain), *f* (frequency domain), *tf* (combination of time and frequency domain) and *s* (short-term measures). The highest value obtained based on individual dataset is in bold text. The level of accuracy achieved differs based on the databases and ECG features utilized. The HCI-lab database with time-domain features, coupled with KNN and DT classifiers, delivered the highest accuracy and F1 score, exceeding 90%. When it comes to Drozy-f, many classifiers had poor accuracy with less than 50%, and the highest accuracy achieved was only 52.63%. Regarding the F1 score, NB performs poorly, mainly when evaluated using the Drozy dataset. This is evident as NB produces a 0% F1 score for Drozy-f, Drozy-s, and Drozy-tf.

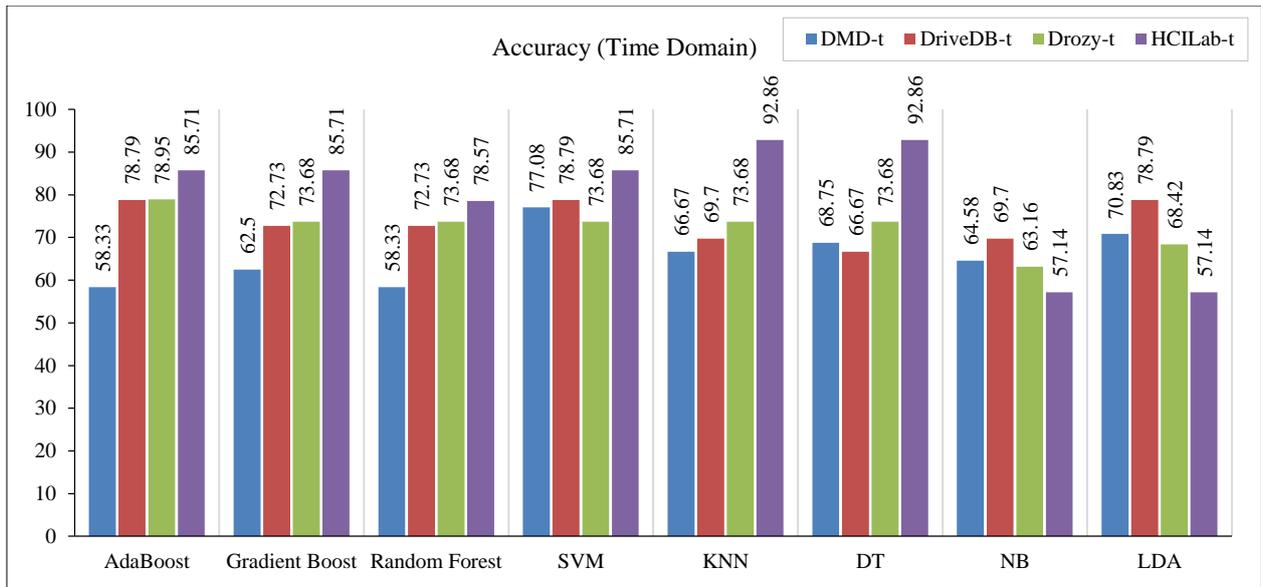


Figure 3. Accuracy of ML using time domain features

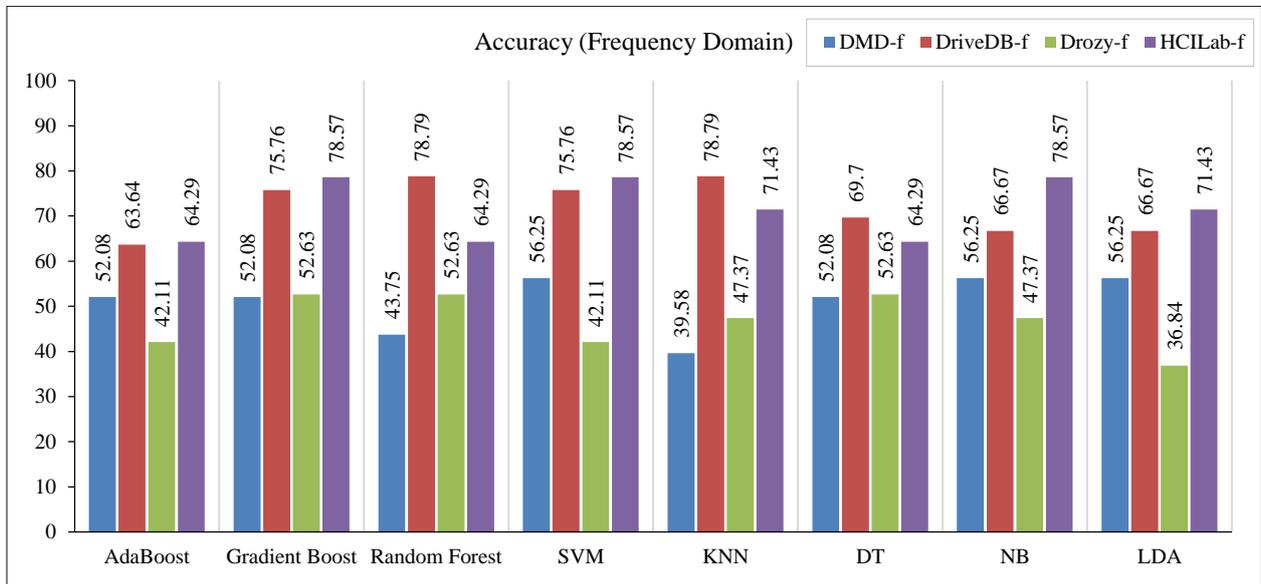


Figure 4. Accuracy of ML using frequency domain features

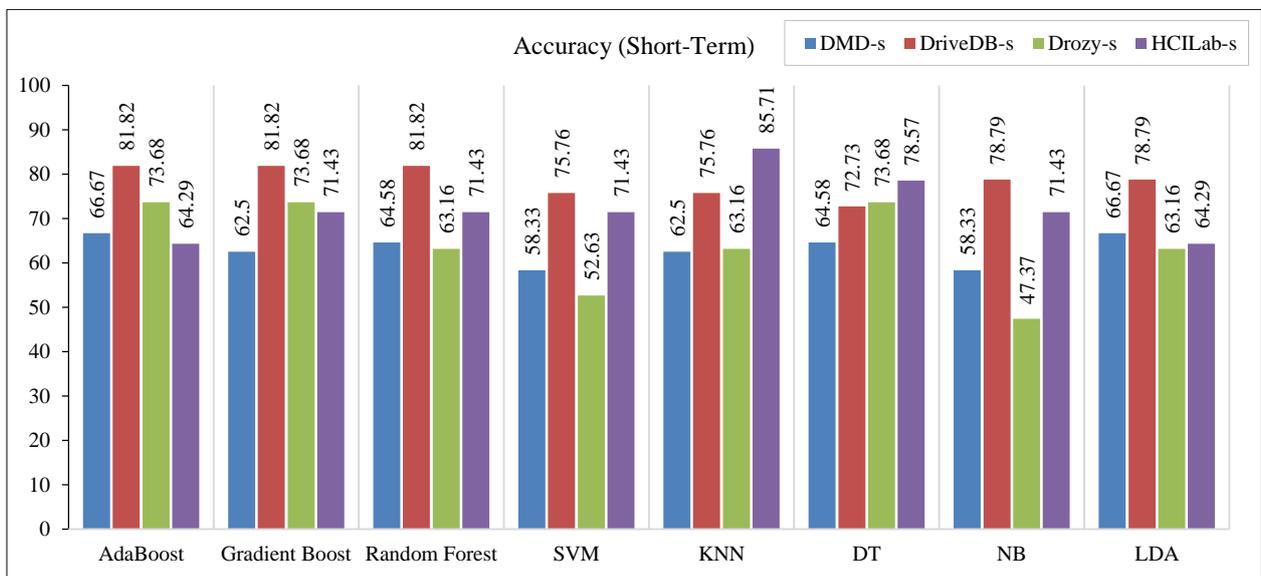


Figure 5. Accuracy of ML using short-term features

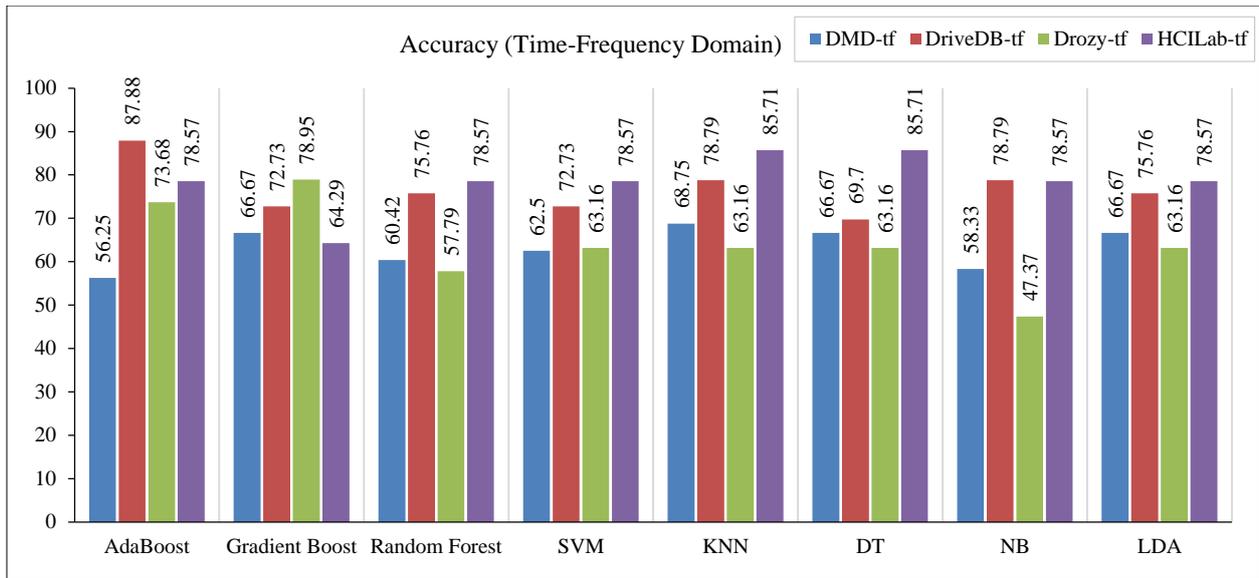


Figure 6. Accuracy of ML using time-frequency domain features

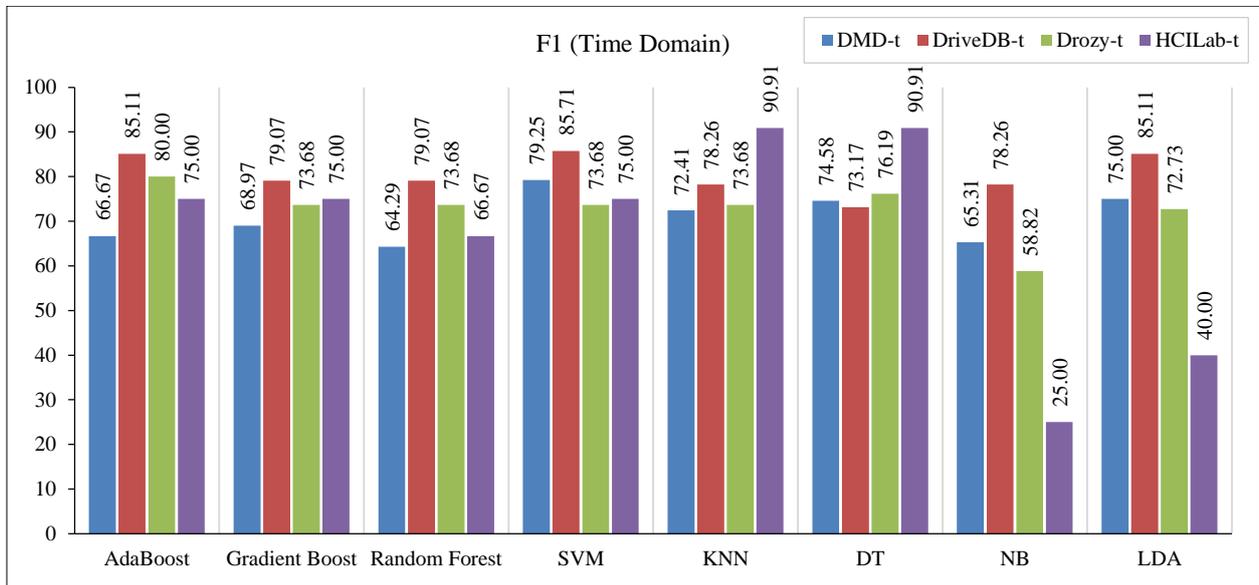


Figure 7. Accuracy of ML using time domain features

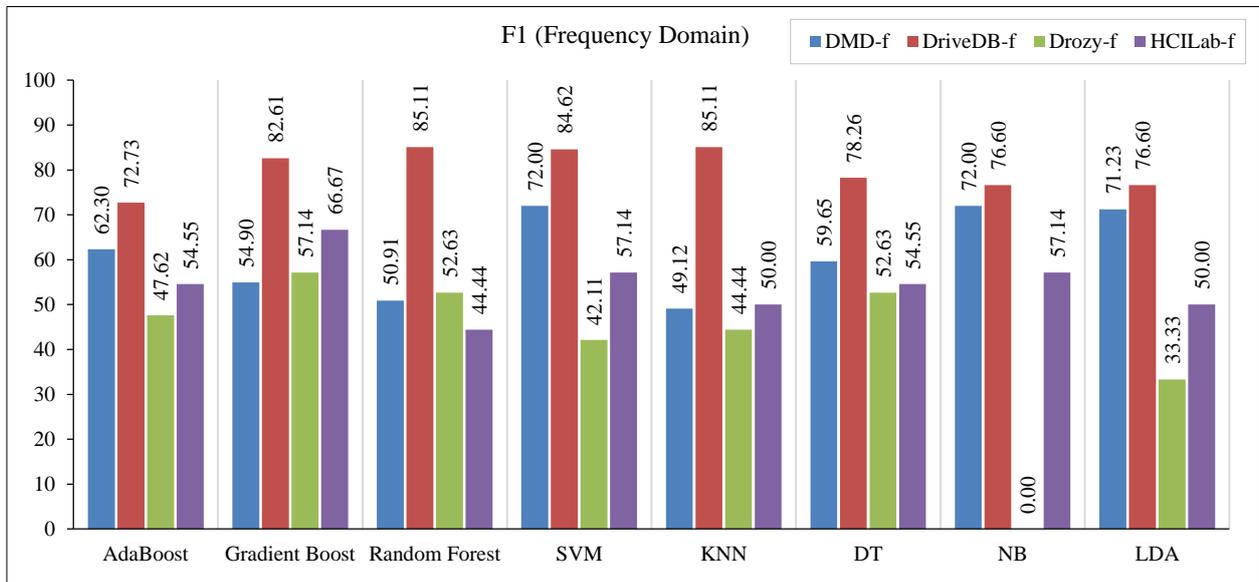


Figure 8. Accuracy of ML using frequency domain features

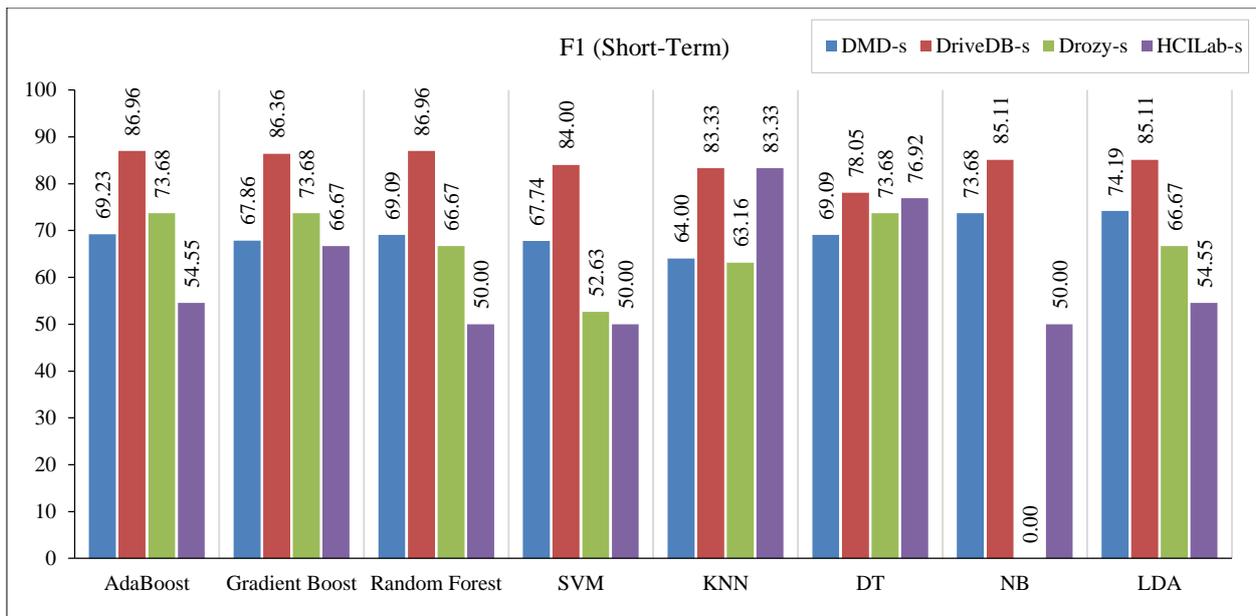


Figure 9. Accuracy of ML using short-term features

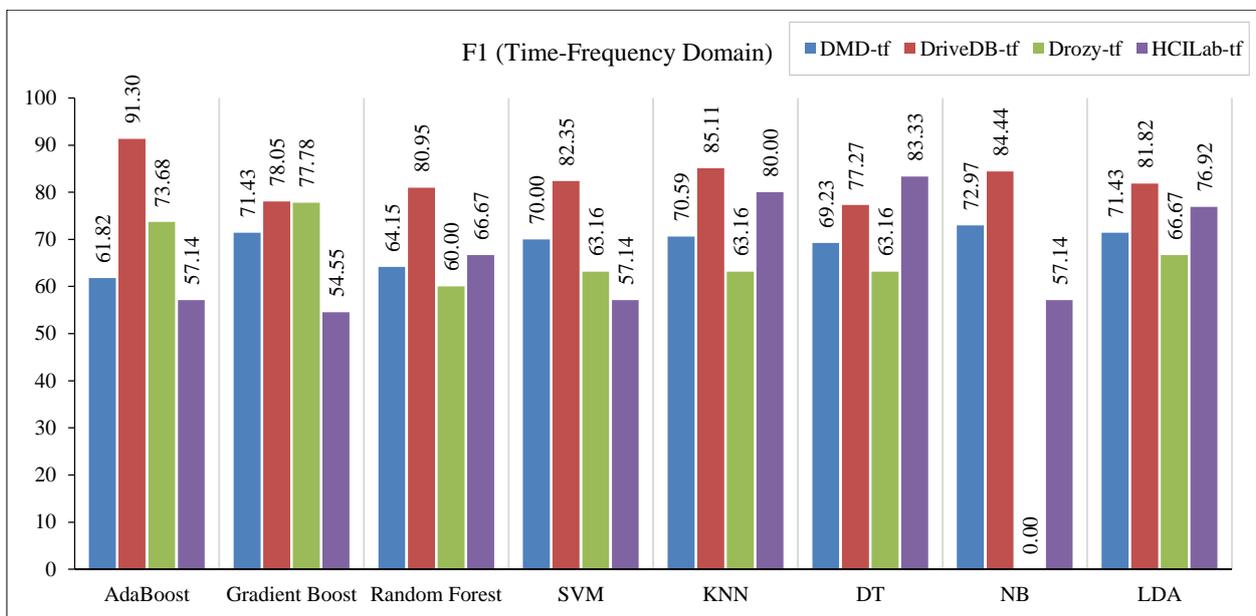


Figure 10. Accuracy of ML using time-frequency domain features

The poor performance can be seen as a result of having an imbalanced dataset. Moreover, we found that since accuracy does not consider how the data is distributed, a high value of accuracy does not mean that the model is actually able to correctly predict the outcomes. For example, the best model for dataset Drozy-f is generated by a decision tree with an accuracy of 52.63. However, the highest F1 score for the same dataset was produced by Gradient Boost, i.e., 57.14%. Hence, in this case, the F1 score is a better metric to be used for imbalanced data such as Drozy-f.

Nevertheless, comparing accuracy and F1 score alone makes it difficult to determine whether the performance of each machine learning classifier is significantly different from another. There is a probability that a classifier produced a higher accuracy and F1 score by random chance since this study only involved a few datasets. Another qualitative way to compare performance is based on the number of times the classifier produces the highest value of accuracy and F1 score compared to other classifiers. However, this approach is very subjective. For instance, in our comparison, AdaBoost produced the highest accuracy (6 times) and F1 score (4 times) compared to other classifiers (refer to Figure 11). Other classifiers have a similar number of occurrences, which makes it not feasible to determine the performance in this manner. Hence, statistical significance tests were applied to provide a better comparison of the classifiers. If the null hypothesis, or assumption, is rejected, it implies that the difference in the classifier's performance is statistically significant.

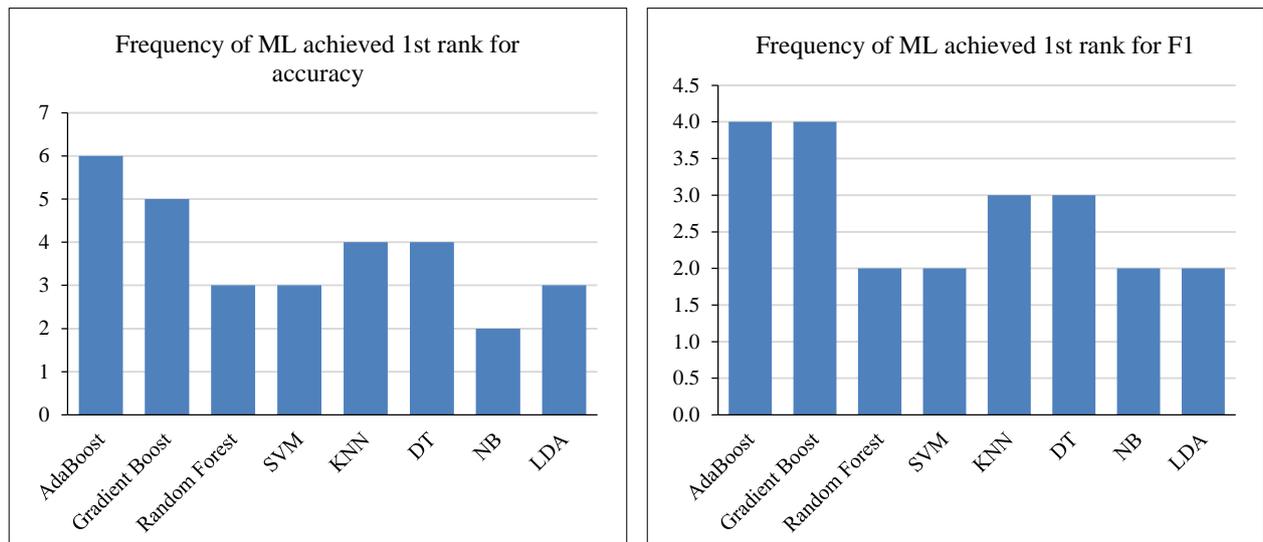


Figure 11. The frequency of ML achieved 1st rank: a) Accuracy, b) F1 score

In this study, we use a non-parametric test known as the Man-Whitney U test, which is equivalent to the Wilcoxon Sum Rank Test. A two-tailed test is performed at a 95% confidence level ($p < 0.05$). The null hypothesis can be rejected when the p-value is less than 0.05, which indicates that there is enough evidence to statistically reject that both classifiers A and B have identical performance. We evaluated the accuracy and F1 score of AdaBoost, Gradient Boost, Random Forest, Support Vector Machine (SVM), Decision Tree (DT), Naïve Bayes (NB), and Linear Discriminant Analysis (LDA) with one another for all datasets listed earlier in Table 3. The p-values are shown in Tables 4 and 5. From the results, we conclude that there is not enough evidence to statistically reject the null hypothesis. Hence, we do not have statistical evidence to claim that any of the investigated machine learning classifiers have different performance from the other classifiers in terms of accuracy and F1 score when tested on the datasets used for this study.

Table 4. Comparison of machine learning classifiers using Man-Whitney U test p-values (accuracy)

Accuracy	AdaBoost	Gradient Boost	Random Forest	SVM	KNN	DT	NB	LDA
AdaBoost	n/a	0.77948	0.56868	0.90448	0.77182	0.6672	0.23014	0.70394
Gradient Boost	0.77948	n/a	0.89656	0.65994	0.48392	0.85716	0.30772	0.92828
RF	0.56868	0.89656	n/a	0.53526	0.33204	1.0000	0.40090	0.88866
SVM	0.90448	0.65994	0.53526	n/a	0.71884	0.50926	0.21498	0.64552
KNN	0.77182	0.48392	0.33204	0.71884	n/a	0.34722	0.10310	0.36812
DT	0.6672	0.85716	1.0000	0.50926	0.34722	n/a	0.38978	0.84930
NB	0.23014	0.30772	0.40090	0.21498	0.10310	0.38978	n/a	0.34722
LDA	0.70394	0.92828	0.88866	0.64552	0.36812	0.84930	0.34722	n/a

Table 5. Comparison of machine learning classifiers using Man-Whitney U test p-values (F1 score)

F1 score	AdaBoost	Gradient Boost	Random Forest	SVM	KNN	DT	NB	LDA
AdaBoost	n/a	0.94420	0.52870	0.90448	0.62414	0.72634	0.18352	0.99202
Gradient Boost	0.94420	n/a	0.33706	0.96012	0.62414	0.71138	0.19020	0.93624
RF	0.52870	0.33706	n/a	0.54850	0.41794	0.21870	0.36282	0.41222
SVM	0.90448	0.96012	0.54850	n/a	0.65994	0.72786	0.22246	0.99202
KNN	0.62414	0.62414	0.41794	0.65994	n/a	0.65272	0.09492	0.70394
DT	0.72634	0.71138	0.21870	0.72786	0.65272	n/a	0.10524	0.68180
NB	0.18352	0.19020	0.36282	0.22246	0.09492	0.10524	n/a	0.20054
LDA	0.99202	0.93624	0.41222	0.99202	0.70394	0.68180	0.20054	n/a

In addition, we further analyzed the impact of different HRV parameters on the machine learning classifiers accuracy and F1 score to predict driver states as either normal or abnormal using the Man-Whitney U test. The null hypothesis can be rejected when the p-value is less than 0.05, which indicates that there is enough evidence to statistically reject that different groups of HRV parameters have identical performance. Based on the significant p-values as underlined in Table 6, we have enough statistical evidence to reject the null hypothesis for time (T)-frequency (F), frequency (F)-

short-term (S), and a combination of time and frequency (TF)-frequency (F). This indicates that the prediction quality is different when HRV analysis applies these three sets: (i) time domain measures or frequency domain measures; (ii) frequency domain measures or short-term measures; and (iii) combining time and frequency domains or only frequency domains. On the other hand, applying only the time domain measure is not different from applying the short-term measures or the combination of time and frequency domain measures. This may provide future researchers with more confidence to further investigate the HRV parameters in the context of monitoring drivers' states.

Table 6. Comparison of Man-Whitney U test p-values on datasets with different HRV parameters

	p-value (accuracy)				p-value (F1 score)			
	T	F	S	TF	T	F	S	TF
T	n/a	<u>0.00024</u>	0.23014	0.4354	n/a	<u>0.00050</u>	0.17384	0.17068
F	<u>0.00024</u>	n/a	<u>0.00600</u>	<u>0.00298</u>	<u>0.00050</u>	n/a	<u>0.01596</u>	<u>0.00528</u>
S	.23014	<u>0.00600</u>	n/a	0.69654	0.17384	<u>0.01596</u>	n/a	0.94420
TF	0.4354	<u>0.00298</u>	0.69654	n/a	0.17068	<u>0.00528</u>	0.94420	n/a

5. Conclusions

In this study, we examined a prediction model for driver condition monitoring systems by combining heart rate variability (HRV) data with supervised machine learning methods. Our work aims to improve driver safety and encourage safe driving habits by combining the predictive power of HRV with the sophisticated capabilities of machine learning algorithms. Our conceptual framework is built on the HRV, which includes the variability in the intervals between successive heartbeats and is impacted by the autonomic nervous system's (ANS) activity. Correlations between HRV and other physiological and psychological disorders have been observed, making HRV a useful indicator of human health. A driver's physiological condition, stress levels, and general well-being can be investigated by examining HRV patterns.

Eight different supervised machine learning classifiers, each with special qualities and abilities, are used in our study. The performance of these classifiers was assessed using four openly available datasets that include non-invasive signals, notably ECG recordings taken from drivers. Analyzing the performance of these classifiers provides opportunities to accurately predict driver states based on HRV characteristics. The findings of this study might influence the creation of cutting-edge driver assistance technologies that can cater to the physiological demands of drivers in real time, thereby enhancing overall traffic safety. Based on heart rate variability (HRV) metrics, which comprise time domain, frequency domain, short-term measurements, and a mix of time and frequency domains, classifiers' effectiveness in predicting normal or abnormal driver states is determined. The databases, extracted features, and classifiers utilized typically have an impact on the performances gained.

For the HCI-lab dataset, KNN and DT had the greatest accuracy and F1 score, with time-domain features over 90%. There is no statistical evidence to reject the null hypothesis, according to Man-Whitney U test hypothesis testing. As a result, the accuracy and F1 score of the machine learning classifiers examined in this study are not significantly different from one another. When several HRV parameters were used on the datasets, there was enough statistical support to reject the null hypothesis.

However, this study may have some limitations because the performance of numerous classifiers, especially the Drozy dataset, may have been impacted by the unbalanced dataset. A future study will investigate sampling techniques like SMOTE and do further research on HRV characteristics to overcome this issue. Additionally, unsupervised learning strategies may be used, particularly when class classification is difficult. To further assess the machine learning models, self-collected data for driver HRV analysis will also be captured using a vehicle simulator and in a real driving situation. The use of unsupervised learning techniques may also be advantageous, particularly when class labeling is difficult. Although the time and frequency domain HRV parameters were the major focus of this work, other HRV parameters or different feature extraction methods may be worth looking into. Research on sophisticated feature selection approaches and dimensionality reduction methods may also be used to determine which HRV characteristics are most useful for predicting driver conditions.

6. Declarations

6.1. Author Contributions

Conceptualization, S.F.A.R. and S.Y.; methodology, S.F.A.R. and S.N.S.I.; software, S.N.S.I.; validation, N.H.K., A.A.A. and M.F.A.A.; formal analysis, S.F.A.R. and S.N.S.I.; investigation, N.H.K. and A.A.A.; resources, M.F.A.A.; data curation, N.H.K. and M.F.A.A.; writing—original draft preparation, S.F.A.R., S.N.S.I., and S.Y.; writing—review and editing, M.F.A.A. and A.A.A.; visualization, S.F.A.R.; supervision, S.F.A.R.; project administration, M.F.A.A.; funding acquisition, S.F.A.R. and A.A.A. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available in article.

6.3. Funding

This research is funded by TM R&D (RDTC/221046).

6.4. Conflicts of Interest

The authors declare no conflict of interest.

7. References

- [1] Aimie-Salleh, N., Aliaa Abdul Ghani, N., Hasanudin, N., & Nur Shakiroh Shafie, S. (2020). Heart Rate Variability Recording System Using Photoplethysmography Sensor. *Autonomic Nervous System Monitoring - Heart Rate Variability*, IntechOpen, London, United Kingdom. doi:10.5772/intechopen.89901.
- [2] Arakawa, T. (2021). A review of heartbeat detection systems for automotive applications. *Sensors*, 21(18), 6112. doi:10.3390/s21186112.
- [3] Bhardwaj, R., & Balasubramanian, V. (2019). Viability of Cardiac Parameters Measured Unobtrusively Using Capacitive Coupled Electrocardiography (cECG) to Estimate Driver Performance. *IEEE Sensors Journal*, 19(11), 4321–4330. doi:10.1109/JSEN.2019.2898450.
- [4] Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., & Schwartz, P. J. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17(3), 354–381. doi:10.1093/oxfordjournals.eurheartj.a014868.
- [5] Castaldo, R., Montesinos, L., Melillo, P., James, C., & Pecchia, L. (2019). Ultra-short term HRV features as surrogates of short term HRV: A case study on mental stress detection in real life. *BMC Medical Informatics and Decision Making*, 19(1), 1-13. doi:10.1186/s12911-019-0742-y.
- [6] de Oliveira Júnior, F. A., Pereira, R. A., Silva, A. S., de Brito Alves, J. L., Costa-Silva, J. H., Braga, V. A., & Balarini, C. M. (2022). Different acquisition systems for heart rate variability analysis may lead to diverse outcomes. *Brazilian Journal of Medical and Biological Research*, 55. doi:10.1590/1414-431X2021e11720.
- [7] Geronikolou, S. A., Chrousos, G. P., & Cokkinos, D. V. (2021). Heart Rate Variability Components in Electromagnetic Hypersensitive Persons. In *Handbook of Computational Neurodegeneration*, 1–10. Springer International Publishing. doi:10.1007/978-3-319-75479-6_54-1.
- [8] Aswathi, C. D., Mathew, N. A., Riyas, K. S., & Jose, R. (2021). Comparison of Machine Learning Algorithms for Heart Rate Variability Based Driver Drowsiness Detection. *2021 2nd Global Conference for Advancement in Technology (GCAT)*. doi:10.1109/gcat52182.2021.9587733.
- [9] Benchekroun, M., Chevallier, B., Istrate, D., Zalc, V., & Lenne, D. (2022). Preprocessing Methods for Ambulatory HRV Analysis Based on HRV Distribution, Variability and Characteristics (DVC). *Sensors*, 22(5), 1984. doi:10.3390/s22051984.
- [10] Bhor, P., Sodhi, G. S., & Sing, D. (2019). Classification of Heart Rate Variability through Machine Learning. *International Journal of Electronic Engineering*, 11(2), 73–80.
- [11] Bousseljot, R.-D. (1995). Use of the PTB's CARDIODAT ECG signal database via the Internet [in German]. *Biomedical Engineering*, 40. doi:10.13026/C28C71.
- [12] Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C. K., & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23), 215–220. doi:10.1161/01.cir.101.23.e215.
- [13] Hantono, B. S., Nugroho, L. E., & Santosa, P. I. (2020). Mental stress detection via heart rate variability using machine learning. *International Journal on Electrical Engineering and Informatics*, 12(3), 431–444. doi:10.15676/ijeiei.2020.12.3.3.
- [14] Hasan, M. M., Watling, C. N., & Larue, G. S. (2022). Physiological signal-based drowsiness detection using machine learning: Singular and hybrid signal approaches. *Journal of Safety Research*, 80, 215–225. doi:10.1016/j.jsr.2021.12.001.
- [15] Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2), 156–166. doi:10.1109/TITS.2005.848368.
- [16] Hejjeel, L. (2004). Suppression of power-line interference by analog notch filtering in the ECG signal for heart rate variability analysis: to do or not to do?. *Medical Science Monitor*, 10(1), MT6-MT13.

- [17] Huang, J., Liu, Y., & Peng, X. (2022). Recognition of driver's mental workload based on physiological signals, a comparative study. *Biomedical Signal Processing and Control*, 71. doi:10.1016/j.bspc.2021.103094.
- [18] Persson, A., Jonasson, H., Fredriksson, I., Wiklund, U., & Ahlstrom, C. (2021). Heart Rate Variability for Classification of Alert versus Sleep Deprived Drivers in Real Road Driving Conditions. *IEEE Transactions on Intelligent Transportation Systems*, 22(6), 3316–3325. doi:10.1109/TITS.2020.2981941.
- [19] Ishaque, S., Khan, N., & Krishnan, S. (2021). Trends in Heart-Rate Variability Signal Analysis. *Frontiers in Digital Health*, 3. doi:10.3389/fdgh.2021.639444.
- [20] Iwamoto, H., Hori, K., Fujiwara, K., & Kano, M. (2021). Real-driving-implementable drowsy driving detection method using heart rate variability based on long short-term memory and autoencoder. *IFAC-PapersOnLine*, 54(15), 526–531. doi:10.1016/j.ifacol.2021.10.310.
- [21] Kakaria, S., Bigné, E., Catrambone, V., & Valenza, G. (2022). Heart rate variability in marketing research: A systematic review and methodological perspectives. *Psychology & Marketing*, 40(1), 190–208. doi:10.1002/mar.21734.
- [22] Iqbal, T., Elahi, A., Wijns, W., & Shahzad, A. (2022). Exploring Unsupervised Machine Learning Classification Methods for Physiological Stress Detection. *Frontiers in Medical Technology*, 4. doi:10.3389/fmedt.2022.782756.
- [23] Koay, H. V., Chuah, J. H., Chow, C. O., & Chang, Y. L. (2022). Detecting and recognizing driver distraction through various data modality using machine learning: A review, recent advances, simplified framework and open challenges (2014–2021). *Engineering Applications of Artificial Intelligence*, 115, 105309. doi:10.1016/j.engappai.2022.105309.
- [24] Liu, S., Koch, K., Zhou, Z., Maritsch, M., He, X., Fleisch, E., & Wortmann, F. (2022). Toward Nonintrusive Camera-Based Heart Rate Variability Estimation in the Car under Naturalistic Condition. *IEEE Internet of Things Journal*, 9(14), 11699–11711. doi:10.1109/JIOT.2021.3131742.
- [25] Lopez-Martinez, D., El-Haouij, N., & Picard, R. (2019). Detection of Real-World Driving-Induced Affective State Using Physiological Signals and Multi-View Multi-Task Machine Learning. 2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW). doi:10.1109/aciiw.2019.8925190.
- [26] Manstetten, D., Beruscha, F., Bieg, H. J., Kobiela, F., Korthauer, A., Krautter, W., & Marberger, C. (2020, October). The Evolution of Driver Monitoring Systems: A Shortened Story on Past, Current and Future Approaches How Cars Acquire Knowledge About the Driver's State. 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services . doi:10.1145/3406324.3425896.
- [27] Massoz, Q., Langohr, T., Francois, C., & Verly, J. G. (2016). The ULG multimodality drowsiness database (called DROZY) and examples of use. 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). doi:10.1109/wacv.2016.7477715.
- [28] Munla, N., Khalil, M., Shahin, A., & Mourad, A. (2015). Driver stress level detection using HRV analysis. 2015 International Conference on Advances in Biomedical Engineering, ICABME 2015, 61–64. doi:10.1109/ICABME.2015.7323251.
- [29] Murugan, S., Selvaraj, J., & Sahayadhas, A. (2020). Detection and analysis: driver state with electrocardiogram (ECG). *Physical and Engineering Sciences in Medicine*, 43(2), 525–537. doi:10.1007/s13246-020-00853-8.
- [30] Nguyen, T. T., Aoki, H., Le, A. S., Akio, H., Aoki, K., Inagami, M., & Suzuki, T. (2021). Driver State Detection Based on Cardiovascular System and Driver Reaction Information Using a Graphical Model. *Journal of Transportation Technologies*, 11(02), 139–156. doi:10.4236/jtts.2021.112009.
- [31] Nunan, D., Sandercock, G. R. H., & Brodie, D. A. (2010). A quantitative systematic review of normal values for short-term heart rate variability in healthy adults. *PACE - Pacing and Clinical Electrophysiology*, 33(11), 1407–1417. doi:10.1111/j.1540-8159.2010.02841.x.
- [32] Oskooei, A., Chau, S. M., Weiss, J., Sridhar, A., Martínez, M. R., & Michel, B. (2020). DeStress: Deep Learning for Unsupervised Identification of Mental Stress in Firefighters from Heart-Rate Variability (HRV) Data. *Studies in Computational Intelligence*, 93–105, Springer, Cham, Switzerland. doi:10.1007/978-3-030-53352-6_9.
- [33] Kim, J. K., & Ahn, J. M. (2019). Digital IIR filters for heart rate variability; a comparison between Butterworth and elliptic filters. *International Journal of Scientific and Technology Research*, 8(12), 3509–3513.
- [34] Rastgoo, M. N., Nakisa, B., Maire, F., Rakotonirainy, A., & Chandran, V. (2019). Automatic driver stress level classification using multimodal deep learning. *Expert Systems with Applications*, 138. doi:10.1016/j.eswa.2019.07.010.
- [35] Riganello, F., Larroque, S. K., Bahri, M. A., Heine, L., Martial, C., Carrière, M., Charland-Verville, V., Aubinet, C., Vanhauzenhuyse, A., Chatelle, C., Laureys, S., & Di Perri, C. (2018). A Heartbeat Away From Consciousness: Heart Rate Variability Entropy Can Discriminate Disorders of Consciousness and Is Correlated with Resting-State fMRI Brain Connectivity of the Central Autonomic Network. *Frontiers in Neurology*, 9. doi:10.3389/fneur.2018.00769.

- [36] Riposan-Taylor, A., & Taylor, I. J. (2018). Personal Connected Devices for Healthcare. *The Internet of Things for Smart Urban Ecosystems*, 333–361, Springer, Cham, Switzerland. doi:10.1007/978-3-319-96550-5_14.
- [37] Schneegass, S., Pfleging, B., Broy, N., Heinrich, F., & Schmidt, A. (2013). A data set of real world driving to assess driver workload. *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. doi:10.1145/2516540.2516561.
- [38] Nunes, C., Beatriz-Afonso, A., Cruz-Jesus, F., Oliveira, T., & Castelli, M. (2022). Mathematics and Mother Tongue Academic Achievement. *Emerging Science Journal*, 6(Special Issue), 137-149. doi:10.28991/ESJ-2022-SIED-010.
- [39] Taylor, P., Griffiths, N., Bhalerao, A., Xu, Z., Gelencser, A., & Popham, T. (2017). Investigating the feasibility of vehicle telemetry data as a means of predicting driver workload. *International Journal of Mobile Human Computer Interaction*, 9(3), 54–72. doi:10.4018/ijmhci.2017070104.
- [40] van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2019). HeartPy: A novel heart rate algorithm for the analysis of noisy signals. *Transportation Research Part F: Traffic Psychology and Behaviour*, 66, 368–378. doi:10.1016/j.trf.2019.09.015.
- [41] Vicente, J., Laguna, P., Bartra, A., & Bailón, R. (2016). Drowsiness detection using heart rate variability. *Medical and Biological Engineering and Computing*, 54(6), 927–937. doi:10.1007/s11517-015-1448-7.
- [42] Wagner, P., Strodthoff, N., Bousseljot, R. D., Kreiseler, D., Lunze, F. I., Samek, W., & Schaeffter, T. (2020). PTB-XL, a large publicly available electrocardiography dataset. *Scientific Data*, 7(1), 154. doi:10.1038/s41597-020-0495-6.
- [43] Wang, K., & Guo, P. (2021). An Ensemble Classification Model with Unsupervised Representation Learning for Driving Stress Recognition Using Physiological Signals. *IEEE Transactions on Intelligent Transportation Systems*, 22(6), 3303–3315. doi:10.1109/TITS.2020.2980555.
- [44] Zontone, P., Affanni, A., Bernardini, R., Brisinda, D., Del Linz, L., Formaggia, F., Minen, D., Minen, M., Savorgnan, C., Piras, A., Rinaldo, R., & Fenici, R. (2020). Comparative assessment of drivers' stress induced by autonomous and manual driving with heart rate variability parameters and machine learning analysis of electrodermal activity. *European Heart Journal*, 41. doi:10.1093/ehjci/ehaa946.3515.