The Journal of Supercomputing (2024) 80:4286–4308
https://doi.org/10.1007/s11227-023-05583-8

An intelligent hybrid classification model for heart disease detection using imbalanced electrocardiogram signals

Shwet Ketu[1] · **Pramod Kumar Mishra**[2]

Accepted: 5 August 2023 / Published online: 12 September 2023

Abstract

Cardiovascular disease (CVD) is among one of the notable menaces to society worldwide. CVD causes the highest number of deaths each year making it one of the most life-threatening diseases across the globe. Most deaths from CVD are sudden therefore patients do not have a chance to get medical assistance in time. Consequently, an immense need for a smart real-time system arises that can be used to monitor heart patients' activities affecting their cardiac health. This system acts as a life-saving tool during serious health emergencies. Data analysis in real-time will proves to be a substantial enhancement in innovative healthcare practices, by which in the near future we can develop an effective, faster, and smarter diagnosis system for doctors. If we talk about real-time data monitoring possibilities, Internet of Things (IoT) empowered systems can provide one of the better solutions. IoT-enabled intelligent healthcare system include a variety of applications, such as Blood Pressure (BP) check, Heart Rate (HR) monitoring, Electrocardiography (ECG) observation, etc. This paper recommends an IoT-enabled ECG monitoring system for data generation (with the help of Node MCU ESP32 and heart rate sensor AD8232) and an intelligent hybrid classification model for data classification. The dataset used has two classes where class 1 represents healthy patients and class 2 represents cardiac ill patients. A comparison among state-of-the-art algorithms and recommended hybrid models has been carried out to establish the accurateness and suitableness of our recommended model. The recommended model attains the highest accuracy of 99.7% under different validation criteria among all the state-of-the-art algorithms, i.e. Adaboost (91.88%), Bagging (92.40%), random forest (92.48%), K-Nearest Neighbor (92.38%), and support vector machine (91.98%). The recommended hybrid model not only handles the complexities of class imbalance for electrocardiogram datasets but will also help in building intelligent and accurate IoT-enabled healthcare systems.

Keywords Intelligent hybrid classification · Cardiovascular diseases (CVD) · Class imbalance problem · Electrocardiography (ECG) · Internet of things (IoT)

Extended author information available on the last page of the article

1 Introduction

According to the World Health Organization (WHO), the approximate count of patients deceased because of cardiovascular disease (CVD) is nearly 17.9 million, accounting for close to 31% of all fatalities [1, 2]. CVD includes various underline diseases, such as raised blood pressure (hypertension), coronary heart disease (heart attack), peripheral artery disease, rheumatic heart disease, cerebrovascular disease (stroke), deep vein thrombosis, heart failure, pulmonary embolism and congenital heart disease [3]. Of these diseases, approximately 85% of deaths are caused by stroke and heart attack. As per the WHO's reports, by 2030, about 23.6 million individuals will die due to CVDs, i.e. primarily from stroke and heart disease [4, 5]. Thus, there is an immense need for continuous monitoring of some essential parameters of the human body, which are critical and should be exhaustively monitored in real-time paradigms.

The enormous growth in the field of Internet of Things (IoT) has facilitated Information Technology (IT) to new heights [6–10]. Rapid development in the empire of IoT-based applications areas makes IoT a rising technology. In the current viewpoints, approximate all the application domains the IoT is getting involved and actively participating in the journey towards a smarter world [11–13]. In the healthcare domain, the traditional procedure was being followed by the patients but after the emergence of IoT in healthcare, the e-health or smart health concept has come into the picture [14–18]. Resultant, a variety of smart devices are being developed for enabling services such as remote monitoring of the patients, unleashing patients' healthy and safe, and empowering doctors to verbalize superlative care [19–21]. This technological advancement will not only reduce the medical overhead but also enable in-time support of the patients at remote locations [14–18]. It also plays a major role in decreasing the total expenditure by minimizing the span of hospital stay with improved treatment outcomes.

In the classification problem, the data with unbalanced nature is one of the biggest issues, and as far as the healthcare domain is concerned it even became more crucial because the medications are totally dependent upon the classification outcome [22–26]. Therefore, in the healthcare domain, the classification of unbalanced datasets is an emerging area of research. Over time a number of researchers have not only suggested their viewpoints in the form of algorithms and theoretical approaches [27, 28] but also developed various class-balancing solutions in the form of hybrid paradigms [29, 30]. As far as data balancing techniques are concerned, two types of data balancing techniques are being widely used where the first is under-sampling and the other is over-sampling [31, 32]. In the under-sampling approach, the class balancing is done by eliminating the data samples from the majority class, whereas, in the over-sampling approach, the class balancing is done by adding up artificial samples to the minority class.

Individuals' well-being is one of the crucial tasks and it becomes more complex when we are dealing with one of the deadliest diseases, i.e. CVD in real-time scenarios. Consequently, there is a need for algorithmic approaches that would play an essential role in reducing the total risk of CVD through its efficient classification.

Keeping these constraints in our mind, we begin the experimental examination with basic classification models that are less accurate and not capable enough to deal with the class imbalance problems. After several trials, we found that the proposed intelligent hybrid classification model is well suited for classifying the imbalanced Electrocardiogram datasets.

The main contributions of the paper are:

- To establish an IoT-enabled ECG monitoring system for data generation with the help of Node MCU ESP32 and heart rate sensor AD8232.
- To propose an intelligent hybrid classification model having the capability of handling the complexities of class imbalance with more accurate results.

The characterization of this paper is as follows: Section two presents a short description of current literature based on algorithmic approaches for the classification of ECG Dataset. In section three, a brief discussion of the methods and materials such as dataset generation and description, proposed epistemology and statistical measures have been presented. The statistical measure-based classification results have been shown in section four. The deeper insights into the classification results have been presented in section five. Section six incorporates the closing remarks along with future routes of the work.

2 Related work

Massive growth in the field of Information technology encourages research to explore the dimensions of recent technologies. It also motivates researchers and groups to build a technological solution for human well-being. In a couple of years, various development not only in the algorithmic perspective but also in system design has been seen [33–37]. If we talk about the healthcare domain, a lot of possibilities are still available, which will catalyze the idea of a smart world. Cardiovascular disease (CVD) is a crucial disease among various life-threatening diseases across the globe, it has gotten the attention of researchers to work on and give their contributions to social well-being. From time to time various algorithmic solutions to the ECG dataset have been suggested but there is still plenty of scope for improvements [38–63]. The quick insights of the current research on ECG datasets are shown in Table 1.

3 Materials and methods

This section introduces the material and methodology that has been used to carry out the experimental evaluation. This section is divided into five subsections, where, the first subsection refers to the hardware setup for ECG data generation. In the second subsection, the dataset description has been presented. The model setup for the classification task has been discussed in subsection three. In the fourth subsection,

Table 1 Insights of the Contemporary Research

Author	Research problem	Method used	Dataset	Nature of data		Accuracy in %
				ECG Signal	Integer/Real	
Dolatabadi et al. [38]	Diagnosis of coronary artery disease (CAD)	Support vector machine (SVM)	STDB dataset	✓		99.20
Tayefi et al. [39]	Prediction of coronary heart disease (CHD)	Decision trees (DT)	Personal health records		✓	94.00
Arabasadi et al. [40]	Detection of heart disease	Hybrid neural network (NN) and genetic algorithm (GA)	Z-Alizadeh sani dataset		✓	93.85
Mustaqeem et al. [41]	Recommender system for cardiac diseases patients	Random forest (RF), support vector machine (SVM), and multilayer perceptron (MLP)	POF hospital		✓	97.80
Boon et al. [42]	Prediction of paroxysmal atrial fibrillation (PAF)	Heart rate variability (HRV) and support vector machine (SVM)	Atrial Fibrillation Dataset	✓		87.70
Mahajan et al. [43]	Detection of congestive heart failure (CHF)	Probabilistic symbol method for pattern recognition and the ensemble method of bagged decision trees	PhysioNet	✓		98.80
Bozkurt et al. [44]	Classification of heart sound for pathology detection	Convolutional neural network (CNN)	PhysioNet	✓		84.60
Sudarshan et al. [45]	Prediction of congestive heart failure (CHF)	Dual-tree complex wavelets transform (DWT)	MIT-BIH + BIDMC CHF	✓		99.86
Aborokbah et al. [46]	Decision-making system for intensive health care (IHC)	Radial basis function (RBF) and linear kernel function (LKF) based support vector machine	Personal health records	✓		87.90
Plawiak et al. [47]	Recognition of cardiac health	Support vector machine with genetic algorithm (10-fold cross-validation)	MIT- BIH	✓		98.85

Table 1 (continued)

Author	Research problem	Method used	Dataset	Nature of data		Accuracy in %
				ECG Signal	Integer/Real	
Miao et al. [48]	Mortality prediction of heart failure	Improved version of random survival forest (RSF)	Personal health records		✓	82.10
Tan et al. [49]	Prediction of coronary artery disease (CAD)	Long Short-term memory network (LSTM) and CNN	PhysioNet	✓		99.85
Dominguez et al. [50]	Classification and recognition heart murmurs	A novel convolutional neural network	PhysioNet	✓		97.00
Shaikh et al. [51]	Arrhythmia classification	FAPAC model	MIT-BIH-ARR dataset	✓		99.97
Alfaras et al. [52]	Arrhythmia classification	Ensemble ESN	MIT-BIH-ARR dataset	✓		98.6
Xie et al. [53]	Arrhythmia classification	Random Forest	MIT-BIH-ARR dataset	✓		96.38
Raj and Ray [54]	Arrhythmia classification	KNN and SVM	MIT-BIH-ARR dataset	✓		99.11
Siavashi and Majidi [55]	Atrial Fibrillation	Decision tree	Personal Health Records		✓	95.00
Zhang et al. [56]	Atrial fibrillation	Convolutional neural network	Personal Health Records		✓	91.60
Kwon et al. [57]	Detection of congestive heart failure (CHF)	Convolutional neural network	Personal Health Records		✓	93.40
Baraeinejad et al. [58]	Prediction of coronary artery disease (CAD)	Artificial neural networks	PhysioNet MIT-BIH dataset	✓		98.70
Shrestha and Yu [59]	Congestive heart failure (CHF	Random forest	PhysioNet MIT-BIH dataset	✓		85.00
Muthu Ganesh and Nithiyanantham [60]	Risk prediction of cardiovascular disease	Novel recurrent neural network named (ERNN)	Personal Health Records		✓	96.00
Sheeba et al. [61]	Risk prediction of cardiovascular disease	Mixed kernel-based extreme Learning machine (MKELM)	Personal health records		✓	99.50
Jansi Rani et al. [62]	Risk prediction of cardiovascular disease	Random forest (RF)	UCI cleveland heart diseases dataset	✓		88.00
Manimurugan et al. [63]	Risk prediction of cardiovascular disease	Recurrent neural networks (RNN)	UCI cleveland heart diseases dataset	✓		99.15

the recommended model has been introduced. Statistical measures for the validation of the classification model have been presented in the last subsection five.

3.1 Hardware setup for ECG data generation

In order to generate the ECG data, we made a setup that mainly consists of a node MCU (ESP32) and a heart sensor (AD8232). In Fig. 1a the graphical representation of the hardware setup has been shown, whereas the nine electrode placement (E1—Fourth intercostal space (at the right sternal border), E2—Fourth intercostal space (at the left sternal border), E3—Intermediate between leads E2 and E4, E4—Fifth intercostal space (at the midclavicular line), E5—Left anterior axillary line (as the same horizontal plane of E4), E6—Left mid axillary line (as the same horizontal plane of E4 and E5), E7—Right arm (inner wrist), E8—Left arm (inner wrist), and E9—Right side of stomach) the human body is shown in Fig. 1b.

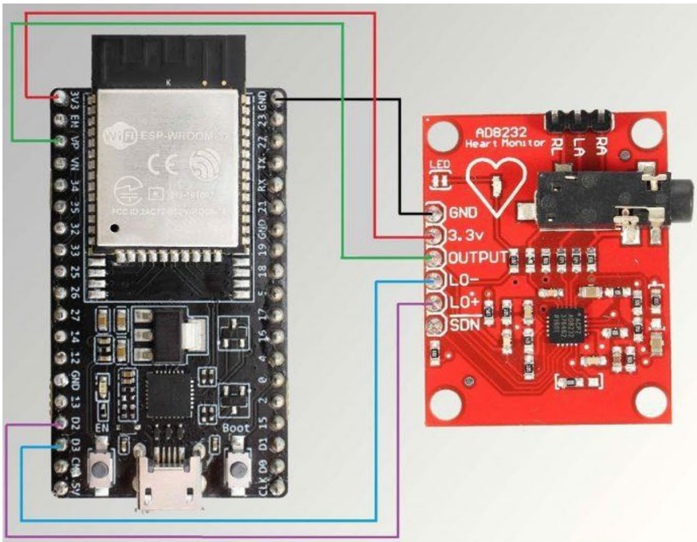
In Table 2, the pin connection among node MCU (ESP32) and heart sensor (AD8232) for the ECG data generation have been shown.

The data has been generated in real-time and stored in cloud storage (Ubidots) over a TCP connection with the help of the HTTP POST command. The generated data is transferred in real-time to the cloud storage by using a Wi-Fi connection. The working steps of the hardware setup have been shown in Fig. 2. The functioning of this hardware setup is as follows:

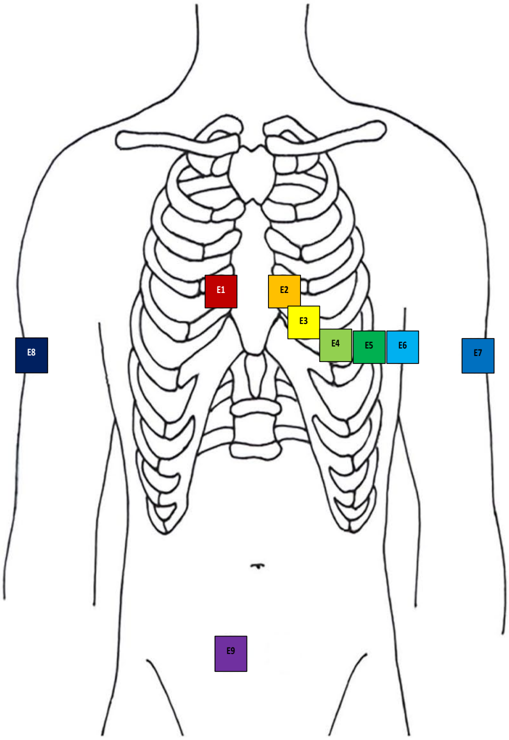
- First of all, the connection between the heart sensor (AD8232) and node MCU (ESP32) is established.
- In the second step, the electrode placement to the human body is performed.
- In the third step, the generated data is visualized on the serial monitor.
- In the fourth step, this generated data is transferred into cloud storage with the help of the ESP32 Wi-Fi module.
- In the last step after this ECG data is extracted from the cloud medium to the local machine for performing further investigation

3.2 Dataset description

For the experimental analysis, the ECG data have been used, which is generated through Node MCU (ESP32) and heart rate sensor (AD8232). Nine sensors (E1–E9) are placed at different body locations and their corresponding readings are observed. This exercise has performed on the 50 volunteer participants over a time span of 150 s. For every second, a tuple consisting of nine attributes is generated by the system and uploaded to the server (Ubidots) over a TCP connection with the help of the HTTP POST command. The generated stream of data is transferred in real-time to the cloud storage by using a Wi-Fi connection. This ECG data has been extracted from the cloud to a local/native machine for evaluation purposes. Based on the current health of the volunteer this dataset has been classified into the two-class where class 1 denotes healthy patients and class 2 represents the cardiac ill patient. This dataset is consisting of 1700 instances of 10 attributes. The visualization of the ECG



(a)



(b)

Fig. 1 Hardware setup for ECG data generation (**a**) Hardware setup (**b**) Electrode placement

Table 2 Pin connection

AD8232	ESP 32
GND	GND
3.3 V	3.3 V
Output	VP
D2	LO+
D3	LO−

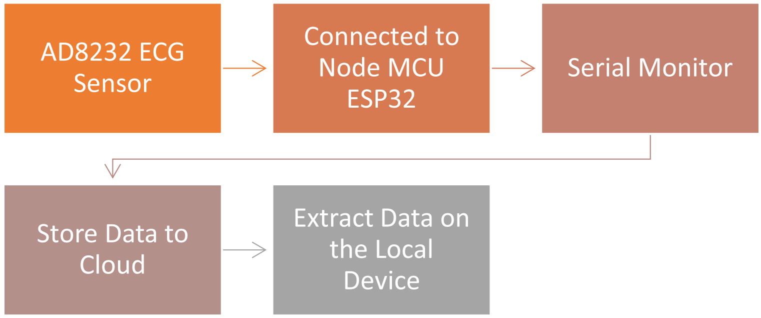


Fig. 2 Working steps of the hardware setup

dataset (nine channels with class level) and their co-relation are presented in respective Fig. 3a, b.

The class-based partitioning of the ECG dataset over nine attributes is shown in Table 3, which consists of the attribute's illustration with the help of range (min and max), means, and standard deviation.

3.3 Model setup

The classification model setup for the experimental analysis of the ECG dataset has been shown in Fig. 4. This setup is comprised of five essential steps. In step one, the ECG data is given as input to the model. In step two, the data preprocessing for the exclusion of unusual objects and missing values has been performed. Step three is consisting of the classification task where the processed data is given out as an input to the classification algorithms (i.e. K-Nearest Neighbor (KNN), support vector machine (SVM), random forest (RF), Adaboost (ADB), and Bagging (BAG)). Performance estimation of the classification algorithm is measured in step four and based on these classification results the identification of the best classification model is identified in step five. All the experimental evaluation has been executed using various evaluation criteria, i.e. 2, 3, 5, and 10-fold on a Dell workstation with a 64-bit Intel Xeon processor running at 3.60 GHz and 32 GB of RAM. Python has been used to implement each of the algorithms being used in the simulation.

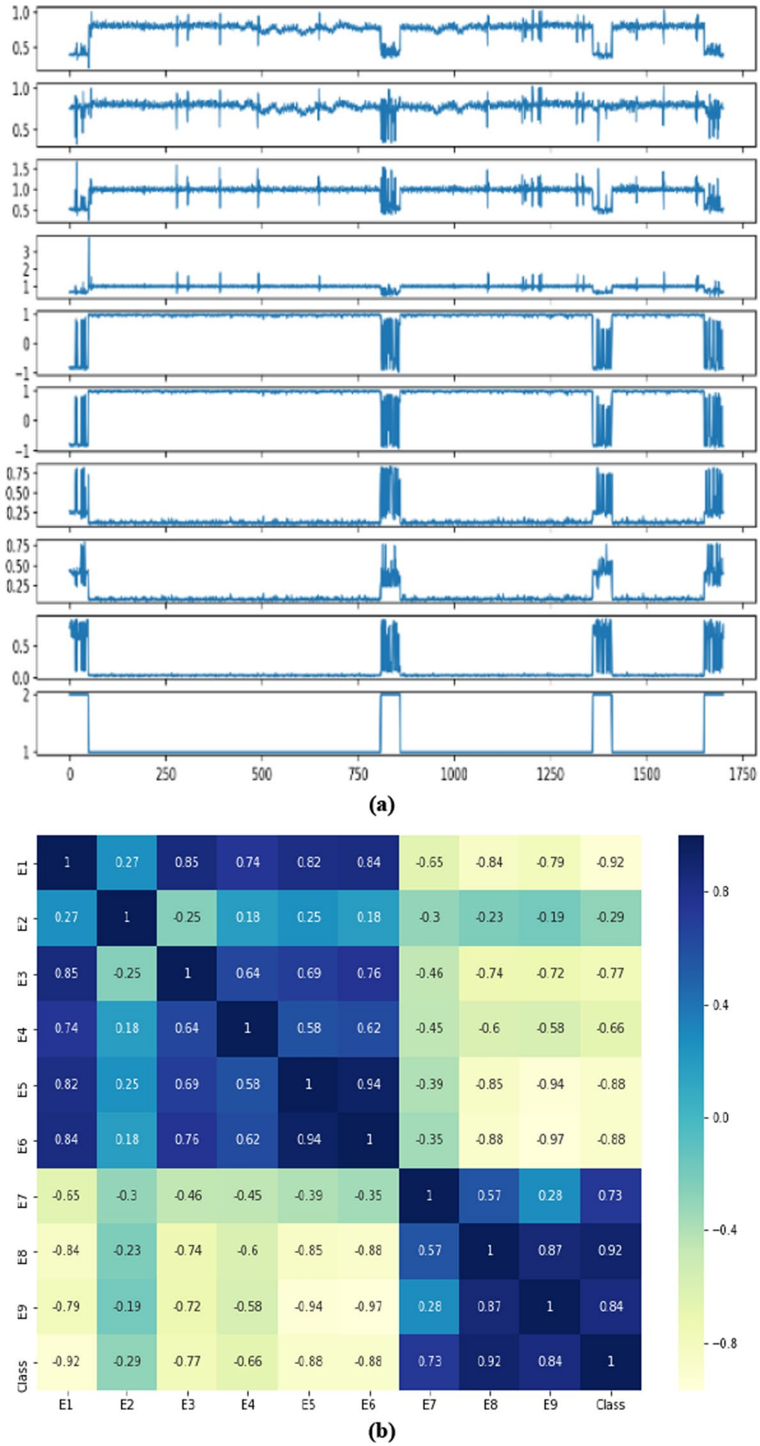


Fig. 3 Dataset description (**a**) Visualization of the ECG dataset (**b**) Co-relation coefficient matrix

Table 3 Class-based distribution of the ECG dataset

Attributes	Class 1				Class 2			
	Range		Mean	Std. Dev	Range		Mean	Std. Dev
	Min	Max			Min	Max		
E1	0.214	1.025	0.792	0.048	0.336	0.572	0.428	0.051
E2	0.522	1.025	0.793	0.045	0.336	0.983	0.738	0.117
E3	0.263	1.58	1.002	0.072	0.387	1.661	0.613	0.225
E4	0.633	3.805	1.005	0.111	0.444	1.295	0.697	0.124
E5	0.842	1	0.978	0.021	−0.962	0.88	−0.364	0.627
E6	0.842	1	0.978	0.021	−0.926	0.91	−0.354	0.681
E7	0.107	0.2	0.126	0.014	0.173	0.827	0.401	0.238
E8	0.067	0.147	0.083	0.011	0.213	0.787	0.421	0.133
E9	0.027	0.08	0.041	0.007	0.08	0.933	0.566	0.315

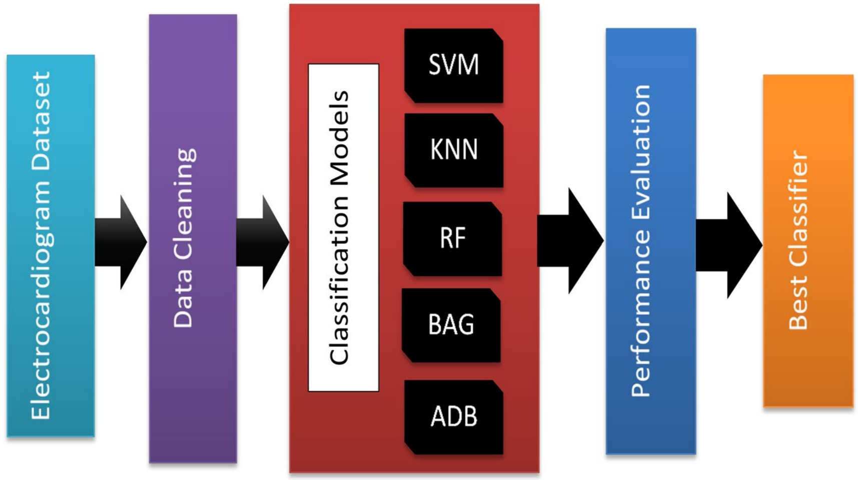


Fig. 4 Classification model setup

3.4 Proposed hybrid classification model

The workflow of the recommended hybrid model is presented in Fig. 5. The recommended hybrid classification model is composed of several steps are:

Step I The raw data is given out as input to the recommended model.
Step II The pre-processing task on the raw ECG dataset is performed to eliminate the missing values and unusual objects from the dataset.
Step III Class balancing has been achieved using SMOTE (Synthetic Minority Oversampling Technique) and which gives a new balanced dataset as output.

Fig. 5 Work-flow of the proposed hybrid model

Step IV This new balanced dataset has been given out as an input to the hyper-tuned random forest algorithms under the various evaluation criteria, i.e. 2, 3, 5, and 10-fold.
Step V The statistical parameters (i.e., accuracy, recall, precision, and f1-score) based on performance evaluation on the hybrid classification model have been performed.

3.4.1 Class balancing using SMOTE

Class balancing is one of the critical matters which should be effectively handled while making the classification. Suppose, we have a binary classification problem where one class holds the majority of samples and the other one has very few data samples. Thus, making the classification based on imbalanced data may give biased results toward the majority class because while making the classification model the majority class contribution will be more as compared to the minority class. Resultantly, the correctness of the classification model will be sacrificed. Therefore, in dealing with the class imbalance problem we have used a SMOTE algorithm which was introduced by Chawla et al. in the year 2002 [64, 65]. The basic principle of this algorithm is to make the class balance by generating artificial samples in the minority class. It uses the k-nearest neighbors (NNs) concept to generate random synthetic samples. The SMOTE-based class balancing result has been shown in Table 4, which contains class-wise distribution with the various SMOTE percentage (i.e. 0, 50, 150, 250, 350, 450, 550, and 650).

The pseudocode of the SMOTE algorithm to solve the class imbalance issue of the ECG dataset is represented in Algorithm 1.

Table 4 SMOTE based class balancing result

Dataset	SMOTE percentage	Class 1		Class 2		Total instances
		Instances	%	Instances	%	
ECG	0	1500	88.24	200	11.76	1700
	50	1500	83.33	300	16.67	1800
	150	1500	75	500	25	2000
	250	1500	68.18	700	31.82	2200
	350	1500	62.5	900	37.5	2400
	450	1500	57.69	1100	42.31	2600
	550	1500	53.57	1300	46.43	2800
	650	1500	50	1500	50	3000

Algorithm 1. Pseudocode of SMOTE

```
Input Parameters:
  D – Input Dataset
  T – Minority Samples
  N – SMOTE Percentage
  k – Nearest Neighbors
Initialization:
      if (N/100 < 1)
      select random samples T′ from T
      for D′_T ← D_T × N̂ and T ← T′ : D_T ← D′_T
      end if
      for j → 1 to |T| do
            knn ← k-nearest neighbors of data point x_j
            N′ = ⌊N/100⌋
            while N’ ≠ 0 do
                select a random point r from knn
                select random value α from [0,1]
                generate a synthetic point s using r and α
                add s to T
                N’ = N’-1
            end while
      end for
      for each attribute in D_T do
      return Synthesized dataset
Output: (N/100) ∗ T – synthetic samples for minority class
```

3.4.2 Hyper-tuned random forest algorithm

The Random forest (RF) algorithm is among the extensively used classification algorithms [66, 67]. Due to its extensive nature, it can be applicable in roughly all application areas. The reason for picking up this algorithm in classification is its extensive coverage and well-established nature. The best parameter for this classification algorithm is achieved by the hyper-tuning selection criteria. The best hyper-parameter is used in the recommended hybrid paradigms. The pseudocode of the hyper-tuned random forest model for the classification of the ECG dataset has been represented in Algorithm 2.

Algorithm 2. Pseudocode of the Hyper-Tuned Random Forest Model

Select the best hyperparameter values from the RandomizedSearchCV and do the classification with the best hyperparameter estimator or score.

Inputs:

X − input dataset
Y − $validation\ samples$
$hyperparameters$ − ($bootstrap$, max $\_depth$, max $\_features$, min $\_sample\_leaf$, min_sample_split, $n\_estimators$)

1. **Function** RandomizedSearchCV ($bootstrap$, max $\_depth$, max $\_features$, min $\_sample\_leaf$, min $\_sample\_split$, $n\_estimators$)
2. Apply randomized search on hyperparameter
3. **return** best hyperparameter estimator
4. **Function** RandomForestClassifier (X_train, Y_train, X_test, Y_test)
5. **Use** RandomizedSearchCV for getting the best hyperparameter estimator
6. Fit the model with X_train, Y_train
7. Now forecast the labels for the X$_test$
8. **return** forecasted values

Output:

Classification results with best hyperparameter estimator

The classification hyperparameters (i.e., min_samples_split, n_estimators, max_features, min_samples_leaf, bootstrap, max_depth) with the various selection criteria and best hyper-parameter settings used for tuning purposes have been presented in Table 5.

3.5 Statistical analysis

For the validation of the classification results, four statistical measures, i.e., accuracy, f1-score, precision, and recall have been used. These statistical measures play an essential role in establishing the accurateness and suitableness of the

Table 5 Classification's hyperparameters

Prediction model	Hyperparameter	Parameter selection	Best hyper-parameter used
Random forest	n_estimators	[100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000]	200
	min_samples_leaf	[1, 2, 4]	2
	min_samples_split	[2, 5, 10]	10
	max_depth	[10, 20, 30, 40, 50, 60, 70, 80, 90, 100, None]	100
	max_features	['auto', 'sqrt']	Auto
	Bootstrap	[True, False]	True

classification model [68]. Statistical measures with their respective mathematical formulation have been shown in Table 6.

4 Result

An accurate model identification in the IoT-enabled smart healthcare environment is among the arduous but innovative tasks. The work primarily aims to create an intelligent hybrid classification model which is proficient in dealing with the class imbalance issue with greater exactness and will play a key role in building the robotics solution for communal well-being. Results are obtained by comparison of five state-of-the-art models namely, ADB, BAG, RF, KNN, and SVM with the proposed model which is also shown in Fig. 6.

To find out the effectiveness of the recommended model, a deep assessment among the five state-of-the-art models under the various evaluation criteria (2, 3, 5, and 10-fold) has been conducted. The validation of the classification results is calculated using four performance measures (namely, accuracy, recall, f1-score, and precision).

The dataset used has two classes where class 1 represents healthy patients and class 2 represents cardiac ill patients. From the empirical evaluation, it is clear that the recommended hybrid model obtained the top accuracy throughout the experiment under various validation measures over the other well-established classification models. The statistical measures-based experimental result is shown in Table 7.

5 Discussion

For the experimental analysis, the ECG data have been used, which has been generated through the heart rate sensor (AD8232) and Node MCU (ESP32). To perform the evaluation this ECG data is been transferred from the cloud to the local machine. This dataset is classified into the two-class, where class 1 denotes healthy patients and class 2 represents the cardiac ill patient. The paper presents a comparison of five state-of-the-art models namely, ADB, BAG, RF, KNN, and SVM with the proposed model. Evaluation is performed against four statistical metrics namely, accuracy, precision, recall, and f1-score. The class-wise visualization of classification

Table 6 Statistical measures

Statistical measure	Mathematical formulation
F1-score	$f1 = \frac{2\times(precision\times recall)}{precision+recall}$
Recall	$Recall = \frac{(TP)}{(TP+FN)}$
Precision	$Precision = \frac{(TP)}{(TP+FP)}$
Accuracy	$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100\%$

FP - false positive, *TP* - true positive, *FN* - false negatives, and *TN* - true negatives

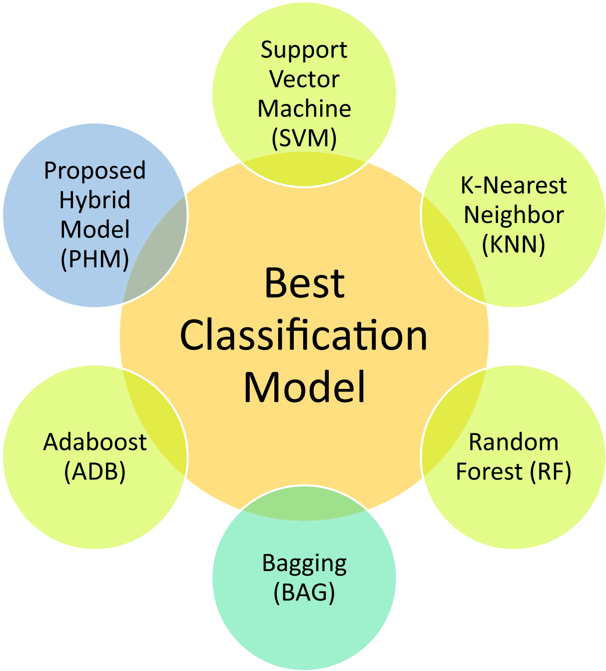


Fig. 6 Classification models a quick look

results with the help of four statistical measures under various validation criteria using cross-validation policy having 2, 3, 5, and 10-fold is shown in Figs. 7a, b, c, and 8a, b, c, respectively. Figure 9 presents the average accuracy of models during the experimental period.

The empirical evaluation shows that the recommended hybrid model is proficient to handle the complexities of class imbalance in the ECG dataset with enhanced performance for both classes, which will give support in building the IoT-enabled smart and accurate healthcare system. A comparison among state-of-the-art algorithms and recommended hybrid models has been carried out to establish the accurateness and suitableness of our recommended model. The recommended model attains the highest accuracy of 99.7% under different validation criteria among all the state-of-the-art algorithms, i.e. Adaboost (91.88%), Bagging (92.40%), random forest (92.48%), K-Nearest Neighbor (92.38%), and support vector machine (91.98%). The recommended hybrid model not only handles the complexities of class imbalance for electrocardiogram datasets but will also help in building intelligent and accurate IoT-enabled healthcare systems.

The dataset has been generated by 50 volunteer participants which are suitable for binary classification problems and are not suitable to cover all types of heart diseases (i.e. for multiclass classification problems). Therefore, in the future this work will be expanded from the data (for adding more feasible attributes) and algorithmic point of view. We will also try to make this problem a multiclass classification problem by generating data related to different types of Cardiovascular diseases.

Table 7 Statistical measures based evaluation result

Classification models		SVM		KNN		RF		BAG		ADB		Hybrid model	
		Class		Class		Class		Class		Class		Class	
		1	2	1	2	1	2	1	2	1	2	1	2
Twofold	Precision	0.92	0.89	0.93	0.89	0.93	0.91	0.93	0.9	0.92	0.87	**1**	**0.99**
	Recall	0.93	0.85	0.93	0.9	0.93	0.9	0.93	0.89	0.92	0.85	**1**	**1**
	f1-score	0.92	0.87	0.93	0.9	0.93	0.9	0.93	0.89	0.92	0.86	**1**	**1**
	Accuracy	92.0		92.4		92.5		92.4		91.8		**99.7**	
Threefold	Precision	0.92	0.89	0.93	0.89	0.93	0.91	0.93	0.91	0.92	0.88	**1**	**0.99**
	Recall	0.93	0.86	0.93	0.89	0.93	0.89	0.93	0.89	0.93	0.86	**0.99**	**1**
	f1-score	0.92	0.87	0.93	0.89	0.93	0.9	0.93	0.9	0.92	0.87	**1**	**1**
	Accuracy	92.0		92.4		92.5		92.5		92.0		**99.7**	
Fivefold	Precision	0.92	0.88	0.93	0.9	0.93	0.9	0.93	0.9	0.92	0.87	**1**	**1**
	Recall	0.93	0.85	0.93	0.9	0.93	0.89	0.93	0.88	0.92	0.84	**1**	**1**
	f1-score	0.92	0.87	0.93	0.9	0.93	0.9	0.93	0.89	0.92	0.86	**1**	**1**
	Accuracy	91.9		92.4		92.4		92.3		91.7		**99.7**	
10-fold	Precision	0.92	0.88	0.93	0.89	0.93	0.91	0.93	0.9	0.92	0.88	**1**	**1**
	Recall	0.93	0.87	0.93	0.89	0.93	0.9	0.93	0.89	0.92	0.87	**1**	**1**
	f1-score	0.92	0.87	0.93	0.89	0.93	0.9	0.93	0.9	0.92	0.88	**1**	**1**
	Accuracy	92.0		92.3		92.5		92.4		92.0		**99.7**	

The results obtained from the proposed model is presented in bold

6 Conclusion

Cardiovascular diseases (CVD) are one of the biggest hazards to human society across the globe. Hence, there is an immense requirement for real-time observation and analysis of cardiac health. Identification of the correct model in IoT-enabled smart healthcare paradigms is an arduous but innovative task. IoT-enabled intelligent healthcare systems include numerous applications like Blood Pressure (BP) check, Heart Rate (HR) monitoring, Electrocardiography (ECG) observation, etc. This paper recommends an IoT-enabled ECG monitoring system for data generation (with the help of Node MCU ESP32 and heart rate sensor AD8232) and an intelligent hybrid classification model. The key intention of this study is to give a smart hybrid classification model for dealing with class imbalance problem with greater exactness and which will play a key role in building the robotics solution for communal well-being. The dataset used has two classes where class 1 represents healthy patients and class 2 represents cardiac ill patients. A rigorous comparison based on various evaluation criteria (2, 3, 5, and 10-fold) among state-of-the-art algorithms and recommended hybrid models have been carried out to establish the accurateness and suitableness of our recommended model. The recommended model attains the highest accuracy of 99.7% throughout the experiment under different validation criteria among all the state-of-the-art algorithms, i.e. Adaboost (91.88%),

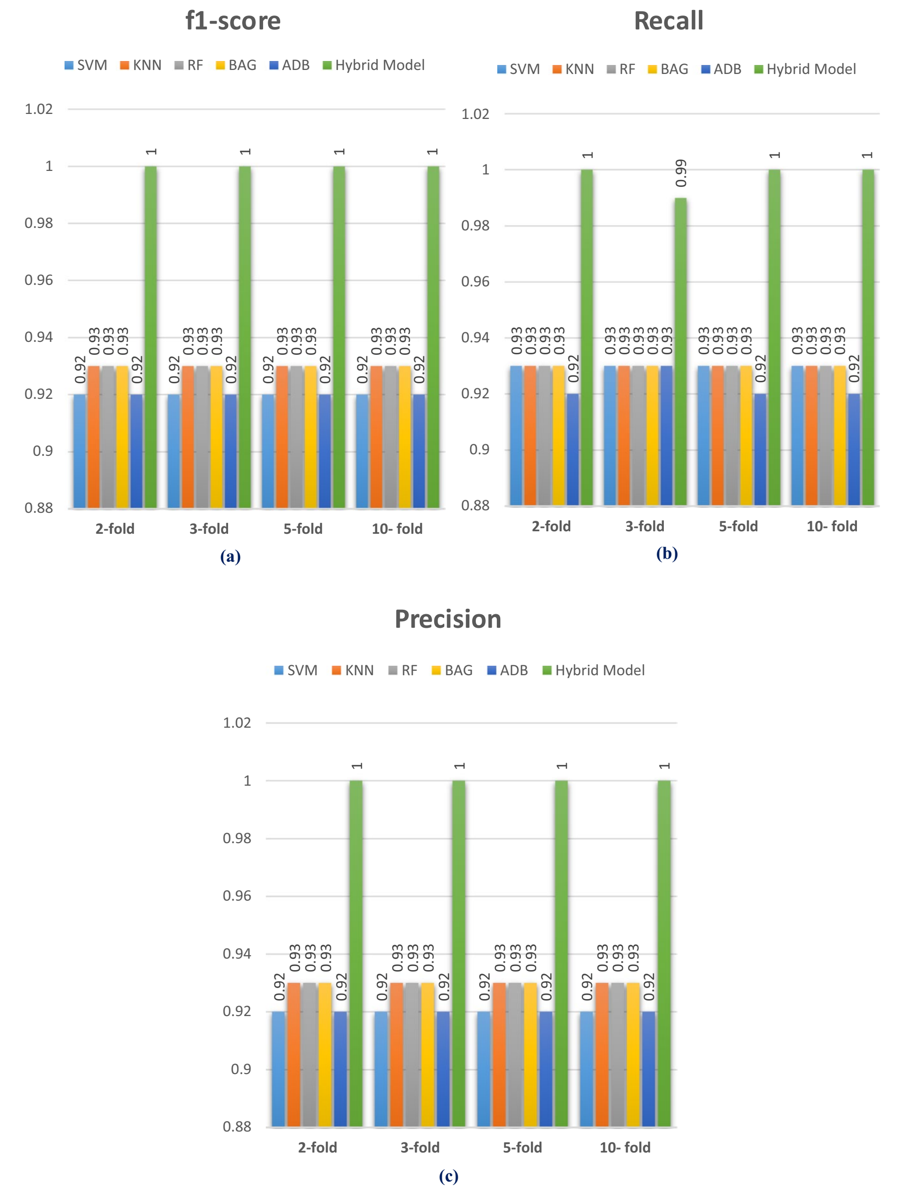


Fig. 7 Classification result of class 1 (**a**) f1-Score (**b**) Recall (**c**) Precision

Bagging (92.40%), random forest (92.48%), K-Nearest Neighbor (92.38%), and SVM (91.98%). The recommended hybrid model not only handles the complexities of class imbalance for electrocardiogram datasets but will also help in building intelligent and accurate IoT-enabled healthcare systems. Thus, accurate classification of cardiovascular health through our recommended model would be useful for

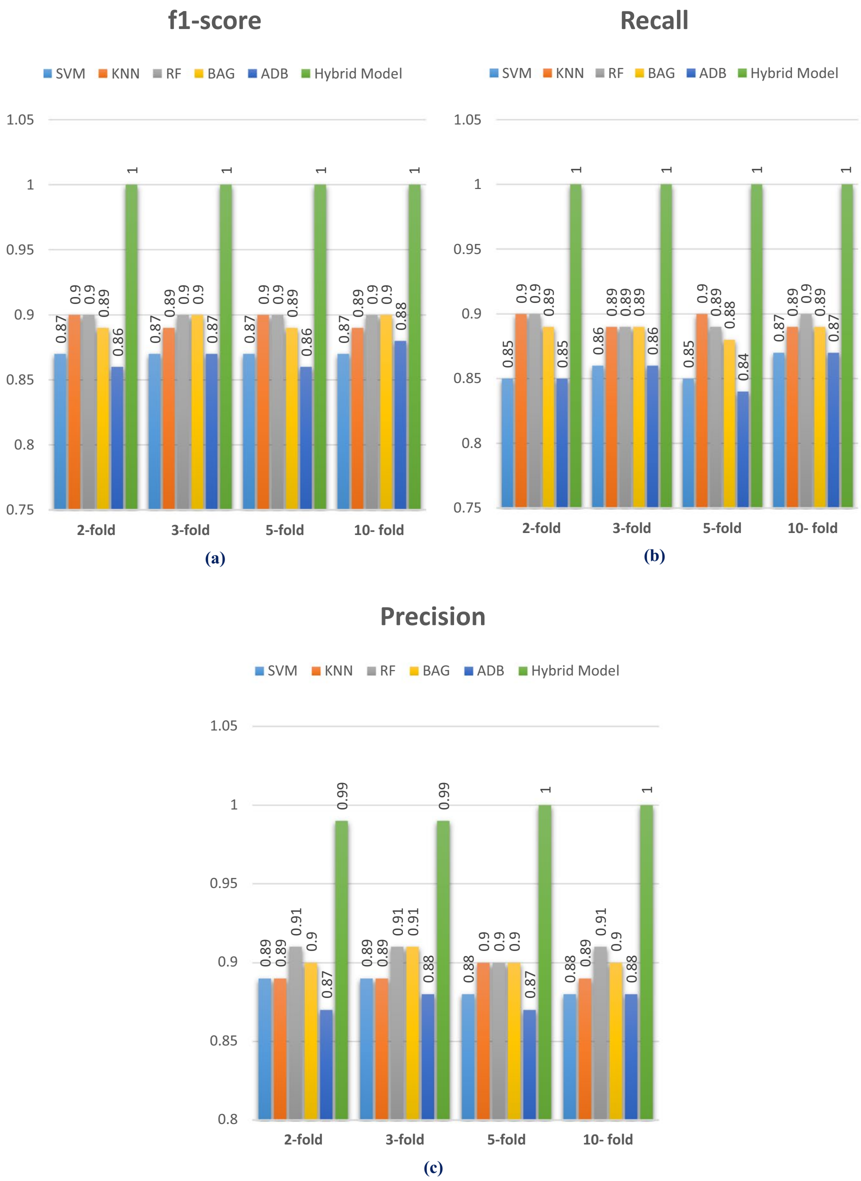


Fig. 8 Classification result of class 2 (**a**) f1-Score (**b**) Recall (**c**) Precision

improving the lifestyle of cardiac patients. This will not only allow patients to be treated from the comfort of their homes but will also reduce the need for hospital visits and reduce the overall expenditure on hospital visits. Furthermore, it would also help in enhancing the capabilities of effective emergency response to any medical emergency.

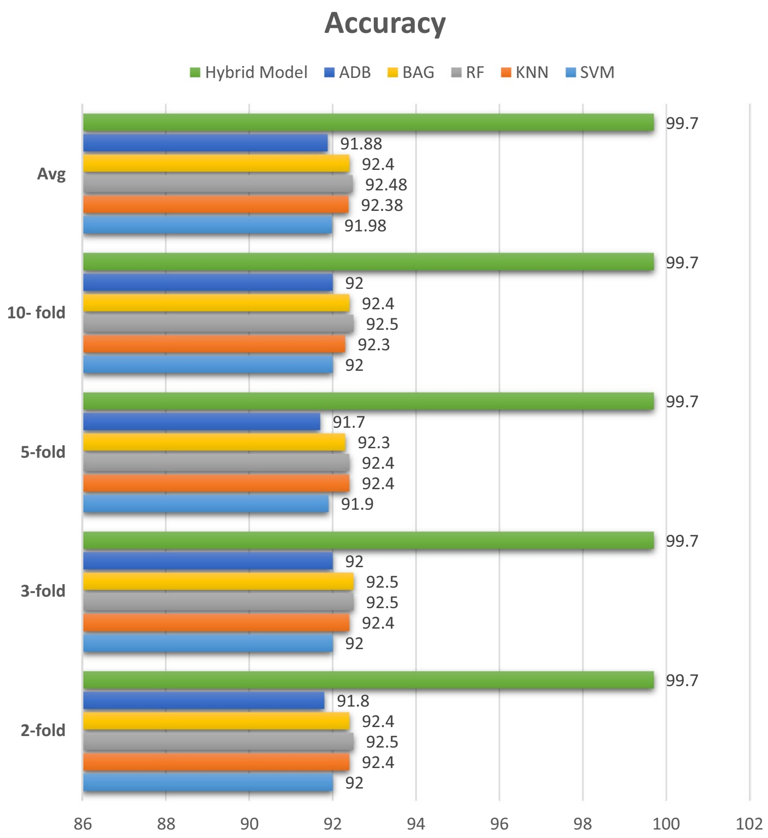


Fig. 9 Results of classification models

In the future, this work will be expanded from the data and algorithmic point of view. We will also try to make this problem a multiclass classification problem by generating data related to different types of cardiovascular diseases. Thus, we can not only detect different types of heart diseases but also classify them correctly. After this, we will try to build wearable devices in the form of a band or chest belt or undergarment which will be a complete cloud-based framework.

Author contributions S.K.: Conceptualization, methodology, writing—original draft preparation, visualization, investigation. P.K.M.: Supervision.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Cardiovascular diseases (CVDs). Available online: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed on 15 August 2020)
2. Foley RN, Parfrey PS, Sarnak MJ (1998) Clinical epidemiology of cardiovascular disease in chronic renal disease. Am J Kidney Dis 32(5):S112–S119
3. Lakka HM, Laaksonen DE, Lakka TA, Niskanen LK, Kumpusalo E, Tuomilehto J, Salonen JT (2002) The metabolic syndrome and total and cardiovascular disease mortality in middle-aged men. JAMA 288(21):2709–2716
4. Anderson KM, Odell PM, Wilson PW, Kannel WB (1991) Cardiovascular disease risk profiles. Am Heart J 121(1):293–298
5. Bonow RO, Smaha LA, Smith SC Jr, Mensah GA, Lenfant C (2002) World Heart Day 2002: the international burden of cardiovascular disease: responding to the emerging global epidemic. Circulation 106(13):1602–1605
6. Ahmed E, Yaqoob I, Gani A, Imran M, Guizani M (2016) Internet-of-things-based smart environments: state of the art, taxonomy, and open research challenges. IEEE Wirel Commun 23(5):10–16
7. Fortino G and Trunfio P (Eds) (2014) *Internet of things based on smart objects: Technology, middleware and applications*. Springer Science & Business Media.
8. Ketu S, Mishra PK (2022) Hybrid classification model for eye state detection using electroencephalogram signals. Cogn Neurodyn 16(1):73–90
9. Ketu S, Mishra PK (2021) Cloud, fog and mist computing in IoT: an indication of emerging opportunities. IETE Tech Rev 36:1–12
10. Ketu S, Mishra PK (2022) A contemporary survey on IoT based smart cities: architecture, applications, and open issues. Wirel Pers Commun 69:1–49
11. Gubbi J, Buyya R, Marusic S, Palaniswami M (2013) Internet of things (IoT): a vision, architectural elements, and future directions. Futur Gener Comput Syst 29(7):1645–1660
12. Chui KT, Alhalabi W, Pang SSH, Pablos POD, Liu RW, Zhao M (2017) Disease diagnosis in smart healthcare: innovation, technologies and applications. Sustainability 9(12):2309
13. Ketu S, and Mishra PK (2020) Performance analysis of machine learning algorithms for iot-based human activity recognition. *In Advances in electrical and computer technologies* (pp. 579–591). Springer, Singapore.
14. Datta SK, Gyrard A, Bonnet C and Boudaoud K (2015) oneM2M architecture based user centric IoT application development. In *2015 3rd International Conference on Future Internet of Things and Cloud* (pp. 100–107). IEEE.
15. Al Mamun MA, Yuce MR (2019) Sensors and systems for wearable environmental monitoring toward iot-enabled applications: a review. IEEE Sens J 19(18):7771–7788
16. Baker SB, Xiang W, Atkinson I (2017) Internet of things for smart healthcare: technologies, challenges, and opportunities. IEEE Access 5:26521–26544
17. Tian S, Yang W, Le Grange JM, Wang P, Huang W, Ye Z (2019) Smart healthcare: making medical care more intelligent. Global Health J 3(3):62–65
18. Ghazal TM, Hasan MK, Alshurideh MT, Alzoubi HM, Ahmad M, Akbar SS, Akour IA (2021) IoT for smart cities: Machine learning approaches in smart healthcare—a review. Futur Internet 13(8):218
19. Kalarthi ZM (2016) A review paper on smart health care system using internet of things. Int J Res Eng Technol 5(03):8084
20. He D, Ye R, Chan S, Guizani M, Xu Y (2018) Privacy in the internet of things for smart healthcare. IEEE Commun Mag 56(4):38–44
21. Ketu S, Mishra PK (2021) Internet of healthcare things: a contemporary survey. J Netw Comput Appl 192:103179
22. Algarni A (2019) A survey and classification of security and privacy research in smart healthcare systems. IEEE Access 7:101879–101894
23. Lee I, Lee K (2015) The internet of things (IoT): applications, investments, and challenges for enterprises. Bus Horiz 58(4):431–440
24. Ketu S, Mishra PK (2022) Empirical analysis of machine learning algorithms on imbalance electrocardiogram based arrhythmia dataset for heart disease detection. Arab J Sci Eng 47(2):1447–1469
25. Ketu S, Mishra PK (2021) Scalable kernel-based SVM classification algorithm on imbalance air quality data for proficient healthcare. Complex Intell Syst 7(5):2597–2615

26. Ketu S (2022) Spatial air quality index and air pollutant concentration prediction using linear regression based recursive feature elimination with random forest regression (RFERF): a case study in India. Nat Hazards 114(2):2109–2138
27. Ngu AH, Gutierrez M, Metsis V, Nepal S, Sheng QZ (2016) IoT middleware: a survey on issues and enabling technologies. IEEE Internet Things J 4(1):1–20
28. Pramanik MI, Lau RY, Demirkan H, Azad MAK (2017) Smart health: big data enabled health paradigm within smart cities. Expert Syst Appl 87:370–383
29. Patel KK, Patel SM (2016) Internet of things-IOT: definition, characteristics, architecture, enabling technologies, application & future challenges. Int J Eng Sci Comput 6(5):10
30. Lima LE, Kimura BYL, Rosset V (2019) Experimental environments for the internet of things: a review. IEEE Sens J 19(9):3203–3211
31. Drummond C, and Holte RC (2003). C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling. In *Workshop on learning from imbalanced datasets II* (Vol. 11, pp. 1–8). Washington DC, Citeseer.
32. Liu AC (2004) The effect of oversampling and undersampling on classifying imbalanced text datasets. The University of Texas, Austin.
33. Majumder BD, Roy JK, Padhee S (2018) Recent advances in multifunctional sensing technology on a perspective of multi-sensor system: a review. IEEE Sens J 19(4):1204–1214
34. Nag A, Mukhopadhyay SC, Kosel J (2017) Wearable flexible sensors: a review. IEEE Sens J 17(13):3949–3960
35. Belhaj Mohamed M, Meddeb-Makhlouf A, Fakhfakh A, Kanoun O (2021) Wireless body sensor networks with enhanced reliability by data aggregation based on machine learning algorithms. Adv Sens Biomed Appl 12:67–81
36. Belhaj Mohamed M, Meddeb-Makhlouf A, Fakhfakh A, Kanoun O (2022) Efficient data aggregation technique for medical wireless body sensor networks. Tm Tech Messen 89(5):328–342
37. Mohamed MB, Meddeb-Makhlouf A, Fakhfakh A, Kanoun O (2022) Secure and reliable ML-based disease detection for a medical wireless body sensor networks. Int J Biol Biomed Eng 16:196–206
38. Dolatabadi AD, Khadem SEZ, Asl BM (2017) Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM. Comput Methods Programs Biomed 138:117–126
39. Tayefi M, Tajfard M, Saffar S, Hanachi P, Amirabadizadeh AR, Esmaeily H, Ghayour-Mobarhan M (2017) hs-CRP is strongly associated with coronary heart disease (CHD): a data mining approach using decision tree algorithm. Comput Methods Programs Biomed 141:105–109
40. Arabasadi Z, Alizadehsani R, Roshanzamir M, Moosaei H, Yarifard AA (2017) Computer aided decision making for heart disease detection using hybrid neural network-genetic algorithm. Comput Methods Programs Biomed 141:19–26
41. Mustaqeem A, Anwar SM, Khan AR, Majid M (2017) A statistical analysis based recommender model for heart disease patients. Int J Med Informatics 108:134–145
42. Boon KH, Khalil-Hani M, Malarvili MB (2018) Paroxysmal atrial fibrillation prediction based on HRV analysis and non-dominated sorting genetic algorithm III. Comput Methods Programs Biomed 153:171–184
43. Mahajan R, Viangteeravat T, Akbilgic O (2017) Improved detection of congestive heart failure via probabilistic symbolic pattern recognition and heart rate variability metrics. Int J Med Inf 108:55–63
44. Bozkurt B, Germanakis I, Stylianou Y (2018) A study of time-frequency features for CNN-based automatic heart sound classification for pathology detection. Comput Biol Med 100:132–143
45. Sudarshan VK, Acharya UR, Oh SL, Adam M, Tan JH, Chua CK, San Tan R (2017) Automated diagnosis of congestive heart failure using dual tree complex wavelet transform and statistical features extracted from 2 s of ECG signals. Comput Biol Med 83:48–58
46. Aborokbah MM, Al-Mutairi S, Sangaiah AK, Samuel OW (2018) Adaptive context aware decision computing paradigm for intensive health care delivery in smart cities—a case analysis. Sustain Cities Soc 41:919–924
47. Pławiak P (2018) Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system. Exp Syst Appl 92:334–349
48. Miao F, Cai YP, Zhang YX, Fan XM, Li Y (2018) Predictive modeling of hospital mortality for patients with heart failure by using an improved random survival forest. IEEE Access 6:7244–7253
49. Tan JH, Hagiwara Y, Pang W, Lim I, Oh SL, Adam M, Acharya UR (2018) Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals. Comput Biol Med 94:19–26

50. Dominguez-Morales JP, Jimenez-Fernandez AF, Dominguez-Morales MJ, Jimenez-Moreno G (2017) Deep neural networks for the recognition and classification of heart murmurs using neuromorphic auditory sensors. IEEE Trans Biomed Circ Syst 12(1):24–34
51. Shaikh A, Al Reshan MS, Sulaiman A, Elmagzoub MA, AlYami S (2023) A fully automatic model for premature ventricular heartbeat arrhythmia classification using the internet of medical things. Biomed Signal Process Control 83:104697
52. Alfaras M, Soriano MC, Ortín S (2019) A fast machine learning model for ECG-based heartbeat classification and arrhythmia detection. Front Phys 7:103
53. Xie T, Li R, Shen S, Zhang X, Zhou B, Wang Z (2019) Intelligent analysis of premature ventricular contraction based on features and random forest. J Healthc Eng 21:5787586
54. Raj S, Ray KC (2018) Sparse representation of ECG signals for automated recognition of cardiac arrhythmias. Exp Syst Appl 105:49–64
55. Siavashi A and Majidi M (2021) Sensing, wireless transmission, and smart processing of heart signals. In *2021 5th International Conference on Internet of Things and Applications (IoT)* (pp. 1–6). IEEE.
56. Zhang H, Zhu L, Nathan V, Kuang J, Kim J, Gao JA, Olgin J (2021) Towards early detection and burden estimation of atrial fibrillation in an ambulatory free-living environment. Proc ACM Interact Mob Wearable Ubiquitous Technol 5(2):1–19
57. Kwon JM, Jo YY, Lee SY, Kang S, Lim SY, Lee MS, Kim KH (2022) Artificial intelligence-enhanced smartwatch ECG for heart failure-reduced ejection fraction detection by generating 12-lead ECG. Diagnostics 12(3):654
58. Baraeinejad B, Shayan MF, Vazifeh AR, Rashidi D, Hamedani MS, Tavolinejad H, Fakharzadeh M (2022) Design and implementation of an ultralow-power ECG patch and smart cloud-based platform. IEEE Trans Instrum Meas 71:1–11
59. Shrestha AP and Yu CH (2022) ECG data analysis with IoT and machine learning. In *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 0323–0327). IEEE.
60. Muthu Ganesh V, Nithiyanantham J (2022) Heuristic-based channel selection with enhanced deep learning for heart disease prediction under WBAN. Comput Methods Biomech Biomed Engin 25(13):1429–1448
61. Sheeba A, Padmakala S, Subasini CA, Karuppiah SP (2022) MKELM: mixed kernel extreme learning machine using BMDA optimization for web services based heart disease prediction in smart healthcare. Comput Methods Biomech Biomed Engin 25(10):1180–1194
62. Jansi Rani SV, Chandran KS, Ranganathan A, Chandrasekharan M, Janani B, Deepsheka G (2022) Smart wearable model for predicting heart disease using machine learning: wearable to predict heart risk. J Ambient Intell Hum Comput 13(9):4321–4332
63. Manimurugan S, Almutairi S, Aborokbah MM, Narmatha C, Ganesan S, Chilamkurti N, Almoamari H (2022) Two-stage classification model for the prediction of heart disease using IoMT and artificial intelligence. Sensors 22(2):476
64. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 16:321–357
65. Fernández A, Garcia S, Herrera F, Chawla NV (2018) SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. J Artif Intell Res 61:863–905
66. Biau G, Scornet E (2016) A random forest guided tour. TEST 25(2):197–227
67. Liaw A, Wiener M (2002) Classification and regression by random forest. R News 2(3):18–22
68. Goutte C and Gaussier E (2005) A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *European Conference on Information Retrieval* (pp. 345–359). Springer, Berlin.

Authors and Affiliations

Shwet Ketu[1] · **Pramod Kumar Mishra**[2]

✉ Shwet Ketu
shwetiiita@gmail.com

Pramod Kumar Mishra
mishra@bhu.ac.in

1 Department of Computer Science & Engineering, Shambhunath Institute of Engineering and Technology, Allahabad, India

2 Department of Computer Science, Institute of Science, Banaras Hindu University, Varanasi, India