Detection of Arrhythmia Using Automated Supervised Learning System For COVID-19

Yashodhan Ketkar[a], Sushopti Gawade[b]

[a]PCE, Panvel, 410206, Maharashtra, India
[b]PCE, Panvel, 410206, Maharashtra, India

ARTICLE INFO

Keywords:
Medical field
Machine Learning
Supervised ML
Prediction

ABSTRACT

COVID-19 disease became a global pandemic in the last few years. This disease was highly contagious, and it quickly spread throughout several countries. Its infection can lead to severe implications in its victims, including cardiovascular issues. This complication develops in some people with a history of cardiovascular illness, whereas it emerges in others after COVID-19 infection. Cardiovascular problems are the primary cause of mortality in COVID-19 patients and are used to predict disease prognosis. Identifying arrhythmia from abnormalities in patient ECG signals is one of the few approaches for detecting cardiovascular disorders. This is a laborious and time-consuming procedure that may be automated using a supervised learning technique. Supervised learning is used in the proposed technique to identify abnormalities in ECG frequency waves. The suggested method will then anticipate the presence of arrhythmia in the patient's ECG and notify medical personnel. The suggested technique detects arrhythmia quicker than existing methods and will aid in decreasing the severity of COVID-19 and its complications in infected individuals.

1. Introduction

COVID-19 has been widespread in recent years. It targets the human respiratory system, causing severe respiratory issues. Depending on a person's condition and the prevalence of comorbidities, this disease can be fatal. Cardiovascular comorbidities are frequent in COVID-19 disease. Cardiovascular comorbidities are also problematic to diagnose in the absence of suitable equipment. Checking for arrhythmia in patients is one approach to detection. Arrhythmia is the irregular beating of the heart. Arrhythmia is detected by examining ECG signals. Because COVID-19 has put a strain on the medical personnel, detection takes longer than usual. Increased Internet connectivity has led to the use of machine learning and artificial intelligence (AI) for service in a variety of sectors. This increases research in the field of machine learning and has an impact on machine learning in a variety of domains. One of them is the medical and healthcare businesses. Machine learning is used to detect and categorize viruses and other microorganisms in patients. In medical applications, machine learning algorithms have already been shown to be quite useful.

The machine learning system may be used to scan these ECG signals and detect them. These signs may be detected considerably faster and more efficiently using supervised learning techniques. In such exact classification problems, supervised algorithms have previously been demonstrated to be quicker than unsupervised techniques. Once taught, this algorithm may also be utilized to make future predictions.

There are several supervised algorithms accessible, allowing us to select the best method for our purposes. This phase can be automated in the case of the general population. A few methods may be pre-programmed into the system, and the computer can then train and pick the best model for the supplied dataset. This will free up medical personnel to focus on patient care and problem-solving.

ketkaryapr19me@student.mes.ac.in (Y. Ketkar); sgawade@mes.ac.in (S. Gawade)

2. Literature Review

Arrhythmia is one of the most common symptoms in patients with COVID19 [2]. Arrhythmia was found in 7% of all 19 Wuhan COVID cases and 14.8% of patients with poor outcomes. Data from 17 studies of 5815 patients suggest that up to 9.3% of confirmed patients detected arrhythmia [16, 13]. Up to 94.4% of all deceased patients had arrhythmia and 95.8% of patients with severe infections had arrhythmia [3, 19]. According to the studies [2, 13, 23], only 8% of patients with arrhythmia had previously had cardiovascular disease, but 56% of the symptoms of arrhythmia recurred after having COVID19 disease.

In the work [22], the author used an ensemble classifier to detect anomalies in the ECG signal. This approach, which combines multiple classifiers for prediction, has proven effective because the accuracy of the ensemble classifier is significantly higher than that of a single classifier. This approach has also been used by other researchers to improve the prediction accuracy of supervised learning models [8, 18, 14]. This study also showed high accuracy and validation scores for anomaly detection [8, 14], while one study made accurate predictions without prior information about [18] in the same class. Was shown to be obtained. You can use the default classifier to automatically get an accurate and accurate prediction of the incoming data stream [9].

To detect the presence of a patient's epileptic seizure, the authors extracted appropriate features and preprocessed the data by applying supervised and unsupervised machine learning classifiers to the patient. Predictions from test data show that supervised learning models are more effective than unstructured learning models [21]. The default classifier can be automated for accurate and accurate predictions [9]. The authors of the [10] study also showed that a commercially available classifier SVM is very efficient in detecting arrhythmia in the ECG signal.

With the ensemble classifier approach, anomalies can be detected by combining multiple classifiers for prediction. The ensemble classifiers can be more effective and accurate than individual supervised learning models [22, 14, 8]. The author of paper [18] discovered that accurate predictions can be obtained with limited or no prior information about past target values of the same class.

Artificial neural networks are also very effective and accurate in detecting arrhythmia from ECG signals. Neural networks can process raw ECG signals and make predictions [7]. Small neural networks are also very efficient in the recognition process [20]. Neural network predictions are relatively accurate and accurate than known classifiers. Lightweight classifiers such as the LDA classifier showed higher accuracy than SVM classifiers for low energy systems [5]. The self-adaptive learning algorithm is also efficient for low energy lightweight systems [12, 17]. The self-learning algorithm makes the system more dynamic and adapts to incoming signals. This dynamic system takes very little time for identification tasks, but classification tasks are more difficult and very time-consuming.

In the work [1], the author uses a two-year dataset collected by Glumo Lake and uses his expertise to train and select models. A mixed approach of data-driven and knowledge-driven modeling is used for the success of the application. As the author of the article [11], he used loo rate and stop criteria for model selection. The author investigated eight different issues and found that a larger loo rate was more desirable. The authors also suggested that modeling difficulties can only be found by careful numerical calculations.

In the article [15], a novel kernel is used to get the dataset description. This approach leads to the discovery of invisible models with minimal human interaction. The author of the article [4] created a new model using the glmulti package. These models are unique and flexible. The model is automatically optimized to provide a multi-model interface. This approach allows you to quickly explore a large set of models for selection purposes.

3. Dataset and Method

3.1. Method

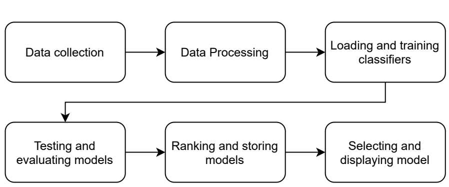

Figure 1: Training and Selection Process

The figure 1, shows the basic architecture of the automated model training and selection system. The data is collected from the user and processed by the application. This data is stored as training and testing datasets.

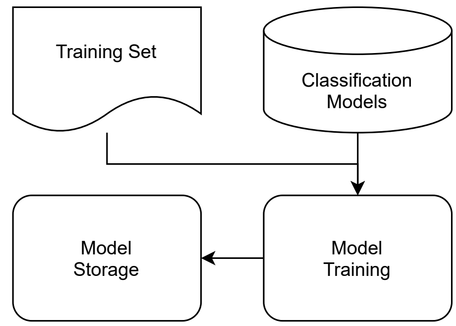

Figure 2: Training Process

Figure 2, shows the training process. In this process, the training dataset is loaded into the system. Premade classification model templates are accessed by the system and trained with provided data. These trained models are stored by the system for the next step.

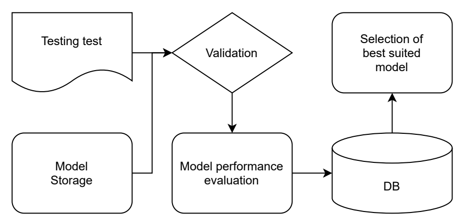

Figure 3: Selection Process

The figure 3, shows the selection process. In this process, the testing dataset is used for the evaluation of the trained models. The trained models are ranked with respect to the performance evaluation. These ranks are used with the help of the tuning parameters to select the best-suited model. This model is stored as the best model for future classification.

3.2. Dataset

The ECG readings in this paper are obtained from the MIT-BIH arrhythmia database. This database is used for automated training and evaluation. This dataset was published in 1999 by MIT-BIH as an open-source database; it consists of training and testing datasets. This database is further divided into four equal parts for analysis. Each training set contains 21888 signals, and the testing set contains 5473 signals.

4. Results And Testing

During the automated training and selection process, the SVM classifier is selected as the best-suited model for dataset 1, dataset 2, and dataset 4. Whereas RF classifier is selected as the best-suited model for dataset 3.

Table 1
Performance of models trained on dataset 1

Metric	KNN	DT	MLP	RF	**SVM**
Accuracy	96.72	94.46	96.89	97.33	**97.40**
F1	89.71	83.43	90.05	91.30	**91.69**
Precision	92.53	81.86	94.71	97.95	**96.43**
Recall	87.06	85.06	85.84	85.50	**87.40**
ROC	92.84	90.68	92.45	92.57	**93.38**
Time(s)	0.457	0.001	0.002	0.015	**0.297**

Table 2
Performance of models trained on dataset 2

Metric	KNN	DT	MLP	RF	**SVM**
Accuracy	96.83	95.04	96.69	97.09	**97.46**
F1	90.28	85.50	89.73	90.75	**92.15**
Precision	94.25	84.90	94.73	98.60	**96.79**
Recall	86.63	86.09	85.23	84.05	**87.93**
ROC	92.77	91.48	92.13	91.90	**93.66**
Time(s)	0.435	0.001	0.003	0.014	**0.295**

Table 3
Performance of models trained on dataset 3

Metric	KNN	DT	MLP	**RF**	SVM
Accuracy	97.07	94.64	96.41	**97.22**	97.44
F1	90.93	84.30	89.34	**91.21**	92.00
Precision	95.82	83.81	90.13	**98.37**	97.81
Recall	86.53	84.80	88.57	**85.02**	86.85
ROC	92.87	90.73	93.29	**92.36**	93.22
Time(s)	0.404	0.001	0.002	**0.017**	0.293

Performance metrics used for evaluation were Accuracy, F1, Precision, Recall, Area under ROC Curve, and Total prediction time. The weightage assigned to these metrics for ranking was 0.65, 0.5, 0.6, 0.6, 0.5, 3 respectively. Where lower value means higher priority.

4.1. Performance Evaluation

The models are tested with a testing dataset obtained from MIT-BIH database. Testing dataset consists of 5473 signals. Figure 4, 5, 6 and 7 shows that model trained with automated system produced satisfactory results. Few models like KNN, RF, and SVM performed better than other models (DT) at the cost of higher prediction time. Whereas MLP models produced good overall results with lower prediction time.

4.2. Performance Error Calculation

For error calculation, best models are tested against training datasets of other models. Figure 8-12 shows the average error introduced when models are tested against training datasets of other models. This chart shows that KNN, MLP, and SVM models introduced minimum errors, whereas DT and RF models introduced large amounts of error. Figure 12 also shows that SVM model produced similar error across all datasets. This smaller difference

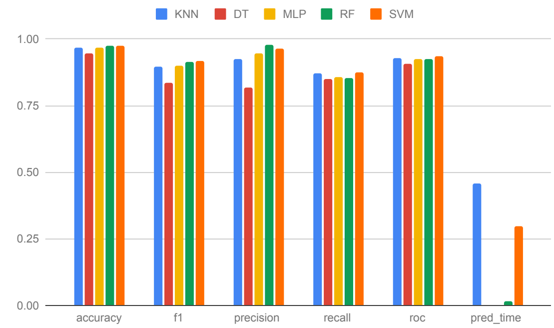

Figure 4: Performance Results Dataset 1

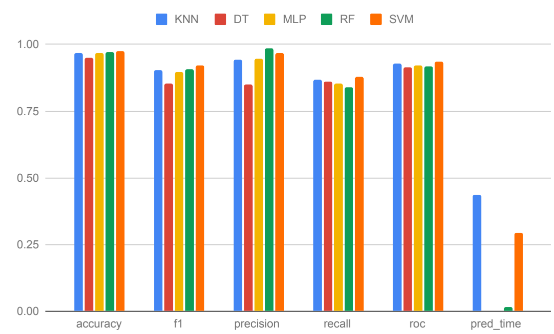

Figure 5: Performance Results Dataset 2

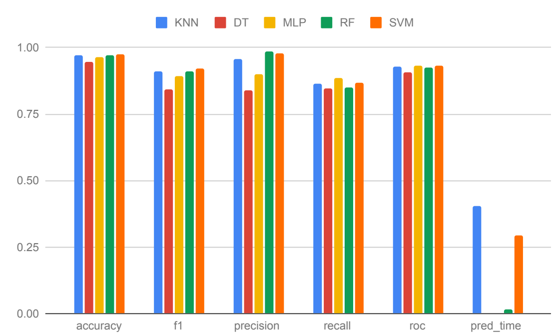

Figure 6: Performance Results Dataset 3

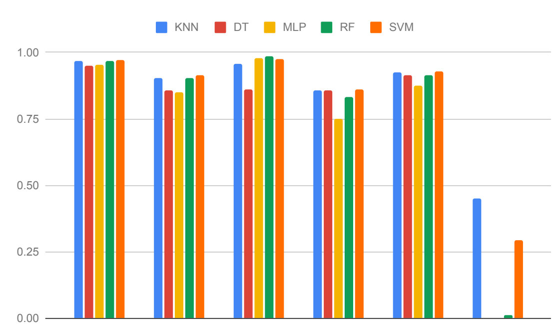

Figure 7: Performance Results Dataset 4

Table 4
Performance of models trained on dataset 4

Metric	KNN	DT	MLP	RF	**SVM**
Accuracy	96.84	94.99	95.34	96.88	**97.15**
F1	90.52	85.73	85.01	90.37	**91.39**
Precision	95.71	85.91	97.83	98.65	**97.41**
Recall	85.86	85.55	75.15	83.37	**86.07**
ROC	92.52	91.28	87.40	91.56	**92.79**
Time(s)	0.452	0.001	0.002	0.014	**0.294**

Table 5
Performance of Decision Tree model trained on dataset 1

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.98	0.94	0.94	0.94
F1	0.96	0.85	0.85	0.84
Precision	0.95	0.83	0.84	0.83
Recall	0.96	0.86	0.85	0.85
ROC	0.97	0.91	0.91	0.90

Table 6
Performance of Decision Tree model trained on dataset 2

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.94	0.98	0.94	0.94
F1	0.84	0.96	0.84	0.85
Precision	0.85	0.96	0.85	0.85
Recall	0.82	0.96	0.84	0.84
ROC	0.89	0.97	0.90	0.90

Table 7
Performance of Decision Tree model trained on dataset 3

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.94	0.94	0.98	0.94
F1	0.84	0.84	0.96	0.83
Precision	0.84	0.83	0.95	0.83
Recall	0.84	0.85	0.96	0.84
ROC	0.90	0.90	0.97	0.90

in error suggests that the SVM classifier can be used for classification tasks of similar nature.

5. Conclusion and Future work

In this paper, we present a novel system. The system provides the end user ability to train the best-suited model for the problem. With the current COVID-19 pandemic, this system can be employed by healthcare professionals for the detection of anomalies such as arrhythmia in patients. The system is tested with the ECG MIT-BIH arrhythmia database. The models trained and selected by the system showed good classification performance, these models also performed satisfactorily against training datasets of other models suggesting good general classification performance.

Future work will involve the use of other freely available datasets to test the general classification performance of

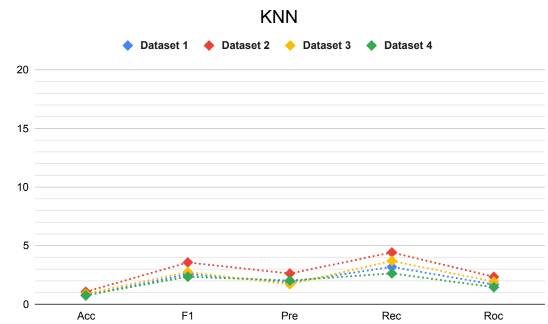

Figure 8: Average Error for KNN model

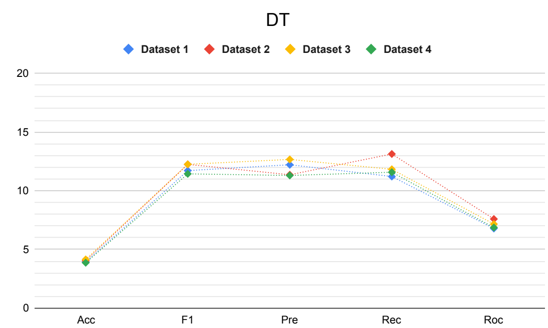

Figure 9: Average Error for DT model

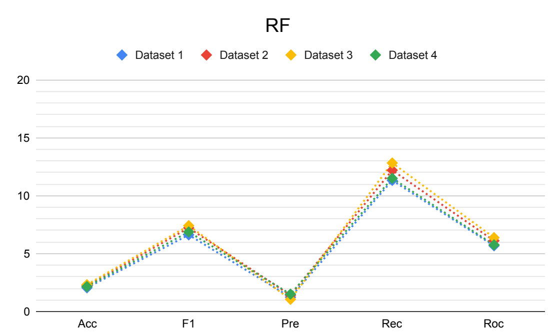

Figure 10: Average Error for RF model

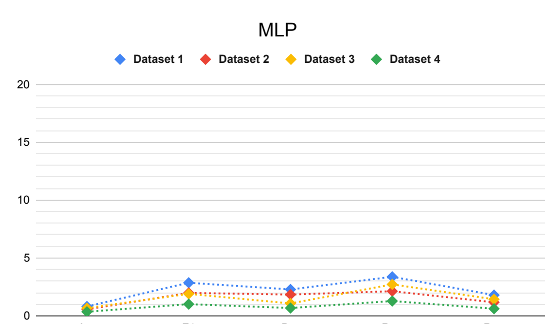

Figure 11: Average Error for MLP model

Table 8
Performance of Decision Tree model trained on dataset 4

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.94	0.94	0.95	0.98
F1	0.85	0.85	0.85	0.96
Precision	0.85	0.85	0.85	0.96
Recall	0.85	0.84	0.85	0.96
ROC	0.91	0.90	0.91	0.97

Table 9
Performance of KNN model trained on dataset 1

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.97	0.97	0.97	0.96
F1	0.93	0.91	0.90	0.90
Precision	0.96	0.95	0.94	0.94
Recall	0.90	0.87	0.87	0.87
ROC	0.94	0.93	0.93	0.93

Table 10
Performance of KNN model trained on dataset 2

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.96	0.97	0.96	0.96
F1	0.90	0.93	0.90	0.89
Precision	0.94	0.96	0.94	0.94
Recall	0.86	0.90	0.86	0.85
ROC	0.92	0.94	0.92	0.92

Table 11
Performance of KNN model trained on dataset 3

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.96	0.96	0.97	0.96
F1	0.90	0.90	0.93	0.90
Precision	0.95	0.95	0.96	0.94
Recall	0.86	0.86	0.89	0.86
ROC	0.92	0.92	0.94	0.92

Table 12
Performance of KNN model trained on dataset 4

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.96	0.96	0.96	0.97
F1	0.90	0.90	0.90	0.92
Precision	0.94	0.95	0.94	0.96
Recall	0.86	0.86	0.87	0.89
ROC	0.92	0.92	0.93	0.94

Table 13
Performance of MLP model trained on dataset 1

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.97	0.96	0.96	0.96
F1	0.92	0.89	0.89	0.89
Precision	0.97	0.95	0.94	0.95
Recall	0.88	0.85	0.85	0.85
ROC	0.93	0.92	0.92	0.92

Table 14
Performance of MLP model trained on dataset 2

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.96	0.97	0.96	0.96
F1	0.90	0.91	0.90	0.90
Precision	0.95	0.97	0.94	0.95
Recall	0.85	0.87	0.85	0.85
ROC	0.92	0.93	0.92	0.92

Table 15
Performance of MLP model trained on dataset 3

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.96	0.96	0.96	0.96
F1	0.89	0.89	0.90	0.88
Precision	0.89	0.89	0.90	0.89
Recall	0.88	0.88	0.91	0.88
ROC	0.93	0.93	0.94	0.93

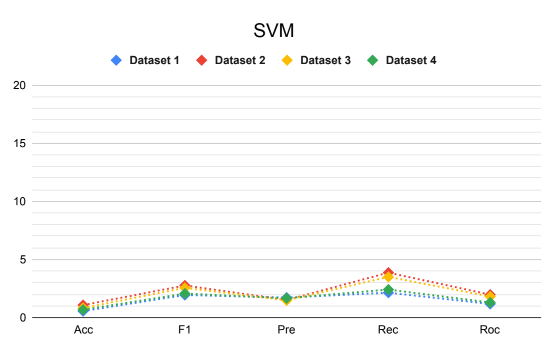

Figure 12: Average Error for SVM model

the system as well as testing the current system in a real-time environment. Also modifying the system to work with non-labeled databases by employing unsupervised learning methods.

References

[1] Atanasova, N., Todorovski, L., Džeroski, S., Kompare, B., 2008. Application of automated model discovery from data and expert knowledge to a real-world domain: Lake glumsø. Ecological Modelling 212, 92–98. doi:10.1016/j.ecolmodel.2007.10.032.

[2] Babapoor-Farrokhran, S., Rasekhi, R.T., Gill, D., Babapoor, S., Amanullah, A., 2020. Arrhythmia in covid-19. SN Comprehensive Clinical Medicine , 1–6.

[3] Beri, A., Kotak, K., 2020. Cardiac injury, arrhythmia, and sudden death in a covid-19 patient. HeartRhythm case reports 6, 367–369.

[4] Calcagno, V., de Mazancourt, C., 2010. glmulti: An r package for easy automated model selection with (generalized) linear models. Journal of Statistical Software 34, 1–29. URL: https://www.jstatsoft.org/index.php/jss/article/view/v034i12, doi:10.18637/jss.v034.i12.

[5] Chen, T., Mazomenos, E.B., Maharatna, K., Dasmahapatra, S., Niranjan, M., 2013. Design of a low-power on-body ecg classifier

Table 16
Performance of MLP model trained on dataset 4

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.95	0.95	0.95	0.95
F1	0.85	0.85	0.85	0.86
Precision	0.97	0.97	0.97	0.98
Recall	0.75	0.75	0.75	0.76
ROC	0.87	0.87	0.87	0.88

Table 17
Performance of RF model trained on dataset 1

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.99	0.97	0.97	0.97
F1	0.98	0.91	0.91	0.91
Precision	0.99	0.98	0.97	0.98
Recall	0.96	0.85	0.85	0.85
ROC	0.98	0.92	0.92	0.92

Table 18
Performance of RF model trained on dataset 2

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.96	0.99	0.97	0.97
F1	0.90	0.97	0.91	0.90
Precision	0.98	0.99	0.97	0.98
Recall	0.83	0.96	0.85	0.84
ROC	0.91	0.98	0.92	0.91

Table 19
Performance of RF model trained on dataset 3

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.96	0.97	0.99	0.97
F1	0.90	0.90	0.97	0.90
Precision	0.98	0.98	0.99	0.98
Recall	0.83	0.84	0.96	0.83
ROC	0.91	0.91	0.98	0.91

Table 20
Performance of RF model trained on dataset 4

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.96	0.97	0.97	0.99
F1	0.90	0.91	0.90	0.97
Precision	0.98	0.98	0.97	0.99
Recall	0.84	0.84	0.85	0.95
ROC	0.91	0.92	0.92	0.97

Table 21
Performance of SVM model trained on dataset 1

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.97	0.97	0.97	0.97
F1	0.93	0.92	0.91	0.91
Precision	0.98	0.97	0.96	0.96
Recall	0.89	0.87	0.87	0.87
ROC	0.94	0.93	0.93	0.93

Table 22
Performance of SVM model trained on dataset 2

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.97	0.98	0.97	0.97
F1	0.91	0.94	0.91	0.91
Precision	0.97	0.98	0.96	0.96
Recall	0.86	0.89	0.86	0.86
ROC	0.92	0.94	0.93	0.92

Table 23
Performance of SVM model trained on dataset 3

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.97	0.97	0.98	0.97
F1	0.91	0.92	0.94	0.91
Precision	0.97	0.97	0.98	0.97
Recall	0.85	0.87	0.89	0.87
ROC	0.92	0.93	0.94	0.93

for remote cardiovascular monitoring systems. IEEE Journal on Emerging and Selected Topics in Circuits and Systems 3, 75–85.

[6] Fazeli, S., . ECG heartbeat categorization dataset. URL: https://www.kaggle.com/shayanfazeli/heartbeat. accessed on 15 Dec 2020.

[7] Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., Tison, G.H., Bourn, C., Turakhia, M.P., Ng, A.Y., 2019. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature medicine 25, 65–69.

[8] Huang, J.S., Chen, B.Q., Zeng, N.Y., Cao, X.C., Li, Y., 2020. Accurate classification of ecg arrhythmia using mowpt enhanced fast compression deep learning networks. Journal of Ambient Intelligence and Humanized Computing , 1–18.

[9] Imbrea, A.I., 2021. Automated machine learning techniques for data streams. arXiv preprint arXiv:2106.07317 .

[10] Jha, C.K., Kolekar, M.H., 2020. Cardiac arrhythmia classification using tunable q-wavelet transform based features and support vector machine classifier. Biomedical Signal Processing and Control 59, 101875.

[11] Lee, J.H., Lin, C.J., 2000. Automatic model selection for support vector machines. online .

[12] Lei, W.K., Li, B.N., Dong, M.C., Vai, M.I., 2007. Afc-ecg: An adaptive fuzzy ecg classifier, in: Soft computing in industrial applications. Springer, pp. 189–199.

[13] Liu, Q., Chen, H., Zeng, Q., 2020a. Clinical characteristics of covid-19 patients with complication of cardiac arrhythmia. The Journal of Infection 81, e6.

[14] Liu, Y., Lou, Y., Huang, S., 2020b. Parallel algorithm of flow data anomaly detection based on isolated forest, in: 2020 International Conference on Artificial Intelligence and Electromechanical Automation (AIEA), IEEE. pp. 132–135.

[15] Malkomes, G., Schaff, C., Garnett, R., 2016. Bayesian optimization for automated model selection, in: Workshop on Automatic Machine Learning, PMLR. pp. 41–47.

[16] Mulia, E.P.B., Maghfirah, I., Rachmi, D.A., Julario, R., 2021. Atrial arrhythmia and its association with covid-19 outcome: a pooled analysis. Diagnosis .

[17] Owis, M.I., Abou-Zied, A.H., Youssef, A.B., Kadah, Y.M., 2002. Study of features based on nonlinear dynamical modeling in ecg arrhythmia detection and classification. IEEE transactions on Biomedical Engineering 49, 733–736.

Table 24
Performance of SVM model trained on dataset 4

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Accuracy	0.97	0.97	0.97	0.97
F1	0.91	0.91	0.91	0.93
Precision	0.97	0.96	0.96	0.98
Recall	0.85	0.86	0.86	0.88
ROC	0.92	0.93	0.93	0.94

[18] Rajak, A., Shrivastava, A.K., Vidushi, 2020. Applying and comparing machine learning classification algorithms for predicting the results of students. Journal of Discrete Mathematical Sciences and Cryptography 23, 419–427.

[19] Ren, H.G., Guo, X., Tu, L., Hu, Q., Blighe, K., Safdar, L.B., Stebbing, J., Weiner, S.D., Willis, M.S., Rosendaal, F.R., et al., 2020. Clinical characteristics and risk factors for myocardial injury and arrhythmia in covid-19 patients. medRxiv .

[20] Sannino, G., De Pietro, G., 2018. A deep learning approach for ecg-based heartbeat classification for arrhythmia detection. Future Generation Computer Systems 86, 446–455.

[21] Siddiqui, M.K., Morales-Menendez, R., Huang, X., Hussain, N., 2020. A review of epileptic seizure detection using machine learning classifiers. Brain informatics 7, 1–18.

[22] Sun, Z., Wang, C., Zhao, Y., Yan, C., 2020. Multi-label ecg signal classification based on ensemble classifier. IEEE Access 8, 117986–117996.

[23] Yarmohammadi, H., Morrow, J.P., Dizon, J., Biviano, A., Ehlert, F., Saluja, D., Waase, M., Elias, P., Poterucha, T.J., Berman, J., et al., 2021. Frequency of atrial arrhythmia in hospitalized patients with covid-19. The American Journal of Cardiology 147, 52–57.