https://doi.org/10.48047/AFJBS.6.12.2024.4914-4924

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

# *Bridging the Gap: Unveiling Practical Implications of Machine Learning in Medical Healthcare*

**S.S.Aravinth[1], P.M. Ashok Kumar [2], Y. Likitha Bhagyasri [3], M. Om Vani Naga Divya [4], D. Sita Samanvitha [5], M. Venkata Naga Sai Vyshnavi [6]**

*[1]Department of Computer Scienceand Engineering, Koneru Lakshmaiah EducationFoundation, Guntur, India. aravinthkrithick@gmail.com*

*[2]Department of Computer Scienceand Engineering, Koneru Lakshmaiah EducationFoundation, Guntur, India. profpmashok@gmail.com*

*[3]Department of Computer Scienceand Engineering, Koneru Lakshmaiah EducationFoundation, Guntur, India. yellalikithabhagyasri@gmail.com*

*[4]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India. mullapudidivya227@gmail.com*

*[5]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India. samanvithadharmavarapus@gmail.com*

*[6]Department of Computer Science and Engineering, Koneru Lakshmaiah EducationFoundation, Guntur, India. vyshnaviminnikanti@gmail.com*

**Article History**
Volume 6, Issue 12, 2024
Received: June 10, 2024
Accepted: July 5, 2024
doi:
10.48047/AFJBS.6.12.2024.4914-4924

**Abstract-**

Computers can now learn autonomously due to advancements in machine learning, a technology widely applied across numerous global industries. This field focuses on developing computer algorithms that enhance data usage through experience. Essentially, machine learning involves enabling computers to acquire knowledge from data. It tackles numerous real-world issues, generating excitement in the medical sector. Despite the thousands of published papers on applying machine learning algorithms to medical data, only a few are truly beneficial for medical treatment. Therefore, one of our goals is to identify potential challenges in this area.

Understanding the Potential and Limitations of Machine Learning in Healthcare: The promise of machine learning in healthcare has garnered significant attention globally, highlighting its potential for transformative applications. However, only a small fraction of the numerous studies showcasing the effectiveness of machine learning algorithms on medical data have translated into actual medical therapies. This gap underscores several challenges, including concerns about data quality and privacy, the complexity of understanding intricate models, legal and ethical considerations, integration into clinical workflows, and the need for collaboration between technological and medical professionals. Addressing these issues requires multidisciplinary efforts to ensure the responsible and effective adoption of machine learning, ultimately enabling it to truly revolutionize healthcare practices.

Several areas of clinical research in machine learning are:

1. Reconstructing diseases
2. Hypothesis testing
3. Recruiting patients
4. Big data
5. Developing diagnostics
6. Improving prognostics
7. Patient monitoring
8. Requiring collaborations

**Keywords:** Medical Healthcare, Machine Learning Algorithm, Heart Disease.

## 1. Introduction

Machine learning is an emerging technology that constructs mathematical models using various algorithms to make predictions based on historical data. The accuracy of these predictive models improves through the application of machine learning. The advantage of machine learning lies in its ability to utilize models and algorithms for predicting outcomes. It can help forecast the presence or absence of locomotor issues, heart diseases, and other medical conditions.

The four main categories of machine learning are:

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning
4. Deep learning

When a machine learning algorithm is trained with data, it generates a machine learning model. This model produces the desired output when provided with appropriate input. Before deployment, these models are tested and trained on sample datasets, a process facilitated by machine learning. The operation of the machine learning algorithm is illustrated in Fig. 1.

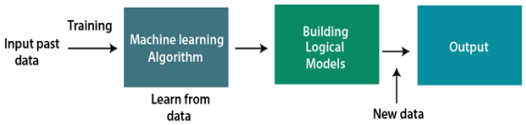


**Figure 1. Working of Machine Learning Algorithm**

As per above Fig. 1, the working of machine learning algorithm involves data input, which is processed through computational models to generate predictions.

The high dimensionality of data is a typical issue in machine learning; the datasets we utilize contain enormous amounts of data, often making it impossible to examine the data in 3D, a phenomenon known as the “curse of dimensionality”. As a result, processing this data requires a large amount of memory. Additionally, overfitting may occur when the data grows exponentially. By using the weightage characteristics, it is possible to reduce dataset redundancy, which in turn aids in lowering execution processing times. Various feature engineering and feature selection strategies can be used to exclude material that is not very important in order to reduce the dimensionality of the dataset.

**Figure 2. Technology in Health Care**

As per above Fig. 2, the integration of advanced technology in health care significantly enhances patient outcomes and operation efficiency.

Providing affordable, high-quality healthcare services is a major challenge for healthcare institutions such as hospitals and clinics. Efficient therapy delivery and precise patient diagnosis are essential for providing high-quality care. There are both numerical and categorical data in the heart disease database that is currently accessible. Cleaning and filtering are conducted to these records prior to additional processing in order to remove unnecessary data from the database. The suggested technique can be used to find specific hidden knowledge, such as heart disease-related patterns and associations from a historical heart disease database. It can also answer difficult concerns regarding the diagnosis of cardiac disease. As a result, as illustrated, in Fig. 2, healthcare professionals may find it useful to make wise therapeutic decisions. The suggested system is highly good in accomplishing the stated mining goals, according to the results.

## 2. Challenges in Healthcare Industry

Heart disease is currently the leading cause of mortality worldwide. According to reports, heart diseases account for 70% of all deaths. The term "heart disease" encompasses a range of disorders that impair the normal functioning of the circulatory system, including the heart and veins. Early detection of cardiac ailments allows patients to receive appropriate and adequate treatment, leading to complete recovery and significantly reduced treatment costs.

As a result, there is a need to develop a predictive framework that can accurately identify the presence or absence of cardiac disorders in patients. Machine learning methods can be used to predict cardiac disease. This article discusses a study in which potential cardiac problems in patients were predicted using machine learning algorithms. The research compares the performance of various machine learning algorithms. The study primarily focuses on different data processing techniques used in predicting heart conditions.

## 3. Problem Statement

Predicting cardiac disease is one of the most challenging tasks in the medical industry. In healthcare, data science plays a crucial role in analyzing vast amounts of data. Given the complexity of predicting cardiac conditions, automating the process is essential to mitigate potential risks and provide timely information to patients. This research employs data processing techniques such as Naive Bayes, Decision Trees, Logistic Regression, and Random Forest to forecast the likelihood of heart problems and classify the patient's risk level.

Doctors maintain extensive records of medical data, which can be analysed to extract vital knowledge. Data mining techniques are used to collect useful and hidden information from this vast amount of data. Consequently, these techniques are employed to analyze large datasets. The goal of this research was to identify the most effective machine learning system for detecting cardiac problems and to provide clinicians with a tool to aid in early diagnosis of cardiac disease. This will facilitate the provision of appropriate therapy to patients and help avoid serious complications.

Coronary Artery Disease (CAD), Congenital Heart Disease (HD), Mitral Valve Prolapse, Arrhythmia, Pulmonary Stenosis, Dilated Cardiomyopathy, Heart Failure, Hypertrophic Cardiomyopathy, and Myocardial Infarction are examples of heart illnesses.

Heart diseases refer to a variety of disorders that affect the heart. According to the World Health Organization, cardiovascular illnesses are the leading cause of death globally, accounting for 17.9 million deaths annually. Several unhealthy behaviors contribute to the increased risk of heart disease, including obesity, high cholesterol, elevated triglyceride levels, and hypertension. The American Heart Association identifies several indicators of heart disease, such as trouble sleeping, an irregular heartbeat, swollen legs, and, in rare cases, rapid weight gain of 1-2 kg per day. These symptoms can resemble those of various other

conditions, particularly in the elderly, making accurate diagnosis challenging and sometimes leading to fatal outcomes.

However, over time, an increasing amount of research data and hospital patient information has become accessible. Numerous publicly available resources provide patient records, enabling studies that leverage diverse computer technologies for accurate patient diagnosis and early detection of potentially fatal diseases. Today, it is widely recognized that artificial intelligence and machine learning are crucial to the medical sector. Various machine learning and deep learning models can be used not only to diagnose illnesses but also to categorize or predict outcomes. Machine learning models simplify the comprehensive examination of genomic data and can be trained to transform and thoroughly analyze medical records, as well as to make predictions about pandemics.

Multiple studies have been conducted utilizing various machine learning models to categorize and forecast diagnoses of cardiac conditions. Melillo et al. employed a machine learning algorithm known as CART (Classification and Regression Trees) to develop an automated classifier for congestive heart failure, distinguishing patients at high risk from those at low risk. The algorithm achieved a sensitivity of 93.3 percent and a specificity of 63.5 percent. Rahhal et al. subsequently proposed a method to enhance the performance of electrocardiograms (ECGs) by selecting optimal features using deep neural networks.

In subsequent research, Guidi et al. introduced a clinical decision support system aimed at the early detection and prevention of cardiac failures. They conducted a comprehensive comparison of various deep learning and machine learning models, focusing particularly on neural networks such as support vector machines, random forests, and CART algorithms. The combination of random forest with CART yielded an impressive accuracy of 87.6 percent, surpassing all other methods used for categorization. Zhang et al. achieved 93.37 percent accuracy in determining the NYHA HF class from unstructured clinical notes by combining rule-based techniques with natural language processing. Parthibian and Srivastsa's SVM algorithms, utilizing common variables such as blood sugar level, patient age, and blood pressure data, achieved a remarkable accuracy rate of 94.60 percent in detecting patients with existing diabetes and predicting heart disease.

Applying feature engineering and selection techniques significantly enhances the performance of classification and prediction in diagnosing heart diseases. Dun et al. experimented with various deep learning and machine learning methods, achieving a notable accuracy of 78.3 percent with neural networks, outperforming other models like logistic regression, SVM, and Random Forest. Singh et al. employed a binary classifier akin to an extreme learning machine and achieved an impressive 100% accuracy in identifying coronary heart disease through generalized discriminant analysis. Yaghoobi et al. successfully classified arrhythmias based on heart rate variability, attaining 100% accuracy using either feature reduction or Gaussian discriminant analysis with a multilayer perceptron neural network. Asl et al. reduced HRV signal characteristics by 15% with Gaussian discriminant analysis, achieving 100% precision with the SVM classifier.

When dealing with high variance or high-dimensional data, employing appropriate dimensionality reduction techniques like PCA enables the retention of crucial information in new components. Many researchers opt for PCA as their primary choice when handling high-dimensional data. In their quest to categorize cardiac arrhythmias, Rajpopal and Ranganathan explored five distinct dimensionality reduction approaches, both linear and nonlinear, utilizing a neural network as a classifier. Fast ICA, employed for independent component analysis, achieved an impressive F1 score of 99.83 percent with at least 10 components. Zhang et al. achieved remarkable results by utilizing PCA and uncorrelated discriminant analysis to identify optimal features for controlling upper limb motions. Avendaño-Valencia et al. aimed to enhance performance by reducing heart sounds using PCA approaches on time-frequency

representations. Kamencay et al. introduced a novel approach to various medical images, achieving an accuracy of 83.6 percent with PCA-KNN, a scale-invariant feature commonly used in medical image scaling, after training on 200 photos. Ratnasri et al. reduced characteristics of X-ray images using a grey-level threshold of 150 based on ROI and PCA.

Previous research has predominantly relied on a dataset containing thirteen features, with each study focusing on classifying patients to determine the presence or absence of heart disease. A recurring trend observed is the frequent utilization of the Cleveland dataset. The outcomes consistently demonstrate high accuracy, with random forests achieving 89.2 percent accuracy, decision trees at 89.1 percent accuracy, ANNs reaching 92.7 percent accuracy, SVMs achieving 89.7 percent accuracy, and a hybrid model combining GA and NN achieving an impressive 94.2 percent accuracy. PCA models, including PCA with regression and PCA1 with NN, attained accuracies of 92.2 percent each. Here, learning three things
primarily involved dimensionality reduction:

1. Choosing the best characteristics
2. Performance validation and
3. The use of six distinct classifiers todetermine the final list of 74 characteristics.

Heart disease is a serious condition with potentially fatal consequences and should not be taken lightly. Research from Harvard Health Publishing indicates that males are at a higher risk of heart disease compared to females, with men having approximately twice the lifetime risk of experiencing a heart attack. Even after considering conventional risk factors such as high blood pressure, high cholesterol, diabetes, body mass index, and physical activity, this elevated risk persists. The dataset used by researchers working on heart disease prediction is considered a benchmark, containing crucial parameters dating back to 1998. Collected in 1988, this dataset spans four databases: Cleveland, Hungary, Switzerland, and Long Beach V. The outcomes of research conducted using this dataset show great promise in advancing our understanding of heart disease.

## 4. An Explanation of the Dataset

| | age | sex | cp | trestbps | chol | fbs |
|---|---|---|---|---|---|---|
| count | 303.00 | 303.00 | 303.00 | 303.00 | 303.00 | 303.00 |
| mean | 54.37 | 0.68 | 0.97 | 131.62 | 246.26 | 0.15 |
| std | 9.08 | 0.47 | 1.03 | 17.54 | 51.83 | 0.36 |
| min | 29.00 | 0.00 | 0.00 | 94.00 | 126.00 | 0.00 |
| 25% | 48.00 | 0.00 | 0.00 | 120.00 | 211.00 | 0.00 |
| 50% | 55.00 | 1.00 | 1.00 | 130.00 | 241.00 | 0.00 |
| 75% | 61.00 | 1.00 | 2.00 | 140.00 | 274.00 | 0.00 |
| max | 77.00 | 1.00 | 3.00 | 200.00 | 564.00 | 1.00 |

| restech | thalach | exang | oldpeak | slope | ca |
|---|---|---|---|---|---|
| 303.00 | 303.00 | 303.00 | 303.00 | 303.00 | 303.00 |
| 0.53 | 149.60 | 0.33 | 1.04 | 1.39 | 0.73 |
| 0.52 | 22.90 | 0.47 | 1.16 | 0.62 | 1.02 |
| 0.00 | 71.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 0.00 | 133.50 | 0.00 | 0.00 | 1.00 | 0.00 |
| 1.00 | 153.00 | 0.00 | 0.80 | 1.00 | 0.00 |
| 1.00 | 166.00 | 1.00 | 1.60 | 2.00 | 1.00 |
| 2.00 | 202.00 | 1.00 | 6.20 | 2.00 | 3.00 |

**Figure 3&4. Dataset Information**

As per the above figures 3&4, the dataset comprises 303 observations detailing various health metrics, including age, sex, chest pain type, resting blood pressure, cholesterol, and fasting blood sugar, along with resting ECG results, maximum heart rate, exercise-induced angina, ST depression, the slope of peak exercise ST segment, and major vessels coloured by fluoroscopy. The statistics highlight significant variability in cholesterol variability in cholesterol levels(126 to 564 mg/dL) and maximum heart rate(71 to 202 bpm). It is therefore necessary to scale the features on the dataset.

## 5. Model Description

In this study, I used four different algorithms, tweaked their various parameters, andcompared the results. 67% of the dataset was used for training, and 33% was used for testing.

### (i) K Neighbours Classifier

An organised database system was built using the gathered data. Pre-processing was carried out by locating the related fields and eliminating any duplicates. The data was then coded in accordance with the domain value after all the missing values had been filled in. A simple supervised machine learning approach that may be applied to both regression and classification problems is the k-nearest-neighbours (KNN) algorithm.

The KNN Algorithm

1. Open the data
2. Initialize K to the chosen number of neighbours
3. For each example in the data

   3.1. Determine the distance, using the data, between the query example and the current example.

   3.2 To an ordered collection, add the example's index and distance.
4. The ordered collection of indices and distances should be sorted by the distances from least to greatest (in ascending order).
5. From the sorted collection, select the first K elements.
6. The K entry labels that have been chosen should be obtained.
7. If the data is regression, return the mean of the K labels.
8. If the categorization is K labels, return the K labels mode.

### (ii) Support Vectors Classifier

The objective of the support vector machine (SVM) algorithm is to find a hyperplane in an N-dimensional space (where N represents the total number of attributes) that effectively separates the data points into distinct categories.

There exist multiple possible hyperplanes that can be selected to separate the two groups of data points. Our aim is to identify a plane with the maximum margin, which is the greatest

distance between data points of both classes. Maximizing the margin distance is crucial for enhancing the classification accuracy of future data points.

Hyperplanes, acting as decision boundaries, are employed to classify data points, with distinct classes on either side. The dimension of the hyperplane is determined by the number of features. For instance, with only two input features, the hyperplane is essentially a line. With three input features, it becomes a two-dimensional plane. However, visualization becomes challenging when the number of features exceeds three.

Support vectors are data points that lie closest to the hyperplane and play a crucial role in determining its orientation and position. By utilizing these support vectors, we aim to maximize the margin of the classifier. Removing support vectors would result in a change in the location of the hyperplane. These support vectors are essential points that enable the construction of our SVM.

**(iii)Decision Tree Classifier**

A decision tree, a type of supervised machine learning algorithm, makes decisions by following a set of rules, akin to human decision-making processes.

Decision trees are indeed referred to as the CART algorithm, which stands for Classification and Regression Tree. This is because decision trees can handle both classification and regression tasks. It's a general term that can describe not only decision trees but also all tree-based methods.

The intuition behind Decision Trees is to utilize the features of the dataset to create yes/no questions, continuously splitting the data until every data point is separated into its respective class.Through this method, the information is being arranged in a tree structure.The tree is expanded by adding a node each time a question is asked. The root node is the name given to the initial node.

By sorting the dataset according to a feature's value, a question's outcome produces new nodes.Leaf nodes are the final nodes to form if the process is ended after a split. The feature space is also branched and partitioned into discontinuous sections each time a question is responded to. All the data points on one branch of the tree correlate to replying "yes" to the question that the preceding node's implicit rule suggested. The remaining data points are contained in a node on the other branch.

**(iv)Random Forest Classifier**

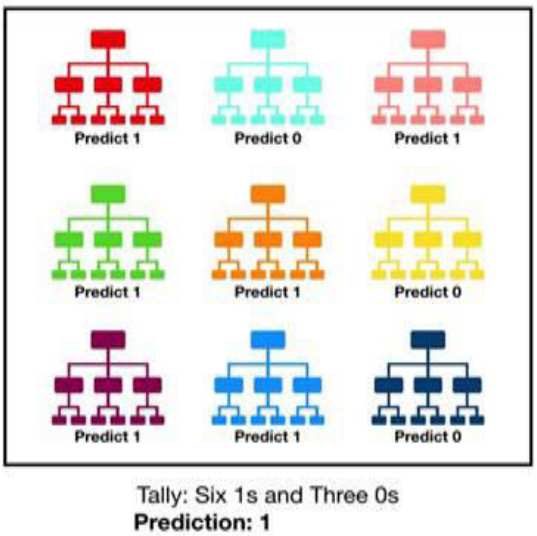


**Figure 5. Working of Random Forest Classifier**

As per above Fig. 5, the Random Forest Classifier operates by constructing multiple decision trees during training and outputting the mode of the classes for classification tasks.

A random forest, as its name suggests, is an ensemble of many decision trees. Each tree in the random forest generates its individual class predictions. Our model then selects the class with the highest number of votes among these predictions as the forecast.

The wisdom of crowds is the fundamental concept behind Random Forest, contributing to its exceptional performance. This model leverages the collective intelligence of multiple decision trees to make predictions.

Any number of the individual component models will not perform as well as a committee made up of numerous suitably uncorrelated models(trees).

The low correlation among models is indeed crucial in ensemble learning methods like Random Forest. Similar to how low-correlated assets, such as stocks and bonds, combine to form a diversified portfolio that outperforms individual investments, uncorrelated models in Random Forest generate ensemble forecasts that surpass the accuracy of individual predictions. The lack of correlation between the models allows them to complement each other, mitigating errors and enhancing overall accuracy. In Random Forest, each decision tree acts as a unique contributor to the ensemble, offering different perspectives on the data and collectively improving predictive performance. This diversity and lack of correlation among models are central to the remarkable effectiveness of Random Forest in various prediction tasks.

## 6. Implementation details

### (i)K Neighbours Classifier

A class is assigned to a given datapoint by this classifier based on the majority class after it has searched for the classes of the K nearest neighbours of the supplied data point. However, the quantity of the neighbours may vary. I calculated the test score for each variant after varying the number of neighbours from 1 to 20.

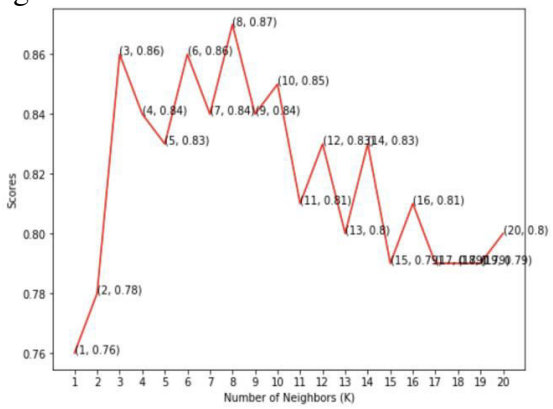


**Figure 6. K Neighbours Classifier Scores for Different K Values**

As per above Fig. 6, as it can be seen, the maximum score of 87% was achieved, when 8 neighbours were chosen.

### (ii)Support Vector Classifier

This classifier aims to construct a hyperplane that can partition the classes as much as possible by adjusting the distance between the data points and the hyperplane. Multiple kernels influence the hyperplane selection. I experimented with four different kernels: sigmoid,poly,rbf,and linear.

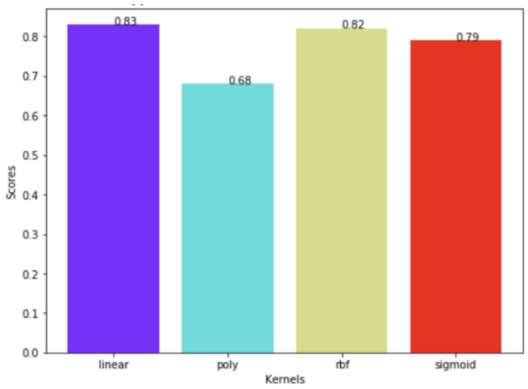


**Figure 7. Support Vector Classifier Scores for Various Kernels**

As per above Fig. 7, the support vector classifier achieves the highest score with the linear kernel (0.83).The polynomial kernel performs the worst with a score of 0.68.

With a score of 83%, the linear kernel demonstrated the best performance for this dataset, as illustrated in Fig.6.

**(iii)Decision Tree Classifier**

This classifier constructs a decision tree and assigns class values to each data item, a process where we specify the maximum number of features that the model will utilize during construction. This range of features typically spans from 1 to 30, representing the total number of features in the dataset after dummy columns have been added.

We will make a line graph to show how the number of features impacts the model scores once we have the findings.

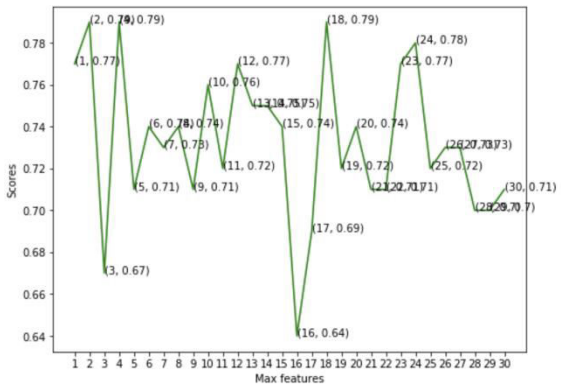


**Figure 8. Scores of the Decision Tree Classifier for Varying Maximum Feature Counts**

As per above Fig. 8, the scores of the Decision Tree Classifier vary significantly with different maximum feature counts, peaking at certain values such as 2,18, and24.This indicates importance of tuning the “max features”parameter to optimize classifier performance.

Figure 7 illustrates that the highest possible score,attained by selecting the maximum number of characteristics to be 2,4, or 18 is 79%.

**(iv)Random Forest Classifier**

This classifier extends the concept of decision trees.A random selection of features selected from the whole set of features makes up each tree in the resultant forest.In this case,the number of trees that will be used to predict the class can be altered. Calculations are made using test data from 10,100,200,500 and 1000 trees.

```
rf_scores = []
estimators = [10, 100, 200, 500, 1000]
for i in estimators:
rf_classifierS= RandomForestClassifier(n_estimators=i, random_state = 0)
rf_classifier.fit(X_train, y_train)
```

rf_scores.append(rf_classifier.score(X_test, y_test))

The results are then displayed on a bar graph to show which scores resulted in the best results.It will be noticed that the X values were not just assigned to the array [10, 100, 200, 500, 1000]. If examined closely, it will show an incomprehensible continuous plot ranging from 10 to 1000. To solve this problem, the X values were initially used as [1, 2, 3, 4, 5], and then xticks were used to rename them.

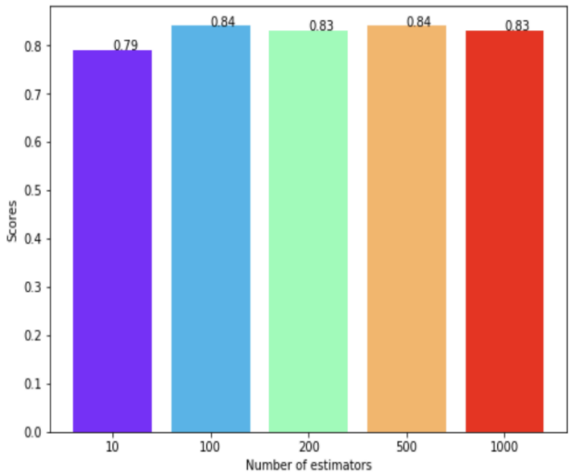


**Figure 9. Scores from the Random Forest Classifier for Various Numbers of Estimators**

As per above Fig. 9, the Random Forest Classifier achieves the highest scores with 100 and 500 estimators. The lowest performance is observed with 10 estimators, resulting in a score of 0.79.

As the bar graph in Fig.8 illustrates,the maximum score of 84% was obtained for both 100 and 500 trees.

## 7. Conclusion

As part of the research,the patient dataset for heart disease was properly processed and analysed.Next,four models with the following maximum scores were trained and tested:

1. K Neighbours Classifier: 87%
2. Support Vector Classifier: 83%
3. Decision Tree Classifier: 79%
4. Random Forest Classifier: 84%

With 8 neighbours, K Neighbours Classifier achieved the highest score of 87%.

## References

[1] Bejnordi, B. E., Veta, M., van Diest, P.J., van Ginneken. B., Karssemeijer, N., Litjens, G., … & Hermsen, M. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA,318(22),2199-2210.

[2] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Sweater, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks.Nature, 542(7639), 115-118.

[3] Gulshan, V., Peng,L., Coram, M., Stumpe, M.C., Wu, D., Narayanswamy, A., … & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22),2402-2410.

[4] Rajkomar,A., Oren, E., Chen,K., Dai,A. M., Hajaj, N., Hrdt, M., … & Dean, J. (2018). Scalable and accurate deep learning with electronic health records.NPJ digital medicine, 1(1), 18.

[5] Ting,D.S.W., Cheung, C.Y.L., Lim, G., Tan, G. S.W., Quang, N.D.,Gan,A.,…& Wong, T. Y. (2017).Development and validation of a deep learning system for diabetic

retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA,312(22),2211-2223.

[6] Vamathevan, J., Clarks, D., CzOdrowski, P., Dunham, I.,Ferran, E.,Lee, G., …&Morgat,A.(2019). Applications of machine learning in drug discovery and development. Nature Reviews Drug Discovery, 18(6), 463-477.

[7] Xiao, C., Choi, E., Sun, J., Hwang, S., Liu, Y., Castro, D., & Ofori-Boadu, L. (2018).Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. Journal of the American Medical Informatics Association,25(10),1419-1428.

[8] Zhang, K., Liu, X., Shen,J.,Li,Z., Sang,Y., Wu, X., … & Wang, C.(2019). Clinically applicable AI system for accurate diagnosis,quantitative measurements, and prognosis of COVID-19 pneumonia using computed
tomography. Cell, 181(6), 1423-1433.