International Conference on Emerging Trends in Engineering, Science and Technology (ICETEST - 2015)

Evaluation of Classification Techniques for Arrhythmia Screening of Astronauts

Deepthi Sa, Aswathy Ravikumar, R. Vikraman Nair*

Dept. of Computer Science
MBCET
Thiruvananthapuram, India
deeptisdeepam@gmail.com, aswathy_2290@yahoo.co.in, rvnair1951@gmail.com

Abstract

Arrhythmia is the major cause of cardiovascular events during space flight. Even though a number of physical tests are conducted to diagnose the disease, in most of the cases the issue remains undetected because of the hidden problems which cannot be pinpointed with regular physical tests. A computation system which can assist in analyzing hidden patterns of physical test is proposed which makes use of data mining and machine learning as the underlying approaches. The present study attempts to evaluate the performance of different individual classifiers such as Naïve Bayes, Support Vector machine (SVM), Classification and Regression Tree (CART), Linear Discriminant Analysis (LDA) and k-nearest neighbor (k-NN). Then the performance of these classifiers is compared with different ensemble techniques such as Majority Voting, Bagging, Dagging and DECORATE (Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples). The performance of the proposed classification methods is analyzed by considering different criteria such as accuracy, sensitivity and specificity. The result shows that among the individual classifiers implemented, k-nearest neighbor is having highest accuracy of around 84.44% only. But Majority Voting, which is an ensemble technique, is having the highest accuracy of 91.11% which is better than the individual classifier.

Keywords: Majority Voting; Bagging; Dagging; DECORATE; Support Vector Machine; K-Nearest Neighbor; Classification and Regression Tree; Discriminant Analysis; Naïve Bayes

*Corresponding author. Tel.:+918281922084.
E-mail address: deepthisdeepam@gmail.com
1. Introduction

Human Space Flight (HSF) aims at space exploration by man. It mainly focuses on exploring and discovering the new world by a crew of astronauts and then returning them securely to a predefined position on ground [1]. Planning phase of the mission is more crucial since the conditions in space are different from those on earth. Astronaut selection is done with utmost care since he/she should be capable of withstanding such conditions on space. Major issues like cardiac problems, bone density decrease, high radiation problems, vision problems and temperature and pressure variations might be faced by astronaut. This paper focuses mainly on the cardiovascular problems and how to reduce the occurrence of cardiovascular problems by proper screening during astronaut selection.

During a long exposure to zero gravity, astronaut faces many cardiovascular problems. Cardiac problems can be categorized into pre-existing cardiac problems, problems that arise during space flight and difficulties faced after space flight because of a long duration mission [2]. If a person with pre-existing problems is identified properly during the selection process itself, either he/she can be eliminated from selection or efficient treatment can be provided to make that person suitable for spaceflight, so that the chance of occurrence of cardiac problems during flight can be avoided to a greater extent.

The main type of cardiac problem during space flight is Arrhythmia. It is an abnormality in heart rhythm i.e. cardiac rhythm can be too fast or too slow. Arrhythmia can be detected by analyzing the Electro Cardio Gram (ECG) signal by considering the QT interval, T interval, heart beat or by Tilt test and Electrophysiology studies. During astronaut selection, these tests are conducted. Still the result might be False Positive (FP) i.e. actually the astronaut might not be fit for flight, but the result says that he is fit. False Positive cases may lead to improper astronaut selection, which increase the probability of cardiac problems during space flight. False Positive result occurs due to the hidden problem that remains undetected with normal physical tests. The methodology to analyze such hidden problems is addressed in the present study. The goal is to arrive at the classification technique yielding maximum accuracy in this domain.

2. Classification Methods

Machine learning and data mining plays an important role in developing the system for astronaut screening for space flight by considering the Arrhythmia disease. Algorithms available in data mining can be categorized mainly into two broad areas namely basic classifiers and ensemble classifiers. Ensemble classifier is obtained by applying some operations on the basic classifier so that the generalization capability of the system gets enhanced. The performance comparison of these basic classifiers with ensemble classifiers is discussed in section 4.

2.1. Basic Classifiers

A number of classifiers are available for the purpose of classification. Classifier selection is based on its performance. Here five different basic classifiers are chosen for the purpose of implementation and comparison. The chosen classifiers are k-nearest neighbor (k-nn), Support Vector Machine (SVM), Decision tree, Discriminant Analysis and Naïve Bayes. Brief review of the classifiers is presented in this section.

2.1.1. K-Nearest Neighbour

The working of k-nn is based on subspace method. It begins with plotting the data points into the provided space. Plotting is done by considering the class labels. K-nn classifier classifies each unlabeled observation by taking majority voting among ‘k’ nearest neighbors. Most commonly used distance metric is Euclidean distance [7]. Equation (1) shows how the computation is done using Euclidean distance.

\[
D(x, y) = \sum_{i=0}^{n} \sqrt{x_i^2 - y_i^2}
\]
In (1), $D(x,y)$ represents the distance between two selected input vectors. $x_i$ and $y_i$ represents the data point contained in the vector space and new data point to be classified respectively.

2.1.2. Support Vector Machine

Support Vector Machine is one of the most popular supervised algorithms used for binary classification. SVM tries to discriminate objects belonging to different classes [8]. Different kernel functions used are linear, polynomial, radial Basis and quadratic.

2.1.3. Naïve Bayes

Naïve Bayes makes use of so called Bayesian theorem as the underlying technique. It applies conditional independence assumption between the features. Let $(x_1, x_2, x_3, \ldots, x_n)$ represent the vector to be classified. It is represented as ‘n’ dimensional space, where $n$ represents the number of features considered [9]. For each class $c$, it assigns probabilities as $p(C_k | (x_1, x_2, x_3, \ldots, x_n))$, which is represented in (2).

$$p(C_k | x) = \frac{p(C_k)p(x|C_k)}{p(x)} \tag{2}$$

where ‘k’ represents the number of classes considered for classification. Equation (2) represents the posterior probability for each class [10]. The main advantage of Naïve Bayes algorithm is its simplicity and it works well when the data are of higher dimensional space.

2.1.4. Decision Tree

Decision tree is a machine learning techniques, which employs binary tree as the underlying logic. Basic working principle behind decision tree is the splitting of dataset into two subsets so that each set contains similar features. Choosing the best splitting criteria is the most important step in decision tree processing [11].

Decision tree is composed of decision nodes and branches. The topmost decision node is termed as root node and the terminal node without any outgoing branches is called the leaf node. The nodes are connected by a structure called branches. Leaf node indicates the possible outcomes or classes. Decisions are represented in each decision node [12].

Different types of decision trees are ID3, C4.5 and CART (Classification and Regression Tree). All these vary only in their splitting criteria. ID3 and C4.5 make use of entropy and information gain as the splitting criteria. Entropy is the measure of uncertainty in the dataset. As entropy is getting higher, there is potential to improve the
classification higher. CART uses GINI index as the splitting criterion [13]. GINI index represents the node impurity measure.

### 2.1.5. Discriminant Analysis

Discriminant analysis is a classification method. It is based on the assumption that different classes generate data based on different Gaussian distributions. To train a classifier, the fitting function estimates the parameters of a Gaussian distribution for each class. To predict the classes of new data, the trained classifier finds the class with the smallest misclassification. There are two types of discriminant analysis. Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA).

### 2.2. Ensemble with single classifier

Ensemble means the combination of two or more classifiers. Ensemble can be homogenous or heterogeneous. Heterogeneous ensemble is also known as hybrid classifier. Homogenous means that each learner present in the ensemble is of same category. They differ only in the data set used for training the learners. Heterogeneous ensemble consists of learners of different categories. Different types of homogeneous and heterogeneous ensemble are discussed in this section.

#### 2.2.1. Bagging

Bagging also termed as bootstrap aggregation is an ensemble technique. The working is very simple. Bagging will first generate random subset from the available training dataset. Random sets are chosen with replacement. For each subset, separate classifier is generated. So there will be ‘n’ classifiers. The classifiers differ only in the dataset given for training. All classifiers make use of same algorithm. It can be SVM, k-nn, Naïve Bayes etc. When a new instance is given for testing, each one will generate results individually. Then voting is done among the results obtained to identify the majority one. That will be considered as the class for the newly given data. The process of combining the results from each classifier to generate the final one is termed as majority voting or combining. In case of bagging, the weights assigned to each classifier don’t change until end of the processing. Each classifier is assigned same weight.

#### 2.2.2. Dagging

Dagging is somewhat similar to Bagging. Dagging, the name is derived from disjoint bagging, which means that, in case of Bagging, different subsets of the training set are chosen randomly and for each subset, a new classifier is built. But the chosen examples are not disjoint i.e. the data are used with replacement. Data can be used repeatedly. The only difference in Dagging is that a data subset can be used for classifier building exactly once, i.e. each subset is disjoint [14]. Subsets are chosen without replacement.

#### 2.2.3. DECORATE

DECORATE (Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples) is the generation of diverse classifiers by means of created artificial training examples which are used for augmenting original training dataset. With the help of created artificial training sets, DECORATE algorithm enables us to create ensembles beyond the constraints imparted by the original training set.

DECORATE uses a base learning algorithm to train a classifier on initial original training set to generate a base ensemble. Using Gaussian data distribution of the original dataset, new artificial example sets are generated. The artificially generated sets are labeled in such a way that it is maximally differing from the labels assigned to it by the base ensemble prediction. A new augmented dataset is created by the union of initial training dataset with artificially generated dataset. The base learning algorithm is again employed to train a classifier on the new augmented dataset to generate a new augmented ensemble which is a union generated with initial base ensemble generates the new ensemble.

In order to validate the new ensemble, it is utilized to classify the original dataset and its classification accuracy is compared with that of the base ensemble. If there is depreciation in the accuracy, new augmented ensemble is removed from the new ensemble. This iterative process is repeated until required number of classifiers is reached or
a maximum allowed iteration step is reached. DECORATE is an iterative process which, after each iteration adds or removes new augmented ensemble from the base ensemble with respect to its efficiency in classifying the original dataset compared to the base ensemble [15].

3. Proposed System

The proposed system for classification of Arrhythmia cases from a given dataset consists of different stages. The input dataset given at the beginning stage passes through each and every stage and each outcome from each stage will be fed into the next stage for processing before it produces the final outcome. The schematic representation of the proposed system is shown in Fig.2.

The dataset used for this particular application i.e. Arrhythmia dataset is collected from University of California Irvine (UCI) repository [16]. The data set is contains only 279 features and its 453 observations. Out of 279 features, 278 are specific to the Arrhythmia disease and the terminal one represents the label or class field, which has different values from 01 to 16. The distribution of 16 different classes present in the dataset is given in Table 1.

| Class code | Class                          | Number of observations |
|------------|--------------------------------|------------------------|
| 01         | Normal                         | 245                    |
| 02         | Coronary Artery Disease        | 44                     |
| 03         | Old Anterior Myocardial Infarction | 15                    |
| 04         | Old Inferior Myocardial Infarction | 15                   |
| 05         | Sinus Tachycardy               | 13                     |
| 06         | Sinus Bradycardy               | 25                     |
| 07         | Ventricular Premature Contraction | 3                    |
| 08         | Supraventricular Premature Contraction | 2               |
| 09         | Left bundle branch block       | 9                      |
| 10         | Right bundle branch block      | 50                     |
| 11         | Degree Atrioventricular block  | 0                      |
| 12         | Degree AV block                | 0                      |

Fig.2. Schematic representation of the proposed system
|   |   |
|---|---|
| 3. | Degree AV block | 0 |
| 4. | Left Ventricule Hypertrophy | 4 |
| 5. | Atrial Fibrillation | 5 |
| 6. | Others | 22 |

Dataset collection is followed by preprocessing. During preprocessing, noises present in the dataset are removed. Missing or inconsistent values like symbols within the dataset will be replaced with appropriate values.

Before actual preprocessing begins, class label conversion is done. The dataset containing 16 different class labels should be transformed into 2 classes in order to simplify the operations on the dataset. The class labels are converted into 0 and 1. ‘0’ indicates absence of the Arrhythmia problem and ‘1’ indicates the presence of the disease. The dataset with class label ‘01’ i.e. normal will be represented as ‘0’ and remaining class labels from ‘02’ to ‘16’ will be represented as ‘1’. So 245 observations will be in class 0 and 207 instances will be in class 1. Remaining operations are performed with the new dataset having 2 class labels.

Data cleaning and data normalization are done as part of pre-processing. Data cleaning involves replacing the missing values by the mean of the attribute containing the cell. Data transformation consolidates the data into an appropriate form which is suitable for data mining. Normalization is performed to scale the data within a specific range. Normalization is done with Principal Component Analysis (PCA) [8].

When the complete data preprocessing has completed, the processed dataset will be divided into two disjoint sets. The size of the dataset is fixed based on 90-10 rule because as the size of the data set used for training increases, the performance is getting enhanced. Training makes the system to learn and produces some inferences from the dataset so that it can provide the outcome for a new observation accurately. Majority Voting helps to predict the result. Majority voting means that, among the five basic classifiers if more than or equal to three classifiers generates same results, then that one will be considered as the final result. Based on the generated outcome, the performance can be analysed. Analysis of the performance of the system is discussed in the next section.

**4. Result Analysis and Discussion**

**4.1. Evaluation Criteria**

The criteria used for evaluation are accuracy, specificity, sensitivity and area under curve are used for evaluating the performance of the implemented system. The actual output and predicted results are compared to create a confusion matrix. From the values generated from the confusion matrix, accuracy, specificity and sensitivity can be calculated.

**4.1.1. Accuracy**

Accuracy is calculated as the ratio of true cases to total cases in the test dataset i.e. the ratio of correctly classified samples (True positive (TP), True Negative (TN)) with the total samples available (TP, TN, False Negative (FN), False Positive (FP)), given by the equation (3):

\[
\text{Accuracy} = \frac{(TP+TN)}{(TP+FP+TN+FN)}
\]  

(3)

Where TP, FN, FP, TN values are obtained from confusion matrix.

**4.1.2. Sensitivity**

Sensitivity is also termed as true positive rate or recall, which gives the ratio of true positive to the sum of true positive and false negative. The values for TP and FN are derived from the confusion matrix.

\[
\text{Sensitivity} = \frac{TP}{(TP+FN)}
\]  

(4)
4.1.3. **Specificity**

Specificity is also termed as true negative rate which gives the ratio of true negative to the total negative cases i.e. proportion of TN to (TN+FP).

\[
\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}
\]  

(5)

4.2. **Performance Analysis**

4.2.1. **Base Learners**

The performance of each base learner in terms of accuracy, sensitivity and specificity is shown in Table 2.

| Algorithms Implemented | Accuracy | Sensitivity | Specificity |
|-------------------------|----------|-------------|-------------|
| KNN K=6                 | 84.44%   | 56.25%      | 100%        |
| Decision Tree (CART)    | 82.22%   | 87.5%       | 79%         |
| IB3                     | 80%      | 68.42%      | 88.46%      |
| Discriminant Analysis   | 80%      | 81.25%      | 79%         |
| Linear Discriminant     | 75.56%   | 69.23%      | 78.12%      |
| Linear                  | 80%      | 81.25%      | 79%         |
| Naïve Bayes             | 75.56%   | 31.25%      | 100%        |
| SVM                     | 73.33%   | 81.81%      | 70.58%      |
| Linear                  | 71.68%   | 58.58%      | 60.52%      |
| Linear                  | 64.44%   | 30%         | 64.44%      |

From the above table, it is clear that the maximum accuracy obtained is only 84.44%, which is for k-nearest neighbor with k=6. Among different SVM kernels (linear, quadratic, polynomial and RBF) linear kernel is having better performance. Linear Discriminant Analysis (LDA) outperforms Quadratic Discriminant Analysis (QDA) in terms of accuracy. Among different decision trees, Classification and Regression Tree has higher accuracy of around 82.22%. So better one among different base classifiers (Linear SVM, k-nn with k=6, Linear Discriminant Analysis, Naïve Bayes and Classification and Regression Tree (CART)) will be chosen for the creation of ensemble classifier.

4.2.2. **Ensemble with single classifier**

Since the performance of the base learners are not much, we can go for ensemble technique. Three homogeneous ensemble techniques used are Bagging, Dagging and DECORATE. The performance of these ensemble based on accuracy are shown in Table 3, 4 and 5. The accuracy of these classifiers is around 89% only. So in order to improve the accuracy level, ensemble with multiple classifiers is employed and result is explained in the next section.

| Ensemble classifier-Bagging | Accuracy | Sensitivity | Specificity |
|-----------------------------|----------|-------------|-------------|
| Bagging - KNN               | 71.11%   | 57.9%       | 80.76%      |
| Bagging - CART              | 82.22%   | 73.7%       | 88.5%       |
| Bagging - Naïve Bayes       | 88.09%   | 87%         | 88.4%       |
| Bagging - SVM               | 82.225%  | 78.9%       | 84.6%       |
| Bagging - LDA               | 66.66%   | 73.7%       | 53.8%       |

| Ensemble classifier-Dagging | Accuracy | Sensitivity | Specificity |
|-----------------------------|----------|-------------|-------------|
| Dagging - KNN               | 77.78%   | 47.4%       | 100%        |
| Dagging - CART              | 88.89%   | 78.9%       | 96.2%       |
| Dagging - Naïve Bayes       | 84.44%   | 84.2%       | 84.6%       |
| Dagging - SVM               | 80%      | 73.7%       | 84.6%       |
| Dagging - LDA               | 82.22%   | 68.4%       | 92.3%       |
| Ensemble classifier-DECORATE | Accuracy | Sensitivity | Specificity |
|-----------------------------|---------|------------|-------------|
| DECORATE - KNN              | 71.11%  | 52.6%      | 84.6%       |
| DECORATE - CART             | 84.44%  | 78.9%      | 88.5%       |
| DECORATE - Naïve Bayes     | 88.89%  | 78.9%      | 96.2%       |
| DECORATE - SVM              | 84.44%  | 78.9%      | 88.5%       |
| DECORATE - LDA              | 62.22%  | 73.7%      | 53.8%       |

4.2.3. Ensemble with multiple classifier

The result shows that Majority Voting is having better performance. Multiple classifiers used are Support Vector Machine, Decision Tree, Naïve Bayes, k-nearest neighbour and Discriminant Analysis. It is having accuracy of around 91.11%, which is higher as compared to other base learners as well as with ensemble with single classifier.

5. Conclusion

This paper presents an automated computation system that can assist in screening of the astronaut by considering the present cardiovascular conditions. The main focus is given to Arrhythmia disease. The system makes use of Majority Voting as the underlying technology. In order to perform Majority Voting, five base classifiers named SVM, k-nn, CART, LDA and Naïve Bayes are used. Performance of the implemented system is compared with other ensemble techniques as well as with base classifiers. Evaluated is done based on accuracy, sensitivity and specificity. The results show that Majority Voting is having highest accuracy of around 91.11%.

References

[1] Future Programme [online] Available : http://www.isro.org/script/futureprogramme.aspx#Human
[2] Thomas Goodwin, Carol Mullinax. Biomedical Research and Environmental Sciences [online]. http://www.nasa.gov/centers/johnson/sld/about/divisions/hacd/project/nxpcm.html
[3] Human Research Program-Cardiovascular System [online].http://www.nasa.gov/externalflash/HRP_Feature/
[4] Areas of Study: Cardiovascular Response (2013 Feb. 6) [online]. http://www.nasa.gov/exploration/humanresearch/areas_study/physics/physiology_cardio.html#VBPPhCJdXKo
[5] European Astronaut Selection. Last update: (22 May 2008) [online]. http://www.esa.int/Our_Activities/Human_Spaceflight/European_Astronaut_Selection/Concept_of_aeromedical_fitness_and_associated_medical_certificate_requirement
[6] Victor A. Convertina. Status of cardiovascular issues related to space flight: Implications for future research directions. Science Direct, Volume 169, Pages S34–S37, October 2009.
[7] Shiliang Sun, Rongquing Huang. An adaptive k-nearest neighbor. IEEE Seventh International Conference on Fuzzy System and Knowledge Discovery, pp.91-94, 2010.
[8] Nitin Aji Bhaskar. Performance Analysis of Neural Network and Support Vector Machine in Detection of Myocardial Infarction. International Conference on Information and Communication Technologies, pp.20-30, 2015.
[9] Yaguang Ji, Songnian Yu and Yafeng Zang. A Novel Naïve Bayes Model: Packaged Hidden Naïve Bayes. Sixth IEEE Joint International Conference on Information Technology and Artificial Intelligence. pp.484-487, 2011.
[10] Yongchuan Tang, Wuming Pan, Haiming Li and Yang Xu. Fuzzy Naïve Bayes Classifier based on fuzzy clustering. IEEE International Conference on System, Man and Cybernetics, Vol.5, 2002.
[11] Amany Abdelhalim and Issa Traore. A new method for learning Decision Trees from rules. International Conference on Machine Learning and Applications, pp.693-698, 2009.
[12] Haijian Li, Honghui Dong, Limin Jia and Moyu Ren. Vehicle Classification with single multi-functional magnetic sensor and optimal MNS based Cart. Journal of Measurement, Vol.55, pp.142-152, 2014.
[13] Leszek Rutkowski, Maciej Jaworski, Lena Pietruczuk and Piotr Duda. CART decision tree for mining data stream. Journal of Information Science, vol.266, pp.1-15, 2014.
[14] D. S. Anyfantis, Karagiannopoulou, S.B. Kotsiantis and P.E. Pintelas. Local Daging od decision stumps for Regression and Classification problems. IEEE Mediterranean Conference on Control and Automation, pp.001-006, 2007.
[15] Bo Sun, Haiyan Chen and Jiandong Wang. An empirical explanation for the effectiveness DECORATE ensemble learning algorithm. Journal of Knowledge –Based System, Vol.78, pp.1-2, 2015.
[16] UCI repository, Last Update (2003) [online], Available: https://archive.ics.uci.edu/ml/machine-learning-databases/arrhythmia