International Journal of Inventive Engineering and Sciences (IJIES)
ISSN: 2319-9598 (Online), Volume-10 Issue-11, November 2023

Heart Disease Prediction Classification using Machine Learning

Shatendra Kumar Dubey, Sitesh Sinha, Anurag Jain

***Abstract:** Heart disease is a leading cause of mortality worldwide, and early detection and accurate prediction of heart disease can significantly improve patient outcomes. Machine learning techniques have shown great promise in assisting healthcare professionals in diagnosing and predicting heart disease. The diagnosis and prognosis of heart disease must be improved, refined, and accurate, because a small mistake can cause weakness or death. According to a recent World Health Organization study, 17.5 million people die each year. By 2030, this number will increase to 75 million.[2] This document explains how to enable online KSRM capabilities. The KSRM smart system allows users to report heart-related problems. This research paper aims to explore the use of machine learning algorithms for effective heart disease prediction classification with Adaboost for improve the accuracy of algorithm.*

***Keywords:** K-NN, SVM, RF, LR, MLP, DT, NB, Adaboost.*

I. INTRODUCTION

Heart disease encompasses a range of conditions that affect the heart's functionality, including coronary artery disease, congestive heart failure, and arrhythmias. Early identification of individuals at risk of heart disease allows for timely intervention and prevention of potentially life-threatening events[3]. Machine learning techniques, such as classification algorithms, can leverage patterns and relationships in large datasets to develop accurate prediction models. The importance of the selection process is that most scientists agree that there are many times when you or someone needs a doctor's help, but sometimes they just can't. KSRM is an online application. Here, users get instant heart disease counselling via online smart system[13][17][18][19][20]. Therefore, we have the right data preparation and we need a good selection process to get a high prediction of heart disease using key features and data mining experts. Although the choice is clearly difficult to combine the right data mining process with the right process. Then it touches on specific points for users to check for various diseases that may be associated with it.

Manuscript received on 12 October 2023 | Revised Manuscript received on 07 November 2023 | Manuscript Accepted on 15 November 2023 | Manuscript published on 30 November 2023.
* Correspondence Author (s)

Shatendra Kumar Dubey*, Ph.D., Research Scholar, Department of Computer Science Engineering, RNTU, Bhopal (M.P), India. E-mail: shatendradubey@gmail.com, ORCID ID: 0000-0002-2319-2039

Dr. Sitesh Sinha, Department of Computer Science Engineering, AISECT University, Bhopal (M.P), India. E-mail: siteshkumarsinha@gmail.com

Dr. Anurag Jain, Department of Computer Science Engineering, REC, Bhopal (M.P), India. E-mail:anurag.akjain@gmail.com

The KSRM model offers seven classification methods: k-Neighbors Neighbor (K-NN), Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), Multi-Layer Perception (MLP), Decision Trees (DT), Random Forest (RF) and Naïve Bayes (NB) to build predictive models. Data preprocessing and feature selection steps are performed before the model is created. Models were evaluated for accuracy, precision, recall, and F1 score. The SVM model performed best with an accuracy of 88%. For the application we use Weka Tool 3.8.6 which has various libraries and headers to make it work accurately and clearly[4].

Accurate diagnosis of heart disease is promising, but it may not be easy. In addition, the combination of key features will increase forecast accuracy. This indicates that extensive testing is required to identify key features to achieve this goal[5]. A comprehensive analysis and comparison of different experimental feature combinations and data mining techniques is not shown[14][21]. Therefore, all attempts should be made to accurately describe data mining techniques and key features to ensure heart disease predictions are accepted and accurate. The app has a lot of content, including content related to heart disease. The KSRM smart system allows users to report heart-related problems [1]. Then it touches on specific points for users to check for various diseases that may be associated with it. The system is available for online cardiology consultation. Today, healthcare providers around the world are using machine learning. In this study, the KSRM model offers seven classification methods: k-Nearest Neighbors (K-NN), Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), Multi-Layer Perception (MLP), Decision Trees (DT)), Random Forest (RF) and Naive Bayes (NB) for building predictive models[7]. In this research I used the Comparative Analysis of Classification Methods and retrieved the data from the UCI repository. After applying the Ada-boost[15] that also work on the stagewise addition method where multiple weak learners used for getting strong learner and improve the result up to 88%

Heart Disease Dataset (HDD) and Predictive Model

The UCI Heart Disease Dataset (HDD) dates back to 1988 and consists of four databases: Cleveland, Hungary, Switzerland, and Long Beach V. It contains 76 features, including the predictive features, but each of the published tests refers to the use of one of the following. 14 of them the "Target" field indicates whether the patient has heart disease. This is an integer value, 0 = no virus, 1 = virus.

Table 1: The Physical Examination Data of Heart Disease Patients

S. N	Attributes	Description	
1	Age	Patients age in years	Continuous
2	Sex	Sex of subject (male- 0,female-1)	Male/Female
3	CP	Chest pain type	Four types (1 typical angina,2 atypical angina, 3 non-anginal, pain 4 asymptomatic)
4	Trest bps	Resting blood pressure	Resting systolic blood pressure (in mm Hg on admission to the hospital)
5	Chol	Serum cholesterol in mg/dl	Serum cholesterol in mg/dl
6	FBS	Fasting blood pressure	Fasting blood sugar>120 mg/dl (0—false; 1—true)
7	Rest ecg	Resting Electrocardiograph	0—normal; 1—having ST-T wave abnormality; 2—left ventricular hypertrophy
8	Thalach	Maximum heart rate achieved	Maximum heart rate achieved
9	Exang	Exercise Induced Angina	Exercise-induced angina (0—no; 1—yes)
10	Old peak	STD depression introduced by exercise.	ST depression induced by exercise relative to rest
11	slope	Slope of Peak Exercise ST segment	slope of the peak exercise ST segment (1—upsloping; 2—fat; 3—down sloping)
12	Ca	Number of major vessels	No. of major vessels (0–3) colored by fuoroscopy
13	Thal	Defect type	Defect types; 3—normal; 6—fxed defect; 7—reversible defect
14	Targets	Heart disease	diagnosis of heart disease status (0—nil risk; 1—low risk; 2—potential risk; 3— high risk; 4—very high risk)

We propose to learn all previous methods and various estimation methods including (but not limited to) k-nearest Neighbours (K-NN), Support Vector Machines (SVM), Random Forests (RF), and Linear Regression (LR). Multi-Layer Perception (MLP), J48 for developing predictive models. We will use medical information in our educational work. The purpose of the above research is to develop a predictive model. The result will be lower defects and lower overhead. The main objective is to improve the prediction rate by applying a strong classifier. The clinical data set has been used for training and testing. The KSRM collects a large amount of medical data for overloading the process. So, we try to improve the overloading issues in KSRM. To minimize the false positive rate as well as the time consumption rate. To deal with all the above issues we tried to apply Naive Bayes, j48, MLP, KNN, SMO, Bayes net, and random forest classifiers for the classification of Heart data set utilizing open source data mining tools like WEKA.

This paper proposed to enhance the efficiency of the prediction of heart disease with the Adaboost boosting Algorithm to improve the accuracy utilizing a dataset

Pre-processing of Data: We need to clean and remove the missing or noise values fromthe dataset to obtain accurate and perfect results, known as data cleaning. Using some standard techniques in Weka 3.8.6, we can fill in missing and noise values[10]. Then we need to transform our dataset by considering the dataset's normalization, smoothing, generalization, and aggregation. Integration is one of the crucial phases in data pre-processing, and various issues are considered here to integrate. Sometimes the dataset is more complex or difficult to understand. In this case, the dataset needs to be reduced in arequired format, which is best to get a good result[6].

Proposed Architecture: The architecture of the proposed work is represented in the following diagram. As we can see the data set used is HDD data set. An intellectual Disease prediction System (KSRM) can be incorporated by means of data set. It is a group of records used to estimate the performance of DPS[8-9]. In the past years heart data set has been utilized by researchers to reveal the applicability and presentation of the knowledge discovery tasks. The process of counting the appropriate or significant variables or attributes for the system creation is known as feature selection[12].

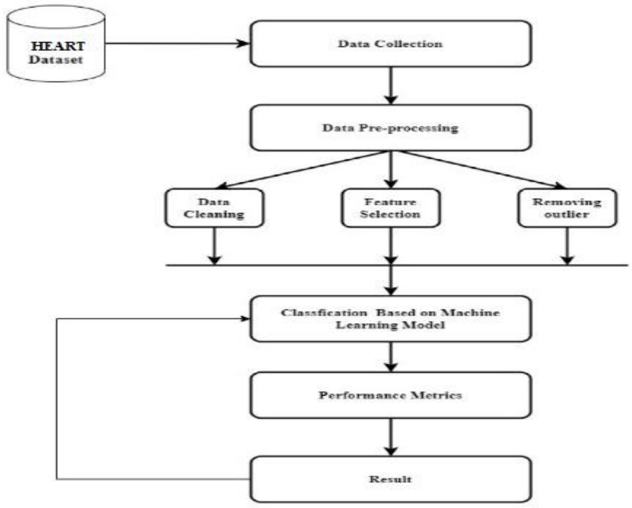

Figure 1. The Architecture of Intellectual Disease Prediction System (KSRM)

Classification is the process of assigning records to one of the two modules. The aim of classification tasks is to correctly classify the data into the exact class depending on the properties of the information. It is a data mining section that can allot records. For instance, a cataloging representation can be used to recognize loan applicants as small, intermediate, or elevated classes[11]. A categorization mission starts with facts set. They are predefined so that the unknown data can be classified easily.

Retrieval Number: 100.1/ijies.B43211213223
DOI: 10.35940/ijies.B4321.11101123
Journal Website: www.ijies.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Algorithm
Input: HDD Dataset
Output: WEKA Compatible. ARFF file in Which all Attacks are Classified Individually.
Step 1: Outlier removal from the dataset.
Step 2: If attack_read == „apache2" then // search for all „apache2" attacks
and replace attack by Cat1 in HEART dataset.
Step 3: else If attack_read == „back" then // search for all „back" attacks replace attack by Cat2 in HEART dataset // repeat this step for all individual attacks in HEART dataset.
Step n: else If attack_read == „xterm" then // search for all „xterm" attacks replace attack by Catn in HEART dataset.
Step n+1: else Mark as Normal in HEART cup dataset.
Step n+2: compile the file and save it as processed HDD. arff After getting HDD.arff, we execute it on the WEKA tool to verify the efficiency of the classification of our proposed method. For this purpose, we used BayesNet, Naive Bayes, MLP, SMO, J48, and Random Forests classifiers. An ensemble classifier is a method that uses or combines multiple classifiers to improve robustness as well as to achieve an improved classification performance from any of the constituent classifiers[16].

II. RESULT ANALYSIS

Analysis of Heart Disease Dataset Figure 2. Target class. Before studying the performance of considering machine learning algorithms in this research, an analysis of thefeatures of the heart disease dataset will be focused on here. The total attributes is 1025, where not having heart disease is 499 (denoted by 0) and having heart disease 526 (represented by 1), see figure 2 So, the percentage of not have heart disease is 45.7%, and the percentage of having heart disease is 54.3%, see Figure 2. It is shown that the rate of heart disease is more than the rate of no heart disease.

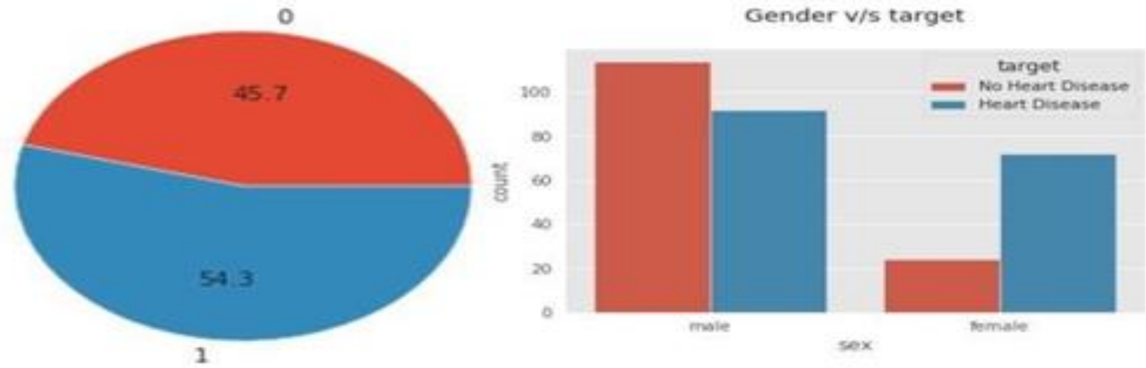

Figure 2. Comparison Chart Gender Vs Target

The accuracy rate is a correct prediction ratio to the total no of given data sets. It can be written an

ACCURACY = TP+TN/TP+TN+FP+FN

Where TP: True Positive TN: True Negative FP: False Positive FN: False Negative After performing the machine learning algorithms for training and testing the dataset,we can find a better algorithm by considering the accuracy rate. The aim of the entire project was to test which algorithm classifies heart disease the best with the proposed optimization methods. The classification experiment in this paper was carried out in a Weka environment. In addition, due to the small number of selected features, 10-fold cross-validation was used. For the purpose of avoiding unstable operationresults, each experiment was run 10 times, and the optimal classification accuracy was selected for comparison. We evaluate the effectiveness of all classifiers in terms of time tobuild the model, correctly classified instances, incorrectly classified instances, and accuracy. Table 2 shows the classification of correct and incorrect instances of normal algorithm and shows the prediction accuracy, Recall, and precision. The SVM shows the maximum accuracy, Recall and Precision in this Evaluation Table.

Table 2: Evaluation Table with All algorithm:

Evaluation Criteria	KNN	SVM	RF	Logistic	MLP	J48	NB
Correctly classified instances	203	227	220	226	211	207	226
Incorrectly classified instances	67	43	50	44	59	63	44
True Positive	115	131	126	131	117	119	131
False Positive	35	19	24	19	33	31	19
False Negative	32	24	26	25	26	32	25
True Negative	88	96	94	95	94	88	95
Prediction Accuracy	75.18	84.07	81.48	83.7	78.14	76.66	83.71
Recall	78.23	84.51	82.89	83.97	81.81	79.33	83.97
Precision	76.66	87.33	84	87.33	78	78.8	87.33

Retrieval Number: 100.1/ijies.B43211213223
DOI: 10.35940/ijies.B4321.11101123
Journal Website: www.ijies.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

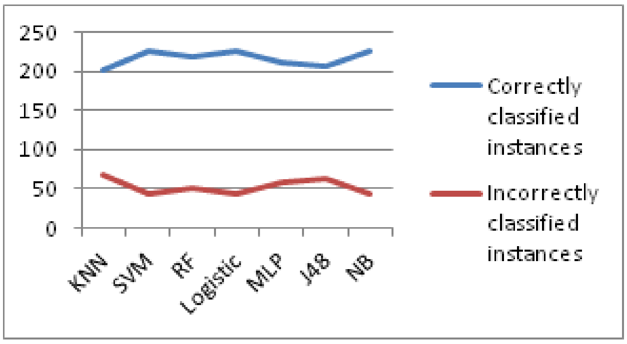

Fig. 3: Correctly and Incorrectly Classified Instances with All Algorithm.

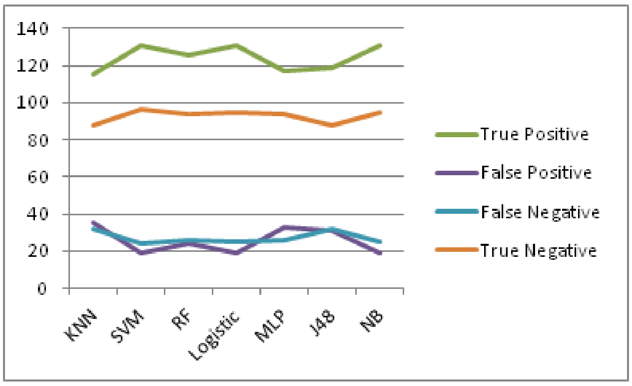

Fig. 4: Confusion matrix with evaluation table with All algorithm.

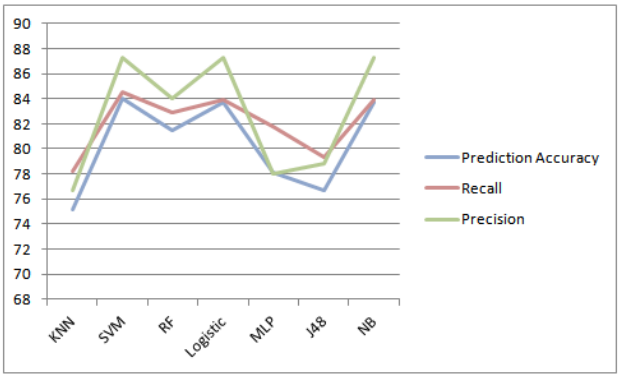

Fig. 5: Prediction Accuracy Recall and Precision with All Algorithm

After the implement of Adaboost on J48 and MLP

Table 3: Evaluation Table Normal Algorithm with Adaboost M1

Evaluation Criteria	AdaboostMLP	Adaboost J48
Correctly classified instances	228	237
Incorrectly classified instance	36	33
True Positive	134	142
False Positive	23	18
False Negative	19	15
True Negative	94	95
Prediction Accuracy	84.44	87.78
Recall	87.58	90.44
Precision	85.35	88.75

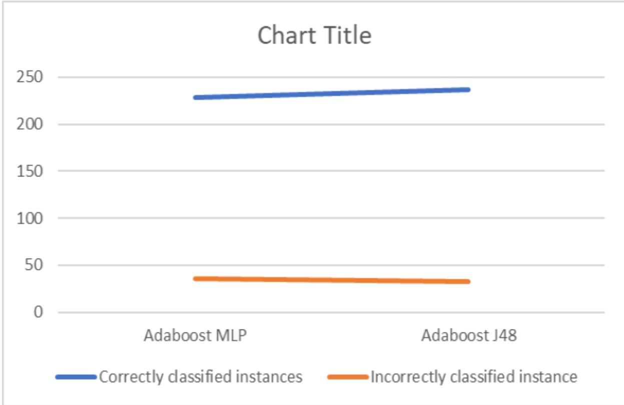

Fig. 6: Correctly and Incorrectly Classified Instances with Adaboost.

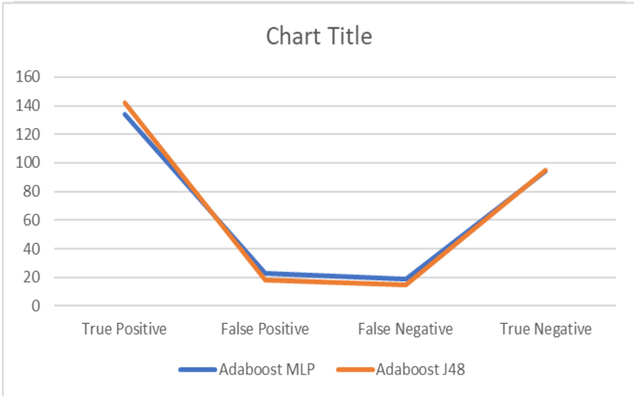

Fig. 7: Confusion Matrix with Evaluation Table with Adaboost

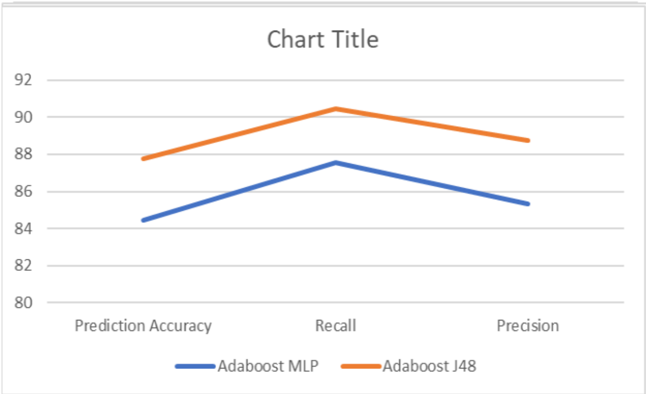

Fig. 8: Prediction Accuracy Recall and Precision with a normal algorithm

Table 4: Accuracy Table for both Algorithm with Adaboost

Comparative Study of Accuracy	
CLASS	Accuracy
MLP	78.14
J48	76.66
ADA MLP	84.44
ADA J48	87.78

Retrieval Number: 100.1/ijies.B43211213223
DOI: 10.35940/ijies.B4321.11101123
Journal Website: www.ijies.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

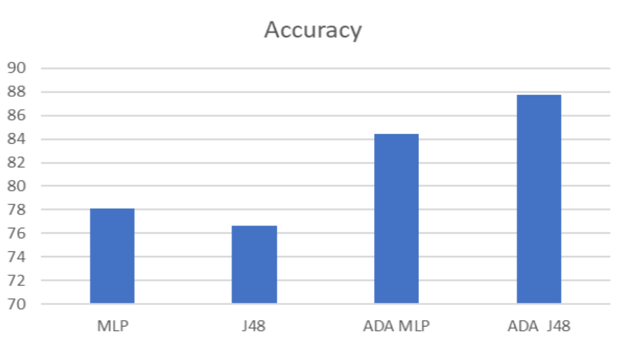

Fig. 9: Prediction Accuracy Recall and Precision with a Normal Algorithm

III. CONCLUSIONS

This research paper aims to contribute to the field of heart disease prediction by evaluating the effectiveness of different machine-learning algorithms. The findings will provide valuable insights into the development of accurate and efficient prediction models that can assist healthcare professionals in identifying individuals at risk of heart disease, enabling early intervention and improved patient outcomes. The research used data mining techniques on UCI-HDD data sets to provide the diagnosis results. The results from the approach were promising. We apply this algorithm to KNN, RF, SMO, Logistic, MLP, J48, and NB on HDD Data Set through the WEKA Tools. After compression of all algorithms accuracy of SMO and Logistic regression gave better results 84.07 and 83.7 respectively. After the Ada Boost on this algorithm of MLP and J48, the Results will be Improved and the Accuracy level will be increased through 84.44 and 87.78.

DECLARATION STATEMENT

Funding	No, I did not receive any financial support for this article.
Conflicts of Interest	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence.
Availability of Data and Material	Not relevant.
Authors Contributions	All authors have equal participation in this article.

REFERENCES

1 Sharma, H., & Rizvi, M. A. (2017). Prediction of heart disease using machine learning algorithms: A survey. International Journal on Recent and Innovation Trends in Computing and Communication, 5(8), 99-104.
2. Balakumar, P., Maung-U, K., & Jagadeesh, G. (2016). Prevalence and prevention of cardiovascular disease and diabetes mellitus. Pharmacological research, 113, 600-609. https://doi.org/10.1016/j.phrs.2016.09.040
3. Gradman, A. H., & Alfayoumi, F. (2006). From left ventricular hypertrophy to congestive heart failure: management of hypertensive heart disease. Progress in cardiovascular diseases, 48(5), 326-341. https://doi.org/10.1016/j.pcad.2006.02.001
4. Yin, K. S. (2019). Network Behavioral Analysis for Detection of Remote Access Trojans (Doctoral dissertation, MERAL Portal).
5. Long, N. C., Meesad, P., & Unger, H. (2015). A highly accurate firefly based algorithm for heart disease prediction. Expert Systems with Applications, 42(21), 8221-8231. https://doi.org/10.1016/j.eswa.2015.06.024
6. Tomar, D., & Agarwal, S. (2013). A survey on Data Mining approaches for Healthcare. International Journal of Bio-Science and Bio-Technology, 5(5), 241-266. https://doi.org/10.14257/ijbsbt.2013.5.5.25
7 Tasnim, F., & Habiba, S. U. (2021, January). A comparative study on heart disease prediction using data mining techniques and feature selection. In 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST) (pp. 338-341). IEEE https://doi.org/10.1109/ICREST51555.2021.9331158
8 Killana Sowjanya, Dr G. Krishna mohan (2020). predicting heart disease using machine learning classification algorithms and along with TPOT, International Journal of Scientific & Technology Research 9(4).3202-3210
9 Gokulnath, C. B., & Shantharajah, S. P. (2019). An optimized feature selection based on genetic approach and support vector machine for heart disease. Cluster Computing, 22, 14777-14787 https://doi.org/10.1007/s10586-018-2416-4
10. Chandra, I., Sivakumar, N., Gokulnath, C. B., & Parthasarathy, P. (2019). IoT based fall detection and ambient assisted system for the elderly. Cluster Computing, 22, 2517-2525. https://doi.org/10.1007/s10586-018-2329-2
11. Tomar, D., & Agarwal, S. (2014). Feature selection based least square twin support vector machine for diagnosis of heart disease. International Journal of Bio-Science and Bio-Technology, 6(2), 69-82. Tomar, D., & Agarwal, S. (2013). A survey on Data Mining approaches for Healthcare. International Journal of Bio-Science and Bio-Technology, 5(5), 241-266. https://doi.org/10.14257/ijbsbt.2014.6.2.07
12. Magesh, G., & Swarnalatha, P. (2021). Optimal feature selection through a cluster-based DT learning (CDTL) in heart disease prediction. Evolutionary intelligence, 14, 583-593. https://doi.org/10.1007/s12065-019-00336-0
13. Ali, F., El-Sappagh, S., Islam, S. R., Kwak, D., Ali, A., Imran, M., & Kwak, K. S. (2020). A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. Information Fusion, 63, 208-222. https://doi.org/10.1016/j.inffus.2020.06.008
14. Hossen, Mohammed Khalid. (2022). Heart Disease Prediction Using Machine Learning Techniques. American Journal of Computer Science and Technology. Vol. 5, No. 3, 2022,. 146-154. 10.11648/j.ajcst.20220503.11.
15. El Hamdaoui, H., Boujraf, S., El Houda Chaoui, N., Alami, B., & Maaroufi, M. (2021). Improving Heart Disease Prediction Using Random Forest and AdaBoost Algorithms. International Journal of Online & Biomedical Engineering, 17(11). https://doi.org/10.3991/ijoe.v17i11.24781
16. Rahman, M. M. (2022). A web-based heart disease prediction system using machine learning algorithms. Network Biology, 12(2), 64.
17. A., T., & C., M. (2020). A Survey on Deep Learning Architectures and Frameworks for Cancer Detection in Medical Images Analysis. In International Journal of Innovative Technology and Exploring Engineering (Vol. 9, Issue 11, pp. 28–34). https://doi.org/10.35940/ijitee.k7654.0991120
18. Kanani, P., & Padole, Dr. M. (2019). Deep Learning to Detect Skin Cancer using Google Colab. In International Journal of Engineering and Advanced Technology (Vol. 8, Issue 6, pp. 2176–2183). https://doi.org/10.35940/ijeat.f8587.088619
19. R, R., & S H, M. (2021). Plant Disease Detection and Classification using CNN. In International Journal of Recent Technology and Engineering (IJRTE) (Vol. 10, Issue 3, pp. 152–156). https://doi.org/10.35940/ijrte.c6458.0910321
20. Wanjau, S. K., Wambugu, G. M., & Oirere, A. M. (2022). Network Intrusion Detection Systems: A Systematic Literature Review o f Hybrid Deep Learning Approaches. In International Journal of Emerging Science and Engineering (Vol. 10, Issue 7, pp. 1–16). https://doi.org/10.35940/ijese.f2530.0610722
21. CNN based Image Classification Model. (2019). In International Journal of Innovative Technology and Exploring Engineering (Vol. 8, Issue 11S, pp. 1106–1114). https://doi.org/10.35940/ijitee.k1225.09811s19

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

AUTHOR PROFILE

Shatendra Kumar Dubey is a research Scholar at Ravindra Nath Tagore University, Bhopal. He is pursuing P. HD in Computer science and Engineering. I was completed my Masters from SATI Vidisha which is affiliated to Rajeev Gandhi Technical University, Bhopal. He has keen interest in Machine Learning, Natural Language Processing and Data Mining. I have more then 20 Research papers in reputed International Journals with good impact factor. In this article we prepared a smart online system to predict the of actual position Hearts Patients with the help of Machine Learning's Algorithm boosting with Ada-Boost with improve the accuracy and efficiency of work.

Retrieval Number: 100.1/ijies.B43211213223
DOI: 10.35940/ijies.B4321.11101123
Journal Website: www.ijies.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)