

# IoT-Based Cardiac Arrest Prediction Through Heart Variability Analysis



Santosh Kumar, Usha Manasi Mohapatra, Debabrata Singh  
and Dilip Kumar Choubey

**Abstract** Current machine learning methods for sudden cardiac arrest have not been tested against physically active heart rates. Developments in wearable technology and advancements in non-intrusive heart rate monitors may allow for a future where people can stream their heart rate readings, with the readings automatically analyzed by robust machine learning algorithms which will alert cardiac arrest risk. This paper presents a new sudden cardiac arrest prediction technique, a random forest classifier implementation, a prospective physical activity heart rate dataset, and an Internet of things solution toward heart rate monitoring and sudden cardiac arrest warning. In this paper, five minutes advance warning is provided with 97.03% accuracy and a 0.9485 *F*-score for the classification of sudden cardiac arrest prediction. The result shows the efficiency of our method compared to other existing methods.

**Keywords** Heart rate variability (HRV) · Random forest · Heart rate monitoring band (HRM) · IoT

---

S. Kumar (✉) · U. M. Mohapatra · D. Singh  
Department of CSIT, Institute of Technical Education and Research S'O'A Deemed to be  
University, Bhubaneswar, Odisha, India  
e-mail: [santoshkumar@soa.ac.in](mailto:santoshkumar@soa.ac.in)

U. M. Mohapatra  
e-mail: [ushamohapatra@soa.ac.in](mailto:ushamohapatra@soa.ac.in)

D. Singh  
e-mail: [debabratasingh@soa.ac.in](mailto:debabratasingh@soa.ac.in)

D. K. Choubey  
Department of Computer Science & Engineering National Institute of Technology,  
Patna, Bihar, India  
e-mail: [dilipchoubey\\_1988@yahoo.in](mailto:dilipchoubey_1988@yahoo.in)

© Springer Nature Singapore Pte Ltd. 2020  
B. Pati et al. (eds.), *Advanced Computing and Intelligent Engineering*,  
Advances in Intelligent Systems and Computing 1089,  
[https://doi.org/10.1007/978-981-15-1483-8\\_30](https://doi.org/10.1007/978-981-15-1483-8_30)

# 1 Introduction

Researchers believe that cardiac arrest may be predicted in advance [1]. Specifically, through heart rate variability (HRV) analysis algorithms, machine learning, the Internet of things (IoT), and big data, we may be able to monitor at-risk individuals and give them advance warning (1–4 h) to get to a hospital. There is a divide between papers focused on electrocardiogram (ECG) features and *HRV* features. Murukesan et al. [2] achieve 96.36% accuracy through the use of a support vector machine (SVM) and *HRV* features. This paper uses a five minute advance warning and two minutes of sample data. This is the baseline that this paper uses. *HRV* is the preferred feature set due to allow for a broader spectrum of potential commercial wearable devices. A significant benefit of such advance warning is corroborating evidence of the need to go to the hospital rather than ignoring symptoms. Developments in wearable technology and advancements in non-intrusive heart rate monitors may allow for a future where people can stream their heart rate readings, with the readings automatically analyzed by robust machine learning algorithms which will alert cardiac arrest risk. This field is new and there is room for progress through additional studies and the development of *HRV* analysis algorithms. This is supported by the increasing number of medical records stored electronically.

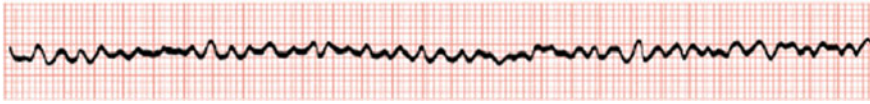
Current machine learning methods for sudden cardiac arrest have not been tested against physically active heart rates. All tested samples come from in-hospital data where the individuals are at a resting state. This paper hopes for better cardiac arrest prediction for the wearable world through *HRV* analysis and develop a solution that can be reliable with high accuracy and F-score.

## 1.1 Heart Rate Variability

Heart rate is a feature which can be derived from the *ECG* signal [3]. This is done by comparing the placement between the *RR* intervals. Each peak on the *R* wave represents a node in the heart rate array. The time between each *R* wave peak is the instantaneous heart rate. It is potentially inconvenient to spend the whole day with 12 electrodes attached to the body, and there is technology to track the heart rate without use of an *ECG* signal. This creates a preference to have a system which relies on more convenient sensors (wearable devices) in comparison to 12 *ECG* electrodes, despite the electrodes having the potential to provide more information.

## 1.2 Arrhythmias

A regular *ECG* signal is called a regular sinus rhythm. There are specific classes of *ECG* signals representative of what is going on inside the heart. There are a couple key vocabulary words that are used to identify these different classes such



**Fig. 1** ECG signal for ventricular tachyarrhythmia

as arrhythmia, tachycardia, bradycardia, atrium, and ventricles [3]. Based on these terms, we can determine an arrhythmia such as ventricular tachyarrhythmia (VT) is a rapid heartbeat that starts in the bottom chambers of the heart [3]. VT is one of the main arrhythmias associated with sudden cardiac arrest. These arrhythmias have certain classifying features. For example, VT is easily identified by the fast oscillatory waves, which represent the rapid twitching of the heart. This is shown in Fig. 1.

### 1.3 Cardiac Arrest

Cardiac arrest occurs when the beating of the heart and all electrical activity stops [3]. This means that blood stops pumping to the body, and it is especially important because brain damage can occur within ten minutes from blood loss to the brain. This is different from a heart attack, which is a physical failure in comparison to an electrical failure. Heart attacks are caused by the blockage of blood flow to the heart. Sudden cardiac arrest (SCA) is cardiac arrest that occurs unexpectedly and can result in death, called sudden cardiac death (SCD). There are three main arrhythmia classes that correlate with cardiac arrest: ventricular tachyarrhythmia (VT), ventricular asystole (VA), and pulse-less electrical activity (PEA). VT is described in the previous section and is the most common predecessor to SCA [3]. VA is represented by a flat line. This is shown in Fig. 2.

### 1.4 Machine Learning

In the history of SCA prediction, various machine learning algorithms are in use. This section highlights the four machine learning classification algorithms explored in this paper. In this case, datasets are broken up into a healthy (non-SCA) class and a SCA



**Fig. 2** ECG signal for ventricular asystole

warning class (five minutes until cardiac arrest) to create a two-class classification problem. Kotsiantis et al. [4] provide a strong review of many classification algorithms. Much of the early work in ECG classification is done with neural networks [5].

**Classification:** Classification is a machine learning problem where a data sample can belong to one of two or more classes [4]. While classification algorithms can deal with more than two classes, this paper focuses on a binary classification problem where the two possible classes are either normal sinus rhythms or the onset of sudden cardiac arrest. Machine learning classifiers use the feature vectors derived from the data samples to learn how these features help identify what class future data samples belong to.

**Support Vector Machine:** A *SVM* is a machine learning algorithm for classification proposed by Cortes and Vapnik in 1995 [6]. The main idea of *SVMs* is to learn a nonlinear function by a combination of linear mappings in high-dimensional feature space. The desired function would be able to map all training data within a certain margin of error. Support vectors are especially good in high-dimensionality spaces because *SVMs* do not depend upon the dimensionality of the input space.

**Decision Tree:** Decision trees build a ruleset to classify input data [7]. Each node in the tree that is built represents a decision. As the input data traverses through this tree, it will eventually arrive to a leaf node which dictates the class of the input data. The trees built have logarithmic run times due to the structure. Unfortunately, the ruleset is prone to overfitting for which there are a number of techniques to avoid.

**Naive Bayes:** Naive Bayes is a classification algorithm based on the Bayes theorem. This classifier uses the theorem to derive the probability that a given feature vector is associated with a specific class. The algorithm naively assumes that there is independence between every pair of features [4, 7]. This assumption creates a weakness in the algorithm, because there will almost never be an independence between every pair of features given enough features. Regardless, the classification algorithm is proven to be strong in smaller training sets.

**Random Forest:** This paper utilizes the random forest algorithm demonstrated by Breiman [8]. The random forests used as classification algorithms are common in machine learning applications, because of providing good results, the ease of use, and the scalability. In many implementations, the single parameter which requires tuning is the number of trees used to build the forest, which usually denotes the trade-off between computational cost and the performance. The depth of individual trees is of less importance, as with big forests the variance of the model imposed by the deep trees is reduced, therefore, the trees do not require pruning.

## 2 Proposed Method

With the increased interest in wearable technology, there will be a market for *HRV* monitoring [9]. With the large amount of cardiac arrests that occur outside of the hospital, constant *HRV* monitoring may be able to prevent a large amount of deaths.

This constant *HRV* monitoring will have less data than typical in-hospital data and will have to be mostly reliant on heart rate data as well as some patient-level data. There has been success in predicting cardiac arrest with in-hospital data, but the potential for success may exist with limited data with physical activity as well. Success, in this case, is defined as a system that is more accurate, in terms of accuracy and *F*-score.

### 2.1 Proposed System

Figure 3 details an outline of the proposed system, similar to the system that builds the physical activity dataset. The wearable device collects heart rate information using its heart rate monitor. This example uses the specific heart rate monitoring band (*HRM*), but the system may be built for any wearable device that has a heart rate monitor or electrodes for *ECG* recording. That data is sent to a smartphone via Bluetooth. An application on the phone holds this data in two-minute intervals. Data may be sent to the phone on each beat or all at once. At the end of each two minute interval, the sample is sent via *HTTP* request to a server hosting the Python machine learning code. Also, on this server is the dataset of labeled heart rate samples, which has the opportunity to grow with each new input sample. The running code will classify the sample and send back a response. If the sample is classified as the onset of sudden cardiac arrest, then the phone and wearable device’s notification system are used to notify the user. Otherwise, the phone sends no warning notification.

The *HRM* is a commercial product available to the public. It features a heart rate monitor, skin temperature sensor, *GPS* tracker, and other sensors. It is visible in Fig. 4. The band communicates via Bluetooth to a synchronized smartphone.

### 2.2 Datasets

The datasets are a prospective physical activity dataset collected via the *HRM*. Prior to this work, there is only one dataset that provides the heart rates of individuals

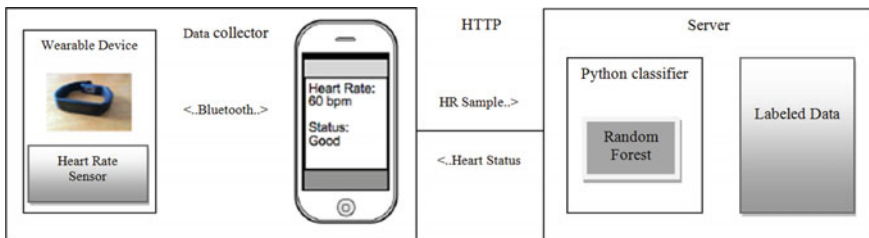


Fig. 3 Proposed system

**Fig. 4** Heart rate monitor band



performing physical activities called ‘PAMAP2’ dataset [10]. The dataset is sampled using an *HR* monitor with a frequency at 9Hz collected in 2012. The dataset is comprised of the heart rates of nine individuals performing 18 different physical activities. The dataset also holds other biometric values such as body mass index, acceleration data, temperature, gyroscopic data, and orientation.

This paper introduces the use of a prospective study that may be representative of future heart rates of people using a commercially available wearable device. The prospective physical activity dataset in this paper is collected from a *HRM*. This dataset features the two-minute samples of the heart rates of five individuals performing ten unique activities. A table is provided with the counts as well in Table 1. There are a total of 105 samples in this dataset.

**Table 1** Physical heart rate activity dataset distribution

Activity	Sample count
Sitting	4
Walking	10
Running	18
Computer work	13
Stretching	3
Car driving	9
Video games	22
Lying	6
Eating	8
Post workout	12
Total	105

### 2.3 Features

The independent variables for this paper are based on common features that act as indicators for cardiac arrest and arrhythmia classes [11, 12]. This includes common *HRV* features such as mean *HR*, minimum *HR*, maximum *HR*, standard deviation of the *HRs*, standard deviation of the beats per minute, median of the *HR*, root mean square of the standard deviation of the *HR*, and the number of outliers. These features are shown in Table 2.

### 2.4 Labels

The dependent variable in this experiment is the class of arrhythmia. This experiment tests using binary classification. The person may have a heart rate that is normal or a heart rate that is exhibiting an onset of sudden cardiac arrest. Being able to classify the rhythm is important so that individuals may take certain precautions or be able to detect and treat specific symptoms.

### 2.5 Accuracy

All data comes labeled regarding the class of the heart rates as well as the time until cardiac arrest [3, 11]. The output of the machine learning algorithms is directly compared to these classes and is judged as wrong or right on a binary scale. Accuracy is used to determine the proficiency of the algorithm. The accuracy of these classifiers is compared against baselines such as the one provided by Murukesan et al. [2]. In all cases, at least five runs were performed to collect an average accuracy.

**Precision, Recall, and *F*-score** Oftentimes, it has been determined that unbalanced datasets such as this one cannot have reliable classification results with accuracy

**Table 2** Features extracted by the system

Number	Features
1	Mean HR
2	Minimum instantaneous HR
3	Maximum instantaneous HR
4	Standard deviation of the HRs
5	Standard deviation of the beats per minute
6	Median of the HRs
7	Root mean square of the standard deviation of the HRs
8	The number of outliers (greater than 50 ms difference)

alone [7]. In this case, a classifier has the potential to nearly always guess that the sample is a normal sinus rhythm, because the number of normal sinus rhythms is twice that of the sudden cardiac arrest onsets. Because of this, the random forest algorithm is also evaluated in terms of precision, recall, and  $F$ -score.

$$\text{Precision} = \frac{\text{No. of true positives}}{\text{No. of true positives} + \text{No. of false positives}} \quad (1)$$

$$\text{Recall} = \frac{\text{No. of true positives}}{\text{No. of true positives} + \text{No. of false negatives}} \quad (2)$$

$$F\text{-Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

## 2.6 Experimental Protocol

The empirical study is broken down into multiple parts. First, data is acquired from the *HRM*. Second, multiple machine learning classifiers are randomly trained upon a 70–30% of the total dataset. Third, the rest of the dataset is used to test against these classifiers and provide an accuracy to demonstrate the proficiency of the algorithm. The first step is the collection and normalization of all data. This creates 105 samples that are provided from the physical database collected via *HRM*. The machine learning algorithms that are compared include a support vector machine with a linear kernel, decision trees, naive Bayes, and random forest. Comparisons are done in terms of accuracy.

## 3 Results and Discussion

### 3.1 Accuracy of Random Forest Versus SVM

The first tables of this section compare the random forest classifier accuracy to the *SVM* classifier accuracy. Table 3 compare the random forest classifier accuracy to the *SVM* classifier accuracy based on the data collected by *HRM* band. The result shows the efficiency of random forest classifier in terms of accuracy over *SVM*.

The rest of the section is devoted to the individual runs of each classifier at 70–30 training-testing percentages. Table 4 shows the accuracies with the physical activity dataset. This table illustrates the high accuracy of random forest over the three other algorithms.



**Table 3** SVM versus random forest classification results

Run 70–30%	SVM accuracy	Random forest accuracy
0	0.9275	0.9824
1	0.9384	0.9602
2	0.9513	0.9527
3	0.9625	0.9824
4	0.9554	0.9756
Average	0.9470	0.9703

**Table 4** Classification results with physical activities

70–30%	SVM	Naive Bayes	Decision tree	Random forest
0	0.8624	0.8139	0.8755	0.9521
1	0.8217	0.8356	0.8692	0.9478
2	0.8448	0.8297	0.8867	0.9502
3	0.8768	0.8215	0.8893	0.9461
4	0.8309	0.8344	0.8972	0.9521
Average	0.8473	0.8270	0.8835	0.9496

### 3.2 Precision, Recall, and F-Score Evaluation

Table 5 details the precision, recall, and *F*-score of the random forest classifier on training-testing dataset splits. These values are the average of five runs each.

The values of *F*-score are detailed in Table 6 and Fig. 5. Similar to the accuracies, the results for random forest are higher than any other classifier at training-testing split.

**Table 5** Random forest accuracy, precision, recall, and *F*-score results

Algorithms	Training-testing split
Accuracy	0.9484
Precision	0.9702
Recall	0.8674
F-score	0.9485

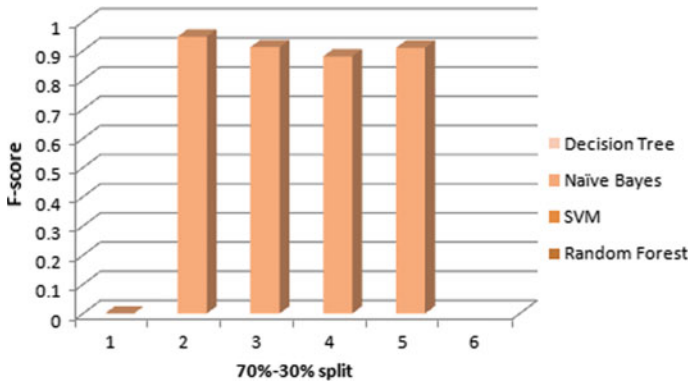


Fig. 5 F-score of each classifier

Table 6 F-score results

Algorithms	Training-testing split
Random forest	0.9485
SVM	0.9124
Naive Bayes	0.8809
Decision tree	0.9102

## 4 Conclusion

This paper presents a new sudden cardiac arrest prediction technique, a random forest classifier implementation for multiple weak learners, a prospective physical activity heart rate dataset, an *IoT* solution toward heart rate monitoring, and sudden cardiac arrest warning. Using a 70–30% training-testing split, the work in this paper is able to achieve 97.03% accuracy with a 0.9485 *F*-score for the classification of sudden cardiac arrest prediction. Comparably, Murukesan et al. [2] achieve 96.36% accuracy, but their approach uses a smaller dataset and does not include physical activity heart rates. This paper’s approach uses *HRV* derived features on two-minute samples of heart rate data. This paper proves that it is possible to classify resting as well as active heart rates against the heart rates of people about to go into sudden cardiac arrest. The heart rate data from a wearable device proves that there is a future for sudden cardiac arrest prediction through wearable devices.

## References

1. Ebrahimzadeh, E., Pooyan, M., Bijar, A.: A novel approach to predict sudden cardiac death (SCD) using nonlinear and time-frequency analyses from HRV signals. *PLoS ONE* 9(2) (2014)

2. L. Murukesan, Murugappan, M., Iqbal, M.: Sudden cardiac death prediction using ECG signal derivative (heart rate variability): a review. In: 2013 IEEE 9th International Colloquium on Signal Processing and its Applications (CSPA). IEEE, New York (2013)
3. Goldberger, A.L., Goldberger, Z.D., Shvilkin, A.: Clinical Electrocardiography: A Simplified Approach E-Book. Elsevier Health Sciences (2017)
4. Kotsiantis, S.B., Zaharakis, I.D., Pintelas, P.E.: Machine learning: a review of classification and combining techniques. *Artif. Intell. Rev.* **26**(3), 159–190 (2006)
5. Dokur, Z., Imez, T.: ECG beat classification by a novel hybrid neural network. *Comput. Methods Progr. Biomed.* **66**(2–3), 167–181 (2001)
6. Cortes, C., Vapnik, V.: Support-Vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)
7. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.*, 2825–2830 (2011 Oct 12)
8. Breiman, L.: Random forests. *Mach. Learn.* **45**(1), 5–32 (2001)
9. Jin, Z. Sun, Y., Cheng, A.C.: Predicting cardiovascular disease from real-time electrocardiographic monitoring: an adaptive machine learning approach on a cell phone. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009. EMBC 2009
10. Reiss, A., Stricker, D.: Introducing a new benchmarked dataset for activity monitoring. In: 2012 16th International Symposium on Wearable Computers (ISWC). IEEE, New York (2012)
11. Lichman, M.: UCI Machine Learning Repository (2013)
12. Joo, S., Choi, K.-J., Huh, S.-J.: Prediction of ventricular tachycardia by a neural network using parameters of heart rate variability. *Comput. Cardiol.* IEEE, New York (2010)