AN EXPLORATION OF HEART DISEASE PREDICTION ALGORITHM USING MACHINE LEARNING TECHNIQUES

Chaitra shree S
M.Tech., Scholar, Dept. of CSE,
The Oxford College of Engineering,
Bengaluru
Chaitrashree1805@gmail.com
Dr. E. Saravana Kumar
Professor, Dept. of CSE,
The Oxford College of Engineering,
Bengaluru
saraninfo@gmail.com
Dr. R CH A Naidu
Professor-HOD, Dept. of CSE,
The Oxford College of Engineering,
Bengaluru
chanr789@gmail.com

Publication History
Manuscript Reference No: IJIRAE/RS/Vol.09/Issue08/SPAUAE10080
Research Article | Open Access
Peer-review: Double-blind Peer-reviewed
Article ID: IJIRAE/RS/Vol.09/Issue08/SPAUAE10080
Received Date: 12, July 2022 | Accepted Date: 25, July 2022 | Available Online: 12, August 2022
Volume 2022 | Article ID SPAUE10080 https://www.ijirae.com/volumes/Vol9/iss-08/01.SPAUAE10080.pdf
Article Citation:Chaitra,Saravana,RchA(2022).An Exploration of Heart Disease Prediction Algorithm using Machine Learing Techniques. International Journal of Innovative Research in Advanced Engineering (Vol. 9, Issue 8, pp. 158–163). AM Publications, India
doi: https://doi.org/10.26562/ijirae.2022.v0908.01
BibTeX key
Academic Editor-Chief: Dr.A.Arul Lawrence Selvakumar
Copyright: ©2022 This is an open access article distributed under the terms of the Creative Commons Attribution License; Which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract: In Digital World, Heart Diseases play a vital role in the medical field. Heart Diseases caused due to pressure or emotional behavior. In developing countries like India, Heart Specialists are needed in semi-urban and rural areas with low financial support. So, Heart Prediction plays an important aspect in using machine learning with a classification algorithm. This paper explores the novelty of researchers' predictive techniques to determine heart diseases using machine learning. This paper mainly explored three concepts of predictive techniques as Feature Selection, Classification Algorithm, and Evaluation of parameters. Classification of Algorithms for predictive techniques for heart diseases that provide effective and efficient performances.
Keyword: Heart Diseases, Feature Selection, Classification Algorithm, Predictive Techniques.

I. INTRODUCTION

With the enhancement of the statistics generation, computer-aided structures generate large quantities Of un cooked information, improving the fashion able middle of power. Acquiring essential information from this shape of information is a hard undertaking for practitioners. Data mining, Artificial Intelligence, device gaining knowledge, and deep gaining knowledge are quite contemporary and promising generations for acquiring relationships orfiguring out notable databases the usage of superiorstatistical approaches. Medical information mining and information exploration represent a quite new and growing area this is a hobby for many researchers.

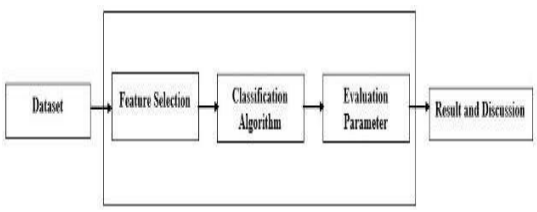

Figure1: Prediction of Heart Diseases.

Embedded strategies integrate the blessings of easy out and wrapper strategies. Here, algorithms consist of their feature choice criteria. By collectively with feature choice to the training process, those strategies decrease computational complexity. This technique is commonly appreciably greater correct at prediction and further green than easy out and wrapper methods. Lasso regression, Ridge regression, and memetic (Dissanayake, 2021) set of regulations are examples of embedded methods. The cause of this has a take a observe is to decide the impact of severing a features election algorithm categorized as an easy out, wrapper, and embedded strategies on enhancing the prediction of coronary heart disease.

Existing Prediction Techniques

The Existing Prediction Techniques to determine the level of heart disease with an appropriate dataset lead to a major role in the medical field. The researchers mainly focused (Rani, 2021) on machine learning techniques to predict the level of heart disease with their novelty approach. The Major unique techniques to predict the level of heart disease as shown below in figure 1. Most Researchers implement artificial intelligence techniques to determine the level of heart disease which acts as a superset of (Yewale, 2022) Machine Learning technique. Trends in the emerging predictive level of heart diseases focused on Machine Learning techniques. This Machine learning technique (Prasanna, 2022) consists of three processes such as input, hidden, and output process.

Input Process: A raw dataset from the heart special is to hospital was taken as input. These dataset consist of several attributes with in consistency. Data Processing techniques are used to optimize the in consistent data to consistent data using (Sowjanya, 2022) SMOTE techniques.

Hidden Process: The Consistency dataset act asinput to the hidden process. These hidden processes were further classified into three processes (Sekar, 2022) Feature Selection, Classification algorithm, and Evaluation parameter based on the confusion matrix.

Output Process: Predicting the level of heart disease (Mohapatra, 2022) with a final report termed to be an output process. This report provides preventive measures for heart patients with physicians' support.

A Probe on Dataset

Consider three researchers, who determine the dataset which acts as input for their novelty to determine the heart disease prediction using machine learning techniques.Table1 represents the dataset considered to predict the level of heart diseases as follows.

Authors	Pooja Rani et al (2021)	Fumin Xu et al(2022)	Norma Latif Fitryani et al(2020)
No. of Parameters	14	14	13
Parameter List	1. Age 2. Sex 3. Types of Chest Pain 4. Resting blood pressure 5. Serum cholesterol 6. Fasting blood Sugar 7. Resting Electrocardiographic result 8. Maximum heart rate 9. Exercise Induced argina 10. ST Depression 11. ST Segment 12. Number of Major Vessels 13. Thallium 14. Target	1. Age 2. Sex 3. Onset time 4. Comorbidity 5. Hypertension 6. Diabetes 7. Coronary Heart Disease 8. Chronic Obstructive Lung Disease 9. Carcinoma 10. Blood Cell Count 11. Inflammatory Marker 12. Liver Functions 13. Renal Functions 14. Chronic Kidney Diseases	1. Age 2. Sex 3. Cp 4. Trestbps 5. Chol 6. Fbs 7. Restecg 8. Thalach 9. Examang 10. Oldpeak 11. Slope 12. Ca 13. thal

Table1: Comparison of Dataset

Attributes were used to determine heart disease prediction by the majority of the researchers using machine learning techniques.

A Review on Feature Selection

In a predictive model, the process of minimizing the number of variables that act as input is called Feature Selection. This Feature Selection process which used to optimize the constructive model cost and also leads to improving the performance of the constructive model. An Effective Three Feature Selection process was to Reduce Noisy data, Improve the accuracy level and reduce the timing of complex algorithms. According to the exploration of heart diseases, two algorithms to be considered as shown in Figure2.

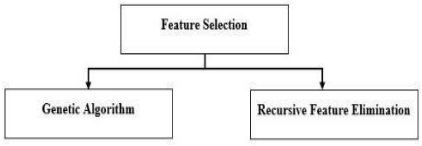

Figure2. Feature Selection

Genetic Algorithm:

A process of Natural selection implemented behind the Genetic algorithm. Multiple solutions are generated from a single source generation. Every Solution is considered to be a chromosome. A Population is termed as a collection of the solution generated by a single source. This Genetic Algorithm (Arabasadi Z, 2017) is an iterative process to filter the optimal solution from the feasible solution. Genetic Algorithms are classified into three processes Selection, Crossover, and Mutation.

Selection operator which used to determine the optimized solution from the feasible solution generated by the Single Generator.

Crossover provides a better individual solution for the future generation by combining the two or more individuals generated by single generator. This Cross over further classified into single point, two point and multi point.

Mutation: Next Generation of an individual may be same as the previous generation, where there generation was not unique term edas Mutations.

Recursive Feature Eliminations

Recursive Feature Elimination is the process of removing the irrelevant features information during the iterative process. A Classifier is processed to train the training data set and determine the features. Each iterative process, the weakest features to be removed and once again the remaining to trained for the next generation. This iterative process (Priscila SS, 2017) would be terminated until the desired number of features achieved. These number of features to be trained and the passed as parameter to the proposed algorithm.

Exploration of Classification Algorithm

Exploration of Classification Algorithm based onthe predication discussed as shown in the following Figure3.

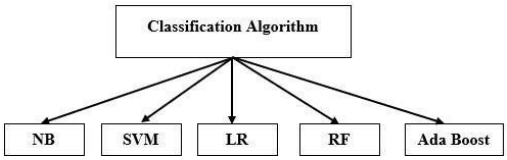

Figure 3. Classification Algorithm based on Predictions

Naïve Bayes Algorithm (NB)

Naïve Bayes prediction algorithm mainly focused with conditional probability. It used to classify certain processor task based on the conditional probability. It used to classify various categorization of documents, filtering based on spam and diagnosis of certain disease such as heart, chronic kidney diseases. An Efficient role of Naïve Bayes algorithm (UN, 2018), feature based on independence used for the prediction process.

Support Vector Machine (SVM)

Support Vector Machine perform the classification based on the hyper plane which act as an interface between the two sides. All the Samples belongs (Rani P, 2020) to some category belongs to one side and other category or group of population belongs to other side as shown in Figure 4.

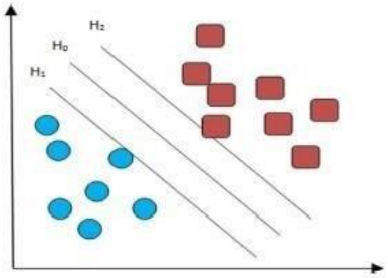

Figure 4.Hyper plane using SVM

It ensures the hyper plane to determine the distance between two classes based on their classification. Support Vector Machine will determine the nearest to hyper plane for further process.

Logistic Regression (LR)

Logistics Regression mainly focused with problem of Binary classification to predict the values. The predictable values classified into two such as “yes”or“no”.Yes, refer to 1 and no refer to 0. It may also have more than two values of possibility. Logistic Regression used to determine the classification based on two or more than two values.

Random Forest (RF)

Random Forest used to combine the process of decision tree to predict and increase the capability. The process of Random Forest (Jabbar MA, Springer) mainly focused with individual and trained independently. Each and Every Tree decision tree trained on sample data. During the construction of Random Forest, feature also randomly selected to be considered.

Ada Boost

A Boosting algorithm based on the adaptive techniques referred as Ada boost. It’s also mainly focused with ensemble techniques. These Techniques used to improve the performance of the weak learner. This algorithm used to train the original dataset with classifier. Classifiers are trained with multiple copies. Each Copy used to determine the error from the previous generation which leads to weak classifier. This Weak Classifier is combined to provide a strong classifier based on the cost functions. At Final prediction, a classifier with higher accuracy considered to be an (Latha CBC, 2019) optimal solution.

Optimization on Parameter Evaluations

Evaluation of the Prediction algorithm mainly focused with (Mienye ID, 2020) confusion matrix. The performance based on the classifier were determined by accuracy, sensitivity, precision, specificity and F-Measure based on confusion matrix with appropriate algorithm represented in the below Figure.5.

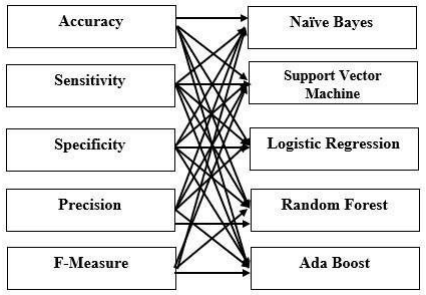

Figure5. Parameters Evaluations

- **Accuracy:** Accuracy can be measured by Division of Corrected Prediction by Total prediction multiples of 100.
- **Sensitivity:** Determined by (True Positive / True Positive + False Negative)*100
- **Specificity:** estimated by (True Negative /True Positive + False Negative)*100
- **Precision:** calculated by (True Positive /True Positive +False Positive) *100
- **F-Measure:** 2*((Sensitivity *Precision)/(Sensitivity + Precision))

Comparison of Results and Discussion

From the parameter evaluation, the following table were generated based on the data set taken from the private hospital as shown in Table 2.

Classifier	Accuracy	Sensitivity	Specificity	Precision	F-measure
NB	85.65	82.64	87.9	86.98	84.84
SVM	84.96	81.33	87.28	86.94	83.89
LR	83.64	80.96	85.48	85.87	82.75
RF	83.66	81.24	85.61	85.87	83.92
Adaboost	82.32	79.64	85.97	84.2	81.24

Table2: Parameter Evaluation.

Accuracy: The Following graph generated to determine the accuracy as shown in Figure 6, from the given table 2.

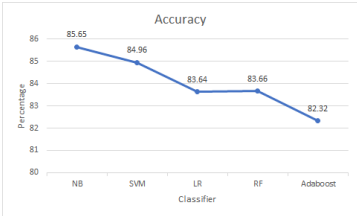

Figure 6: Accuracy

Naïve Bayes Algorithm provides more accuracy than the remaining algorithm. Sensitivity: The Following graph generated to determine the Sensitivity as shown in Figure7, from the given table2.

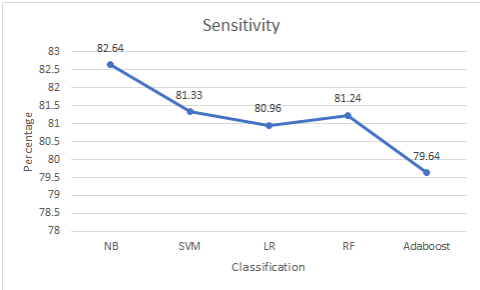

Figure 7: Sensitivity which higher efficiency with NaïveBayes algorithm and lower efficiency may be Ada boost Algorithm.

Specificity: The Following graph generated to determine the Specificity as shown in Figure8, from the given table2.

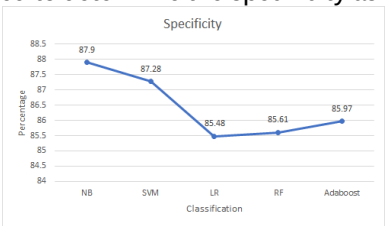

Figure8: Specificity

Lower specificity with logistic Regression and Higher specificity with Naïve Bayes Algorithm Precision: The Following graph generated to determine the Precision as shown in Figure 9,from the given table2.

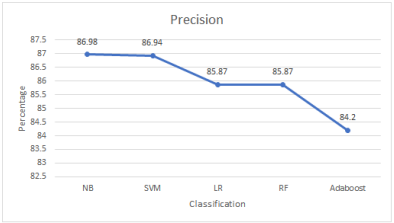

Figure 9: Precision

Precision is with little difference between each algorithm. The difference between the algorithms is a minimum of 2%. F-Measure: The Following graph was generated to determine the F-Measure as shown in Figure10, from the given table2.

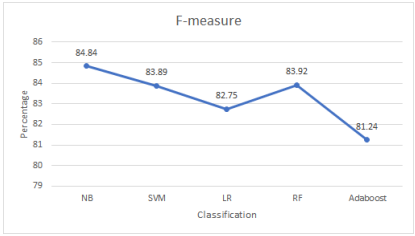

Figure10: F-Measure

F-Measure efficiency different with minimum or less than % when compared with different types of classifiers. The Performance of the prediction based on the classifier was analyzed from the above graphs resulting that, NB,RF, and Ada boost predicting that performance was static. NB,RF, and Ada boost performance resulted that no changes during the parameter evaluations based on the confusion matrix. The Performance of LR and SVM results that the performance changes based on each parameter.

CONCLUSION

This paper concludes with three processes of predictive techniques as Feature Selection, Classification Algorithm, and Evaluation of Parameters. Feature Selection plays significant role in determine the population, chromo some, cross over and mutation. Classification of Algorithm such as NB, RF,LR,SVM and ada boost plays their novelty in determine the prediction of heart diseases using Machine learning techniques. All the classification of algorithm in predicting the heart disease was unique and performance was approximately same. This paper concludes that, the various classifier predictive technique was unique and similar performance in predicting the heart diseases.

REFERENCE

1. Arabasadi Z,A.R.(2017).Computer aided decision-making for heart disease detection using hybrid neural network Geneticalgo.
2. Dissanayake, K. a. (2021). Comparative Study on Heart Disease Prediction Using Feature Selection Techniques on Classification Algorithms. Applied Computational Intelligence and Soft Computing.
3. Jabbar MA,D.B. (Springer). Prediction of heart disease using random forest and feature subset selection. Innovations in bio-inspired computing and applications,2016.
4. Latha CBC,J.S. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. InfMed Unlocked.
5. Mienye ID,S.Y. (2020). Improved sparse auto encoder based artificial neural network approach for prediction of heart disease.InfMed Unlocked.
6. Mohapatra,D.S. (2022). Distribution preserving train-test split directed ensemble classifier for heart disease prediction. International Journal of Information Technology.

7. Prasanna, K. a. (2022). Building an Efficient Heart Disease Prediction System by using Clustering Techniques. In High Performance Computing and Networking, Springer, Singapore.
8. PriscilaS S,H.M. (2017). Improving the performance of entropy ensembles of neural networks(EENNS) on classification .International Journal of Pure Applied Mathematics.
9. Rani P, K. R. (2020). Taxonomy of machine learning algorithms and its applications. Journal of The ror Nanoscience.
10. Rani, P. R. (2021). A decision support system for heart disease prediction based upon machine learning. Journal of Reliable Intelligent Environments, 263-275.
11. Sekar, J. P. (2022). An efficient clinical support system for heart disease prediction using TANFIS classifier. Computational Intelligence.
12. Sowjanya, A. M. (2022). Effective treatment of imbalanced datasets in healthcare using modified SMOTE coupled with stacked deep learning algorithms. Applied Nano science.
13. UN,D.(2018).Prediction system for heart disease using Naive Bayes and particle swarm optimization. Biomedical Resource.
14. Yewale,D.a.(2022). Comprehensive review on machine learning approach for heart disease prediction: Current status and future prospects. In AIP Conference Proceedings.