IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 7, July 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007096 |

Heart Disease Prediction Using Majority Voting

Suryanshoo Kaushal, Dr. L Anand

Department of Networking and Communications, School of Computing, SRMIST, Kattankulathur (T.N), India

ABSTRACT: Many healthcare companies have grown into major businesses that generate large amounts of health-care data on a daily basis that can be used to extract data for predicting ailments that a patient may face in the future. The patient's prescription history as well as their health information are examined. This hidden information can be used to make better decisions in the future when treating the patient. This is an area where there is still a lot of room for improvement. The accuracy of algorithms including Logistic Regression, Decision Trees,
Random Forest, Support Vector Machines (SVM), Nave Bayes, and K-Nearest Neighbour (KNN) is investigated inthis paper. The following characteristics, such as hypertension, hyperglycemia, smoking, BMI, work type, and so on, are used to forecast whether patients are at risk for heart disease. The goal is to develop a disease prediction model that is based on machine learning. Any single machine learning analysis of healthcare data cannot be relied on to produce accurate results. As a result, the goal is to investigate machine learning-based strategies for forecasting cardiac disease and to ensemble the best learners. The Logistic Regression, SVM, and KNN algorithms outperform the others. To forecast heart disease, our model uses a weighted average technique and majority voting.The findings are compared to diagnostic reports generated by medical laboratories. With an accuracy of 98.12 percent, the Ensemble-based classifier trumps the other classifier.

I. INTRODUCTION

Machine learning is a type of artificial intelligence that allows computers to learn without having to be programmed. In the process of training and prediction, specific algorithms are used. The training data is fed into an algorithm, which then runs through it and makes predictions on a new set of test data. Machine learning is a topic of study divided into three subcategories: reinforcement learning, unsupervised learning, and supervised learning.
The system is taught using well-labelled data in supervised learning. Following that, the machine is fed training data, which the algorithm analyses and delivers the desired result. Unsupervised learning,on the other hand, has nolabels. The algorithm reacts to the data without any prior training or instruction. The system then sorts the unsorted data into categories based on patterns, similarities, and differences. There is no prior training of the data. The reinforcement agent chooses the optimum course of action for completing the task.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007096 |

II. LITERATURE REVIEW

S.NO	TITLE	AUTHOR	DESCRIPTION
1.	Prediction of heart disease using Data Mining techniques.	*S. Kiruthika Devi, S. Krishnapriya, Dristipona Kalita*	The research paper aims at hybridization of various data mining techniques so as to improve the prediction of a heart disease. It is inferred that a hybrid model helps to make quick decisions and precise predictions, also that data mining is one of the best approaches used so as to predict heart disease.
2.	An extensive Investigation on Coronary Heart Disease using various Neuro Computational methods.	*D. Rajeswara Rao, Dr JVR Murthy*	The paper aims at the study of different methods by researchers with high accuracy in predicting the heart diseases. There are numerous systems in ANN ideas which are likewise contributing themselves in yielding most elevated expectation precision over medical information.
3.	C4.5 classification algorithm with backtrack pruning for prediction of heart disease.	*Jothikumar R, Siva Balan RV*	The paper focuses on prediction of the heart disease by implementing Backtrack Pruning technique. This technique is applied to construct a decision tree using non categorical attributes and concept of information entropy. The information gain and entropy are calculated for pruning the decision tree.
4.	Survey on Data Mining Algorithms in disease prediction.	*V. Kirubh, S. Manju Priya*	The paper aims to analyze application of data mining in medical domain and algorithms used to predict diseases. it is observed that results vary for different disease diagnosis based on tools and techniques used.
5.	Comparative study on prediction of Heart Disease using Cluster and Rank based approach.	*K. Aravinathan, Dr. M. Vanitha*	The paper introduces an approach which uses NN algorithm to improve the accuracy of the process. They used ANN to find missing historical data dependencies. They concluded that ANN works best.
6.	Efficient Heart Disease prediction system.	*Purushottama,c*, Prof. (Dr.) Kanak Saxena, Richa Sharma*	The aim of the paper was to propose a system that made use of KEEL principles or rules for prediction of heart disease. KEEL is an open source Java programming apparatus to implement developmental process for Data Mining issues.

The majority of the study articles we looked at had flaws in them or were slower. The cardiac disease must be classified with highest accuracy and low background error in our use case. We'll start with fundamental categorization and work our way up to dynamic classification. We now understand how the algorithm for Majority Voting works. We've also learned about the major strategies Majority employs to forecast heart disease. In majorityvoting, heart disease detection is done as a regression problem, which yields the input's class probabilities. A votingensemble is a machine learning model that integrates predictions from a number of different model.

IV. METHODOLOGIES

The figure depicts the design of our model. The information of patients is obtained by the hospital dataset, which undergoes a series of pre-processing steps like cleaning of data, transformation of data in appropriate form, reduction of data etc. Label encoder, which converts data into numerical values which are better understood by themodels, is majorly

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007096 |

used for pre-processing. Raw data also holds a major part of the pre-processing process. Now the data set is divided into training and testing data sets. The trained data is then analysed by using various supervised machine learning algorithms like Decision Tree, Naive Bayes, Random Forest, Support Vector Machine, Logistic regression andK-Nearest Neighbour. A comparative analysis is performed for the mentioned algorithms so as to figure out the best and most accurate out of all. On the basis of the comparison results, the top most accurate algorithms are used and implemented in the Ensemble method. The ensemble model makesuse of a voting classifier so that the model with highest accuracy can be obtained from it. The testing data is responsible forpredicting the disease by selecting the best model by accuracy.

I. Algorithm

1: Select a dataset that interests you, or develop your
2: Create a data set for training. Assigning pathways and defining categories (labels), will be part of preparingour dataset for training.
3: Collect data for training, the data should be consistent and apply data features to get the best results. 4: Rearrange the Dataset
5: Implement various algorithms to understand which model best supports our use case. 6: Compare them based on various parameters such as ROC Curve, confusion matrix etc. 7: Separate X and Y forCNN.
8:Compile, define, and train the CNN Model9: Model Accuracy and Score
Fig 5 : Union Over Intersection

II. Architecture Diagram

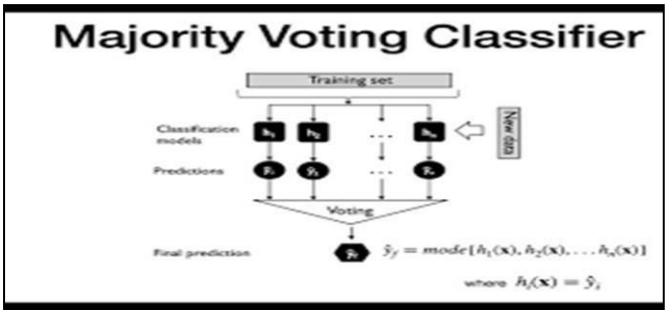

Fig 1 Architecture Diagram for Majority Voting Classifier

The purpose of the project is to increase the accuracy of the most efficient technique, for the prediction of heart stroke disease, to such an extent that it is higher than any other algorithms used. To achieve such a model we have used the Ensemble Learning approach, which combines models toproduce better results than single models would.
Majority Voting (Stacking), which is one of the methods used under Ensemble, is used to aggregate the results ofeach classifier passed to the voting classifier and calculates the output on the basis of the most number of votes. It forms a single combined model that predicts the output based on overall majority.
This approach gives us the highest accuracy results among other ensemble approaches, mainly boosting and bagging.
After performing the ensemble approach, an accuracy of 98.12% was achieved which hasn't been obtainedpreviously.

III. **Modules**

List of Modules:

- Data validation and pre-processing techniques (Module-01)
- Exploration data analysis of visualisation and training a model by given attributes (Module-02)
- Performance measurements of Supervised Machine Learning Techniques (Module-03)
- Performance measurements of Ensemble learning algorithms (Module-04)
- GUI based prediction of heart stroke disease (Module-5)

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007096 |

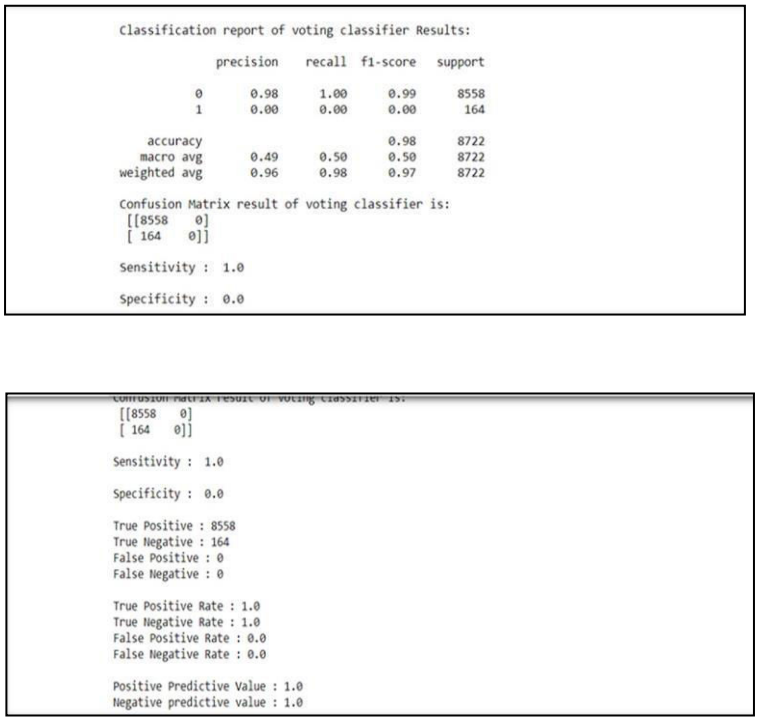

Fig 2. Outcome of voting classifier

V. UML DIAGRAM TO DEPICT WORKFLOW

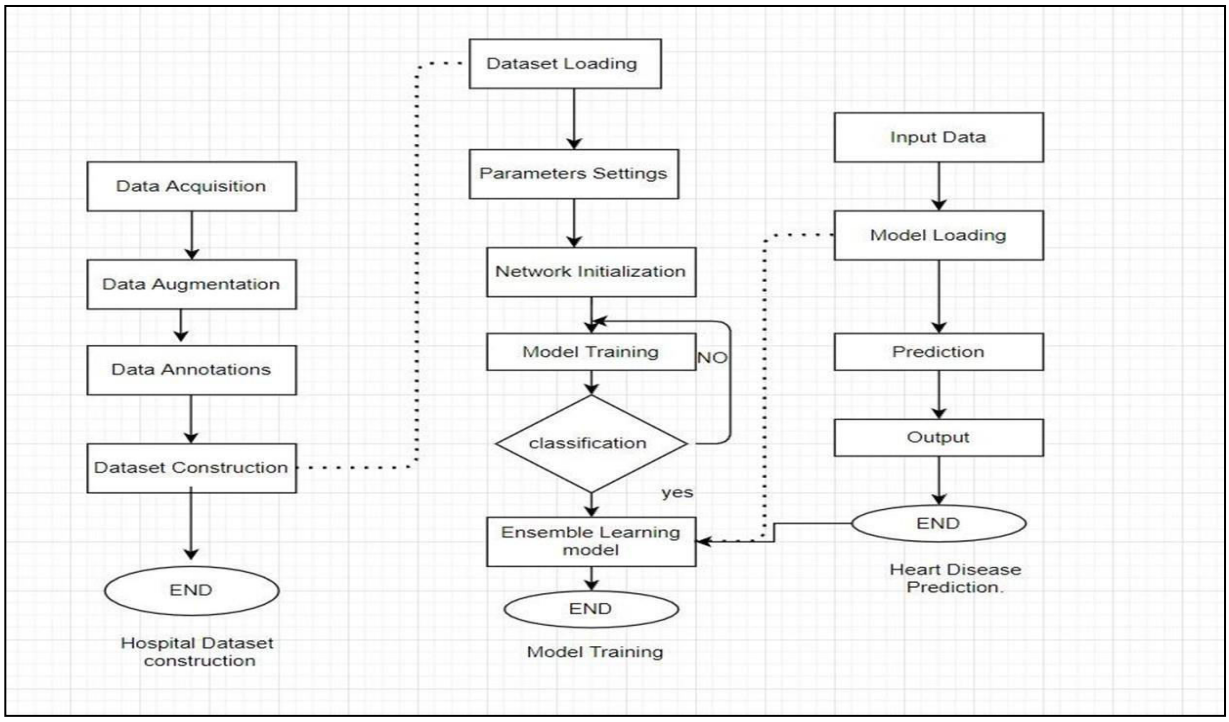

Fig.2 UML Diagram for disease classification

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007096 |

VI. RESULT

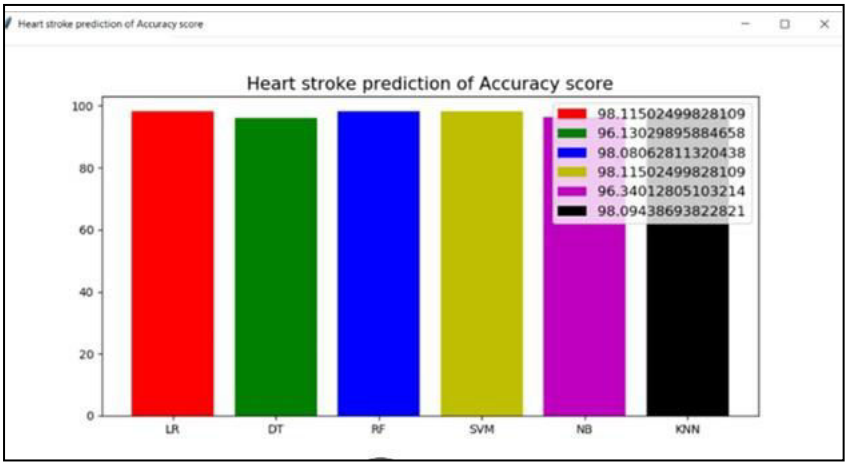

Fig 3. Graphical Representation of Accuracy Score:-

	Precision	Recall	F1-score	Support	Sensitivity	Specificity	Accuracy
LR	0.96	0.98	0.97	8722	0.999883	0.0	98.115024
DT	0.96	0.96	0.96	8722	0.976162	0.048780	96.078701
RF	0.96	0.98	0.97	8722	1.0	0.0	98.101265
SVM	0.96	0.98	0.97	8722	1.0	0.0	98.115024
NB	0.97	0.96	0.97	8722	0.979317	0.0134146	96.264446
KNN	0.96	0.98	0.97	8722	0.999766	0.0	98.104705

Fig 4. Comparison of algorithms on the basis of the following parameters

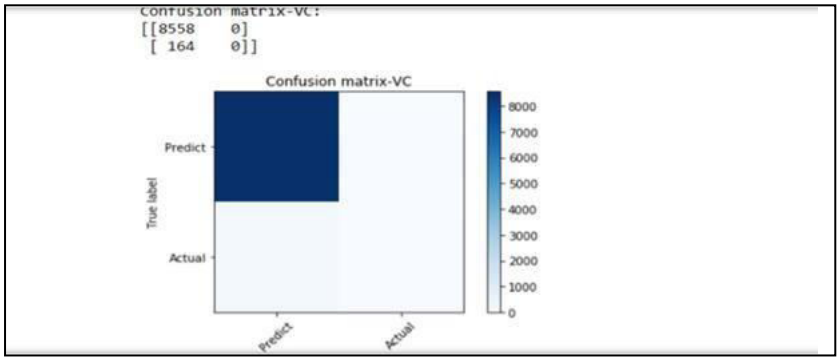

Fig 5. Confusion Matrix of Voting Classifier

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007096 |

Inferences

Our effort seeks to forecast heart stroke disease with the greatest degree of accuracy. We found the finest possible technique to achieve the same after conducting extensive research. Our preferred method was ensemble learning. Our project was unique in that it chose the best algorithms and then generated a model that improved our accuracyto 98.12 percent.

We were successful in completing our job by the time our report was due. In terms of strategy, we found that Majority Voting (Stacking) was the most effective. Anaconda and Jupyter Notebook were among the software requirements.

The nicest thing for us as partners was that we always managed to evenly split any tasks assigned to us and assistedeach other in resolving any issues that arose. We didn't feel the need to seek outside assistance.

We were able to finish our research, although we ran into several problems, such as selecting the most acceptable data set with a sufficient number of data labels. Additionally, working with such a massive data collection proved to be extremely taxing.

Because machine learning is such a broad topic, in-depth research took a long time and included a lotof theory.

Learning the functions and methods took some time because we were unfamiliar with numerous ideas and syntax in Python.

VII. CONCLUSION

There are a variety of algorithms that can be used to evaluate heart disease, and we evaluated a variety of algorithmsfor our use case and compared them on a variety of parameters such as confusion matrix, ROC Curve, and so on. We found that Majority Voting produced the best results because it combines a variety of algorithms and uses themfor prediction; we defined the weights of these algorithms according to their accuracy and precision, which increased the overall accuracy and precision of the final model.

REFERENCES

1. Nikhar, S., &Karandikar, A. M. (2016). Prediction of heart disease using machine learningalgorithms. International Journal of Advanced Engineering, Management and Science.
2. Thanigaivel, R., & Kumar, D. K. R. (2015). Review on heart disease prediction systems using datamining techniques. Asian Journal of Computer Science and Technology (AJCST) Vol, 3, 68-74.
3. Rao, D. R., & Murthy, J. V. R. (2016). An Extensive Investigation on Coronary Heart Disease using Various Neuro Computational Models. Global Journal of Computer Science and Technology.
4. Sharma, P., Saxena, K., & Sharma, R. (2016). Heart disease prediction system evaluation using C4. 5 rulesand partial tree. In Computational Intelligence in Data Mining—Volume 2 (pp. 285-294). Springer, NewDelhi.
5. Golande, A., Pavan Kumar T. (2019). Heart Disease Prediction Using Effective Machine Learning Techniques. International Journal of Recent Technology and Engineering (IJRTE), 8(1S4), 944-950.
6. Hazra, A., Mandal, S.K., Gupta, A., Mukherjee, A., Mukherjee, A. (2017). Heart Disease Diagnosisand Prediction Using Machine Learning and Data Mining Techniques: A Review. Advances in Computational Sciences and Technology, 10(7), 2137-2159.
7. Karthikeyan, T., Kanimozhi, V. (2017). Deep learning approach for prediction of heart disease using data mining classification algorithm deep belief network. International Journal of Advanced Research in Science, Engineering and Technology, 4(1), 3194-3201.
8. Karthikeyan, T., Kanimozhi, V. (2017). Deep learning approach for prediction of heart disease using data mining classification algorithm deep belief network. International Journal of Advanced Research in Science, Engineering and Technology, 4(1), 3194-3201.
9. Madhukesh D, Manish Kumar Giri, Nitesh Jaiswal, Nirmith S Rai and Chaithra (2019). Effective Heart Disease Prediction using Machine Learning. International Journal for Scientific Research and Development, 7(3), 920-924.
10. S.Kannadhasan and R,Nagarajan, Performance Improvement of H-Shaped Antenna With Zener Diode for Textile Applications, The Journal of the Textile Institute, Taylor & Francis Group, DOI: 10.1080/00405000.2021.1944523
11. Geetha, M., Ganesan, R., Sai, T.T. (2019). Research on the Machine Learning Algorithms on Heart Condition Predictions. International Journal of Innovative Technology and Exploring Engineering,

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details