International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

Classification and Identification of Cardiac Arrhythmia

Annapoorna M S[1], Chaitra Gajanan Bhat[2], Chaitra N[3], Joshna P N[4], Sudarsanan D[5]

[1-4]UG student, [5]Assistant Professor, Dept. of Information Science and Engineering, Sapthagiri College of Engineering, Bengaluru, Karnataka, India

---

Abstract: *Cardiac arrhythmia is a heart condition with an irregular heartbeat that can be too fast, too slow or unstable. Many machine learning algorithms have successfully applied to classifying different types of heart arrhythmia. Using ECG data set, design of an intelligent system to identify cardiac arrhythmias is considered. This is proposed to name the arrhythmia recognition method that involves phases of data preprocessing, extraction and classification of features. In the paper, to reduce the size of input space feature extraction methods like principle component analysis (PCA) and random forest algorithms are used. The concept behind this analysis is to figure out a reduced space for the function to work correctly by the classifier. After the extraction of features, K-Nearest Neighbor (KNN), Weighted KNN, Support Vector Machine (SVM), Logistic Regression and Naive Bayes Classifier are used for classification purposes. The proposed algorithms were applied and estimated using the UCI ECG dataset. The proposed approach has provided an attractive accuracy of classification.*

Keywords: Cardiac Arrhythmia, UCI ECG dataset, Data Preprocessing, Feature Extraction, Classification.

1. INTRODUCTION

Over the years, clinical biomedical datasets are being accumulated. Advances in methods for collecting and storing data. The Clinical History studies have demonstrated the recognition of the disease and its probable results by analyzing the health records electronically. Such reports contain useful information on the major risk factors and other details associated to the causation of disease [5,6]. Hardware Development and Computer technology has helped to boost work in biomedicine and lots of clinical studies. Applications of specific machine intelligences and algorithms of machine learning on the dataset for cardiac arrhythmias will help precise detection, outlook and medical decision taking. Cardiac arrhythmia or irregular heartbeat are the circumstances in which the heartbeat occurs. Tachycardia where the beat of the heart is too slow and the heart rate exceeds 100 beats per minute. Bardycardia where the beat of the heart is too high and rate is below 60 beats per minute. Statistics indicate that large numbers of people suffer from stroke, cardiac arrest or many heart similar disorders. Tachycardia and bradycardia Rae cardiovascular diseases, which are one amongst Machine learning is a methodology that can discover previously unknown regularities and patterns from different datasets, in the expectation that machines will contribute to the sometimes repetitive and error-prone cycle of information acquisition from empirical evidence, and help clarify and codify knowledge. In this work, we present a technique for feature extraction which we call the Principal Component Analysis and the Random Forest technique. The main intention of this work is to make cardiac predictions more accurate. In addition, five classification methods, namely linear SVM, KNN, Weighted KNN, Logistic Regression and Naive Bayes classifier, are used for evaluating the output of the feature selection methods.

2. LITERATURE SURVEY

The motivation behind the paper [1], is to specialize in the analysis of ECG signals. The authors have used various schemes of information mining like naive Bayes, J48, and OneR techniques to define ECG medical database arrhythmia. The research done aimed at classifying the arrhythmias and evaluating the efficiency of various machine learning methods. Naive Bayes and OneR have produced the most robust precision. Demerit: J48 algorithm couldn't perform efficiently.

The paper [2]described about the Modular Neural Network (MNN) model, used to differentiate arrhythmia from natural and anomalous groups. The model was built by variating the number of invisible layers from 1 to 3 and using six different measures for efficiency checking.: The results showed the accuracy of the test classification up to 82.22 %.

A prediction system [3] using clustering and regression was developed by Prathibhamol Cp et al. The dataset is split into disassociate clusters by using the clustering method known as Density-Based Spatial Clustering of Applications with Noise (DBSCAN). From this the lesser-instance clusters are considered for training. Then there was the implementation of multiclass logistic regression. Demerit: The accuracy clad to be 80%.

The paper [4] aims to use various algorithms for machine learning, such as Naive Bayes, SVM, Random Forests and Neural Networks to predict and classify arrhythmia into different categories. Experimented with two separate selection strategies for the filter functionality. Demerit: The accuracy is only 77.4%.

3. PROPOSED SYSTEM

The proposed system classifies the arrhythmia using the ECG and human health parameters dataset taken from the UCI repository. The proposed model first selects the foremost distinctive features using an enhanced feature selection method**.** Classification process is performed after selecting the important features.

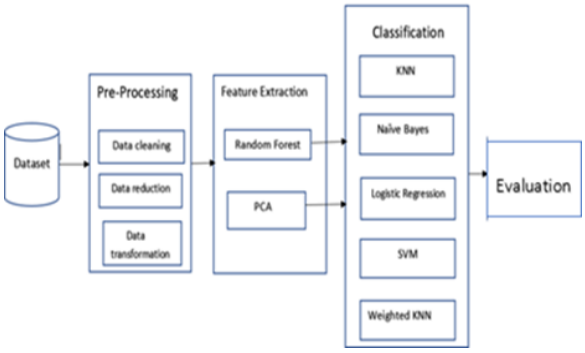

Fig. 1: Proposed system Architecture

3.1 Data Preprocessing

It involves handling of missing data, categorical data and dividing the dataset into the dataset for training and testing. The data set has been obtained from the UCI-machine learning repository which contains patient ECG data. Every record carries 279 characteristics.

3.2 Feature Extraction

It was infeasible to remove all the features from the dataset out of the 279 features available. That is because many features used some knowledge which is not available to the physicians when reviewing patient ECG data. Therefore, with the aid of Principal Component Analysis (PCA) and Random Forest algorithm, the dataset was narrowed.

I) Principal Component Analysis (PCA)

The main component analysis is a method of removing variables that has the greatest effect on the final decision to make as much information as possible available. Reporting can be achieved by low-dimensional substance representation in addition to simplicity and efficiency compression. It finds its application in the identification of statistical patterns in image processing, in addition to that dimensionally. PCA is user friendly and quick. It also applies to raising the number of characters obtained by extracting an iris image from a person. Within this paper, the objective of a PCA is to minimize the dataset that contains large quantities of dimensions and to consider small size features. The main component in a dataset is a combination of the initial normalized linear predictors.

II) Random Forest Algorithm

Random forest could be a supervised learning technique employed for both classification yet as regression. It gives outstanding performance on variety of experimental problems, because it's not sensitive to noise within the dataset, and it is not subjected to overfitting. This technique construct decision tree on data samples to urge prediction from each of them and by using voting it'll selects the simplified solution.

Step 1: From the given dataset select random samples.

Step 2: Create a option tree for each sample, then from each decision tree get the predicted result.

Step 3: carry out voting on each predicted outcome.

Step 4: Lastly, select the outcome of the pre-eminent vote prediction.

3.3 Classification Methods

As already mentioned, SVM, Naïve Bayes, KNN, Weighted KNN and logistic regression are the algorithms chosen to classify cardiac arrhythmia data.

I) Support Vector Machine (SVM)

Support vector machines are used for nonlinear regression and pattern classification. It's a mathematically rigorous, machine learning technique to make a linear binary classifier. It creates a hyperplane during a high dimensional space which will accurately slice a dataset into two segments in step with the required objectives. The algorithms for developing the classifier is mathematically challenging though. SVM's are popular since they're state-of-the-art for several practical

problems, like identifying spam emails and other text mining applications.

II) Naive Bayes Classifier

Naive-bayes (NB) technique is a supervised learning technique that uses probability-theory-based analysis. It is a machine-learning technique that computes the probabilities of an instance belonging to each one of many target classes,, given he prior probabilities of classification using individual factors. Naïve bayes technique is used often in classifying text documents into one of multiple predefined categories.

III) K- Nearest Neighbor (KNN)

K-nearest neighbor (KNN) algorithm may be a style of supervised ML algorithm which will be used both for classification and for predictive regression problems. It uses 'function similarly' to predict new datapoint values which further means a price are going to be assigned to the new datum, backed by how closely it fits the points inside the training set.

IV) Weighted KNN

Weighted KNN is a weighted variant of neighbors nearest to K. It forecasts the class of an object focus on two or more indicator statistical variables. In this technique, a weight is given for the nearest k points using a function called the kernel function. The function behind weighted knn is to give more weight to the nearby points, and less weight to the farther points. Any function can be used for the weighted knn classifier as a kernel function, whose value decreases as the distance increases. Inverse distance unction is used as simple function.

V) Logistic Regression

Logistic regression (LR) can be a classification technique, accustomed predicts the likelihood of a target variable, the character of the variable quantity is split, implying that only two classes can be achieved. The variable quantity means it's a binary in nature with data coded as either 1 (which means yes) or 0 (which means no). LR is a most effective Machine learning algorithms used for various classification problems. It is best suited to binary classification of knowledge set with categorical dependent.

4. RESULTS

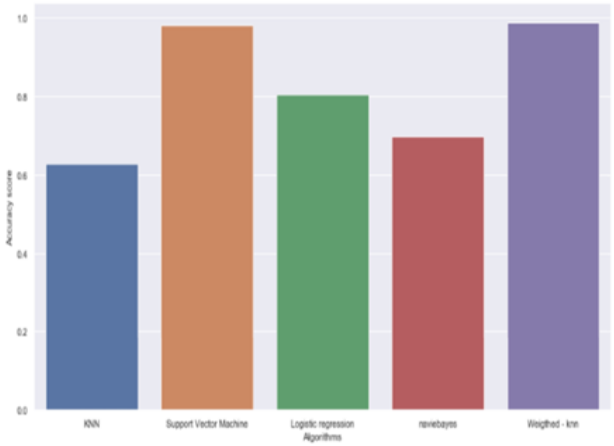

Fig 2: Accuracy Score of classifier with random forest.

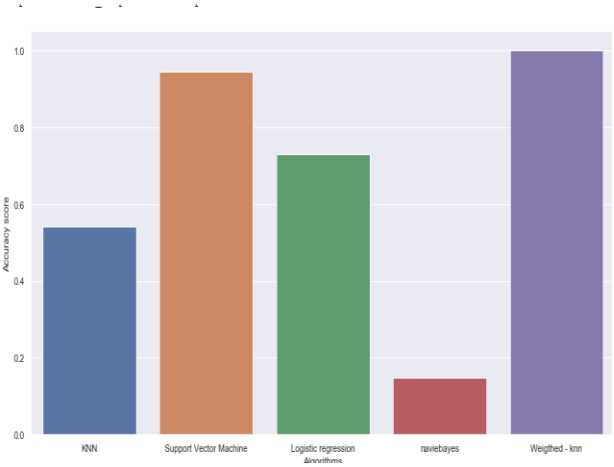

Fig. 3: Accuracy score of classifiers with PCA

Fig 2 and Fig 3 illustrates the accuracy score of each classifier with PCA and Random Forest.

In the proposed system we've got used five different algorithms with feature extraction methods to test which algorithm will give more accuracy so it'd help to detect arrhythmia at early stage.

5. CONCLUSION

For the diagnosis of cardiac arrhythmia we used 5 algorithms. Those were Naïve bayes, Support Vector Machine, Weighted Knn, K- nearest neighbor and Logistic Regression. Upon cross-validation and checking of the dataset, weighted Knn classifier finds the highest accuracy. The accuracy obtained was 98.9 percent. So we used the

weighted Knn classifier in our approach to achieve the best possible outcomes for arrhythmia diagnosis.

6. FUTURE WORK

The results suggest that the machine learning methods can help in detecting heart arrhythmia. It is hoped that further exploration of the data will follow more interesting results. Future work involves repeating the experiment with other algorithms of the machine learning.

REFERENCES

[1] Soman T, Bobbie P. O: "classification of arrhythmia using machine learning techniques". WSEAS Trans. Comput.4, 548-552 (2005).

[2] Jadhav S. M, Nalbalwar S. L, Ghatol A. A: "ECG arrhythmia classification using modulsrneural network model". In: 2010 IEEE EMBS conference on Biomedical Engineering and Science (IECBES), pp, 62-66 (2010).

[3] Prathibhamol Cp, Anjana Suresh, Gopika Suresh. "Prediction of cardiac arrhythmia type using clustering and regression approach (P-CA-CRA)".2017 International Conference on Advances in computing, Communications and Informatics (ICACCI).

[4] Gupta V, Jawa V: "Classification of arrhythmia using conjuction of machine learning algorithms and ECG diagnosis criteria". Int. J. Biol. Biomed.I, 1-7 (1026).

[5] Esteban S., Tablado M. R., Peper F. E., Mahumud Y. S., Ricci R. I., Kopitowski K. S., Terrasa S. A.,:" Development and validation of various phenotyping algorithms for diabetes mellitus using data from electronic health records". Comput. Methods programs Biomed. 152, 53-70 (2017).

[6] Wang Y., Li P. –F., Tian Y., Ren J. –J., Li J. –S.,:" A shared decision-making system for diabetes medication choice utilizing electronic health record data". IEEE J. Biomed. Heal. Informatics.21, 1280-1287 (2017).

[7] Rashad Ahmed, Samer Arafat.: "Cardiac Arrhythmia Classification Using Hierarchical Classification Model". 2014 6th International Conference on CSIT 987-1-4799-3999-2.