(RESEARCH ARTICLE)

Check for updates

Heart Disease Prediction using SVM and RF

Rashad Bakhshizada [1, 2, *]

[1] Department of Physics, Missouri University of Science and Technology, Rolla, MO, USA.
[2] Faculty of Physics, University of Warsaw, Warsaw, Poland.

Abstract

Heart disease is one of the top priority health issues in the world, and early detection is paramount to enhancing the efficacy of treatment and reducing complications. Recently, machine learning has shown promises in predicting heart disease. The present study examines predictions based on random forests, and support vector machines on realistic data sets. The Boruta algorithm is used to find all the significant attributes from the dataset with respect to an outcome variable. From 13 input attributes, 6 attributes are selected including age, chest pain, maximum heart rate, ST depression induced by exercise relative to rest, number of major vessels colored by flourosopy, Thallium stress test. In this research paper, the presence of heart disease is predicted by employing Random Forest, Support Vector Machine. The F1- Score, recall and precision score of Random forest classifiers are 84.60%, 85.12% and 84.82%, respectively.

1. Introduction

Technology integration is becoming an integral part of our daily life as a result of the technological advancement of various technologies [1]. This results in less manual work and aids in ubiquitously interconnecting everyone, and IoT plays a major role, offering smooth and seamless ubiquitous services for everyone [2,3]. In general, the IoT refers to the networking of physical devices that are smart and interconnected [4] and comprises sensors, software, and network connectivity that enables it to collect and exchange data [5,6]. Currently, the IoT is shaping and transforming both the business and consumer worlds, finding its way into every global business and consumer domain. Apart from this, it is also being delivered in many other domains, including healthcare, smart cities, agriculture, the military, and so on [4,5,6,7,8]. Hence, the IoT may significantly enhance the way people interact with the world. Based on recent reports, the IoT market size was valued at USD 761.4 billion in 2020 and is projected to reach USD 1386.06 billion by 2026, which signifies its importance as a dominant technological paradigm towards improving the well-being of billions of people all around the world [7,8].

When it comes to the IoT in healthcare, or what is well known as IoMT, it refers to a wide variety of IoT devices whose main purpose is to facilitate and aid in fundamental patient care [7,8]. the global IoT in healthcare market size is USD 71.84 billion in 2020 [9] and the market is projected to grow from USD 89.07 billion in 2021 to USD 446.52 billion by 2028 [9]. As of now, healthcare providers are utilizing various IoMT based applications and services for patient treatment, disease management, medical diagnosis, to improve patient care, and lower the costs of care, where they are capable of collecting various information such as vital body parameters from patients and monitor pathological details by implantable medical sensors or small wearable sensors that are worn by the patient. With the aid of IoMT devices, patient condition can be monitored remotely and in real-time, and the captured data can then be analyzed and transmitted to the cloud data storage or the medical data centers for further processing and storage before offering

* Corresponding author: Rashad Bakhshizada

services to various stakeholders such as physicians and other related medical staff, caregivers, and insurance service providers. In general, IoMT applications include solutions that are designed for remote health monitoring, emergency patient care, healthcare management, the monitoring of elderly patients, clinical decision support systems, wireless capsule endoscopy, and so on.

2. Literature Review

Ahmad and Polat proposed a heart disease prediction using the Jellyfish Optimization Algorithm. This shows better results in terms of accuracy and speed toward early diagnosis, thus helping doctors with early and efficient decisions in treatment [1]. Bahuguna et al. (2023) present a statistic analysis coupled with machine learning models for heart disease prediction, giving a robust comparison of algorithms and highlighting their effectiveness to identify key risk factors for heart disease [2]. Biswas et al. (2023) explore early-stage prediction, focusing on various feature selection techniques in heart disease. Their findings highlight the prime importance of feature selection for the optimization of predictive accuracy and model performance [3]. Kaushik et al. (2024) use AI-powered techniques for skin cancer detection, a method that could inspire similar approaches in heart disease prediction, emphasizing the interdisciplinary applications of AI in healthcare [4]. Boukhatem et al. (2022) utilize machine learning algorithms for heart disease prediction, providing a practical solution for healthcare providers [5]. Their work highlights how datadriven approaches can assist in early diagnosis and timely intervention. Pushpavathi et al. (2021) analyze feature ranking methods in predicting heart failure. Their research shows that ranking features based on their relevance significantly enhances prediction accuracy, offering useful insights for clinical decision-making [6]. Patidar et al. (2022) conduct a comparative analysis of various machine learning algorithms for heart dis- ease prediction. Their findings indicate that ensemble methods outperform individual classifiers in terms of prediction accuracy and generalization [7]. Shaikh et al. (2023) employ ensemble learning techniques to predict heart disease, highlighting the advantages of combining multiple models to increase accuracy and reduce bias, offering a promising solution for heart disease prediction [8]. Moushi et al. (2023) enhance the explainability of heart disease prediction models using interpretable machine learning techniques. Their work improves the transparency of model predictions, which is crucial for gaining trust in clinical settings [9]. Prakash (2024) explores the use of blockchain technology in supply chain management, providing a novel framework that could also have applications in healthcare for improving data security and patient confidentiality in heart disease prediction systems [10]. Ogunpola et al. (2024) propose predictive models based on machine learning for detecting cardiovascular diseases, highlighting the potential for using these models to identify at-risk individuals and provide early preventive care [11]. Kumari and Mehta (2021) introduce a novel approach for predicting heart disease using machine learning, focusing on algorithmic improvements to enhance the precision of predictions and provide more reliable outcomes for healthcare professionals [12]. Hasan (2021) explores different machine learning algorithms for heart disease prediction, comparing their performance. The study suggests that decision trees and random forests provide the best balance of accuracy and interpretability [13]. Bhatt et al. (2023) focus on heart disease prediction using advanced machine learning techniques, underscoring the importance of selecting the right models to achieve high accuracy and reduce the risks of misdiagnosis in clinical practice [14]. Yadav et al. (2021) optimize feature selection for heart disease prediction, demonstrating that reducing unnecessary features can enhance model efficiency and accuracy, making it a significant contribution to healthcare data science [15]. Rathore (2021) offers insights into the realistic application of the M/M/S queueing model for heart disease prediction. While primarily focused on theoretical analysis, the study adds a unique perspective on applying mathematical models to predict healthcare outcomes [16]. Hassan et al. (2022): The study utilizes machine learning classifiers to accurately predict coronary heart disease, emphasizing the potential of advanced algorithms in improving medical diagnostics and personalized healthcare approaches [17]. Sengar and Pandey examines how job satisfaction influences academic faculty performance in private colleges and universities, offering insights into improving workplace morale and educational outcomes in Indore [18]. Benjamin et al. provide a statistical report on heart disease and stroke, which serves as a foundation for understanding the broader epidemiological context of heart disease, essential for designing effective prediction models [19]. Pathan et al. investigate the impact of feature selection on heart disease prediction accuracy, concluding that better feature selection strategies result in more precise and actionable insights for clinicians [20]. Nissa et al. explore boosting ensemble techniques for heart disease prediction, demonstrating their effectiveness in improving model accuracy and robustness, especially in datasets with imbalanced classes [21]. Rathore et al. explore AI-enhanced healthcare in the context of Environment 4.0, focusing on intelligent well-being and smart ecosystems that leverage AI for proactive health monitoring and disease prediction [23]. Kaushik et al. focuses on using facial recognition for enhancing home security in smart environments, demonstrating the potential of deep learning for real- time identification and access control [24]. Rathore et al. present a predictive model for Black Friday sales using machine learning, enabling retailers to forecast demand, optimize inventory, and improve customer targeting in ecommerce [25]. Two types of SVM model such as linear SVM and polynomial SVM is used in [29]. Accuracy, precision, recall and F1 score has been evaluated for comparing linear SVM and polynomial SVM. Polynomial SVM provides better accuracy than linear SVM [29]. Mondal *et al.* [26] provide a comprehensive analysis of how machine learning transforms drug discovery workflows, highlighting

advances in supervised and unsupervised learning approaches that enable high-throughput screening, property prediction, and compound prioritization — thereby addressing key bottlenecks in pre-clinical research and enhancing the reproducibility [26] and efficiency of drug development processes [26]. This synthesis demonstrates the growing impact of ML on biopharmaceutical innovation and supports the deployment of advanced computational frameworks in translational research [26]. IoMT generates vast quantities of sensitive patient and device data, necessitating robust mechanisms for data integrity, privacy, and trust [30, 31]. Uddin *et al.* [27] examine the integration of blockchain technology within IoMT ecosystems to address these challenges, presenting frameworks that enhance decentralized authentication, immutable record keeping, and secure data sharing across heterogeneous healthcare infrastructures. This shows how combining blockchain with IoMT not only mitigates security vulnerabilities but also supports scalable and interoperable health systems, thereby enabling more trustworthy and patient-centric care delivery. Some highlights the effectiveness of data-driven and machine learning–based approaches in improving system transparency, analytical decision-making, and predictive performance across diverse application domains [33–35].

3. Dataset

The data are retrieved from the Heart disease data set from kaggle database [32]. The dataset provides the patients' information that consists of 303 records with 14 attributes.

Data Preprocessing: Data is often incomplete and inconsistent. Inconsistent and incomplete dataset significantly influence the performance of machine learning algorithms. The heart disease data set has been checked for missing values and null values as null value considerably effect on the conclusion that drawn from the data . There were no missing values in our dataset. As the percentage of missing value in our dataset is zero, we further moved to the feature selection process. We also scaled the features to speed up the training of the classifiers. Feature scaling is a process during the data preprocessing step. It is a method to normalize the independent variable or features of data within a particular range.

Feature Selection: The major purpose of selecting feature is to eliminate unrelated and redundant attributes. An irrelevant feature reduces the accuracy of the algorithm. Our research work used Feature selection technique to select the most important features. The Boruta algorithm is used to find all the significant attributes from the dataset with respect to an outcome variable [5]. From 13 input attributes, 6 attributes ('age', 'cp', 'thalach', 'oldpeak', 'ca', 'thal') (age, chest pain, maximum heart rate, ST depression induced by exercise relative to rest, number of major vessels colored by flourosopy, Thallium stress test) are selected as most important attribute. We have split the data into 70% as training and 30% as testing.

4. Results

The data set contains 303 records. It includes both positive and negative observations. We have divided these 303 records of the data into training and testing. 30% of records that is 91 records have been used for testing and remaining 212 records for training. Evaluating model is one of the core mechanisms while building machine learning model to check the efficacy of the model. We have also evaluated and compared the performance of the supervised machine learning algorithms using the metrics such as accuracy, precision, F1-score and AUC. One of the popular methods of estimating the performance of the algorithms is to use confusion matrix that contains information pertaining to predicted (columns) and actual class (rows) which can be visualized easily. It is a matrix representation of the prediction results. It is used to describe the performance of the classifier.Confusion matrix gives the counts of test records correctly and incorrectly predicted by the model.
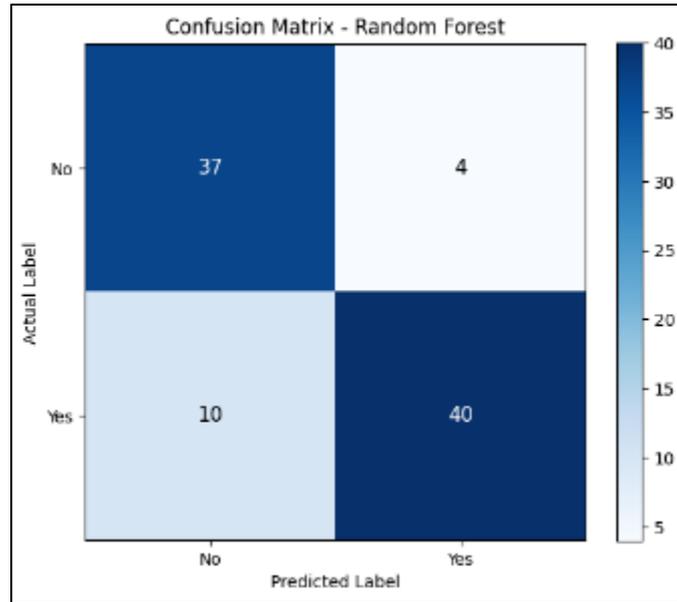
Figure 1 Confusion Matrix of Random Forest

```
Classification Report:

              precision    recall  f1-score   support

          No     0.7872    0.9024    0.8409        41
         Yes     0.9091    0.8000    0.8511        50

    accuracy                         0.8462        91
   macro avg     0.8482    0.8512    0.8460        91
weighted avg     0.8542    0.8462    0.8465        91
```

Out of those 91 records, the random forest classifier predicted "Yes" 44 times and "No" 47 times

- True Negatives (TN): 37 people were correctly predicted "No" and also in actual they do not have the heart disease.
- True Positives (TP): 40 people were correctly predicted "Yes" and also in actual they have the heart disease.
- False Negatives (FN): 10 people were incorrectly predicted "No" while they actually have heart problem.
- False Positives (FP): 4 people were incorrectly predicted "Yes" while in actual they do not have heart problem

Figure 2 Confusion Matrix of SVM

```
Classification Report:

              precision    recall  f1-score   support

          No     0.7955    0.8537    0.8235        41
         Yes     0.8723    0.8200    0.8454        50

    accuracy                         0.8352        91
   macro avg     0.8339    0.8368    0.8344        91
weighted avg     0.8377    0.8352    0.8355        91
```
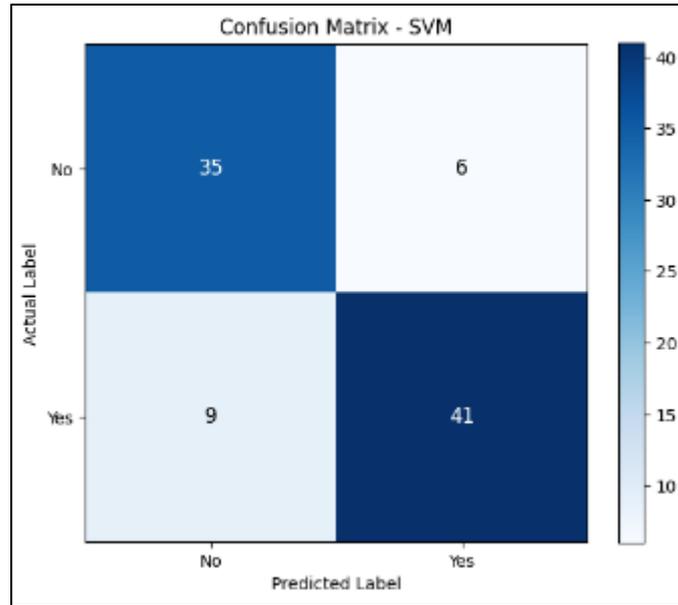
Out of those 91 records, the SVM predicted "Yes" 47 times and "No" 44 times.

True Negatives (TN): 35 people were correctly predicted "No" and also in actual they do not have the heart disease.

True Positives (TP): 41 people were correctly predicted "Yes" and also in actual they have the heart disease.

False Negatives (FN): 9 people were incorrectly predicted "No" while they actually have heart problem.

False Positives (FP): 6 people were incorrectly predicted "Yes" while in actual they do not have heart problem

In [28], using Linear SVM and Polynomial SVM Accuracy is 83% and 80%, respectively, where here SVM and RF gives accuracy of 83.52% and 84.62%, which is around 3% improvement from Linear SVM and Polynomial SVM [28].

5. Conclusion

The objective of this research work is to compare the performance of different supervised machine learning algorithms that can be employed in automated heart disease prediction systems. In this research work, heart disease is predicted using RF and SVM. The classifiers are implemented using Python on Heart disease dataset form Kaggle website. From the experimental results it is found that the random forest classifier is more accurate compared to SVM. The work is done with small size dataset. With more data better models can be built. The future work of this research is to use different Machine learning techniques for heart disease prediction with large size dataset and also to combine some of the machine learning techniques to build reliable and accurate prediction model for heart disease in real time.

References

[1] A. A. Ahmad and H. Polat, "Prediction of Heart Disease Based on Machine Learning Using Jellyfish Optimization Algorithm," Diagnostics, vol. 13, no. 14. MDPI AG, p. 2392, Jul. 17, 2023. doi: 10.3390/diagnostics13142392.

[2] A. Bahuguna, M. Gupta and R. Kumar, "Statistical Analysis and Prediction of Heart Disease Using Machine Learning," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-5, doi: 10.1109/ICCCNT56998.2023.10308166.

[3] N. Biswas et al., "Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques," BioMed Research International, vol. 2023, no. 1. Wiley, Jan. 2023. doi: 10.1155/2023/6864343.

[4] P. Kaushik, Y. Chopra, A. Kajla, M. Poonia, A. Khan, and D. Yadav, 'AI-Powered Dermatology: Achieving Dermatologist-Grade Skin Cancer Classification,' 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), Gwalior, India, 2024, pp. 1-6, doi: 10.1109/IATMSI60426.2024.10502664.

[5] C. Boukhatem, H. Y. Youssef and A. B. Nassif, "Heart Disease Prediction Using Machine Learning," 2022 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 2022, pp. 1-6, doi: 10.1109/ASET53988.2022.9734880.

[6] T. P. Pushpavathi, S. Kumari and N. K. Kubra, "Heart Fail- ure Prediction by Feature Ranking Analysis in Machine Learn- ing," 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2021, pp. 915-923, doi: 10.1109/ICICT50816.2021.9358733.

[7] S. Patidar, D. Kumar, and D. Rukwal, "Comparative Analysis of Machine Learning Algorithms for Heart Disease Prediction," Advances in Transdisciplinary Engineering. IOS Press, Nov. 09, 2022. doi: 10.3233/atde220723.

[8] Y. Shaikh, V. K. Parvati and S. R. Biradar, "Heart Disease Prediction using Ensemble Learning," 2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC), Dharwad, India, 2023, pp. 1-5, doi: 10.1109/ICAISC58445.2023.10200375.

[9] O. M. Moushi, N. Ara, M. Helaluddin and H. S. Mondal, "Enhancing the Accuracy and Explainability of Heart Disease Prediction Models through Interpretable Machine Learning Techniques," 2023 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), Dhaka, Bangladesh, 2023, pp. 1-6, doi: 10.1109/ICICT4SD59951.2023.10303572.

[10] A. Prakash, "Blockchain Technology for Supply Chain Management: Enhancing Transparency and Efficiency," International Journal for Global Academic & Scientific Research, vol. 3, no. 2. International Consortium of Academic Professionals for Scientific Research, pp. 01–11, Jul. 02, 2024. doi: 10.55938/ijgasr.v3i2.73.

[11] A. Ogunpola, F. Saeed, S. Basurra, A. M. Albarrak, and S. N. Qasem, "Machine Learning-Based Predictive Models for Detection of Cardiovascular Diseases," Diagnostics, vol. 14, no. 2. MDPI AG, p. 144, Jan. 08, 2024. doi: 10.3390/diagnostics14020144.

[12] A. Kumari and A. K. Mehta, "A Novel Approach for Prediction of Heart Disease using Machine Learning Algorithms," 2021 Asian Conference on Innovation in Technology (ASIANCON), PUNE, India, 2021, pp. 1-5, doi: 10.1109/ASIANCON51346.2021.9544544.

[13] R. Hasan, "Comparative Analysis of Machine Learning Algorithms for Heart Disease Prediction," ITM Web of Conferences, vol. 40. EDP Sciences, p. 03007, 2021. doi: 10.1051/itmconf/20214003007.

[14] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques," Algorithms, vol. 16, no. 2. MDPI AG, p. 88, Feb. 06, 2023. doi: 10.3390/a16020088.

[15] D. P. Yadav, P. Saini and P. Mittal, "Feature Optimization Based Heart Disease Prediction using Machine Learning," 2021 5th In- ternational Conference on Information Systems and Computer Net- works (ISCON), Mathura, India, 2021, pp. 1-5, doi: 10.1109/ISCON52037.2021.9702410.

[16] R. Rathore and S. Pratap Singh Rathore, "Machine Learning Applications in Human Resource Management: Predicting Employee Turnover and Performance," International Journal for Global Academic & Scientific Research, vol. 3, no. 2. International Consortium of Academic Professionals for Scientific Research, pp. 48–59, Jul. 02, 2024. doi: 10.55938/ijgasr.v3i2.77.

[17] Ch. A. ul Hassan et al., "Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers," Sensors, vol. 22, no. 19. MDPI AG, p. 7227, Sep. 23, 2022. doi: 10.3390/s22197227.

[18] S. Sengar and M. Pandey, "To Study the Effect of Job Satisfaction on the Performance of Academic Faculties Working in Private Colleges and Private Universities in Indore," International Journal for Global Academic & Scientific Research, vol. 3, no. 1. International Consortium of Academic Professionals for Scientific Research, pp. 13–24, Apr. 04, 2024. doi: 10.55938/ijgasr.v3i1.69.

[19] E. J. Benjamin et al., "Heart Disease and Stroke Statistics—2019 Update: A Report From the American Heart Association," Circulation, vol. 139, no. 10. Ovid Technologies (Wolters Kluwer Health), Mar. 05, 2019. doi: 10.1161/cir.0000000000000659.

[20] M. S. Pathan, A. Nag, M. M. Pathan, and S. Dev, "Analyzing the impact of feature selection on the accuracy of heart disease prediction," Healthcare Analytics, vol. 2. Elsevier BV, p. 100060, Nov. 2022. doi: 10.1016/j.health.2022.100060.

[21] N. Nissa, S. Jamwal, and M. Neshat, "A Technical Comparative Heart Disease Prediction Framework Using Boosting Ensemble Techniques," Computation, vol. 12, no. 1. MDPI AG, p. 15, Jan. 16, 2024. doi: 10.3390/computation12010015.

[22] S. P. S. Rathore, R. R. Palle, H. Bhardwaj, D. Singh, R. Bharti and Priyanka, "AI-Enhanced Healthcare: Forging the Path towards Intelligent Well-being and the Cutting Edge Environment 4.0," 2023 International Conference on Smart Devices (ICSD), Dehradun, India, 2024, pp. 1-6, doi: 10.1109/ICSD60021.2024.10751255.

[23] P. Kaushik, H. Kumar, N. Tyagi, Shubhang, Jatin and Mrinal, "Improving Home Security through Facial Recognition in a Smart Environment," 2023 International Conference on Smart Devices (ICSD), Dehradun, India, 2024, pp. 1-6, doi: 10.1109/ICSD60021.2024.10751272.

[24] S. P. S. Rathore, V. M. Bihade, S. Sane, N. V. Limbore, R. Lenka and Priyanka, "Smart Model For Black Friday Sales Prediction," 2023 International Conference on Smart Devices (ICSD), Dehradun, India, 2024, pp. 1-5, doi: 10.1109/ICSD60021.2024.10751174.

[25] R. Rathore, H. Yennapusa, S. Kumar, M. Mahawar, S. Miglani and G. Sharma, "Cutting-Edge Skin Cancer Detection in a Smart Ecosystem," 2023 International Conference on Smart Devices (ICSD), Dehradun, India, 2024, pp. 1-5, doi: 10.1109/ICSD60021.2024.10751528

[26] KALLOL, KANTI MONDAL, LUCKY MICHAEL DANIEL, and MANDAL PABITRA. "How ML transforms drug discovery." GLOBAL JOURNAL OF ENGINEERING AND TECHNOLOGY ADVANCES Учредители: Global Scholarly Communication 21.1 (2024): 197-203.

[27] Uddin, Md Bahar, et al. "Blockchain Integration in IoMT for Secure Healthcare: Challenges, Integration, and Solutions." 2025 6th International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE, 2025.

[28] Hoque, R., Billah, M., Debnath, A., Hossain, S. S., & Sharif, N. B. (2024). Heart disease prediction using SVM. International Journal of Science and Research Archive, 11(2), 412-420.

[29] Mondal, K. K., Rahman, R., Bipasha, Y. A., & Kadir, A. (2023). Polygenic Hazard Score and Amyloid PET Imaging Mediation Analysis in Alzheimer's Disease Diagnosis. International Journal of Science and Research Archive, 2023, 10(02), 1451-1457.

[30] Mahmud, T. (2023). Applications for the Internet of medical things. International Journal of Science and Research Archive, 10(02), 1247-1254.

[31] Tanvir Mahmud, "ML-driven resource management in cloud computing", World Journal of Advanced Research and Reviews, 2022, 16(03), 1230-1238.

[32] https://www.kaggle.com/ronitf/heart-disease-uci?select=heart.csv

[33] Bipasha, Y. A. (2023). Blockchain technology in supply chain management: transparency, security, and efficiency challenges. International Journal of Science and Research Archive, 10(01), 1186-1196.

[34] Bipasha, Y. A. (2022). Market efficiency, anomalies and behavioral finance: A review of theories and empirical evidence. World Journal of Advanced Research and Reviews, 15(02), 827-839.

[35] Das, D., Billah, M. M., Nath, A. D., Sharif, N. B., & Mondal, K. K. (2022). Breast Cancer Classification using LGBM and SVM. International Journal of Science and Research Archive. International Journal of Science and Research Archive, 2022, 07(02), 876-881.