

## Intelligent Heart Disease Prediction System Using DBSCAN and Fuzzy Logic with Automated Patient Report Generation

M. S. Midhuna<sup>1</sup>, B.Tech Artificial intelligence and data science, Amrita college of engineering and technology

J. Sarrvesh Maniyan<sup>2</sup>, B.Tech Artificial intelligence and data science, Amrita college of engineering and technology

M. Harishma<sup>3</sup>, B.Tech Artificial intelligence and data science, Amrita college of engineering and technology,

Mrs R Saranya<sup>4</sup>, Assistant professor, B.Tech Artificial intelligence and data science, Amrita college of engineering and technology, Nagercoil.

saranya@amrita.edu.in

### Abstract:

Heart disease continues to be a major cause of mortality worldwide, highlighting the urgent need for effective early detection and preventive healthcare systems. This research paper proposes a new system called the Intelligent Heart Disease Prediction System that combines the use of modern technology such as DBSCAN Clustering and Fuzzy Logic Classification. This system is designed to provide high-quality, reliable risk assessments based on multiple factors like age, cholesterol levels, dietary habits, physical activity level, and medical history in order to determine a person's likelihood of developing heart disease. A secure web-based interface is used to collect patient data, which is accessible and offers privacy of data. The collected information is stored in an SQLite database and undergoes preprocessing, including handling missing values and normalisation, to enhance prediction accuracy. DBSCAN clustering has been applied to identify latent patterns and to cluster groups of patients with similar profiles, without the use of a pre-defined label. Fuzzy logic is then used to categorise patients into different risk levels by effectively managing uncertainty and imprecision within the medical data. The system produces automated PDF reports featuring in-depth risk analysis, symptom evaluation and tailored recommendations. Interactive dashboards provide a visual representation of patient trends and historical data. Overall, the proposed system supports efficient, remote heart disease prediction, facilitating early diagnosis, reducing healthcare costs, and improving patient outcomes.

**Keywords:** Heart Disease Prediction, DBSCAN Clustering, Fuzzy Logic, Machine Learning, Healthcare Analytics, Data Preprocessing, Risk Assessment, Web-Based Healthcare Systems, Automated Report Generation, Data Visualization Systems.

### 1. Introduction

Heart disease has emerged as one of the most serious global health concerns, contributing significantly to mortality rates and long-term medical complications. Poor diet, lack of

physical exercise, smoking, and stress are some of the unhealthy lifestyle habits that have significantly increased the risk of cardiovascular diseases in all age groups. Much attention should be paid to the early detection and proper prediction to prevent the number of fatal outcomes and increase the rates of survival [1]. Older diagnostic methods depend more on clinical knowledge and laboratory analysis and manual interpretation of patient information, which may be time-consuming, costly and subject to human error. As technology has improved, and especially in the area of data science and artificial intelligence, there have been intelligent systems created to assist healthcare providers in diagnosing diseases more effectively. Machine learning algorithms allow handling a significant amount of medical data to find concealed patterns and forecast possible risks. In this context, the research proposed is based on the creation of a smart system that predicts heart diseases using the methods of clustering and fuzzy logic. Such a combination improves the accuracy of prediction and promotes effective decision-making, which leads to enhanced preventive healthcare strategies [2].

The proposed system is a web-based system which enables the users to enter their health and lifestyle information easily. The system works with this information by employing very complex computational algorithms to come up with the forecast on the basis of heart disease likelihood [15]. In addition, healthcare professionals are usually exposed to large volumes of patient data, and it is hard to interpret and analyse [3]. An intelligent system can contribute regarding the creation of timely predictions and the identification of high-risk individuals in order to reduce the

workload and improve the quality of diagnosis. Accessibility and efficiency are also prioritised in the system, allowing the user to track their health remotely and take preventive measures [4].

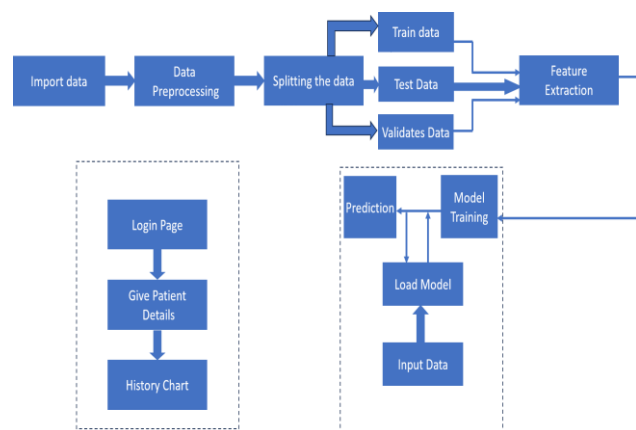
Heart disease prediction remains a complex problem due to the involvement of multiple interrelated risk factors and the inherent uncertainty present in medical data. Conventional ways of diagnostic procedures are usually manual in their interpretation and this can result in delays, inconsistency and inaccuracy. The second serious issue is the inability to find open platforms that enable people to test their health at a distance [5]. The majority of diagnostic instruments need hospitalisation and special devices, which is not as convenient in continuous monitoring. Also, the current systems tend to offer a poor level of insight by showing the results of prediction but not the analysis or suggestions. The need to develop systems that can effectively handle large data and provide a high level of data security and privacy is also on the increase. Unless the data is properly handled, the accuracy of healthcare applications can be undermined.

To overcome these difficulties, the proposed system combines DBSCAN clustering and fuzzy logic, which allows making correct predictions, taking into account the uncertainty and making it accessible to the users [6]. The developed system makes accurate, timely, and reliable predictions that may assist clinicians with early diagnosis or preventative care [7]. The clustering technique assists researchers with the interpretation of the relationships between different health parameters and the clustering of similar patients together. Furthermore, fuzzy logic can also be used to classify

patients into various risk categories by managing uncertainty and vagueness [8]. The research also develops a web-based user interface, which helps clinicians with data entry and result retrieval. Also, the web interface has automated PDF reports of specific patient conditions, and it includes data visualisation tools to assist healthcare professionals in their analysis and decision-making processes [9]. This paper produces a web application capable of predicting heart disease using advanced computational methods, based on patient data regarding health and lifestyle. The user will enter information into the system and receive predictions, as well as a report summarising his or her results [10]. The steps involved in this process include collecting data, performing data preprocessing and clustering, performing classification, and visualising the results of data analysis. A secure database will be used to store records of patients so that data can be protected [14]. Additionally, this system will be capable of supporting multiple users and providing real-time predictions that can be applied in real-life medical practice. However, since this system is based only on the input data and will not replace a professional medical diagnosis, it is designed to assist healthcare providers and those with knowledge about their health so that they may take preventative action [11]. The system assists individual people to take care of their own health and lead healthier lifestyles through personalised advice and automatic reports [12]. The system is also helpful to healthcare professionals as it offers useful insights and helps them to work less. The presence of data visualisation tools also increases the analysis of patient data and, therefore, enables making better decisions [13].

## 2. System Architecture

The system architecture in Figure 1 represents the general framework and the relation between the different components of a heart disease prediction system. The user interface is going to be built primarily using web technologies like HTML, CSS, and JavaScript. Next, input data will be protected in the SQLite database. Records of patients will be stored, including prediction results and previous patient records, and accessed for future purposes. This next step in patient record processing is known as the preprocessing module so that the data is valid, such as managing nulls, removing any inconsistencies, and normalising each patient input record to allow for valid analysis of patient profiles. Thus, pre-processed patient records will be submitted to a DBSCAN clustering module for clustering similar profile records into clusters with the purpose of identifying underlying patterns.



**Figure 1:** Proposed Architecture

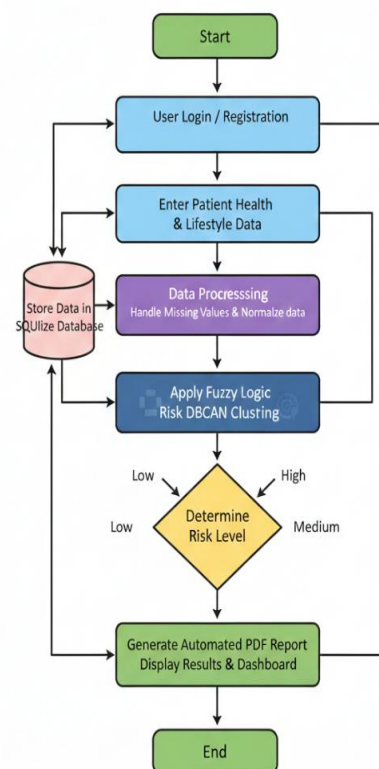
After clustering is done, fuzzy logic classification will be performed on the clustered records to assess the evaluated clusters of records based on a defined set of rules to determine a heart disease risk level for each patient. Moreover, the system has a report generation module that gives detailed

PDF reports which include prediction results, analysis and personal recommendations. The module of visualisation is also useful in enhancing understanding as it shows the data in charts and graphs where the users and healthcare professionals can easily analyse the trends and make an informed decision.

### 3. Methodology

Heart disease prediction is critical to cardiovascular disease research because of the increasing number of cases and the need for earlier detection. Historical methods are based on a manual review of patient records and also use basic computational techniques for prediction. The use of machine learning technologies to provide greater accuracy in predicting cardiovascular disease is possible with machine learning but is limited because of the data requirements and are not designed to accommodate unknowns in the patient records. Additionally, these machine learning methods are not able to provide real-time processing or remote monitoring capabilities. In order to address these issues, the proposed system combines both DBSCAN clustering and fuzzy logic methods to identify patterns and deal with uncertainties. The proposed system includes an interface that supports a Web-based user interface, automatic report generation, data visualisation, and improved accuracy, efficiency, and usability in predicting heart disease. System workflow in Figure 2 starts with user authentication, where people register or log in so that they can have safe access to the application. After authentication, a user is then presented with a structured input interface in which he/she is required to provide his/her health and lifestyle information. The system also validates input to prevent entry of erroneous, incomplete and excessive data. Once validated, the data is

saved in the SQLite database to be processed. The stored data is then fed over to the preprocessing module where the data is then cleaned up, and the missing values are addressed, and the data normalised to ensure that it is consistent and adds more accuracy to the analysis. The processed information is sent to the DBSCAN clustering module that recognises patterns and clusters similar patient profiles according to their characteristics. It is a step to be taken to realise the connection between various health parameters and identify possible risk patterns. Fuzzy logic classification of post-clustering heart disease risk is applied in the system. The system determines the risk as low, medium or high, depending on preset criteria. The user is then presented with the prediction outcomes on the interface. Also, the system produces a more detailed PDF report with the analysis and individual recommendations.



**Figure 2: System Workflow**

### **3.1 Data Collection and Input Acquisition**

The proposed system methodology starts with the collection of patient health and lifestyle information by safe and user-friendly web-based interface. This is the entry point of the system where users can enter important parameters, including age, cholesterol level, blood pressure, diet, smoking, physical activity, and medical history. An interface was built using advanced web technologies so that everyone can use it easily. To ensure that the data entered was complete, accurate, and in a valid range before it is sent to the backend server, there were validation mechanisms put in place. A SQL database is used to store valid entries of patient information, prediction results, and previous entries that can be used at a later time. The process of collecting data is very critical in deciding how effective the system is since the better the data input, the better the predictions are made.

### **3.2 Data Preprocessing and Normalisation**

After data collection, the next step in the methodology involves preprocessing the input data to enhance its quality and suitability for analysis. The preprocessing module addresses these problems by undertaking a number of tasks, which include data cleaning, lack of values, normalisation and transformation. Handling of missing data will include options like the use of mean substitution, and the assignment of default values, in order to ensure that the records, which have missing data, do not impact the analysis. Normalisation of the data inputs, meaning that input variables are adjusted into a standard range, becomes crucial for some algorithm types such as DBSCAN as they are distance-based clustering methods. This means no one feature will weigh more than the others during the clustering algorithm.

The various pre-processing modules utilise the capabilities of well-established Python Libraries like NumPy and Pandas to implement these functions on large scale datasets. This stage increases the accuracy and reliability of the next stage of clustering and classification greatly by improving the consistency and structure of the data.

### **3.3 DBSCAN Clustering for Pattern Identification**

After the preprocessing of data, the system uses the DBSCAN algorithm to detect hidden patterns and cluster similar patient profiles. DBSCAN is an unsupervised algorithm, meaning that it uses density to group data points rather than be able to utilise known labels. The algorithm considers two key parameters, including epsilon ( $\epsilon$ ), which is used to determine the radius of a neighbourhood, and minimum points (MinPts), which is used to determine the minimum number of data points required to form a cluster. DBSCAN uses the concentration of the points to identify patients who would be clustered with similar characteristics, and the noise and outliers that cannot be classified within any cluster are identified. In comparison to the classical clustering algorithms, DBSCAN does not require a priori knowledge of the number of clusters prior to its use; therefore, it is better adaptable and flexible to realistic data. The step is critical in improving the accuracy of general prediction.

### **3.4 Fuzzy Logic-Based Risk Classification**

When patients are grouped according to heart disease risk through clustering, the system then uses fuzzy logic to place them into various risk categories. Fuzzy logic is designed to deal with uncertainty and

uncertainty in medical data since it is often impossible to accurately classify an individual using traditional binary classifications. The fuzzy inference process occurs in three phases, which consist of fuzzification, rule evaluation, and defuzzification. At the fuzzification stage, patient-specific input parameters (cholesterol levels, blood pressure level, and age), are represented in fuzzy sets such as low, normal, or high with the aid of membership functions. The rule evaluation phase utilises predetermined fuzzy rules based on known medical associations for each input to process the input values. Finally, during the defuzzification phase, an actual number is derived from the fuzzy output which represents the risk level. This method allows for more versatility in classification and provides a more understandable reference point for both the user and healthcare professionals when assessing risk levels.

### **3.5 Prediction Output, Reporting, and Visualisation**

The last phase of the methodology is the creation of prediction results and the presentation of such results in an understandable and informative way. The system identifies the risk level of heart disease after the classification process and classifies the risk to be low, moderate, or high. These results are presented at a convenient interface and this is what simplifies their comprehensibility. The system also focusses on the key contributing factors in addition to simple delivery of the prediction, enabling the user to identify the potential risk to the health and avoid it. Automated report generation is also available through the system that creates comprehensive PDF reports containing patient information, prediction results, analysis, and customised lifestyle

recommendations. The visual representation of the data can be used by teams of users and health care providers to monitor and identify patterns, and to track changes over time.

#### **Algorithm: Intelligent Heart Disease Prediction Using DBSCAN and Fuzzy Logic**

##### **Input:**

Patient dataset DDD containing health and lifestyle parameters

##### **Output:**

Heart disease risk level (Low / Medium / High)

##### **Step 1: Data Collection**

Collect patient input data through a web interface (age, BP, cholesterol, lifestyle factors).

##### **Step 2: Data Storage**

Store the collected data securely in the SQLite database.

##### **Step 3: Data Preprocessing**

Handle missing values

Remove inconsistencies

Normalize input features

##### **Step 4: Apply DBSCAN Clustering**

Define parameters:  $\epsilon$  (epsilon), MinPts

Group similar patient data based on density

Identify clusters and remove noise/outliers

##### **Step 5: Feature Pattern Extraction**

Extract meaningful patterns from clustered data for classification.

##### **Step 6: Fuzzy Logic Classification**

Convert inputs into fuzzy sets (Low, Medium, High)

Apply fuzzy rules based on medical knowledge

Perform inference and defuzzification

##### **Step 7: Risk Prediction**

Classify patient into risk category: Low / Medium / High

##### **Step 8: Result Generation**

Display prediction results to the user interface

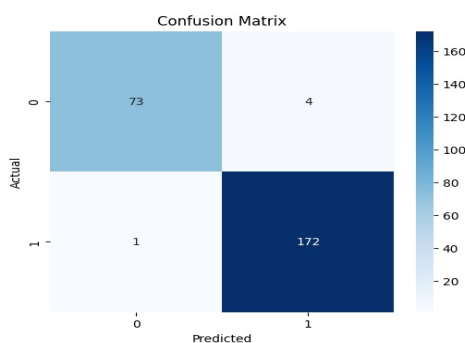
**Step 9: Report Generation**  
 Generate automated PDF report with analysis and recommendations

**Step 10: Data Visualization**  
 Display trends and history using dashboard charts

**End**

#### 4. Result and discussion

The confusion matrix provided in the Figure 3 shows the capability of the proposed heart disease prediction system to classify patients correctly. The matrix reveals four significant values which are true positives, true negatives, false positives and false negatives. The model accurately forecasted 73 cases of non-heart diseases (true negatives) and 172 cases of heart diseases (true positives), which reflects a high level of classification. There were only 4 cases that were wrongly classified as heart disease when they were not (false positives) and none that were wrongly classified as non-heart disease when it was actually present (false negatives). Such low misclassifications indicate the strength of the system. Such a small number of the false negatives is especially significant in medical practice, where false alarms about heart disease can be fatal.



**Figure 3:** Confusion Matrix

The classification report in Figure 4 confirms the efficiency of our proposed system by providing evaluation metrics such as

precision, recall and F1 score. The precision for class 0 (non-disease) was 0.99 and the recall was 0.95, and a very small number of them were incorrectly classified. For class 1, the precision of our model was 0.98, with a recall of 0.99, which is a very good indication of its ability to correctly identify true cases of heart disease. The overall accuracy of the system is 0.98, which indicates that our model performs consistently well across all classes. The macro and weighted average scores of 0.98 show that all class matrices are performing well without having a positive or negative bias towards any class. These results demonstrate that by using DBSCAN clustering and fuzzy logic as a means to improve classification performance, we can identify complex patterns in data and produce useful predictions in healthcare systems.

Classification Report:

	precision	recall	f1-score	support
0	0.99	0.95	0.97	77
1	0.98	0.99	0.99	173
accuracy			0.98	250
macro avg	0.98	0.97	0.98	250
weighted avg	0.98	0.98	0.98	250

**Figure 4:** Classification Report

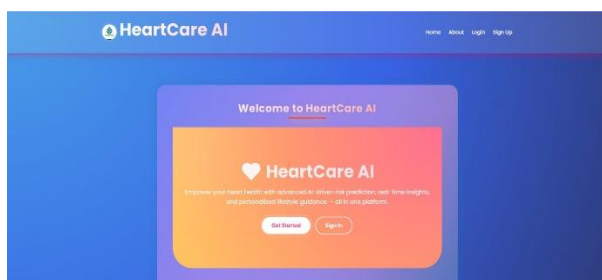
The interface of the HeartCare AI system of prediction results in Figure 5 is simple and demonstrates the outcome of the prediction in a clear and understandable way. The result presented above show that the system is able to identify a high-risk state and indicate the percentage of risks and immediate suggestions. This interface does not just inform the user about his or her health status but also offers him or her an actionable recommendation such as seeing cardiologists. Furthermore, the list of recommended specialists and common symptoms is provided by the system, which increases the

awareness and decision-making of the user. The design ensures that the non-technical users will not get lost in the results. It is not a simple classification tool. However, a mix of prediction output and medical recommendations will provide the user with valuable healthcare information.



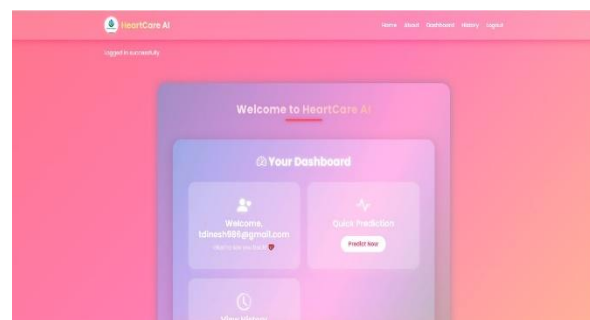
**Figure 5: Prediction Output Interface Analysis**

The home page (Figure 6) features a simple layout with well-placed buttons such as login, create account, and access dashboard, all of which help support the clarity of purpose of the HeartCare AI system to provide users with AI-generated heart disease predictions and individualised information on individual health. The use of visual appeal and organisation promote interaction and accessibility for the user as they are able to access the various functions of the application, such as registering and logging in. Additionally, the home page is designed so that users, whether they have experience with technology or not, will find the application user-friendly.



**Figure 6: Home page interference analysis**

The dashboard interface in Figure 7 acts as a central platform for users to access key components such as fast prediction capabilities and data analysis of their health history and trends. The features shown are user-specific, providing a personal experience and the rapid prediction feature allows for instant analysis of possible risks associated with their current situation, while the history feature allows the user to easily locate prior predictions and track changes over time, which assists users in analysing the changing trends associated with their health and taking preventative action as well. This interface is fairly straightforward and straightforward for users to navigate, so key components are easily accessible to users. Through integrated service delivery and one place for all processes, the system offers users greater productivity and pleasure by combining features from packaged software and customised services in one place.



**Figure 7: Dashboard Interface Analysis**

The input form interface is a key component within the overall component of the system as it gathers required health and lifestyle data to predict accurately shown in Figure 8. A variety of parameters are included in this form, including, but not limited to, age, sex, weight, blood pressure, cholesterol levels, glucose levels, smoking and drinking patterns and physical activity. The use of validation upon completion and drop-down lists to

ensure correct input reduce errors and/or incomplete information, providing for higher quality data collection. The input form design allows all end users to complete their information quickly and accurately without any confusion regarding what is being requested. The ability to collect data in a consistent and effective manner through the completion of the input form will directly impact the prediction model's overall effectiveness as the collection of better-quality data will produce more accurate outputs.



**Figure 8:** Input from Interface Analysis

## 5. Conclusion

The development of the intelligent heart disease prediction system using DBSCAN clustering and fuzzy logic represents a significant contribution to modern healthcare technology. This study aims to create an effective system for analysing health and lifestyle data from patients to assess their risk of heart disease. The proposed computational approach is sophisticated and incorporates a web-based interface to create the potential for a practical and accessible healthcare application that can be applied to real-world problems. The proposed method is effective in addressing the weaknesses of the traditional means of diagnosing a problem, as it presents an automated and data-driven approach. Data preprocessing makes clean and consistent input data, and DBSCAN

clustering provides the opportunity to detect hidden patterns and relationships in complex datasets. The application of fuzzy logic increases the system's ability to compensate for uncertainty and imprecision when making predictions, which provides a higher level of accuracy and understanding for the user. Features such as automated report generation, data visualisation, and a web-based design increase the ease of use and facilitate informed decision-making. The web-based design allows for remote access, which supports ongoing monitoring of health and prevention measures. Overall, the system is highly accurate, efficient, and reliable, making it a valuable tool for early diagnosis. This study leads to intelligent healthcare development and helps to achieve better patient outcomes and quality of life.

## REFERENCE

1. Ali, Md Liakot, Muhammad Sheikh Sadi, and Md Osman Goni. "Diagnosis of heart diseases: A fuzzy-logic-based approach." *Plos one* 19, no. 2 (2024): e0293112.
2. Shafiq, Sana, Wajahat Akbar, Altaf Hussain, Tariq Hussain, Abdullah Soomro, Irshad Khan, Muhammad Inam Ul Haq, and Fazal Adnan. "FuzzDL-HeartPredict: Heart Attack Risk Prediction using Fuzzy Logic and Deep Learning." *ICCK Transactions on Advanced Computing and Systems* 1, no. 2 (2025): 63-77.
3. Mani, Dilli Babu, and Sambath Munusamy. "Fuzzy Rule Based-Model for Proficient Heart Disease Prediction in Data Lake." *Revue d'Intelligence Artificielle* 37, no. 4 (2023).
4. Coffey, Michael, Fiona Lugg-Widger, Ben Hannigan, Viktoriya Velikova, and Anthony Byrne. "Severe mental illness and last

- year of life: Identifying service use from a National Health Service digital dashboard in Wales, UK." *Journal of Mental Health* 34, no. 5 (2025): 549-555.
5. Salehi, Amirreza, and Majid Khedmati. "Identifying at-risk patients for congenital heart disease using integrated predictive models and fuzzy clustering analysis: A cross-sectional study." *Heliyon* 10, no. 20 (2024).
  6. Saryazdi, Mohammad Dehghani, and Ali Mostafaepour. "Identification and validation of key predictive factors for heart attack diagnosis using machine learning and fuzzy clustering." *Engineering Applications of Artificial Intelligence* 142 (2025): 109968.
  7. Arief Kanza, Rafly, M. Udin Harun Al Rasyid, and Sritrusta Sukaridhoto. "Efficient early detection of patient diagnosis and cardiovascular disease using an IoT system with machine learning and fuzzy logic." *International Journal of Computing and Digital Systems* 16, no. 1 (2024): 183-199.
  8. Ramsingh, J., and V. J. A. S. C. Bhuvaneshwari. "An integrated multi-node Hadoop framework to predict high-risk factors of Diabetes Mellitus using a Multilevel MapReduce based Fuzzy Classifier (MMR-FC) and Modified DBSCAN algorithm." *Applied Soft Computing* 108 (2021): 107423.
  9. Muhyi, Diyar Fadhil, and Oğuz Ata. "Integrative machine learning approaches for enhanced cardiovascular disease prediction: a comparative analysis of XGBoost and ANFIS algorithms." *The Journal of Supercomputing* 81, no. 12 (2025): 1213.
  10. Khan, Hassan, Muhammad Dawood Majid, Hammad Ali, Abdullah Qureshi, Muhammad Umer Ali Awan, and Qasim Afraz. "Intelligent Hybrid System for Automated Heart Disease Prediction and Identification." *VFAST Transactions on Software Engineering* 13, no. 2 (2025): 308-320.
  11. Yang, Wen-Chieh, Jung-Pin Lai, Yu-Hui Liu, Ying-Lei Lin, Hung-Pin Hou, and Ping-Feng Pai. "Using medical data and clustering techniques for a smart healthcare system." *Electronics* 13, no. 1 (2023): 140.
  12. Manivannan, D., N. K. Manikandan, S. Sathiya Priya, and S. Bharathi. "A Smart Healthcare Application for Early Heart Disease Prediction: Leveraging Feature Optimization and Deep Learning Techniques." In *2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, pp. 823-829. IEEE, 2024.
  13. Jaafar, Abdul Hassan Majli, O. A. Mukhlus, Liwaa Ali Hussein, Karrar Hayder Shakir, Muhammad Musa Jaafar, Mohammed I. Habelalmateen, and Alaa Mohammed. "AI-Based Smart Prediction of Clinical Diseases: Random Forest and Naive Bayes Approach." In *2024 International Conference on IoT, Communication and Automation Technology (ICICAT)*, pp. 623-628. IEEE, 2024.
  14. Jokardarabi, Mohammad, Hasan Khademi Zare, Seyedeh mahdieh Namayandeh, Mohammad S. Owlia, and Mohammad mehdi Lotfi. "Development of a Decision-Maker System to Predict In-Hospital Mortality of Acute Myocardial Infarction Patients; an Artificial Intelligent Approach: Yazd Cardiovascular Disease Registry Data Use." (2022).
  15. Latha, S., Dhanalakshmi Samiappan, P. Muthu, and R. Kumar. "Fully Automated Integrated Segmentation of Carotid Artery Ultrasound Images Using DBSCAN and Affinity Propagation: S. Latha et al." *Journal of Medical and Biological Engineering* 41, no. 2 (2021): 260-271.