ISSN: 2584-0495 Vol. 2, Issue 12, pp. 1462-1472

International Journal of Microsystems and IoT

ISSN: (Online) Journal homepage: https://www.ijmit.org

A Fuzzy Inference System for Disease Diagnosis Through Blood Report Analysis

Sai Varshitha Gondela, A. Sai Charan, D. Rajeswara Rao, Akhila Gudala, and *Ashutosh Satapathy

Cite as: Sai Varshitha, G., A. Sai, C., D. Rajeswara, R., , Akhila, G., & Ashutosh, S. (2024). A Fuzzy Inference System for Disease Diagnosis Through Blood Report Analysis. International Journal of Microsystems and IoT, 2(12), 1462–1472. https://doi.org/10.5281/zenodo.15463072

Published online: 24 December 2024

DOI: https://doi.org/10.5281/zenodo.15463072

International Journal of Microsystems and IoT
Vol.2, Issue 12, pp.1462-1472; DOI: https://doi.org/10.5281/zenodo.15463072

A Fuzzy Inference System for Disease Diagnosis Through Blood Report Analysis

Sai Varshitha Gondela[1], A. Sai Charan[2], D. Rajeswara Rao[1], Akhila Gudala[1], and *Ashutosh Satapathy[1]

[1]Department of Computer Science and Engineering, Siddhartha Academy of Higher Education, Deemed to be University, Vijayawada, Andhra Pradesh, India.
[2]Lead Data Analyst, Sutherland Global Services, Hyderabad, Telangana, India.

ABSTRACT
Limited education and complexity of medical reports make it difficult to understand them. Seventy-two percent require help in understanding blood test results as research confirms. A user-friendly GUI application has been developed for this purpose, which can simplify the process of interpreting medical reports. This particular study includes medical reports like Complete Blood Count, Widal test, rapid test, and serology test. It is achieved by making use of Fuzzy Inference System that accepts user inputs and diagnoses diseases based on reported results. The fuzzy model dataset consists of 33 rows and 12 attributes. The main goal is to make medical report more understandable so that people can make better-informed decisions about their health care options. Initiatives such as this one aim at advancing healthcare literacy levels thereby improving health outcomes as well as allowing individuals to become proactive regarding their wellbeing. This fuzzy model has an impressive accuracy rate of 95.84%. This article presents a new way being used to increase the accessibility of medical report interpretation to the masses. Through employment of fuzzy logic with a friendly user interface, complex medical jargons are broken down into simple terms that could be understood easily hence enhancing patient engagement and health literacy by providing clear actionable insights from intricate medical data sets.

KEYWORDS
Fuzzy Inference System, Blood Report Analysis, Red Blood Cells, White Blood Cells, Platelets, Complete Blood Count, Gaussian Membership function

1. INTRODUCTION

In healthcare diagnostics, blood report analysis is important due to its ability to provide insight into the general health of a person and detect underlying medical conditions. Deriving crucial information from parameters such as red cell counts; white cell counts and platelet levels calls for professional involvement in this field. This study focuses on enhancing the accessibility and interpretability of medical reports using fuzzy inferencing. Blood tests, particularly those analyzing Red Blood Count (RBC), White Blood Count (WBC), and platelets, provide vital data for disease diagnosis and treatment monitoring, with the Complete Blood Count (CBC) serving as a primary health indicator [1]. Physicians rely on blood test results, comparing them against established reference ranges that consider factors such as age and gender. The proposed model aims to simplify the interpretation of these results, empowering individuals to understand their health status and take proactive measures for better healthcare outcomes. By leveraging fuzzy logic principles, this model seeks to enhance diagnostics and facilitate personalized healthcare management [2].

Fuzzy inferencing and fuzzy logic principles hold profound relevance to the proposed model focused on blood report analysis for disease diagnosis [3]. Fuzzy inferencing is a robust methodology for drawing conclusions and making predictions based on imprecise or uncertain data, a characteristic often encountered in medical diagnostics. The proposed model leverages fuzzy inferencing techniques to interpret blood test results, transforming crisp data into fuzzy values through membership functions that define the degree to which an input variable belongs to a specific set. Using fuzzy sets and rules, it regulates the link between different blood parameters, allowing a prediction of the likelihood or seriousness of an ailment based on input values. It is a robust mechanism for simulating data-driven systems. The main components of the fuzzy reasoning technique (Figure 1) are membership functions, fuzzy sets and rules, rule evaluation and aggregation, defuzzification [4]. Membership functions delineate how much an input variable aligns with a particular set. They represent linguistic terms, transforming crisp data into fuzzy values. A membership function can define how critical the disease is based on some parameters. Fuzzy sets are collections of values that describe the boundaries of membership functions.

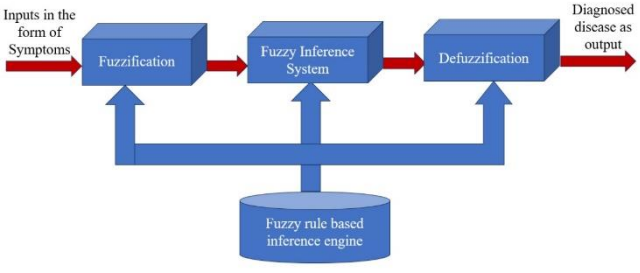

Fig. 1 Fuzzy rule-based inference model

In fuzzy inferencing, rules govern the relationships between fuzzy sets. These rules determine the output based on the input values. During the evaluation phase, each rule is assessed to determine its level of applicability or activation. Aggregation techniques such as maximum or minimum help

combine the outputs of different rules into a single value. This value represents the overall inference [5]. Defuzzification involves transforming the fuzzy output into a precise value. Disease diagnosis is a crucial process in healthcare that involves identifying and understanding the underlying causes of a person's health issues. One common aspect is that blood tests help reveal potential irregularities and determine the individual's overall health. Fuzzy logic is not just a mathematical framework, but a versatile tool that allows for greater flexibility by accommodating imprecision and uncertainty. The model built was based on fuzzy logic, a principle that works on the idea that things can be true to a certain degree rather than purely true or false [6]. Its adaptability is demonstrated by its main characteristics such as linguistic variations, fuzzy operations and applications of fuzzy logic. Fuzzy logic uses the concept of linguistic variables which help in expressing imprecise or uncertain data in terms used in daily language. This is very important for the purposes of calculating and making decisions based on inputs that are fuzzy. Control systems, decision processes and pattern recognition among others are some of the interesting places where it has found application (see Figure 2). This flexibility therefore makes it a flexible tool that can be studied and tailored by researchers depending on their need.

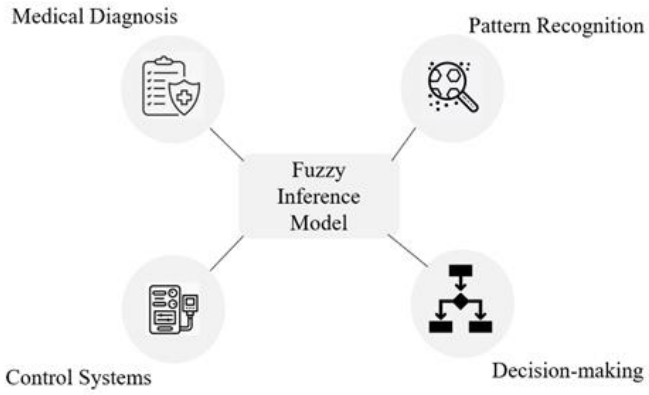

Fig. 2 Various applications of the fuzzy inference model

In the forthcoming sections, we delve into the existing research conducted on blood report analysis. Furthermore, we present the development of a Fuzzy Inference System designed for disease prediction utilizing blood report analysis. The outcomes of this system are rigorously validated using data sourced from multiple authentic outlets and previous implementations in the field.

2. LITERATURE REVIEW

In 2023, Shubin et al. introduced a cost-effective diagnostic model aimed at enhancing the efficiency of medical processes concerning Chronic Hepatitis B Virus (HBV) related diseases [7]. The model, founded on clinical blood tests, discerned critical states by identifying essential items and their corresponding values. Central to its architecture were the utilization of total protein tests and blood clotting tests. Employing a dataset sourced from the Department of Infectious Diseases and Hepatobiliary Surgery, the model achieved a precision rate of 84.47%. Similarly, Shankar et al. in 2019 proposed a diagnostic model for diabetes leveraging fuzzy logic in conjunction with grey wolf optimization [8]. Utilizing the PIMA dataset from the UCI machine learning repository, which encompasses pertinent data for diabetes prediction, the model attained an accuracy level of 81.1585%. In a study dating back to 2010, Ali et al. devised a Fuzzy Expert System tailored for heart disease diagnosis [9]. This model integrated multiple parameters including blood pressure, blood sugar levels, heart rate, age, and sex as inputs, employing the Mamdani inference method. Utilizing datasets from the Long Beach and Cleveland Clinic Foundation, the model exhibited an accuracy of 91.58%. Furthermore, Ghosh et al. proposed a monitoring model in 2020 that incorporates fuzzy logic to address blood pressure (BP) and blood sugar control in patients suffering from both conditions. The model generates alarms to warn patients of their stability or changes in their condition. The authors justify the importance of their model through fuzzy inference rules, first-order logic, and analytical reports [10]. This model demonstrated an accuracy rate of 92.5%.

In 2019, Sunita et al. introduced a model aimed at early-stage diagnosis of leukemia, a critical step for improving the prognosis of the disease [11]. Employing a Support Vector Machine (SVM) and an Extreme Learning Machine (ELM), the model focused on detecting blood or bone marrow cancer based on cell count data extracted from clinical reports. Utilizing the publicly available ALL-IDB1 dataset, the model achieved an accuracy of 92.2448% with ELM and 86.3636% with SVM. In 2022, Bashir and his colleagues propose a model for detecting kidney disease [12]. The fuzzy membership functions used in this model for data classification were exploited to develop diagnosis outcome. By applying the cognitive dataset, the model achieved an accuracy of 92%. Similarly, in 2015, Rian et al. presented a methodology for the early detection of Diabetes Mellitus [13]. Empowered by a fuzzy hierarchical model, the proposed approach mimicked the diagnostic process of medical professionals, drawing conclusions from patient reports to determine the presence or absence of the disease. With guidance from two medical doctors to enhance system efficiency, the model utilized data from Eastern Jakarta Hospital, achieving an accuracy level of 87.46%.

In 2023, Yan et al. introduced a model designed for early disease prediction within the context of smart healthcare systems [14]. Their proposed approach utilized a fuzzy expert model integrating fuzzy clustering to eliminate outliers, a fuzzy logic classifier to address missing data, and a forest optimization algorithm. Employing datasets related to diabetes and heart disease, the model achieved an accuracy rate of 85%. Similarly, Bahani et al. proposed a model focusing on heart disease diagnosis with an emphasis on early detection [15]. Their methodology utilized fuzzy clustering to establish a fuzzy rule-based system, aiming to mitigate erroneous decision-making in disease diagnosis by leveraging accurate decisions derived from trained data. Utilizing datasets from Cleveland and Combined human heart disease repositories, the model achieved accuracies of 84.84% with the Naive Bayes (NB) classifier, 84.19% with Support Vector Machines (SVM), and 83.87% with Artificial Neural Networks (ANN). In another domain, Anand et al. in 2023 presented a model for Malaria Parasite detection and counting using YOLO-mp [16]. In

response to the persistent challenge of malaria-related mortality, their model leveraged Rapid Diagnostic Test (RDT) data from the Makerere AL Lab, achieving an impressive accuracy of 94%.

Furthermore, Basso et al. proposed a fuzzification algorithm designed to automatically process data values based on their distribution, incorporating variables from datasets or expert-provided information [17]. This enhancement facilitated accurate data analysis and informed decision-making, culminating in an accuracy rate of 88%. Khan et al. demonstrated an effective method for distinguishing between normal and diseased serum utilizing a Support Vector Machine (SVM) algorithm. Their model achieved a diagnostic accuracy of approximately 98% and a precision of around 97% [18]. In a comprehensive survey, Kumar et al. examined the use of AI techniques in diagnosing a plethora of illnesses such as Alzheimer's, tumors, diabetes, heart disease, TBs and CVS (Cerebrovascular Diseases), high blood pressure among others. For example, the study reported that the system achieved 95.7% accuracy in diagnosing diabetes using Convolution Neural Network (CNN) [19]. Maini et al. developed a machine learning-based prediction system for the early diagnosis of cardiovascular diseases (CVDs), which is accessible via the Internet. Their study highlighted the effectiveness of Logistic Regression (LR) in predicting heart disease risk, with the model correctly classifying 470 out of 501 medical records, resulting in a diagnostic accuracy of 93.8%. Additionally, the model demonstrated a sensitivity of 92.8% and a specificity of 94.6% [20]. Another approach by Subhadra et al., was to apply Multilayer Perceptron Neural Network (MLP) with backward propagation algorithm for training as a heart disease prediction system. Their model had five neurons in the hidden layer and showed a high accuracy level of 93.39% [21].

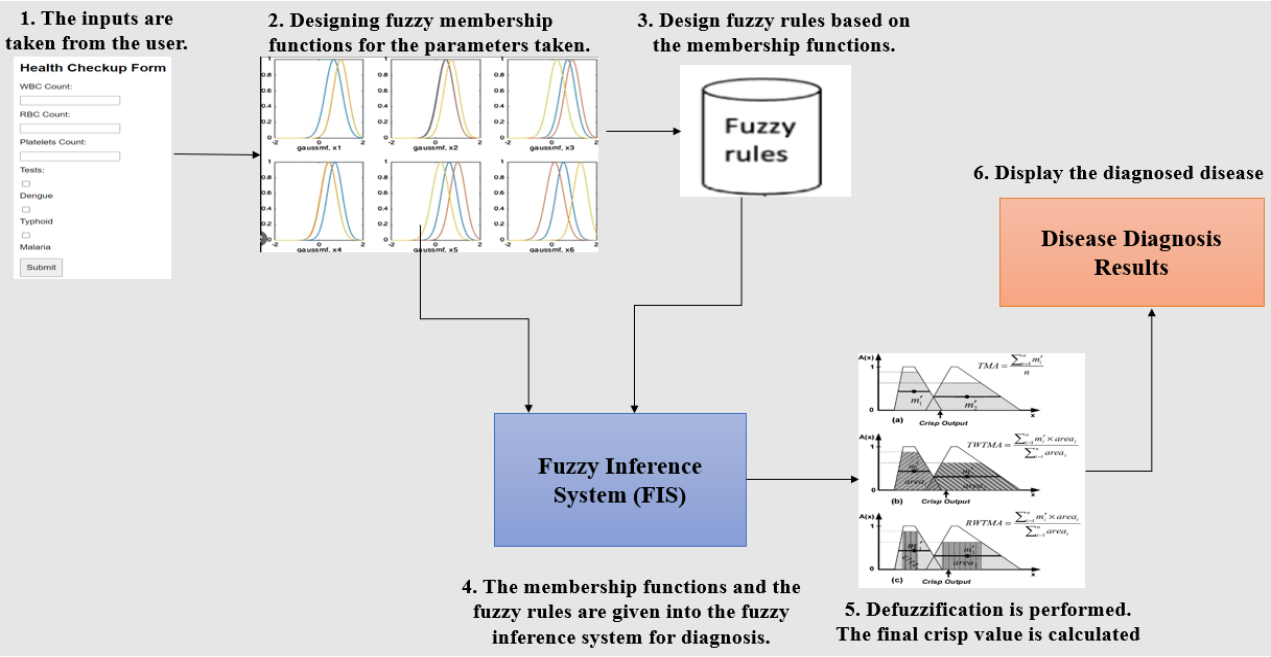

Fig. 3 Process flow diagram of FIS for disease diagnosis.

3. METHODOLOGY

3.1. Proposed Workflow Diagram

The first step is for the diagnostic solution to get specific inputs from the user, such as WBC count, RBC count, and Platelet count. A membership function is used to process these input parameters which compute membership values between 0 and 1. These numerical quantities are then used in fuzzy rules that enable fuzzy decision-making. It involves Fuzzy Inference System (FIS) where it takes blood test data as inputs from a user and crisp input values are converted into fuzzy values using membership functions. These membership functions define how much an input value belongs to a particular fuzzy set [22]. After this stage, each fuzzy rule is evaluated based on how its antecedents (conditions) are satisfied by fuzzified input values. They finally convert the fuzzy output back into crisp values via defuzzification process. The resultant crisp value represents the diagnostic decision which is then shown to the user. This multi-stage process is shown in figure 3 that explains how information flows between different components. At first, crisp input values are changed into fuzzy ones by means of membership functions. Membership functions are vital since they quantify the degree of membership for each input within a fuzzy set. Then comes to applying the fuzzy rules whereby their evaluations depend on whether or not their antecedents have been satisfied by the fuzzified inputs. The last phase includes converting the fuzzy output back into a crisp value thus presents an explicit and understandable diagnostic decision or prediction.

The flow of control in a system is illustrated by the Activity Diagram and refers to the steps required for executing a use case. It is significant for activity diagrams to focus on its flow condition and sequence. Figure 4 presents an activity diagram where a user uploads inputs into web application. It outlines how a user, an application, a database, and FIS interact with

one another. This shows an interaction between a user and the application, database and FIS system. Firstly, user interacts with application upon which he or she either signs up for an account or logs in using his/her credentials. Depending on what action they take, the app may display either login or home page. If credentials have been entered by a person trying to log in, their verification against database takes place at this point. After successful login, users are taken to a homepage where they can input blood test results values through forms provided. The user enters their blood test details into the system which then transmits them to FIS. FIS likely uses fuzzy logic rules during analysis of these blood tests data that was put by the patient. Fuzzy Logic as machine learning mechanism handles uncertainties and approximates real world situations using membership functions. Every point of data has membership functions that assign degrees of membership [23]. Upon completion of computations, inferred disease details are passed by FIS back to application. Lastly it displays disease details to users as part of its procedure.

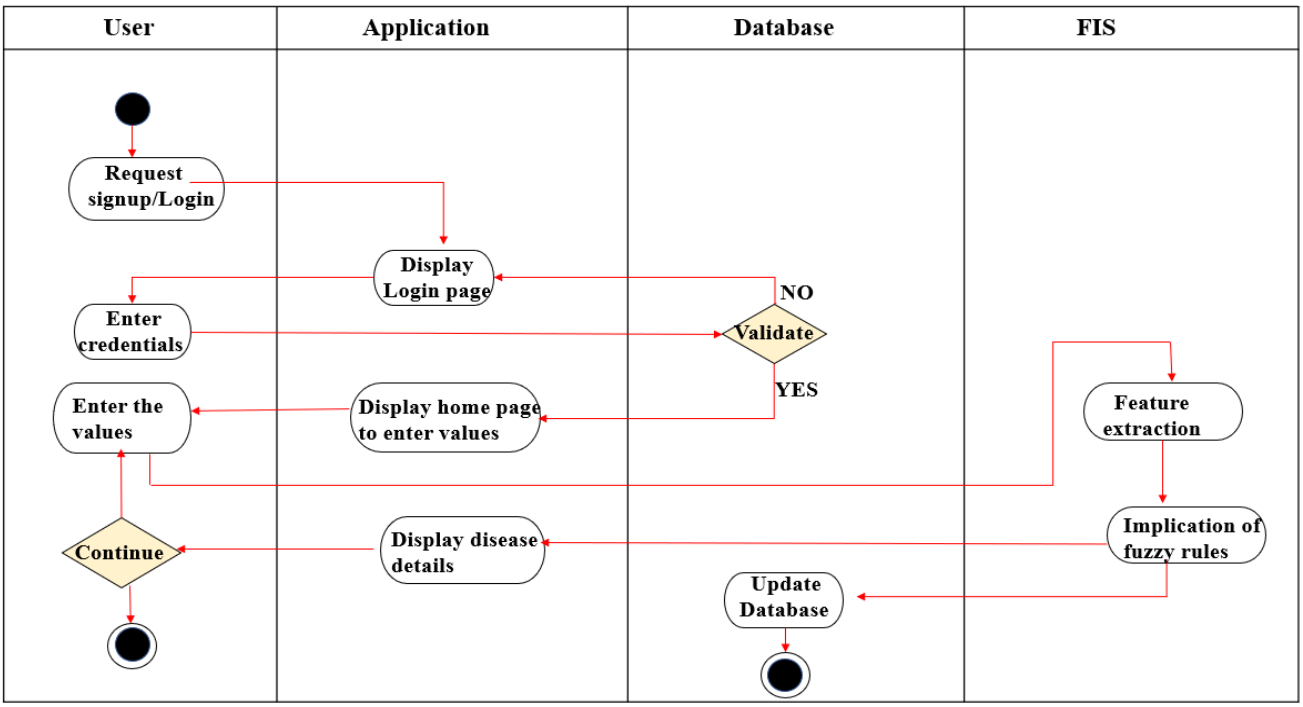

Fig. 4 Activity diagram for blood report analysis system

3.2. Gaussian Membership Function

Gaussian membership functions are integral to fuzzy logic, providing a sophisticated means of representing uncertainty. These bell-shaped curves depict the likelihood of an element belonging to a fuzzy set, with the membership degree diminishing gradually from the center of the curve outward. Their smooth and continuous nature makes Gaussian membership functions particularly effective for modeling the inherent variability found in blood test results [24]. In blood test analysis, Gaussian membership functions enhance the system's ability to manage uncertainty in individual data points, thereby facilitating more precise and reliable diagnostic inferences. For designing the fuzzy rules in this study, Gaussian membership functions were adopted [25]. Initially, triangular, and trapezoidal membership functions were considered, but they led to significant defuzzification errors. This issue arose because WBC, RBC, and Platelet counts do not have distinct boundary limits to define higher and lower values. The uniform distribution characteristic of the Gaussian bell curve makes it an ideal fit for this application. Figure 5 illustrates the output curves associated with the variable diagnosis. These curves demonstrate the effective application of Gaussian membership functions in generating accurate diagnostic outcomes. The adoption of Gaussian membership functions thus ensures that the system can handle the continuous nature of blood test parameters, reducing errors and improving the overall reliability of the diagnostic process.

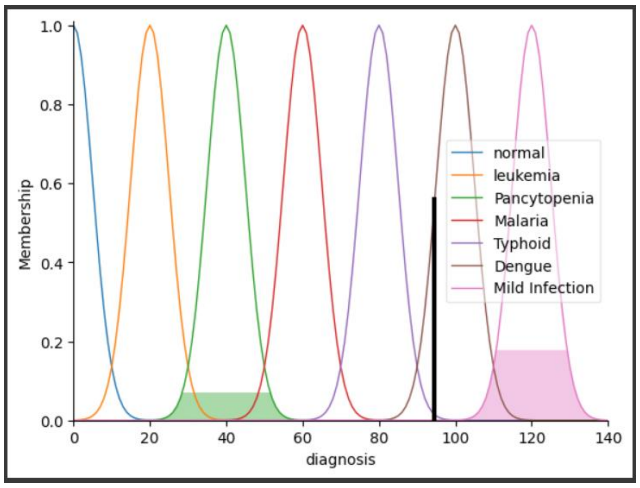

Fig. 5 Gaussian Membership functions of diagnosis fuzzy sets.

This study utilizes fuzzy sets and variables whose descriptions are provided in Table 1. Each fuzzy variable has a mean denoted by μ and a standard deviation denoted by σ. The

values are employed to construct bell curves representing Gaussian membership functions.

Table. 1 Description of fuzzy variables and sets

S. No	Fuzzy Variables	Representation	Fuzzy Sets	Representation of fuzzy numbers	Fuzzy gaussian numbers
1	WBC	wbc	Low	low	μ = 2000, σ = 1000
			Normal	normal	μ = 7000, σ = 1500
			Medium	medium	μ = 30000, σ = 10000
			High	high	μ = 100000, σ = 25000
2	RBC	rbc	Low	low	μ = 1.85, σ = 0.925
			Normal	normal	μ = 4.25, σ = 0.275
3	Platelets	platelets	Low	low	μ = 75000, σ = 37500
			Normal	normal	μ = 225000, σ = 37500
4	Output	Diagnosis	Normal	normal	μ = 0, σ = 5
			Leukaemia	leukaemia	μ = 20, σ = 5
			Pancytopenia	pancytopenia	μ = 40, σ = 5
			Malaria	malaria	μ = 60, σ = 5
			Typhoid	typhoid	μ = 80, σ = 5
			Dengue	dengue	μ = 100, σ = 5
			Mild Infection	mild infection	μ = 120, σ = 5

The generalized Gaussian membership function is given by:

$$\boldsymbol{\mu}(\mathbf{x}) = \begin{cases} e^{-\frac{1}{2}\left(\frac{|x-c|}{\alpha}\right)^{m}}, & \text{If } x \in \text{ some range} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

The standard Gaussian membership function is given by:

$$\boldsymbol{\mu}(\mathbf{x}) = \begin{cases} e^{-\frac{(x-c)^2}{2\sigma^2}}, & \text{If } x \in \text{ some range} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Equation (1) indicates the generalized Gaussian membership where $\mu(x)$ is the membership value at point x, c represents the average of the Gaussian distribution, α represents the spread of the membership function and m regulates its shape. Equation (2) suggests standard Gaussian membership where $\mu(x)$ is the membership value at point x, c denotes center of a Gaussian distribution and σ refers to variance.

3.3. Fuzzy Rules

In this diagnostic model, a total of 16 fuzzy rules have been established. Of these, five rules specifically address the diagnosis of five distinct diseases, one rule pertains to normal health conditions, and the remaining ten rules are designated for mild infections, which represent cases not classified under the specific diseases. Table 2 provides a detailed depiction of these fuzzy rules, illustrating the various diseases that can be diagnosed based on abnormal levels of WBCs, RBCs, and platelets. According to these rules, if the levels of WBCs, RBCs, and platelets are all within the normal range, the system concludes that the patient is in a normal health condition. These fuzzy rules are crucial for the accurate identification and differentiation of health conditions, leveraging the nuances of abnormal blood parameters to offer precise diagnostic insights. This structured approach ensures comprehensive coverage of potential health scenarios, from specific diseases to mild infections, thereby enhancing the robustness and reliability of the diagnostic system.

4. EXPERIMENTAL SETUP

4.1. Data Collection

A dataset was created through consultation with medical professionals and by sourcing information from reputable websites, including MedlinePlus [26] for diagnostic test information, Medscape [27] for standards and interpretations, and Verywell Health [28]. Medical reports pertinent to the diseases of interest were gathered initially. These reports underwent meticulous examination to identify crucial diagnostic details. This information was structured within the dataset like a table, with identified details listed as columns and each medical report represented as a row. The rows were

categorized into male and female groups. Subsequently, standard values were established for each detail, tailored to specific diseases and gender. The utmost attention was given to ensuring the accuracy of collected information and adherence to ethical guidelines, particularly concerning patient confidentiality. Thorough documentation was maintained, delineating the data's sources and the measures taken to ensure its reliability. Plans are in place for regular dataset updates to uphold its utility, with collaboration with medical professionals to validate its suitability for medical applications.

4.2. Data Preparation

Table. 2 Fuzzy rules of FIS for disease diagnosis.

S. No	Fuzzy Rules
1	If (WBC is normal), (RBC is normal), and (platelets are normal), then the patient is normal.
2	If (WBC is high), but (RBC is low) and (platelets are low), then the patient has Leukaemia.
3	If (WBC is low), (RBC is low), and (platelets are low), then the patient has Pancytopenia.
4	If (WBC is normal) and (platelets are normal), but (RBC is low), then the patient has Malaria.
5	If (WBC is medium), (RBC is normal), and (platelets are normal), then the patient has Typhoid.
6	If (WBC is normal), (RBC is normal), and (platelets are low), then the patient has Dengue.
7	If (WBC is low), (RBC is low), but (platelets are normal), then the patient has Mild Infection.
8	If (WBC is low), (platelets are low), but (RBC is normal), then the patient has Mild Infection.
9	If (WBC is low), (RBC is normal), and (platelets are normal), then the patient has Mild Infection.
10	If (WBC is high), (RBC is low), and (platelets are normal), then the patient has Mild Infection.
11	If (WBC is high), (RBC is normal), and (platelets are normal), then the patient has Mild Infection.
12	If (WBC is high), (RBC is low), but (platelets are normal), then the patient has Mild Infection.
13	If (WBC is normal), (RBC is low), and (platelets are normal), then the patient has Mild Infection.
14	If (WBC is medium), (RBC is low), and (platelets are normal), then the patient has Mild Infection.
15	If (WBC is medium), (RBC is normal), and (platelets are low), then the patient has Mild Infection.
16	If (WBC is medium), (RBC is low), and (platelets are normal), then the patient has Mild Infection.

5. RESULTS AND DISCUSSION

5.1. Performance Metrics

Performance metrics are crucial for evaluating the effectiveness of a fuzzy inference system used in disease diagnosis based on WBC, RBC, and platelet counts. These metrics help determine the accuracy, precision, and reliability of the model in identifying various health conditions. By assessing these metrics, we can ensure that the system provides accurate and dependable diagnostic results, thereby improving patient outcomes and enhancing the overall effectiveness of medical diagnostics.

Data collection is followed by data preparation and data pre-processing. Data from multiple sources have various standards and are usually in the form of unstructured data. This information contains text data, missing values, and null values. Nevertheless, fuzzy analysis requires numerical and precise data. As a result, it was decided to process the data in accordance with these requirements. At first, data is arranged into rows and columns: rows stand for parameters that need to be taken into account while columns represent diseases that require diagnostics. Remember that not all parameters may be applicable to every disease. Missing values are excluded from the dataset whereas null values too are removed to ensure accuracy in real-time situations.

1. **True Positive (TP_i):** The count of items that were correctly classified as class *i*. It is computed by tallying the number of cases where the real class is *i* and so is the predicted class.

$$\mathbf{TP_i} = \sum_j 1(y_j = i \text{ and } \hat{y}_j = i) \quad (1)$$

2. **False Positive (FP_i):** The number of instances incorrectly classified as class *i*, though they belong to other classes. It is calculated by counting instances where the actual class is not *i*, but the predicted class is *i*.

$$\mathbf{FP_i} = \sum_j 1(y_j \neq i \text{ and } \hat{y}_j = i) \quad (2)$$

3. **True Negative (TN_i):** The number of instances rightly classified as not in class *i*. It is calculated by counting where neither the actual class nor the predicted class is *i*.

$$\mathbf{TN_i} = \sum_j 1(y_j \neq i \text{ and } \hat{y}_j \neq i) \quad (3)$$

4. **False Negative (FN_i):** The number of instances belonging to class i but incorrectly classified as another class. It is calculated by counting instances where the actual class is *i*, but the predicted class is not *i*.

$$\mathbf{FN_i} = \sum_j 1(y_j = i \text{ and } \hat{y}_j \neq i) \quad (4)$$

Here, in the given context, y_j represents the true class label for the j^{th} example or instance, $\hat{y}_j$ represents the predicted class label for the j^{th} example or instance and 1 is an indicator function which returns 1 if condition is true otherwise 0.

5. **Accuracy**: General correctness of a model which is a ratio between all correct predictions (TP and TN) and total population (instances).

$$\mathbf{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (5)$$

6. **Precision:** Positive predictive value, or precision, represents how many actual positive cases were detected correctly.

$$\mathbf{Precision} = \frac{TP}{(TP + FP)} \quad (6)$$

7. **Sensitivity (Recall):** Recall rate or true positive rate shows how well algorithm detects positive cases among all possible ones.

$$\textbf{Recall} = \frac{TP}{(TP + FN)} \quad (9)$$

8. **Specificity:** It refers to ability to recognize negative events out of all available ones which are truly such in reality.

$$\textbf{Specificity} = \frac{TN}{(TN + FP)} \quad (10)$$

9. **F1-Score:** A measure that balances precision and recall using their harmonic mean; it proves useful when there is a significant discrepancy between classes being studied.

$$\textbf{F1-Score} = \frac{2TP}{(2TP + FP + FN)} \quad (11)$$

To sum up, computation of performance measures from the confusion matrix gives a comprehensive understanding of how well the diagnostic model works. Accuracy provides a general indication of how right it is in its predictions, precision shows the dependability of positive predictions; recall evaluates its ability to capture actual positives; specificity gauges its capacity to spot negatives, while F1-score takes care of the precision-recall trade-offs. All these parameters help assess accuracy and reliability for various diseases with which this model is concerned.

5.2. Analysis of Results

The multi-class classification is illustrated in Figure 6, where each class undergoes testing against 25 test cases. This multiclass classification involves five classes, each representing a disease for diagnosis. Correct results are positioned off the diagonal, and instances where a different class is output are recorded within their respective classes. Consequently, a comprehensive 5x5 confusion matrix is formulated, encompassing rigorous testing of all classes, and considering real-time cases [30]. TP is computed by extracting the diagonal elements for each class. TN comprises all values in the table, excluding the corresponding row and column of the class. FP encompasses all values in the column corresponding to the class, excluding TP. FN includes all values in the row corresponding to the class, excluding TP. The mathematical formulas used for computing TP, FP, TN, and FN are depicted in equations (3)-(6) respectively. After all observations were recorded, a confusion matrix was built from them. General mathematical operations were performed to calculate TP, TN, FP, and FN over all test cases considered. Additionally, accuracy, precision, f1-score, specificity, etc., were computed using equations (7)-(11), and results are noted in Table 4. These performance metrics were individually calculated for five diseases that the model can diagnose.

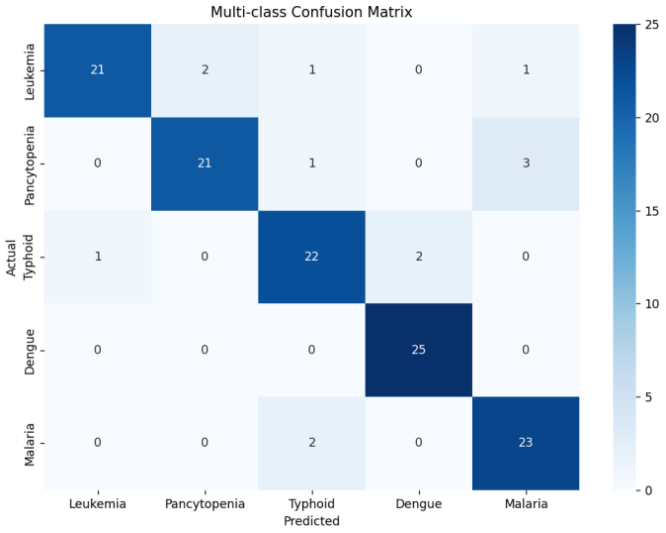

Fig. 6 Multi-class confusion matrix by FIS for disease diagnosis.

Table. 4 Performance metrics of individual diseases.

S. No	Disease	Accuracy	Precision	F1-Score	Sensitivity	Specificity
1	Leukaemia	96.00%	95.45%	89.36%	84.00%	99.00%
2	Pancytopenia	95.20%	91.30%	87.50%	84.00%	98.00%
3	Typhoid	94.40%	84.62%	86.27%	88.00%	96.00%
4	Dengue	98.40%	92.59%	96.15%	100.0%	98.00%
5	Malaria	95.20%	85.19%	88.46%	92.00%	96.00%

Each disease class is described in Table 4 concerning accuracy, precision, F1-score, sensitivity (recall), and specificity. These metrics are calculated using standard formulas. TP, TN, FP, and FN values are derived from a pre-computed confusion matrix for each category. The overall performance of the model can be seen in Table 5, which illustrates various comprehensive performance measures, including its accuracy, precision, F1 score sensitivity, and specificity. The model was subjected to 125 test cases for testing purposes, and it achieved an accuracy rate of 95.84%, which is very impressive. Furthermore, as an indicator showing how well positive predictions were made among all such predictions being made, precision stands high at 89.83%. There should be no doubt about this model's capability since it has performed excellently on important evaluation criteria, such as having an f1 score of 89.54%, recall of 89.6%, and specificity equaling 97.4%. This table also shows some important things about the blood report analysis system, which might include, among others, that it can detect abnormalities in blood tests with up to 95.84% accuracy levels when identifying anomalies or patterns within them (Table 5).

Table 5. Overall performance metrics of FIS

Metrics	Value
Accuracy	95.84%
Precision	89.83%
F1-Score	89.54%
Sensitivity	89.6%
Specificity	97.4%

This level of precision underscores the system's potential to significantly enhance patient care by streamlining blood report interpretation, reducing human error, and providing timely and accurate insights to healthcare professionals. However, to fully unlock the system's potential, future research should focus on expanding the dataset, integrating with electronic health records, developing personalized risk

prediction models, and improving the user interface. These research directions are crucial for the ongoing development and potential of the system.

Table. 6 Comparison between performance of FIS with existing implementations.

S. No.	Literature	Classifier/ Model	Accuracy/ Precision/ F1-Score/ Sensitivity/ Specificity
1	Shubin et al. [7]	Network models	Precision 84.47%
2	Shankar et al. [8]	Fuzzy logic with grey wolf optimization	Accuracy 81.16%
3	Ali et al. [9]	Mamdani inference model	Accuracy 91.58%
4	Ghosh et al. [10]	Fuzzy inference model	Accuracy 92.5%
5	Sunita et al. [11]	ELM, SVM	Accuracy 92.24% Accuracy 86.36%
6	Basheer et al. [12]	Fuzzy model	Accuracy 92%
7	Rian et al. [13]	Fuzzy hierarchical model	Accuracy 7.46%
8	Yan et al. [14]	Fuzzy clustering model and a forest optimization algorithm	Accuracy 85%
9	Bahani et al. [15]	NB Classifier	Accuracy 84.84%
10	Anand et al. [16]	YOLO-mp	Accuracy 94%
11	Basso et al. [17]	Fuzzification algorithm	Accuracy 88%
12	Khan et al. [18]	SVM	Accuracy 98% Precision 97%
13	Kumar et al. [19]	CNN	Accuracy 95.7%
14	Maini et al. [20]	LR	Accuracy 93.8% Sensitivity 92.8% Specificity 94.6%
15	Subhadra et al. [21]	MLP	Accuracy 93.39%
16	Tamer et al. [22]	Adaptive Neuro-Fuzzy Inference System (ANFIS)	Accuracy 97%
17	Kemal et al. [23]	PCA, ANFIS	Accuracy 89.47%
18	Hammam et al. [24]	ANFIS	Accuracy 95% Sensitivity 75% Specificity 97.25%
19	Faisal et al [25]	ANFIS	Accuracy 86.13% Sensitivity 87.5% Specificity 86.7%
20	Proposed Work	FIS	Accuracy 95.84% Precision 89.63% F1-Score 89.54% Sensitivity 89.6% Specificity 97.4%

Table 6 provides a comprehensive comparison of the current and previous projects, outlining the methodologies and results of various existing implementations. It also illustrates the contrast in accuracy, precision, F1-score, specificity, and sensitivity scores between the proposed work and other implementations. The proposed blood report analysis system using fuzzy logic demonstrates superior performance compared to existing models, achieving an impressive 95.84% accuracy in diagnosing diseases such as leukemia, pancytopenia, malaria, dengue, and typhoid, based on WBC, RBC, and platelet counts. The system's precision (89.83%) and F1-score (89.54%) highlight its reliability. Smoothness in modeling and strong diagnostic reasoning are ensured by this as it deals with blood test result variations using Gaussian membership functions. This makes it easy for anyone to input values obtained from blood tests because of its friendly interface hence making them accessible and reducing chances of errors attributable to humans. The precision rate recorded by a low-cost model that diagnoses diseases related to HBV was 84.47% [7] while fuzzy logic achieved 81.16% accuracy during diabetes detection with grey wolf optimization being used [8]. Furthermore, a Mamdani inference-based fuzzy expert system attained an accuracy level of 91.58% when diagnosing heart problems [9]. A monitoring model in 2020 that incorporates fuzzy logic to address blood pressure (BP) and blood sugar control in patients suffering from both conditions attained accuracy of 92.5% [10]. Models using SVM and ELM for leukemia diagnosis achieved 92.24% and 86.36% accuracies, respectively [11]. Another system using fuzzy membership functions for kidney disease detection reached 92% accuracy [12].

Further comparisons include a fuzzy hierarchical model for early diabetes detection achieving 87.46% accuracy [13], a fuzzy expert model for early disease prediction in smart healthcare with 85% accuracy [14], and fuzzy clustering for heart disease diagnosis achieving up to 84.84% accuracy with Naive Bayes [15]. A YOLO-mp model for malaria detection reported 94% accuracy [16], while a fuzzification algorithm achieved 88% accuracy [17]. Diagnostic accuracy of 98% was noted for separating normal and diseased serum using SVM [18], and AI techniques for diagnosing various diseases achieved 95.7% accuracy for diabetes using CNN [19]. Logistic regression for heart disease prediction reached 93.8% accuracy [20], and MLP for heart disease prediction achieved 93.39% accuracy [21]. Overall, the innovative application of fuzzy logic in the proposed blood report analysis system positions it as a leading solution in healthcare diagnostics, surpassing many existing models in both accuracy and reliability.

5.3. GUI Implementation

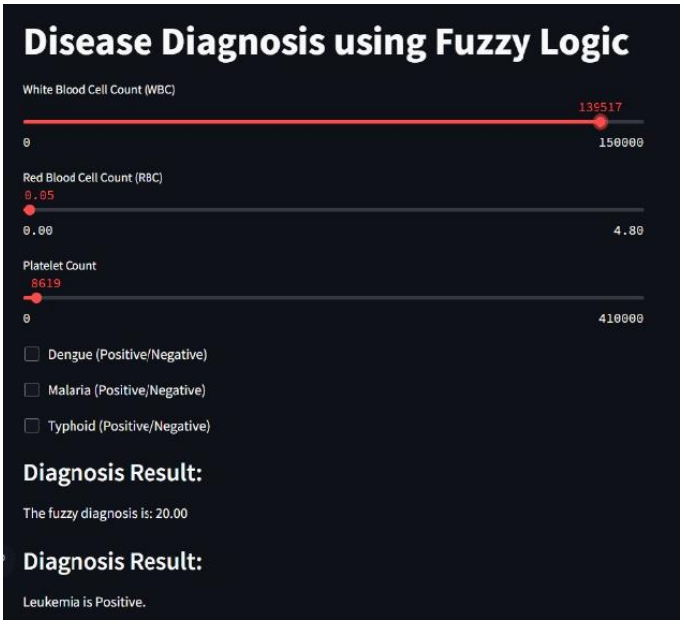

Fig. 7 GUI of inputs from a blood report and its diagnosis result.

The web application allows users to select WBC, RBC, and platelet counts from blood reports using three sliders presented as shown in Fig. 7. The sliders, labeled WBC, RBC, and Platelets, are used by the application to enable the selection of appropriate WBC, RBC, and platelet counts from the patient's blood report by moving the sliders to the desired values. The user can select the diseases by checking the boxes next to each. It shows the test results that can indicate a patient's disease. Medical professionals play an active role in diagnosing by selecting the desired values and clicking the Submit button. This action triggers the display of a diagnosis and the ticking of checkboxes for corresponding diseases. The report interpretation area then provides insights into the findings or results. Successful deployment of the GUI application in Streamlit cloud has been achieved. The application (see Figure 7) indicates low RBC and platelets when WBC is high, which suggests Leukemia. The entered WBC value was 139517, RBC was 0.05, and Platelet count was 8619, leading to negative findings for Dengue, Malaria, and Typhoid and a positive finding for Leukemia with a fuzzy diagnosis score of 20, which identified the disease as Leukemia.

6. CONCLUSION

In this work, an FIS has been proposed that leverages fuzzy rules for disease diagnosis, representing a significant advancement in automated health assessments. By incorporating fuzzy logic rules, the model achieves a commendable accuracy of 95.84% in interpreting blood test parameters, including WBC, RBC, and platelet values. By doing so, this method copes with the natural uncertainties and approximations of medical data to diagnose diseases more effectively. The interface is designed for people, making it easy for them to enter their blood test numbers without any hassle. It makes use of fuzzy logic which does not only guarantee accuracy when diagnosing illnesses but also provides clear and quick health findings. This model represents an important step towards connecting clinical records with useful healthcare details; it demonstrates great potential to improve diagnosis in medicine by involving patients more closely and using advanced technology as well. However, the model is constrained only to a few parameters, which might need to be more effective in some real-time conditions. Future work, including the blood report analysis project, can be expanded to incorporate voice-based input and output, enhancing user convenience and accessibility. Introducing multi-linguistic support will make the system more inclusive and globally applicable. Additionally, exploring integration with natural language processing (NLP) to understand diverse medical reports better can be a valuable avenue.

REFERENCES

1. Karthikeyan, S., de Herrera, A. G. S., Doctor, F., & Mirza, A. (2021). An ocr post-correction approach using deep learning for processing medical reports. IEEE Transactions on Circuits and Systems for Video Technology, 32(5), 2574-2581.
2. Salahuddin, Z., Woodruff, H. C., Chatterjee, A., & Lambin, P. (2022). Transparency of deep neural networks for medical image analysis: A review of interpretability methods. Computers in biology and medicine, 140, 105111.
3. Chuang, Y. H., Su, J. H., Han, D. H., Liao, Y. W., Lee, Y. C., Cheng, Y. F., ... & Wang, C. C. (2022). Effective Natural Language Processing and Interpretable Machine Learning for Structuring CT Liver-Tumor Reports. IEEE Access, 10, 116273-116286.
4. Taira, R. K., Bashyam, V., & Kangarloo, H. (2007). A field theoretical approach to medical natural language processing. IEEE Transactions on Information Technology in Biomedicine, 11(4), 364-375.
5. Pandey, B., Pandey, D. K., Mishra, B. P., & Rhmann, W. (2022). A comprehensive survey of deep learning in the field of medical imaging and medical natural language processing: Challenges and research directions. Journal of King Saud University-Computer and Information Sciences, 34(8), 5083-5099.
6. Das, M., Sambodhi, P. P., Khare, A., & Naik, S. A. (2022, November). Challenges of Medical Text and Image Processing. In 2022 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC) (pp. 1-6). IEEE.
7. Hou, G., Chen, Y., Liu, X., Zhang, D., Geng, Z., & Si, S. (2024). A Resilience approach for diagnosing and predicting HBV-related diseases based on blood tests. Engineering, 32, 174-185.
8. Manikandan, K. (2019). Diagnosis of diabetes diseases using optimized fuzzy rule set by grey wolf optimization. Pattern Recognition Letters, 125, 432-438.
9. Muhammad, L. J., & Algehyne, E. A. (2021). Fuzzy based expert system for diagnosis of coronary artery disease in Nigeria. Health and technology, 11(2), 319-329.
10. Ghosh, G., Roy, S., & Merdji, A. (2020). A proposed health monitoring system using fuzzy inference system. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 095441192090801.
11. Chand, S., & Vishwakarma, V. P. (2019, September). Leukemia diagnosis using computational intelligence. In 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT) (Vol. 1, pp. 1-7). IEEE.
12. Khalil, A. M., Zahran, A. M., & Basheer, R. (2023). A novel diagnosis system for detection of kidney disease by a fuzzy soft decision-making problem. Mathematics and Computers in Simulation, 203, 271-305.
13. Lukmanto, R. B., & Irwansyah, E. J. P. C. S. (2015). The early detection of diabetes mellitus (DM) using fuzzy hierarchical model. Procedia Computer Science, 59, 312-319.
14. Yan, F., Huang, H., Pedrycz, W., & Hirota, K. (2024). A disease diagnosis system for smart healthcare based on fuzzy clustering and battle royale optimization. Applied Soft Computing, 151, 111123.
15. Bahani, K., Moujabbir, M., & Ramdania, M. (2021). An accurate fuzzy rule-based classification systems for heart disease diagnosis. Scientific African, 14, e01019.
16. Koirala, A., Jha, M., Bodapati, S., Mishra, A., Chetty, G., Sahu, P. K., ... & Hukkoo, A. (2022). Deep learning for real-time malaria parasite

detection and counting using yolo-mp. IEEE Access, 10, 102157-102172.
17. Fernandez-Basso, C., Gutiérrez-Batista, K., Morcillo-Jiménez, R., Vila, M. A., & Martin-Bautista, M. J. (2022). A fuzzy-based medical system for pattern mining in a distributed environment: Application to diagnostic and co-morbidity. Applied Soft Computing, 122, 108870.
18. Khan, S., Ullah, R., Shahzad, S., Anbreen, N., Bilal, M., & Khan, A. (2018). Analysis of tuberculosis disease through Raman spectroscopy and machine learning. Photodiagnosis and photodynamic therapy, 24, 286-291.
19. Kumar, Y., Koul, A., Singla, R., & Ijaz, M. F. (2023). Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. Journal of ambient intelligence and humanized computing, 14(7), 8459-8486.
20. Maini, E., Venkateswarlu, B., Maini, B., & Marwaha, D. (2021). Machine learning–based heart disease prediction system for Indian population: An exploratory study done in South India. medical journal armed forces india, 77(3), 302-311.
21. Subhadra, K., & Vikas, B. (2019). Neural network based intelligent system for predicting heart disease. International Journal of Innovative Technology and Exploring Engineering, 8(5), 484-487.
22. Khairuddin, S. H., Hasan, M. H., Hashmani, M. A., & Azam, M. H. (2021). Generating clustering-based interval fuzzy type-2 triangular and trapezoidal membership functions: A structured literature review. Symmetry, 13(2), 239.
23. Bhadani, V., Singh, A., Kumar, V., & Gaurav, K. (2024). Nature-inspired optimal tuning of input membership functions of fuzzy inference system for groundwater level prediction. Environmental Modelling & Software, 105995.
24. Jayaram, M. A., & Chandana, M. (2024). Design of flexible pavements through fuzzy inference system with genetic algorithm optimized rule base. International Journal of Transportation Science and Technology, 13, 284-301.
25. Hosseini, S., Gordan, B., & Kalkan, E. (2024). Development of Z number-based fuzzy inference system to predict bearing capacity of circular foundations. Artificial Intelligence Review, 57(6), 146.
26. MedlinePlus - Health Information from the National Library of Medicine. (1998, October 22). MedlinePlus. Retrieved August 3, 2024, from https://medlineplus.gov/.
27. Latest Medical News, Clinical Trials, Guidelines - Today on Medscape. (2024, May 31). Retrieved June 3, 2024, from https://www.medscape.-com/.
28. Know more. Feel better. (2020, January 29). Verywell Health. Retrieved August 3, 2024, from http://www.verywell.com/.
29. Khatter, H., Gupta, A. K., Garg, R. R., & Sain, M. (2022). Analysis of the S-ANFIS algorithm for the detection of blood infections using hybrid computing. Electronics, 11(22), 3733.
30. Cao, J., Zhou, T., Zhi, S., Lam, S., Ren, G., Zhang, Y., ... & Cai, J. (2024). Fuzzy inference system with interpretable fuzzy rules: Advancing explainable artificial intelligence for disease diagnosis—A comprehensive review. Information Sciences, 662, 120212.
31. Faisal, T., Taib, M. N., & Ibrahim, F. (2012). Adaptive neuro-fuzzy inference system for diagnosis risk in dengue patients. Expert Systems with Applications, 39(4), 4483-4495.
32. Polat, K., & Guneş, S. (2007). An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. Digital Signal Processing, 17(4), 702-710.
33. Deif, M., Hammam, R., & Solyman, A., (2021). Adaptive Neuro-Fuzzy Inference System (ANFIS) for Rapid Diagnosis of COVID-19 Cases Based on Routine Blood Tests. International Journal of Intelligent Engineering and Systems, vol.14, no.2, 178-189.
34. Walia, N., Singh, H., & Sharma, A. (2015). ANFIS: Adaptive neuro-fuzzy inference system—A survey. International Journal of Computer Applications, 123(13), 32.

AUTHORS:

Sai Varshitha Gondela is a final-year B. Tech student specializing in Computer Science and Engineering at VRSEC, Vijayawada, recognized for her outstanding academic performance. She has worked on diverse projects in Blockchain, AI/ML, and cybersecurity, and her research has been published in IEEE Conferences. She was selected for the prestigious Women Engineering (WE) program by NSE Talent Sprint and Google, showcasing her talents in a highly competitive environment. As a Grace Hopper Celebration (GHC) Scholar, she has participated in the largest global gathering of women in tech. She holds certifications in Blockchain, Ethical Hacking, Security, AI, and Data Science. Her passion for learning is reflected in her active participation in hackathons, coding competitions, and technical expos, both nationally and internationally. Her research focuses on Blockchain, AI-based security, Data Science, and AI/ML applications.

E-mail: saivarshithagondela4126@gmail.com

A. Sai Charan is currently a Lead Data Analyst at Sutherland Global Services, with expertise in advanced statistical modeling, machine learning, and natural language processing. A graduate from NIT Bhopal, Sai Charan has a strong academic foundation and has contributed to impactful projects, including predictive analytics, operational efficiency optimization, and customer satisfaction modeling. With a passion for uncovering insights from data, Sai Charan aspires to bridge the gap between analytics and strategic decision-making, aiming to make meaningful contributions in data science.

D. Rajeswara Rao is an academician who has taught for over 25 years. He has completed two PhDs in Machine Learning and Soft Computing and a Master's in Computer Science and Engineering. Currently, he is working as a Professor and Head of the Department of Computer Science and Engineering at Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India. Data science and machine learning optimization algorithms are some areas that interest him most when it comes to his research work. He has guided numerous Ph.D scholars and has published extensively in international conferences, journals, and patents. He has acted as program chair for many international conferences, among other duties like being NAAC Convenor, and he has shown commitment towards quality education service delivery leadership qualities through various administrative positions held at different times.

E-mail: rajeshpitam@gmail.com

Akhila Gudala is a final-year B. Tech student majoring in Computer Science at VRSEC College, where she ranks among the top students. Enthusiastic about diverse tech stacks, she is currently part of an ISRO project in the domain of Remote Sensing and Edge Computing. She has honed her expertise through certifications in Cloud Computing, AI, and Linux, complemented by a strong understanding of data analysis gained through her coursework. Her keen interest in problem-solving and critical thinking drives her research in AI, Remote Sensing, and Edge Computing. Her commitment to continuous learning is evident through her participation in competitive programs and real-world projects, further solidifying her position as a top student in her field.

E-mail: gudalaakhila123@gmail.com

Ashutosh Satapathy is currently working in the Department of Computer Science and Engineering as an assistant professor at VR Siddhartha Engineering College. He completed his Ph.D. at Vellore Institute of Technology, India, and his Master of Technology (M. Tech) in Information Security and Computer Forensics at SRM University, India. During his Ph.D. program, he completed a research internship at Information and Communications Research Laboratories, ITRI-Taiwan. Prior to this, he worked as a lecturer at Odisha University of Technology and Research (formerly College of Engineering and Technology, Bhubaneswar). He has published 19 research articles, including international journal articles, book chapters, and conference proceedings, all indexed in SCIE (WoS) or Scopus.

E-mail: ashutosh.satapathy1990@gmail.com