Arrythmia Classification Using MATLAB® Classification Learner App

Cesar N. Silva, Fernanda F. Lopes, Jefferson A. Matos[a] and Maria Claudia F. Castro[b]
Electrical Engineering Department, Centro Universitario FEI, São Bernardo do Campo, Brazil

Keywords: Electrocardiogram (ECG), Cardiac Arrhythmia, Classification, Machine-Learning.

Abstract: Vital sign monitoring is becoming a part of our daily lives, emerging as a trend of smart wearable devices used to manage health. Cardiac arrhythmia is any variation in the normal heartbeat rhythm, causing the heart to beat improperly. This work presents a study on the classification of cardiac arrhythmias in 4 classes, Normal (N), Supraventricular Ectopic (SVE), Ventricular Ectopic (LV), and Fusion of Normal and Ventricular (F). Using the MIT-BIH Arrhythmia Database and the Classification Learner App from MATLAB® for training, it was possible to investigate 24 models, where the Subspace KNN Ensemble obtained the best accuracy (74.4%) and was later used for implementation in the suggested user interface application.

1 INTRODUCTION

According to the World Health Organization (WHO, 2021), cardiovascular diseases (CVDs) are the leading global cause of death, taking an estimated 17.9 million lives each year, with more than 75% occurring in low- and middle-income countries (LMICs). While intensive global efforts to prevent cardiovascular disease are underway, cardiac arrhythmias remain neglected, especially in LMICs (Mkoko et al., 2020).

Cardiac arrhythmia refers to any variation in the normal heartbeat rhythm, causing the heart to beat too fast (Tachycardia), too slowly (Bradycardia), or erratically. An arrhythmia occurs when the sinus node, known as the natural pacemaker, develops an abnormal rhythm, the normal conduction pathway is interrupted, or when another part of the heart takes over as the pacemaker (Humphreys et al., 2011; American Heart Association (AHA), 2016). When the heart does not beat properly, it can not effectively pump blood, and the organs, such as the brain, lungs, and even the heart may be damaged or shut down. Thus, arrhythmias should be diagnosed and treated as early as possible to reduce the risk of sudden death.

Currently, vital sign monitoring is becoming a part of our daily lives, emerging as a trend of smart wearable devices used to manage health. Their adoption has further accelerated with the growth of telehealth during the COVID-19 pandemic. The most widely used tool for monitoring and diagnosing heart function, such as arrhythmia, is the Electrocardiogram (ECG), a graphical representation of the heart's electrical activity. For an early diagnosis, an efficient, intelligent, and robust automated arrhythmia classification system must be incorporated into smart wearable devices (Bayoumy et al., 2021).

To cope with such challenges, several works have been carried out on arrhythmia classification. Machine-learning-oriented techniques are adopted, requiring at least five steps: ECG signal conditioning such as amplification and denoising, feature extraction, feature selection, classification, and performance analysis (Mohebbanaaz et al., 2020).

ECG signal features mainly depend on time interval, amplitude, and segment duration. The most common are morphological information such as amplitudes and intervals identification of peaks P, R, T, and QRS complex, as well as information about the RR range/interspace, which is the distance between peaks of two successive R waves in the ECG signal (de Albuquerque et al., 2018; Celin and Vasanth, 2018; Kuila et al., 2020; Mohebbanaaz et al., 2022).

Recently, a great interest has been in the application of classification algorithms based on Deep Learning (Zhang et al., 2020; Hassan et al., 2022; Irfan et al., 2022) with accuracies up to 99.35%. However, other techniques have also been used to classify arrhythmias, such as Decision Trees (Mohebbanaaz et al., 2022), Random Forest (AbdElMoneem et al., 2020), K-Nearest Neighbor (Mustaqeem et al., 2018; AbdElMoneem et al., 2020), Ensemble Clas-

[a] https://orcid.org/0000-0002-8800-713X
[b] https://orcid.org/0000-0002-2751-0014

Silva, C., Lopes, F., Matos, J. and Castro, M.
Arrythmia Classification Using MATLAB® Classification Learner App.
DOI: 10.5220/0011666300003414
In *Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2023) - Volume 4: BIOSIGNALS*, pages 220-225
ISBN: 978-989-758-631-6; ISSN: 2184-4305

sifiers (Shalini et al., 2019; AbdElMoneem et al., 2020), Support Vector Machines (de Albuquerque et al., 2018; Mustaqeem et al., 2018; AbdElMoneem et al., 2020), and others (de Albuquerque et al., 2018). In these cases, accuracies range from 70 to 98%.

Most works comprise multiclass classification to differentiate among up to 16 types of arrhythmias. However, given that atrial fibrillation (AF) is the most common heart arrhythmia, its detection has received specific attention, either in a simple recognition system or to classify it into subtypes (Celin and Vasanth, 2018; Horoba et al., 2019; Ganapathy et al., 2021; Ramesh et al., 2021; Sager et al., 2021; da Silva et al., 2021; Fuadah and Lim, 2022).

In this context, this work presents a study on the detection and classification of cardiac arrhythmias using the MATLAB® Classification Learner application. Four classes were defined according to the Association for the Advancement of Medical Instrumentation (AAMI): Normal (N), Ectopic Supraventricular (SVE), Ventricular Ectopic (VE), and Fusion of Normal and Ventricular (F). As a suggestion, a low-cost device for ECG acquisition and a user interface for communication with the health professional is also presented.

2 MATERIALS AND METHODS

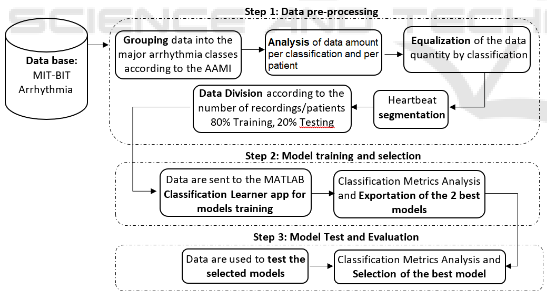

Figure 1: Data processing flowchart.

The flowchart for conducting the research is shown in Figure 1. Firstly, the MIT-BIH dataset was prepared for model training by performing signal extraction, using the WFDB package, class preparation, data balance, and segmentation. Then ECG signals were partitioned into two sets of records in order to separate the patients into training/validation (80%) and testing (20%) groups. The structured data was exported to the MATLAB® Classification Learner to the training process and investigation of the best two models exported back to the MATLAB® algorithm to the test phase and final metric analysis.

2.1 Dataset

The MIT-BIH Arrhythmia Database, which is publicly available online at physionet.org (Goldberger et al., 2000), is a well-known and worldwide used standard dataset for arrhythmia detectors evaluation (de Albuquerque et al., 2018; Celin and Vasanth, 2018; Kuila et al., 2020; Hassan et al., 2022; Irfan et al., 2022; Mohebbanaaz et al., 2022). It was collected by Boston's Beth Israel Hospital (BIH) Arrhythmia Laboratory between 1975 and 1979 (Moody and Mark, 2001). The dataset contained 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects, both male and female (25 and 22, respectively) of different age groups (between 23 and 89 years). The analog records were digitized according to a sampling rate of 360 Hz, filtered using a 0.1–100 Hz bandpass filter, the heartbeats were marked and manually classified by experts in 16 classes of arrhythmia.

Following the recommendation of ANSI/AAMI standard EC57 (ANSI/AAMI, 2020), the 15 arrhythmia classes reported in the database's annotations were grouped into 5 classes, as depicted in Table 1. The 5th class (D), with an Unknown or with a pacemaker, was discarded.

Table 1: MIT-BIH database classes grouped according to AAMI Standard.

AAMI Group	MIT-BIH Class	MIT-BIH Class (Description)
N	N	Normal beat
	L	Left bundle branch block beat
	R	Right bundle branch block beat
	e	Atrial escape beat
	j	Nodal (junctional) escape beat
SVE	A	Atrial premature beat
	a	Aberrated atrial premature beat
	J	Nodal (junctional) premature beat
	S	Supraventricular premature beat
VE	V	Premature ventricular contraction
	E	Ventricular escape beat
F	F	Fusion of ventricular and normal beat
D	f	Fusion of paced and normal beat
	/	Paced beat
	Q	Unclassifiable beat

However, as the database is unbalanced, with the most typical classes having much more examples, the results could be biased. Data balancing was performed through the resampling method with downsampling for the majority class. Each one of the considered classes had 2296 samples. Based on that, data records were partitioned into two groups for training and testing according to Table 2.

Before training, the annotated R wave peaks (Figure 2) were taken into consideration, a window of 300 samples around the peaks (P-149 to P+150 samples) was segmented. No further pre-processing was done, nor feature was extracted, other than the data window (Singh et al., 2019).

Table 2: Data records partitionnig.

Training					Testing	
101	106	108	109	111	100	103
112	114	115	116	118	105	113
119	121	122	123	124	117	200
201	202	203	205	207	209	212
208	210	214	215	219	213	
220	221	222	223	228		
230	231	232	233	234		

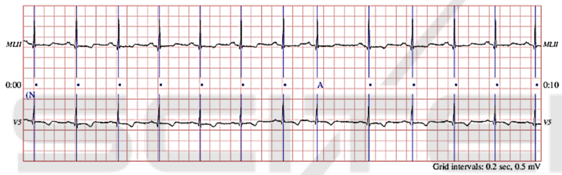

Figure 2: ECG example data.

2.2 Classification Learner App

The classification learner application provided by MATLAB® is a toolbox that allows interactive data analysis training classifiers with several machine-learning models, such as Decision Trees, Discriminant Analysis, Support Vector Machines (SVM), Nearest Neighbor Classifiers (KNN), and Ensemble Classifiers. The app provides classifier performance metrics, such as validation accuracy, confusion matrix, receiver operating characteristic curve (ROC), and the area under the ROC curve (AUC), among other resources.

In this research, a set of 24 classifiers were adopted: Fine, Medium, and Coarse Decision Trees; Linear and Quadratic Discriminant Analysis; Gaussian and Kernel Naive Bayes; Linear, Quadratic, Cubic, and Gaussian SVM; Fine Medium, Coarse, Cosine, Cubic, and Weighted KNN; Boosted Trees, Bagged Trees, Subspace Discriminant, Subspace KNN, and RUSBoost Trees Ensemble Classifiers. Default classifier parameters, and the k-fold cross-validation method, using k=10, were applied. Above mentioned criteria were used to evaluate the performances of different classifiers to assess the two classifiers that had the highest validation accuracies. Those models were exported for testing and final evaluation.

3 RESULTS

Table 3 shows validation accuracies for all classifiers. Most SVM and KNN classifiers reached accuracies above 90%. Cubic SVM and Subspace KNN Ensemble Classifiers reached accuracies of 94%, being the best, while Naive Bayes Classifiers presented had the worst results. Figures 3 and 4 show the confusion matrices of the best two classifiers, from which it can be noted that the major misclassification occurred for the F class.

These two trained models were exported back to the MATLAB® algorithm for testing. Figures 5 and 6 show the resulting confusion matrices, with mean accuracies of 67.1% for the Cubic SVM and 74.4% for the Subspace KNN Ensemble. In both cases, major misclassification occurred between the SVE e N classes.

Table 3: Validation Accuracies.

Classifier	Accuracy (%)
Cubic SVM	**94.1**
Subspace KNN Ensemble	**94.0**
Quadratic SVM	93.4
Fine KNN	93.2
Weighted KNN	92.2
Medium Gaussian SVM	91.8
Cosine KNN	91.7
Bagged Trees Ensemble	91.3
Fine Gaussian SVM	91.0
Medium KNN	90.8
Cubic KNN	90.8
Linear SVM	82.9
Boosted Trees Ensemble	81.6
Fine Tree	81.4
RUSBoost Trees Ensemble	81.0
Coarse KNN	78.5
Quadratic Discriminant	78.4
Coarse Gaussian SVM	78.3
Subspace Discriminant Ensemble	78.0
Linear Discriminant	77.6
Medium Tree	75.3
Coarse Tree	65.0
Kernel Naive Bayes	61.8
Gaussian Naive Bayes	55.9

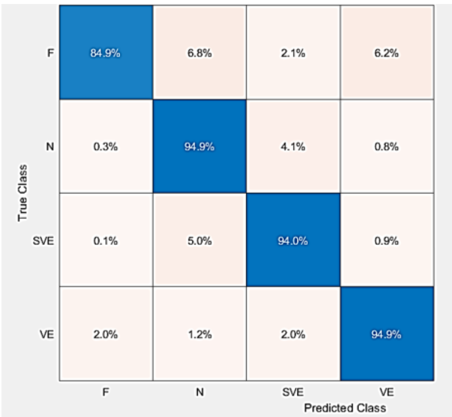

Figure 3: Validation Confusion Matrix for Cubic SVM.

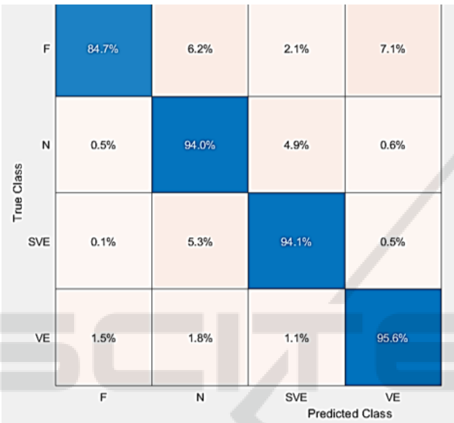

Figure 4: Validation Confusion Matrix for Subspace KNN Ensemble.

3.1 Hardware and User Interface

From a practical point of view, we propose the use of a low-cost device for capturing the ECG signal, consisting of an AD8233 module connected to an Arduino, and a simple interface developed on the MATLAB® Appdesigner (Figure 7). After the acquisition, and classification, the interface will inform, through an e-mail, the detection of possible arrhythmia to a doctor or accredited person.

4 DISCUSSION

The monitoring of vital signs by wearable devices can contribute to the decentralization of health care, allowing self-management and anticipation of emergency care. Therefore, even if the final diagnosis is the healthcare professional's responsibility, machine-learning techniques can automatically recognize and classify specific patterns in these signals, indicating to

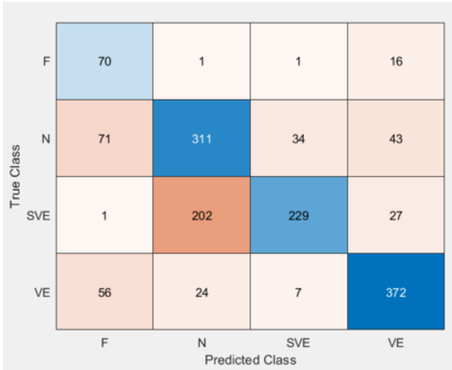

Figure 5: Test Confusion Matrix for Cubic SVM.

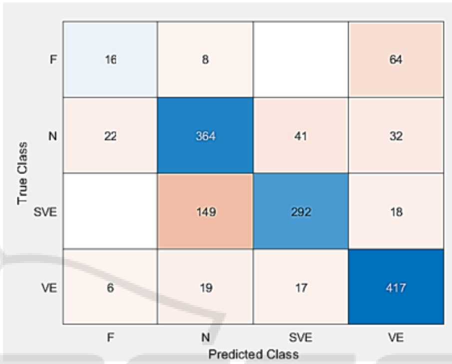

Figure 6: Test Confusion Matrix for Subspace KNN Ensemble.

the user. Several works in the literature are dedicated to applying machine-learning techniques to recognize cardiac arrhythmias, mostly with accuracies between 70% and 98%.

This work presented a study with 24 classifiers using the Classification Learner application from MATLAB® and the MIT-BIH Arrhythmia Database, which is one of the most used databases. However, although the validation results were promising, showing accuracies of 94% for Cubic SVM and Subspace KNN Ensemble, the test phase results showed lower accuracies (74%), with most misclassifications between SVE e N classes.

The ECG data were digitized according to a sampling rate of 360 Hz and filtered using a 0.1–100 Hz bandpass filter. Hence, noises like power line interference, baseline drifts, motion artifacts, and electromyography noise can be added, and thus the lack of pre-processing can be affected classification performance. From a practical point of view, despite the need for an acquisition system to implement some filtering in the signal to eliminate noise, the literature consulted did not clearly show this type of processing.

Another fact was using the raw QRS complex in-

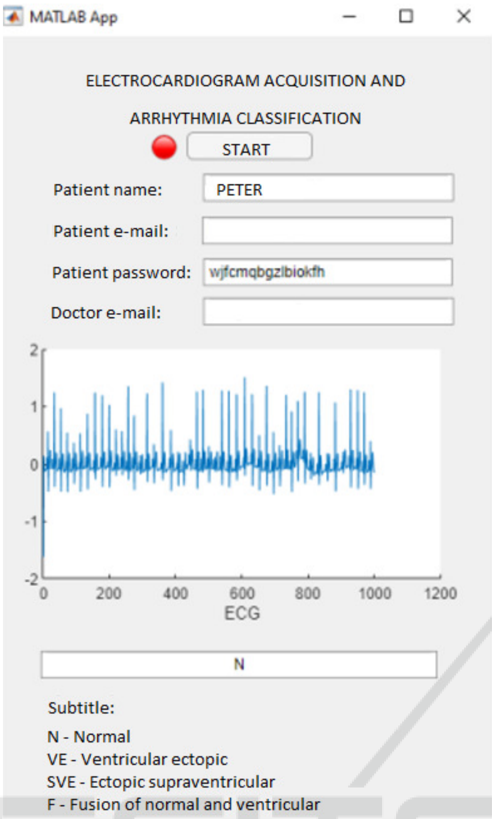

Figure 7: Hardware Design.

stead of a set of extracted features. Results showed that the investigated models could not deal with it. The literature showed that the only model able to deal with a raw signal are those based on Deep Learning due to a proper structure (Zhang et al., 2020; Hassan et al., 2022; Irfan et al., 2022). Other models need a set of features able to discriminate different classes (de Albuquerque et al., 2018; Celin and Vasanth, 2018; Kuila et al., 2020; Mohebbanaaz et al., 2022). Furthermore, despite reducing the number of classes, arrhythmias groups may interfere in this discrimination, especially without specific features as applied in this study.

Moreover, datasets are usually unbalanced. The results could be biased because the class with non-ectopic data has much more samples than the others. It is critical to balance the dataset or approximate it to ensure that each class receives the same priority (AbdElMoneem et al., 2020; Hassan et al., 2022). However, most of the literature work did not mention any data balance. The use of a resampling approach can accomplish balancing. Otherwise, as this work implemented data balance only using the down-sampling technique for the majority class, the number of resulting samples could be not enough for the machine-learning approach, being responsible for the reached accuracies. The use of the up-sampling technique in combination with down-sampling would increase the number of available samples, improving results.

Despite that, for practical use, some improvements are expected, such as the accuracy increase for the arrhythmia classification, and for a smart device, the classifier must be embedded instead of being through a PC interface.

5 CONCLUSION

This work showed a study of 24 models for cardiac arrhythmia classification using the Classification App from MATLAB® and suggested a low-cost device for capturing the ECG signal with a simple interface developed on the MATLAB® Appdesigner, allowing rapid health professional communication for practical use. Results were promising; however, more attention should be given to the extraction of features in order to increase classification accuracy and to the implementation of an embedded system.

ACKNOWLEDGEMENTS

The authors would like to thank Centro Universitário FEI for project support.

REFERENCES

AbdElMoneem, S. S., Said, H. H., and Saad, A. A. (2020). Arrhythmia disease classification and mobile based system design. *Journal of Physics: Conference Series*, 1447(1):012014.

American Heart Association (AHA) (2016). What is an arrhythmia? Available online: https://www.heart.org/en/health-topics/arrhythmia/about-arrhythmia.

ANSI/AAMI (2020). *ANSI/AAMI EC57:2012/(R)2020 - Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms*.

Bayoumy, K., Gaber, M., Elshafeey, A., Mhaimeed, O., Dineen, E. H., Marvel, F. A., Martin, S. S., Muse, E. D., Turakhia, M. P., Tarakji, K. G., and Elshazly, M. B. (2021). Smart wearable devices in cardiovascular care: where we are and how to move forward. *Nat Rev Cardiol.*, 18:581–599.

Celin, S. and Vasanth, K. (2018). Ecg signal classification using various machine learning techniques. *J Med Syst*, 42.

da Silva, L. F., Queiroz, J. A., Vanessa, C., Barros, A. K., Lopes, G. C., and Cabral, L. (2021). Separation method of atrial fibrillation classes with high order statistics and classification using machine learning. In *BIOSIGNALS*.

de Albuquerque, V. H. C., Nunes, T. M., Pereira, D. R., da S. Luz, E. J., Menotti, D., Papa, J. P., and Tavares, J. M. R. S. (2018). Robust automated cardiac arrhythmia detection in ecg beat signals. *Neural Comput & Applic*, 29:679–693.

Fuadah, Y. N. and Lim, K. M. (2022). Optimal classification of atrial fibrillation and congestive heart failure using machine learning. *Frontiers in Physiology*, 12.

Ganapathy, N., Baumgärtel, D., and Deserno, T. M. (2021). Automatic detection of atrial fibrillation in ecg using co-occurrence patterns of dynamic symbol assignment and machine learning. *Sensors*, 21(10).

Goldberger, A., Amaral, L., Glass, L., JM, H., Ivanov, P., Mark, R., Mietus, J., Moody, G., Peng, C., and Stanley, H. (2000). Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation.*, 101(23):E215–20.

Hassan, S. U., Zahid, M. S. M., Abdullah, T. A., and Husain, K. (2022). Classification of cardiac arrhythmia using a convolutional neural network and bi-directional long short-term memory. *DIGITAL HEALTH*, 8:20552076221102766.

Horoba, K., Czabanski, R., Wrobel, J., Matonia, A., Martinek, R., Kupka, T., Kahankova, R., Leski, J. M., and Graczyk, S. (2019). Recognition of atrial fibrilation episodes in heart rate variability signals using a machine learning approach. In *2019 MIXDES - 26th International Conference "Mixed Design of Integrated Circuits and Systems"*, pages 419–424.

Humphreys, M., Warlow, C., and McGowan, J. (2011). *Arrhythmias and their Management*, chapter 10, pages 132–155. John Wiley & Sons, Ltd.

Irfan, S., Anjum, N., Althobaiti, T., Alotaibi, A. A., Siddiqui, A. B., and Ramzan, N. (2022). Heartbeat classification and arrhythmia detection using a multi-model deep-learning technique. *Sensors*, 22(15).

Kuila, S., Dhanda, N., and Joardar, S. (2020). Feature extraction of electrocardiogram signal using machine learning classification. *International Journal of Electrical and Computer Engineering (IJECE)*, 10(6):6598–6605.

Mkoko, P., Bahiru, E., Ajijola, O. A., Bonny, A., and Chin, A. (2020). Cardiac arrhythmias in low- and middle-income countries. *Cardiovasc Diagn Ther.*, 10(2):350–360.

Mohebbanaaz, Kumari, L. V. R., and Sai, Y. P. (2022). Classification of ecg beats using optimized decision tree and adaptive boosted optimized decision tree. *Signal, Image and Video Processing*, 16:695–703.

Mohebbanaaz, Sai, Y. P., and kumari, L. R. (2020). A review on arrhythmia classification using ecg signals. In *2020 IEEE International Students' Conference on Electrical,Electronics and Computer Science (SCEECS)*, pages 1–6.

Moody, G. B. and Mark, R. G. (2001). The impact of the mit-bih arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine.*, 20(3):45–50.

Mustaqeem, A., Anwar, S. M., and Majid, M. (2018). Multiclass classification of cardiac arrhythmia using improved feature selection and svm invariants. *Computational and Mathematical Methods in Medicine*, 2018:7310496.

Ramesh, J., Solatidehkordi, Z., Aburukba, R., and Sagahyroon, A. (2021). Atrial fibrillation classification with smart wearables using short-term heart rate variability and deep convolutional neural networks. *Sensors*, 21(21).

Sager, S., Bernhardt, F., Kehrle, F., Merkert, M., Potschka, A., Meder, B., Katus, H., and Scholz, E. (2021). Expert-enhanced machine learning for cardiac arrhythmia classification. *PLoS One*, 16(12):e0261571.

Shalini, B., Nandini, V., Sandhya, M., and R, B. (2019). Prediction and classification of cardiac arrhythmia. *International Research Journal of Engineering and Technology (IRJET)*, 6(6):572–576.

Singh, V., Tewary, S., Sardana, V., and Sardana, H. K. (2019). Arrhythmia detection - a machine learning based comparative analysis with mit-bih ecg data. In *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, pages 1–5.

WHO (2021). Cardiovascular diseases (cvds). fact sheets. Available online. https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). Accessed Oct/2022.

Zhang, J., Liu, A., Gao, M., Chen, X., Zhang, X., and Chen, X. (2020). Ecg-based multi-class arrhythmia detection using spatio-temporal attention-based convolutional recurrent neural network. *Artificial Intelligence in Medicine*, 106:101856.