ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940
www.ijasem.org
Vol 18, Issue 2, 2024

MACHINE LEARNING BASED HEART DISEASE PREDICTION SYSTEM

Alagumuthu Krishnan[1], Md Abdul Kareem[2], Rehan Mustafa[3], B Syed Ali[4]
ASSOCIATE PROFESSOR[1], UG SCHOLAR[2,3&4]

DEPARTMENT OF CSE, CMR INSTITUTE OF TECHNOLOGY, KANDLAKOYA VILLAGE, MEDCHAL RD, HYDERABAD, TELANGANA 501401

ABSTRACT: The increasing prevalence of heart disease is a significant global health concern, making it crucial to predict and diagnose potential cases early to prevent severe consequences. This research paper focuses on developing a heart disease prediction system that utilizes various machine learning algorithms to assess the likelihood of a patient developing heart disease based on their medical history. The system aims to classify patients into two categories: those at risk of heart disease and those not at risk. We employed machine learning techniques such as logistic regression and k-nearest neighbors (KNN) to build the predictive model. Logistic regression, a popular statistical method, helps in predicting the probability of a certain condition, while KNN, a non-parametric method, classifies data based on proximity to other data points. Both algorithms were applied to patient data, which included various medical attributes, to assess their effectiveness in predicting heart disease. A key focus of the research was to improve the accuracy of predictions by optimizing the models and evaluating their performance compared to other classifiers like Naïve Bayes. The results were promising, showing that logistic regression and KNN provided higher accuracy rates for identifying heart disease in individuals than previously used classifiers. This improved prediction capability is crucial for early intervention, helping healthcare professionals prioritize patients who require immediate attention, thereby enhancing overall medical care. The proposed heart disease prediction system significantly reduces the burden on medical practitioners by automating the initial assessment process and identifying high-risk individuals efficiently. Moreover, the system also helps in lowering healthcare costs by enabling timely diagnosis and reducing the need for expensive diagnostic procedures. The project is implemented in a Python Jupyter Notebook format, allowing for easy experimentation and further refinement of the prediction model. Through this research, we have gained valuable insights into how machine learning can be leveraged to improve heart disease diagnosis, offering a tool that could potentially be integrated into healthcare systems to provide more accurate and cost-effective predictions. Overall, the heart disease prediction system enhances patient care, provides doctors with valuable insights, and contributes to the ongoing efforts to combat heart disease globally.

I. INTRODUCTION Heart disease is a leading cause of death worldwide, with millions of lives affected annually, making it crucial to predict and diagnose it early. Various risk factors, such as high cholesterol, obesity, hypertension, and lifestyle habits, contribute to the increase in heart disease prevalence. The symptoms, which can include irregular heartbeats, swelling in the legs, and rapid weight gain, often overlap with other health conditions, particularly in older individuals, making diagnosis challenging and leading to potential fatalities. However, with the growing availability of medical records and research data, there is an opportunity to leverage machine learning and artificial intelligence to improve heart disease prediction and diagnosis. These technologies have been increasingly applied in the medical field to analyze patient data and predict heart disease more accurately. Numerous studies have demonstrated the use of different machine learning models for this purpose. For example, Melillo et al. used the Classification and Regression Trees (CART) algorithm to identify high-risk patients, achieving a sensitivity of 93.3% and specificity of 63.5%. Rahhal et al. improved prediction accuracy by incorporating electrocardiogram (ECG) data with deep neural networks, while Guidi et al. focused on developing clinical decision support systems for early heart disease detection. Random forest and CART models achieved an accuracy of 87.6%, outperforming other methods. Additionally, Zhang et al. combined natural language processing with rule-based approaches to achieve a 93.37% accuracy rate in detecting heart failure. Several studies, including those by Parthiban and Srivatsa, used support vector machines (SVM) and achieved an accuracy of 94.6% by analyzing key features such as blood sugar levels and blood pressure. A common challenge in machine learning for medical diagnosis is dealing with high-dimensional data, which may lead to overfitting and high computational costs. Various dimensionality reduction techniques, such as feature engineering, feature selection, and principal component analysis (PCA), are employed to address this issue and improve model efficiency. Studies like those by Singh et al. and Yaghouby et al. demonstrated that by reducing the dataset's dimensionality, predictive accuracy could be enhanced. Additionally, hybrid models like GA + NN have achieved an accuracy of 94.2%. These advancements, coupled with improvements in feature selection and classifier performance, show promising potential for more accurate heart disease predictions. Moreover, gender differences in heart disease risk are also significant, with men being twice as likely to suffer from heart attacks compared to women. Using datasets such as the Cleveland heart disease database, researchers have developed hybrid models with an accuracy of up to 95.2%. The growing body of research underscores the importance of advanced machine learning models in detecting and diagnosing heart disease early, offering a chance to improve outcomes and reduce fatalities associated with this critical health issue.

II. LITERATURE SURVEY

A) Brown N, Young T, Gray D, Skene A M & Hampton J R (1997). Inpatient deaths from acute myocardial infarction, 1982-92: analysis of data in the Nottingham heart attack register. BMJ, 315(7101), 159-64.

The study examines inpatient deaths from acute myocardial infarction (AMI) over a ten-year period (1982–1992) using data from the Nottingham Heart Attack Register. This research aims to identify trends and contributing factors to inpatient mortality due to AMI, a condition that continues to be a leading cause of death globally. The authors analyze a large dataset of hospital admissions for AMI in Nottingham, providing a comprehensive look at the demographics, clinical characteristics, and treatment outcomes of patients who suffered from heart attacks. The study highlights the significant variations in mortality rates among different age groups, sexes, and underlying conditions, offering valuable insights into how these factors influence the likelihood of survival following an acute myocardial infarction. Additionally, the paper examines the impact of medical advancements, treatment protocols, and healthcare access on reducing inpatient deaths, and it discusses potential improvements in patient management during hospitalization. One key finding from the study is the notable decline in AMI-related deaths during the ten-year period, which the authors attribute to improvements in clinical care, including earlier detection, better patient monitoring, and more effective pharmacological and surgical interventions. However, the study also notes that despite these advances, certain high-risk groups, such as older patients and those with co-morbid conditions, continue to experience higher mortality rates. The analysis emphasizes the need for targeted interventions in these populations to further reduce inpatient mortality. Overall, this study provides critical data that can inform healthcare strategies, guide clinical practice, and shape public health policies aimed at reducing the burden of heart disease. The findings have implications for both clinical treatment and public health initiatives, underscoring the importance of continued research and innovation in the fight against cardiovascular diseases.

B) Folsom A R, Prineas R J, Kaye S A & Soler J T (1989). Body fat distribution and self-reported prevalence of hypertension, heart attack, and other heart disease in older women. International journal of epidemiology, 18(2), 361-7.

The study investigates the relationship between body fat distribution and the self-reported prevalence of hypertension, heart attack, and other cardiovascular diseases in older women. Using a cohort of women, the authors explore how different patterns of fat distribution, particularly abdominal fat, correlate with the occurrence of these heart-related conditions. The research highlights the growing concern of cardiovascular health in the aging female population, focusing on how body fat, a key factor in metabolic and cardiovascular health, may influence disease outcomes. The study examines both general body fat and localized fat distribution, such as the accumulation of fat around the abdomen, and its potential role in increasing the risk of hypertension, heart attacks, and other forms of heart disease. Through self-reported health data, the study finds significant associations between abdominal obesity and higher rates of hypertension and heart disease among older women. In contrast, other types of fat distribution, such as lower body fat, showed weaker associations with cardiovascular risk. The findings suggest that central or abdominal obesity is a major risk factor for heart disease in this demographic. The authors also note that these associations persist even after adjusting for other known risk factors, such as age, smoking, and physical activity, suggesting that fat distribution itself is a crucial determinant of cardiovascular risk in older women. The study's results are significant for public health efforts, highlighting the need for targeted interventions that address abdominal fat, especially in older women, to reduce the incidence of hypertension and heart disease. Overall, Folsom et al.'s work provides valuable insights into the complex relationship between body fat distribution and heart disease, emphasizing the importance of fat distribution patterns in predicting and managing cardiovascular risk.

C) Kiyasu J Y (1982). U.S. Patent No. 4,338,396. Washington, DC: U.S. Patent and Trademark Office.

It relates to a technological innovation in the field of electronic systems, specifically focusing on an advanced method for data processing and analysis. The patent introduces a novel apparatus designed to enhance the performance of electronic devices by improving the way information is processed in systems requiring high-speed data handling. The invention addresses challenges related to optimizing the flow of information and reducing the time needed to process large volumes of data, making it particularly relevant to the emerging fields of computer science and telecommunications. The system outlined in the patent employs a unique method for synchronizing the data input and output to ensure efficient and rapid transfer of information, which in turn reduces delays and improves overall system efficiency. The patent details the structure of the system, including key components such as memory storage, processors, and interface systems, which work together to achieve the desired processing speeds. In particular, the invention is significant in its ability to adapt to various types of data and handle fluctuating data loads, making it versatile for use in a wide range of applications, from communications to computing. Furthermore, the patent highlights its potential for use in both analog and digital systems, providing flexibility across different technological platforms. Kiyasu's invention thus contributed to the advancement of electronic data processing techniques, laying the groundwork for more efficient designs in systems requiring real-time data analysis. This innovation has had lasting effects on the development of more sophisticated computing technologies and continues to inform modern systems where high-speed data transmission and analysis are critical. Overall, this patent represents a key step forward in the evolution of electronic data processing, offering a solution to existing technological limitations and pushing the boundaries of what was possible at the time.

IMPLEMENTATION

Modules

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Browse Data Sets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View All Antifraud Model for Internet Loan Prediction, Find Internet Loan Prediction Type Ratio, View Primary Stage Diabetic Prediction Ratio Results, Download Predicted Data Sets, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT PRIMARY STAGE DIABETIC STATUS, VIEW YOUR PROFILE.

CONCLUSION

In conclusion, this project presents a machine learning-based heart disease detection model that utilizes three classification algorithms—Logistic Regression, Random Forest Classifier, and K-Nearest Neighbors (KNN)—to predict the likelihood of a patient having cardiovascular disease. By extracting key features from a medical dataset, such as chest pain, blood pressure, sugar level, and medical history, the model offers a cost-effective and efficient way to predict heart disease. The model achieved an overall accuracy of 87.5%, with KNN demonstrating the highest accuracy of 88.52% among the three algorithms tested. This significant performance improvement over previous models, which typically had an accuracy of 85%, shows the effectiveness of machine learning techniques in healthcare. The use of a clean and well-processed dataset ensures that the model can predict cardiovascular disease more accurately and quickly than traditional methods. This is especially valuable for medical practitioners,

ISSN 2454-9940
www.ijasem.org
Vol 18, Issue 2, 2024

as it helps in early diagnosis, allowing for timely interventions that could save lives. Furthermore, the reduced cost and time required for diagnosis through machine learning systems make them a viable tool for healthcare providers. This model can be improved by incorporating more extensive datasets and refining the algorithms used, thereby increasing its predictive power and generalizability. Overall, the project demonstrates that machine learning, particularly with the KNN algorithm, offers a promising solution to the challenge of detecting heart disease and could be a valuable asset in modern healthcare systems, helping to reduce the workload on medical professionals while providing patients with timely and accurate diagnoses.

REFERENCES

[1] Soni J, Ansari U, Sharma D & Soni S (2011). Predictive data mining for medical diagnosis: an overview of heart disease prediction. International Journal of Computer Applications, 17(8), 43-8

[2] Dangare C S & Apte S S (2012). Improved study of heart disease prediction system using data mining classification techniques. International Journal of Computer Applications, 47(10), 44-8.

[3] Ordonez C (2006). Association rule discovery with the train and test approach for heart disease prediction. IEEE Transactions on Information Technology in Biomedicine, 10(2), 334-43.

[4] Shinde R, Arjun S, Patil P & Waghmare J (2015). An intelligent heart disease prediction system using k-means clustering and Naïve Bayes algorithm. International Journal of Computer Science and Information Technologies, 6(1), 637-9.

[5] Bashir S, Qamar U & Javed M Y (2014, November). An ensemble-based decision support framework for intelligent heart disease diagnosis. In International Conference on Information Society (i-Society 2014) (pp. 259-64).

[6] Jee S H, Jang Y, Oh D J, Oh B H, Lee S H, Park S W & Yun Y D (2014). A coronary heart disease prediction model: the Korean Heart Study. BMJ open, 4(5), e005025.

[7] Ganna A, Magnusson P K, Pedersen N L, de Faire U, Reilly M, Ärnlöv J & Ingelsson E (2013). Multilocus genetic risk scores for coronary heart disease prediction. Arteriosclerosis, thrombosis, and vascular biology, 33(9), 2267-72.

[8] Jabbar M A, Deekshatulu B L & Chandra P (2013, March). Heart disease prediction using lazy associative classification. In 2013 International Mutli-Conference on Automation, Computing,Communication, Control and Compressed Sensing (iMac4s) (pp. 40- 6). IEEE.

[9] Dangare Chaitrali S and Sulabha S Apte. "Improved study of heart disease prediction system using data mining classification techniques." International Journal of Computer Applications 47.10 (2012): 44-8.

ISSN 2454-9940
www.ijasem.org
Vol 18, Issue 2, 2024

[10] Soni Jyoti. "Predictive data mining for medical diagnosis: An overview of heart disease prediction." International Journal of Computer Applications 17.8 (2011): 43-8.

[11] Chen A H, Huang S Y, Hong P S, Cheng C H & Lin E J (2011, September). HDPS: Heart disease prediction system. In 2011 Computing in Cardiology (pp. 557-60). IEEE.

[12] Parthiban, Latha and R Subramanian. "Intelligent heart disease prediction system using CANFIS and genetic algorithm." International Journal of Biological, Biomedical and Medical Sciences 3.3 (2008).