WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Heart Disease Identification in E-Healthcare by using Machine Learning

Prof. Vidyadhar Hanji [a], Varun Patil [b], Santosh Mane [c], Ramesh Harobidi [d], Viresh Ganachari [e]

[a] *Department of Computer Science and Engineering, Angadi Institute of Technology and Management, Belagavi-590009, India*
[b] *Department of Computer Science and Engineering, Angadi Institute of Technology and Management, Belagavi-590009, India*
[c] *Department of Computer Science and Engineering, Angadi Institute of Technology and Management, Belagavi-590009, India*
[d] *Department of Computer Science and Engineering, Angadi Institute of Technology and Management, Belagavi-590009, India*
[e] *Department of Computer Science and Engineering, Angadi Institute of Technology and Management, Belagavi-590009, India*

A B S T R A C T

Heart disease is one of the most common and complex diseases in the world. It is very important that the eHealth system be able to identify suitable patients for this condition. We proposed a system for accurate diagnosis of heart disease using machine learning techniques. The system was developed using various features such as ANN, SVM and LR. We were able to improve accuracy and reduce the time spent on manual analysis. The system has many performance metrics that can be used to improve the efficiency of the system.

Keywords: Heart Disease identification, Parameters, Feature selection, medical data analysis.

1. Introduction

Heart disease is one of the leading causes of death worldwide, and early detection and analysis are critical to improving patient outcomes. With the increasing availability of electronic health records (EHRs) and wearable devices, there is an opportunity to use machine learning algorithms to predict heart disease risk and provide personalized care to patients. Here is a step-by-step process for implementing such a system.

Identifying heart disease using machine learning in e-healthcare involves using a dataset of patient health data to train a machine learning algorithm and adapting it to input characteristics such as age, gender, medical history, and lifestyle habits. including predicting the presence or absence of heart disease based on Clinical data such as blood pressure, cholesterol levels, and electrocardiogram (ECG) recordings. Once the model is trained and evaluated, it can be deployed in production as a web application or mobile app that allows users to enter health data and take heart disease risk assessments.

In today's world, many machine learning techniques are being developed to reap the benefits of using machine learning to detect heart disease in eHealth. This includes improving the accuracy and reliability of risk assessment, personalized care, and early detection of heart disease. By identifying patients at high risk for heart disease, providers can intervene early and provide targeted treatment to improve patient outcomes.

Overall, machine learning heart disease detection in eHealth is an interesting research area with the potential to improve patient outcomes and reduce healthcare costs. With advances in machine learning algorithms and increased availability of health data, we can expect more innovative approaches to heart disease detection and treatment in the coming years.

In this study, we proposed a machine learning-based method for identifying heart disease. Use algorithms such as SVM, ANN, decision trees, logistic regression, and Naive Bayes used to identify heart disease. The first step is to collect a dataset of patient health information. This includes demographic information, medical history, lifestyle, and clinical data such as blood pressure, cholesterol levels, and electrocardiogram (ECG) recordings. Once the data are collected, they need to be preprocessed to remove discrepancies, missing values, or outliers. This includes data cleansing, feature selection and engineering, and scaling. The preprocessed data is split into two sets: training and testing. The training set is used to train machine learning algorithms and the test set is used to evaluate performance.

2. Literature Survey

In [1] the authors suggested: They used KNN, a genetic, naive Bayes algorithm. Identifying raw health data processing of cardiac information can save lives in the long term and help determine heart disease abnormalities early. A hybrid random forest heart disease prediction model using a linear model improves performance levels with an accuracy of 88.70%.

In [2], we used the author's proposed machine learning and BPNN classifiers to test which classifiers are more effective in diagnosing heart disease. Based on the metrics that evaluate model performance, SVM performed exceptionally well on all features, achieving an accuracy of 86%. Their experimental results show that the performance of his BPNN-based diagnostic system for diagnosing heart disease is more effective than the accuracy (classification accuracy) of BPNN 93.

In [3] the author introduces this system. The system consists of her two main components: feature selection using the RFRS subsystem and data classification using the classification system. It is mainly based on his KNN and SVM algorithms. This system has some weaknesses. The number of nearest neighbors (k) and weight threshold (θ) are not stable in the rescue algorithm. This system gives an accuracy of about 91.5%.

In [4] the author proposed a method. The system used sequential backward selection of features and the ANN algorithm. Here we used a minimal data set of 70% on average. This system can achieve accuracies of up to 90%.

In [5] the author defined: This system is based on artificial neural networks. This system provides a decision support system for identifying her three major heart conditions:

Mitral valve stenosis, aortic valve stenosis, vertical septal defect. Using real medical data, we conducted a series of tests to examine the performance of the proposed solution with approximately 90% accuracy.

In [6], the author's proposed probabilities obtained from applying the Cleveland algorithm are Bayesian derived from a published medical study called CADENZA, applying his algorithm to the same three patient tests his group. were compared with the probabilities obtained from Both algorithms overestimated the probability of disease in Hungarian and American centers.

In [7], the author's classification of heart disease is of added value to physicians. This chapter aims to assist physicians in making decisions to classify healthy patients from those with coronary artery disease (CAD) using a generic modified decision tree using a genetic algorithm. Performance analysis of the proposed method is compared with data mining approaches and classification based on probabilistic rules. The five machine learning algorithms include K Nearest Neighbors (KNN), Artificial Neural Networks, Support Vector Machines (SVM), Decision Trees, and Modified Genetic Algorithms Decision Trees. Analyzes were conducted in terms of accuracy, performance, and sensitivity.

In [8] the authors suggested that the diagnosis is used to ascertain whether a patient has heart disease. The dataset was obtained from UCI Machine Learning. This dataset was split into training, validation, and test datasets and fed to the network. The intelligent system was modeled with feedforward multilayer perceptron and support vector machines. The detection rates obtained from these models were later compared to determine the best model for intelligent systems given their importance in the medical field. The obtained results are 85% and 87.5% for feedforward multilayer perceptron and support vector machines, respectively.

In [9], this author's contribution proposes a very fast feature selection method based on conditional mutual information. Selecting features that maximize mutual information with the class to conditionally predict ensures selection of features that are individually informative and pairwise weakly dependent on features already selected. We show that this feature selection method outperforms other classical algorithms and that a simple Bayesian classifier constructed with features selected in this way has similar errors to modern methods such as boosting and SVM. indicates that the rate is achieved.

In [10], the author's defined the proposed method is based on the fusion of a filter and a wrapper method, i.e., the Conditional Mutual Information Maximization (CMIM) method and the support vector machine recursive feature elimination, respectively. The performance of the proposed method was evaluated against four state-of-the-art feature selection methods, minimum redundancy maximum relevancy, fast correlation based feature selection, CMIM, and Relief, using four classifiers, support vector machine, naive Bayes, linear discriminant analysis, and k nearest neighbors on five different SNP data sets obtained from the National Center for Biotechnology Information gene expression omnibus genomics data repository.

In [11], the author's proposed a model using this paper presents a machine-learning-based model for Parkinson's disease recognition. Specifically, a hybrid feature selection algorithm has been designed by integrating the Relief and ant-colony optimization algorithms to select relevant features for training the model. Moreover, the support vector machine has been trained and tested on the selected features to achieve optimal classification accuracy. Additionally, a K-fold cross-validation technique was employed to optimally evaluate the hyperparameter values of the model. Experimental results on real-world datasets, i. A dataset on H. Parkinson's disease shows that the proposed system accurately detects Parkinson's disease and outperforms the first competitor by achieving an accuracy of 99.50 for selected traits. I'm here. Due to the high performance achieved by the proposed method, it is highly recommended for PD detection.

In [12], the author's article proposes a new estimation method in linear models. Lasso minimizes the remaining sum of squares if the sum of the absolute values of the coefficients is less than a constant. Due to the nature of this constraint, it tends to produce coefficients that are strictly 0, thus producing an interpretable model. Our simulation studies suggest that lasso exhibits some of the favorable properties of both subset selection and ridge regression. Generate interpretable models such as subset selection and demonstrate the stability of ridge regression. It is also interestingly related to the recent work on adaptive function estimation by Donoho and Johnstone.

3. System design

3.1 System Architecture

The system is built using machine learning algorithms and is the most used language for recommendation engines of all time. We used a K-neighbors classifier, a decision tree classifier, a random forest classifier, and a support vector classifier.

The system starts with a user uploading data (parameters), the uploaded data is preprocessed, after preprocessing the data is matched against a dataset, after matching, the system predicts heart disease, then analyze the results. A graphical representation of the results is then displayed with the modified graph data.

Finally, users can view and consult doctors from different doctors added by the admin.

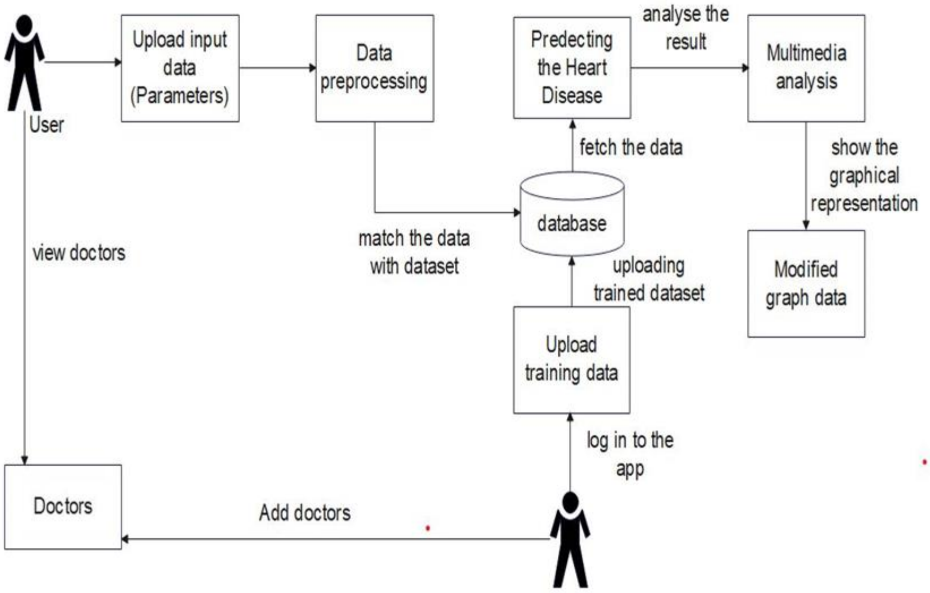
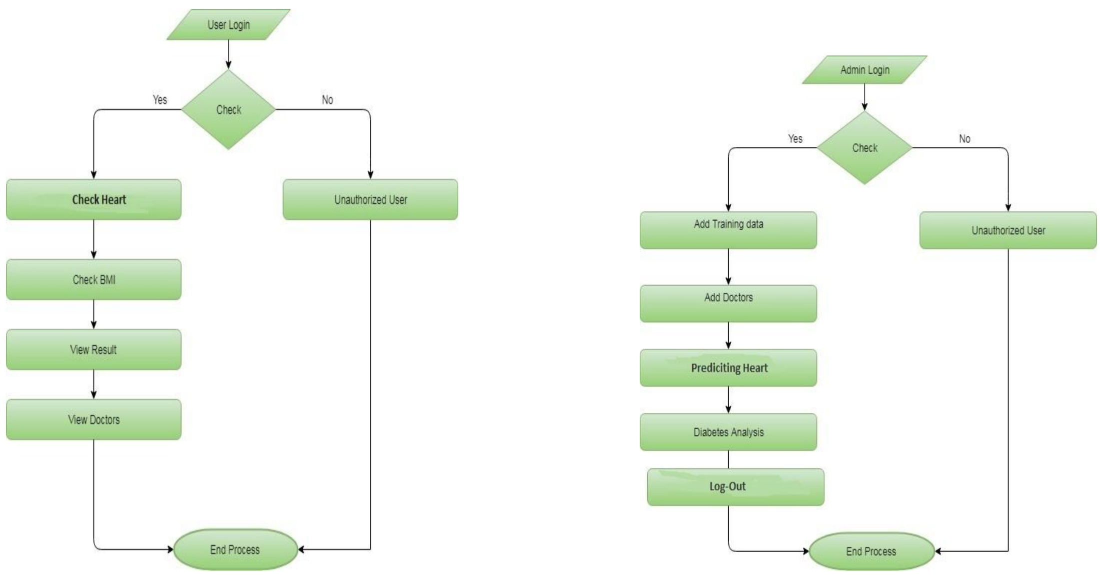


Fig 1. Architecture of proposed Heart Disease identification.

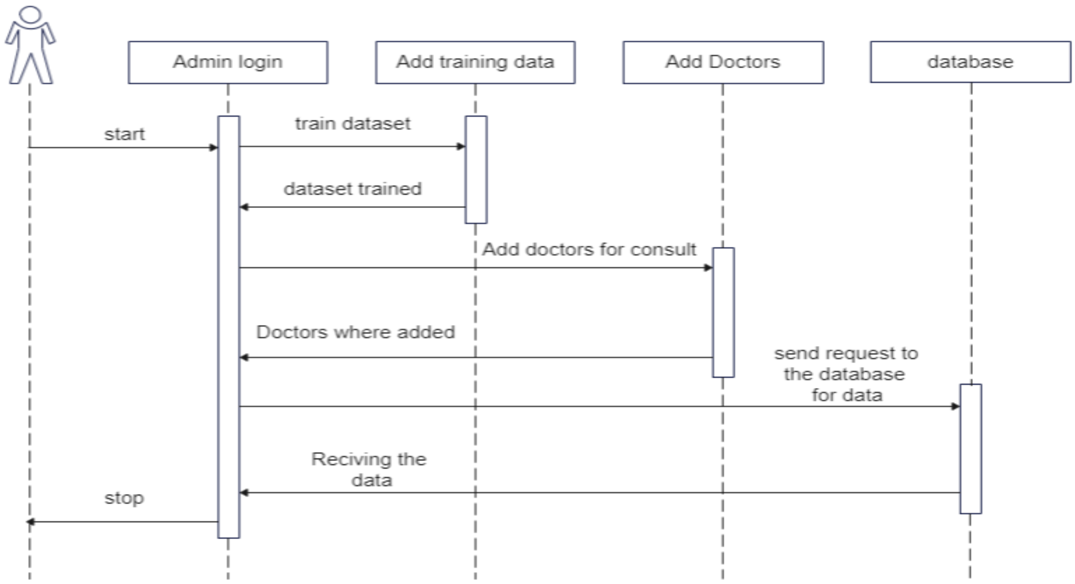
3.2 Data Flow



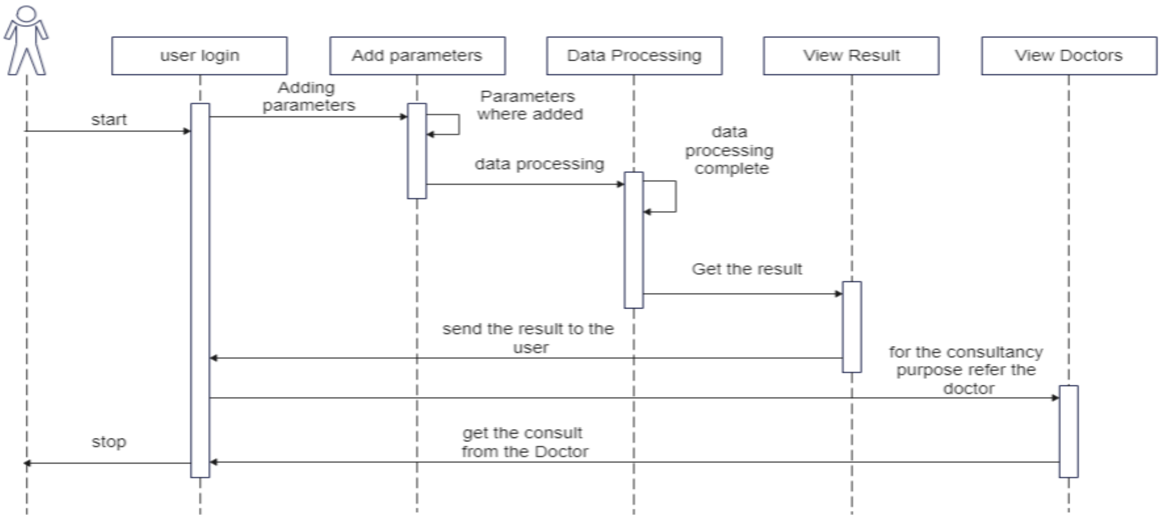
a) USER **b) ADMIN**

Fig.2- Data flow Diagram of proposed Heart disease Prediction System.

3.3 Sequence Diagram



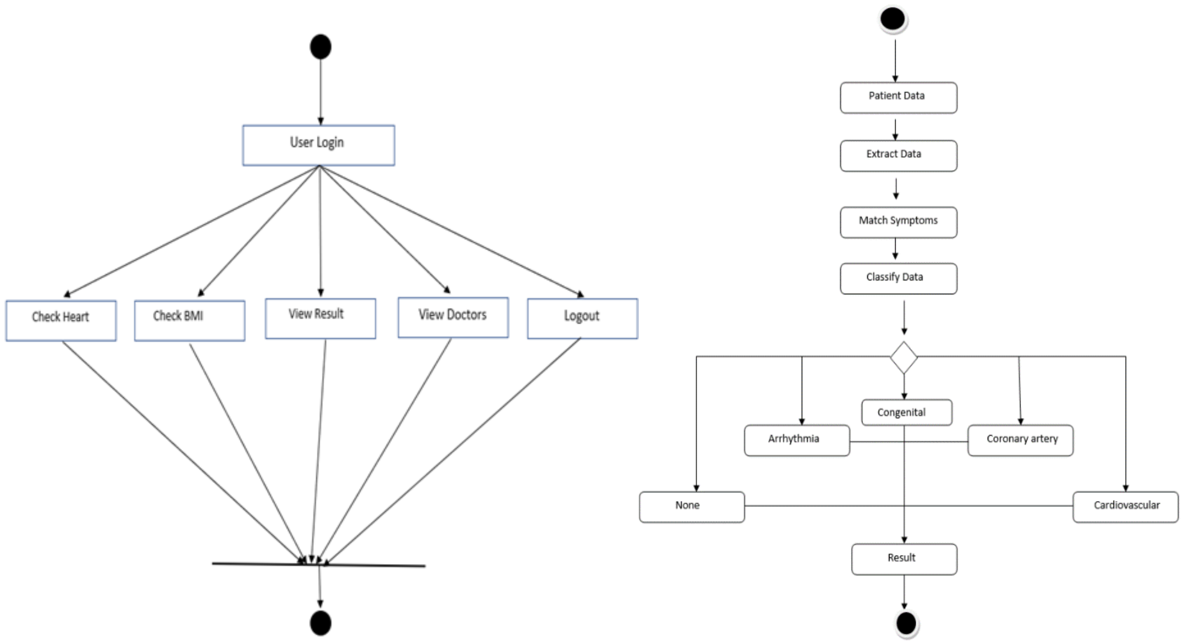
a) USER



***b)* ADMIN**

Fig.3- Sequence Diagram of proposed Heart disease Prediction System.

3.4 Activity Diagram



a) **USER** **b) ADMIN**

Fig.4- Activity Diagram of proposed Heart disease Prediction System.

4. Implementation

1. Collect data:- The first step is to collect a dataset of patient health information. This includes demographic information, medical history, lifestyle, and clinical data such as blood pressure, cholesterol levels, and electrocardiogram (ECG) recordings.

2. Data preprocessing:- Once the data are collected, they need to be preprocessed to remove discrepancies, missing values, or outliers. This includes data cleansing, feature selection and engineering, and scaling.

3. Data division:- The preprocessed data is split into two sets:training and testing. The training set is used to train machine learning algorithms and the test set is used to evaluate performance.

4. Select a machine learning algorithm:-Several machine learning algorithms can be used to identify heart disease, including logistic regression, decision trees, random forests, and support vector machines (SVM). Algorithms are chosen based on the size and complexity of the dataset and the desired performance metric.

5. Training a machine learning model:-The training data set is used to train a selected machine learning algorithm to classify heart disease based on input features.

6. Evaluate performance:-Once the model is trained, it is evaluated against the test dataset using performance metrics such as precision, accuracy, recall, and F1 score.

7. Deployment model:-Once the model performs satisfactorily, it can be deployed in production as a web application or mobile app that allows users to enter health data and take heart disease risk assessments. The application can also send notifications to patients and healthcare providers when risks are detected.

8. Continuous learning:-Machine learning models can improve over time by integrating new data and retraining the model. This can be done using techniques such as transfer learning and online learning to improve model accuracy and reliability.

5. Results and Screenshot

The figure 5 shows the output screen of the proposed system. It is the home page in which we get registration and login button. We use Django framework for front-end. Once we start the Wamp server, we get this page http://127.0.0.1:8000/home in our localhost port

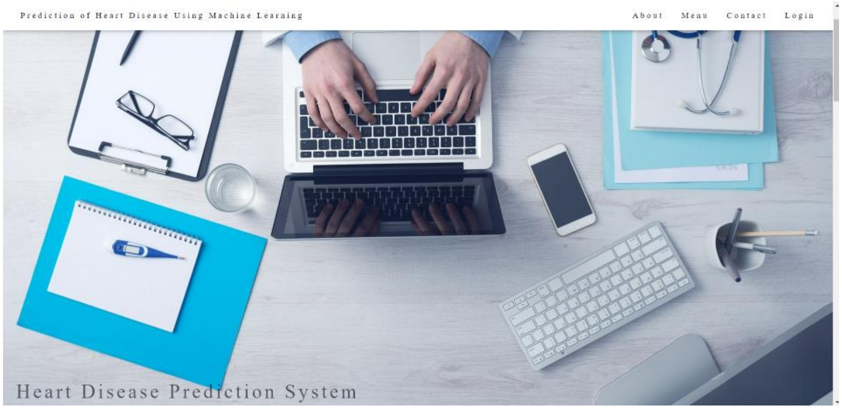


Fig.5- Home Page

The figure 6 shows the Login page for the Heart disease detection.

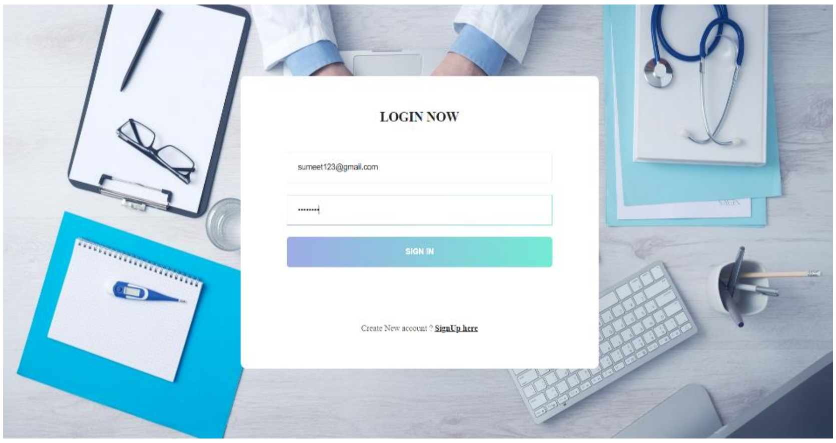


Fig.6- Login Page

The figure 7 shows the admin login

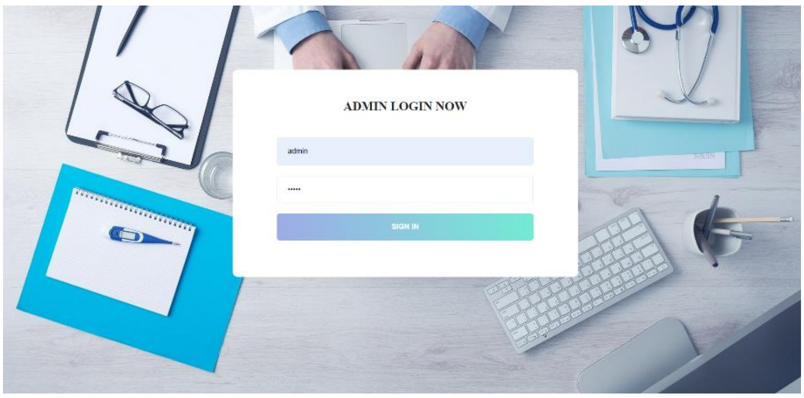
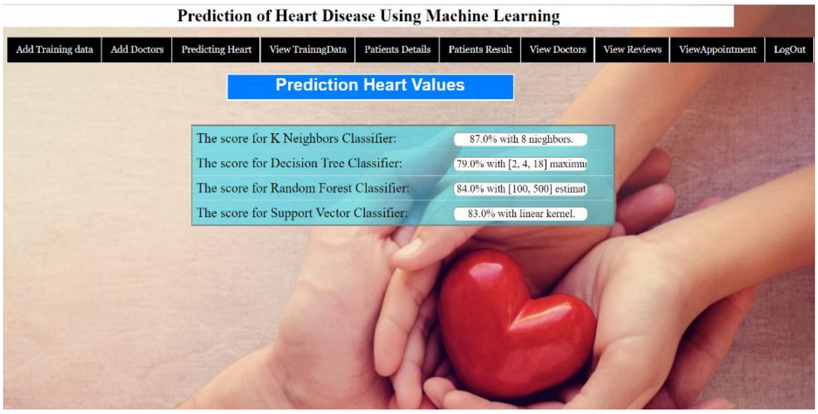


Fig.7- Admin Login.

The figure 8 Comparion of all algorithms and results.



5. Conclusion

Identifying raw health data processing of cardiac information can save lives in the long term and help detect abnormalities in heart disease early. In this work, we used machine learning techniques to process the raw data to enable new and novel identifications related to heart disease. Predicting heart disease is difficult and very important in medicine. However, if the disease is detected early and preventive measures are taken as soon as possible, mortality can be greatly reduced. It would be highly desirable to further extend this work to focus investigations on real-world datasets, rather than solely on theoretical approaches and simulations. A proposed hybrid HRFLM approach is used. It combines properties of random forests (RF) and linear methods (LM). HRFLM was found to be fairly accurate in predicting heart disease. Future courses of this research can be done in a variety of combinations, from machine learning techniques to better predictive techniques. In addition, new feature selection methods can be developed to broadly recognize important features and improve the performance of cardiac disease prediction

References

[1] S. Mohan, C. Thirumalai, Gautam. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," IEEE Access, Vol. 7, pp. 81542-81554, 2019.

[2] A. Ur. Haq, J. Li, J. Khan, M.H. Memon, S. Parveen, W. Akbar, T. Ahmad, S. Ullah, "Determining the predictive power of machine learning classifiers for designing cardiac disease detection systems," Proc.16th int. The law of the conference wavelet to compute. media technology. Inf. Process, December 2019, pp. 130-138。

[3] X. Liu、X. Wang、Q. Su、M. Zhang, Y. Zhu, Q. Wang, and Q. Wang, "A hybrid classification system for the diagnosis of cardiac disease based on the RFRS method," Comput. Mathematics. Methods Med., Vol. 2017, pp. 1-11, January 2017.

[4] S. Ghwanmeh, A. Mohammad, A. Al-Ibrahim, "Innovative Artificial Neural Network-Based Decision Support System for Cardiac Diagnosis," vol. 8, No. 2, pp. 150-154, 2011.

[5] A. U. Haq、J. Li、M. H. Memon、M. H. Memon, J. Khan, and SM Marium, "A Cardiac Disease Prediction System Using Machine Learning Models and Sequential Backward Selection Algorithms for Feature Selection," Proc.IEEE 5th Int. conf. Technology March 2019, p. 1-4.

[6] R. Detorano, A. Janosi, W. Steinbrun, M. Pfisterer、J.-J. Schmid, S. Sandhu, K. H. Guppy, S. Lee and V. Froelicher, "International application of new probabilistic algorithms for the diagnosis of coronary artery disease," Amer. J. Cardiol. , vol. 64, no. 5, S. 304–310, 8. 1989.

[7] R. Shivaranjani, V. S. Naresh and N.V. Murthy, "Predicting Coronary Artery Disease Using Decision Trees Based on 4 Genetic Algorithms," Intel. Decision Making Support system, application. Signal Processing, vol. 4, p. 71, October 2019.

[8] E.O.Oranii, O.K.Oidtun, K. Adnan, "Diagnosis of Cardiac Disease Using Neural Network Arbitration," Int. J. Intel. system, vol. 7, no. 12, p. 72, 2015.

[9] F. Fleuret, "Fast binary feature selection with conditional mutual information", J. Learning to do. Res., Vol. 5, pp. 1531-1555, Nov. 2004.

[10] R. Alzubi, N. Ramzan, H. Alzubi, A. Amira, “Hybrid Functional Selection Method for SNPs in Complex Diseases,” IEEE Access, vol. 6, pp. 1292–1301, 2018.

[11] A. Ul Haq, J. Li、Z. Ali、J. Khan、M. H. Memon、M. Abbas und S. Nazir, “Recognition of Parkinson's Disease Using a Hybrid Feature Selection Approach,” J. Intell. Fuzzy Systems, BD. 39, S. May 1-21, 2020 Doi:10.3233/JIFS-200075.

[12] R. Tibshirani, “Regression reduction and selection with the lasso,” J. Roy. statistics Soc. , B, Methodol. , vol. 58 No. 1, S. 267–288, January 1996.