iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 Issue: V Month of publication: May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52709

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Rule Based Analysis of Disease Detection

Dr. Shivani Agarwal[1], Kanhaiya Lal[2], Kuldeep Singh Bartwal[3], Mayank Verma[4], Pranjal Srivastava[5]
[1]*Assistant Professor, Department of Information Technology, AKGEC, Ghaziabad, U.P., India,*
[2, 3, 4, 5]*UG Final Year Students, Department of Information Technology, AKGEC, Ghaziabad, U.P., India*

Abstract: We employ fuzzy sets and fuzzy logic for illness diagnosis. By utilizing a fuzzy logic framework, uncertainty in data and decision-making processes may be made up for. Diagnostic models that can manage enormous volumes of complex and raw medical data may be made using fuzzy logic. Fuzzy logic has a number of advantages when it comes to disease detection, having the ability to deal with incomplete and inaccurate data, the ability to incorporate expert knowledge and feedback, the potential to improve diagnostic accuracy and decrease the number of false positives and false negatives, as well as the capacity for handling incomplete and inaccurate data. The method known as "multiple disease detection using fuzzy rules" combines fuzzy logic with rule-based systems to identify various diseases based on symptoms given by the user. The system accepts user-provided symptoms, converts them into fuzzy sets using fuzzy logic. The fuzzy rules are then evaluated using these fuzzy sets, and the degree of integration of each illness is then determined. The level of integration shows the extremely exact likelihood that a user accurately predicts when they will experience a specific circumstance.
Keywords: fuzzy classification rule learning by clustering, artificial intelligence, fuzzy rule-based classification systems, fuzzy classification rule learning, heart disease prediction

I. INTRODUCTION

One of the main causes of death globally and a serious public health concern, heart disease is expanding quickly [1]. Early detection and diagnosis of heart disease are crucial for boosting the likelihood of successful treatment and lowering the risk of additional health problems. The accuracy and speed of heart disease diagnosis were recently demonstrated to be much improved by recent developments in machine learning approaches. One machine learning method that has attracted interest recently is fuzzy rule-based categorization, which can deal with ambiguous and imprecise data. In this research proposal, we propose to investigate the use of fuzzy rule-based categorization to diagnose heart disease and create a model that can accurately and quickly detect the condition.
Artificial intelligence (AI) technologies known as fuzzy rule-based systems (FRBS) can be used to quickly and accurately diagnose illnesses in their early stages [2,5]. Fuzzy logic, a mathematical method that enables non-specific or ambiguous data to be analysed in a very effective manner, is used by FRBS. Fuzzy if-then rules known as FRBS connect input variables like the results of a medical test or a patient's symptoms to an output variable like the presence or absence of a disease. These guidelines were created using subject-matter expertise and extensive experience. Additionally, the system may improve its performance over time by learning through a variety of ways, including data-driven and data mining techniques.
FRBS can assess patient information or symptoms, such as blood pressure, cholesterol levels, and other tests, in order to detect heart illness. A diagnostic or risk evaluation for the patient may be generated by the system, enabling healthcare professionals to immediately inform treatment decisions and improve patient outcomes early on, thus lessening the patient's suffering. Combining FRBS's perspective technique of sickness detection and diagnosis with standard medical procedures is one option. More research and development into these systems is very likely to enhance medical outcomes and lower the toll that heart disease has on both individuals and society [8]. The objective is to create a classification model with fuzzy rules for detecting heart disease that can accurately distinguish between healthy and unhealthy individuals utilizing various input characteristics. to assess the model's performance on real-world heart illness datasets and show how well it can accurately and successfully identify heart disease. To assess the generated model's superiority in terms of accuracy, sensitivity, specificity, and efficiency [11,9] in comparison to existing heart disease detection models that employ different algorithms. In order to better understand how the model makes decisions based on input data, it is important to look into the model's interpretability for heart disease detection.

A. An Uncertainty-Based Classification System

Fuzzy logic is used to do classification problems through a type of machine learning model known as a fuzzy rule-based system [6,14]. It is a development of the conventional, logically sound rule-based categorization scheme. In contrast to conventional Boolean logic, fuzzy logic is used to define the classification rules in a FRBCS.

The categorization choice in Instead of using a simple true/false dichotomy, fuzzy logic is built on degrees of membership conclusion, allowing for more supple and subtle rules. This makes the model useful for situations where the input data may be noisy or incomplete since it can handle uncertainty and imprecision in the data.
Based on the input data, FRBCS creates a series of fuzzy rules that are then combined to provide a categorization conclusion. The structure used to specify the rules is often an if-then one, where the "if" section describes the circumstances in which the rule applies and the "then" part describes the output categorization.

The following stages are commonly involved in creating a FRBCS:
1) Defining the fuzzy sets of the input factors.
2) Defining the fuzzy sets of the output factors.
3) Developing a set of fuzzy rules that are depending on the input and output factors.
4) Putting rules together and coming up with a categorization result via fuzzy inference.

FRBCS is useful for a variety of classification applications, including data mining, image recognition, and natural language processing. To create efficient categorization rules and get precise results, you must have a solid grasp of fuzzy logic and its applications.

B. Fuzzy Classification Rules Learning Through Clustering
By first grouping the data into groups and then deriving the rules from the cluster centroids, a method known as fuzzy classification rules learning by clustering may be used to create fuzzy classification rules from data [8,13]. When there is doubt or ambiguity in the data as well as when it is complicated or high-dimensional, this technique might be helpful.

The following stages are commonly taken while learning fuzzy classification rules by clustering:
1) Employing the fuzzy clustering method, to divide the input data into groups.
2) Calculating each cluster's centroid.
3) Fuzzy rules are derived from the cluster centroids by labeling the centroid coordinates with language words like "low", "medium", or "high".
4) By describing the fuzzy sets for each input variable, which are based on the linguistic words given to the centroid coordinates, the rule conditions are defined.
5) By mentioning the output variables and their fuzzy sets, the rule conclusions are defined.

After been learnt, the fuzzy rules may be used to categorize problems. To establish the most likely categorization, the fuzzy rules are applied to fresh input data during the classification process. In order to achieve this, it is determined to what extent each rule's antecedent (condition) includes the input data.. These degrees of membership are then combined using fuzzy logic to get the overall degree of membership in each class. The categorization choice is then made based on which output class has the highest level of membership.

II. METHODOLOGY
This project's primary goal is to use a system of rule-based classifiers to forecast the chance of getting future illnesses. Rule-based classifiers are a particular kind of classifier that bases its decisions on different "if-else" rules [5,9,13]. Since the principles are simple to understand, this method is frequently used to create descriptive models. Because they rely on a set of established rules that are based on medical knowledge and research, rule-based classifiers are efficient at analysing illnesses. These guidelines aid in recognising the traits and signs of many diseases and can be applied to forecast a patient's propensity to get a certain illness. The benefit of employing rule-based classifiers is that they offer a clear and comprehensible model that can be used to demonstrate how the predictions were made [14]. This method is very helpful in the medical industry since it enables medical professionals and patients to comprehend the elements that increase the risk of contracting a disease.
Rule-based classifiers are useful for building descriptive models as well since they provide precise instructions for identifying and diagnosing disorders. This strategy streamlines the diagnosis process so that doctors may diagnose patients more quickly and correctly. The goal of this project is to create a system of classifiers that use rules that can predict the chance of getting future diseases. This approach provides a straightforward and accessible model that is based on knowledge and research from the medical community, making it a crucial tool for diagnosing and treating illnesses.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

The following technique is suggested for this study:

1) *Data gathering:* Kaggle and the UCI Machine Learning Repository are only two of the locations from which we will collect real-world datasets related to heart disease [11]. These datasets will contain a range of input variables, including blood pressure, cholesterol levels, age, gender, and electrocardiogram (ECG) signals which make it straightforward to detect heart disease.
2) *Data Pre-processing:* To deal with outliers, the data is normalized, missing values are removed, the datasets that have been acquired will first go through pre-processing. Pre-processing is one of the most important stages in machine learning since it ensures that the data is clean and ready for modeling to avoid any further problems.
3) *Feature Selection:* Using a variety of methods [8], relevant features will be chosen from the pre-processed data. The process of feature selection aids in choosing the most crucial characteristics, it decreases the input data's dimensionality and helps in the identification of heart disease.
4) *Development of a Fuzzy Rule-Based Classification Model:* Using the chosen characteristics and fuzzy logic system, we will create a fuzzy rule-based classification model to identify heart condition. Applications for medical diagnostics can benefit from fuzzy logic, a mathematical framework that can manage ambiguous data and data that is subject to question.
5) *Model Evaluation:* Using several performance measures, the constructed model will be assessed using real-world heart disease datasets 8 that were taken into account from a variety of sources. These measures assist in identifying the model's efficacy in identifying cardiac disease and reveal its strengths and flaws.
6) *Model Comparison:* In this stage, we'll compare the performance of the constructed model to other already-in-use models for detecting heart disease that use methods like logistic regression, decision trees, and support vector machines. Through this comparison, it will be determined whether the developed model is preferable in terms of accuracy, sensitivity, specificity, and efficiency.
7) *Model Interpretation:* By examining the fuzzy rules and membership functions, we will look at the interpretability of the developed model. We are able to use experience rather than knowledge to solve mathematical issues by using membership functions. This study will aid in raising awareness of the model's decision-making process based on input data as well as how the model may be further developed and processed in accordance with requirements.

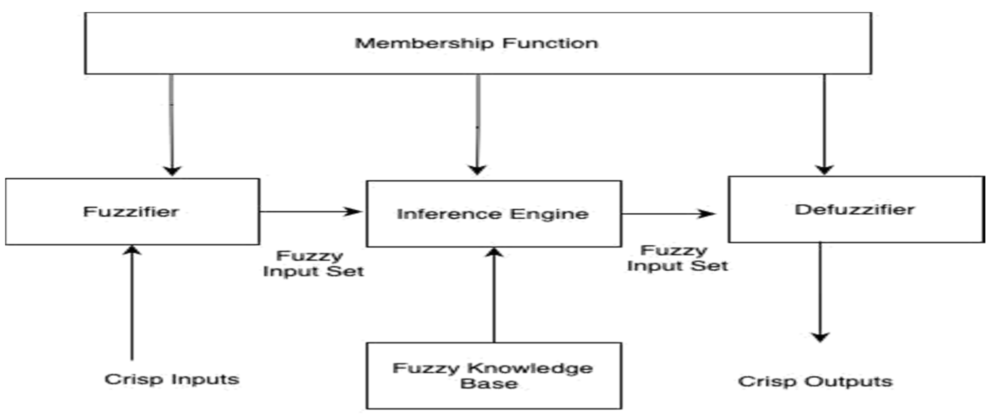

Fig 1: System Architecture of Fuzzy Logic

III. RESULT AND DISCUSSION

The architecture overview, its operation, and the specifics of how various functions are implemented are the main topics of this talk. In this article, we'll talk about the system's architecture, operation, and methods for creating and integrating its many parts. We will also give thorough descriptions of each function we are developing.

1) *Input layer:* The input layer is where the fuzzy system's inputs are sent. These inputs, which show how a system or process is currently functioning, can be expressed as linguistic or numerical variables. Before these inputs may be used in the fuzzy system, they may occasionally need to be pre-processed or normalized.
2) *Layer of fuzzification:* This layer assigns membership degrees to each input depending on its linguistic values, transforming the crisp inputs into fuzzy sets. The extent to which a given input is a member of a certain fuzzy collection is determined by the membership function. Depending on the characteristics of the input variables, multiple membership functions, including triangular, trapezoidal, and Gaussian, can be chosen.

3) *Rule Base:* The rules that specify how the inputs and their fuzzy sets should be merged to produce an output are included in this layer. The conditions under which certain actions should be conducted are often specified in the rules as IF-THEN statements. For instance, "TURN ON THE AIR CONDITIONING IF the temperature is high AND the humidity is low." Experts in the relevant fields can manually develop the rule foundation, or machine learning techniques can automatically generate it from data.
4) *Inference Engine:* To produce a fuzzy output, the inference engine layer applies the rules to the input variables and their fuzzy sets. The rules can be combined to produce the result using a variety of inference techniques. The degree to which each rule is met is calculated by the inference engine, which then combines the findings to produce a single fuzzy output.
5) *Defuzzification Layer:* This layer transforms the output from fuzzy into a clear value that may be utilized for control or decision-making. The centroid approach, which determines the fuzzy output set's gravitational centre, is the defuzzification technique that is most frequently employed. Depending on the application, other techniques can also be employed, including the max-min and weighted average approaches.
6) *Output Layer*: The output layer, which can be utilized for further processing or shown to the user, reflects the fuzzy system's ultimate product. The output, which indicates the course of action depending on the inputs, can be either a numerical number or a language variable.

A. Membership Function

Based on the current issue and the selected fuzzy set, fuzzy membership functions are produced. These functions gauge how much an entity and the set resemble one another while also representing the fuzzy set. Membership functions can be designed in a variety of forms, including bell-shaped, triangular, trapezoidal, linear, and Gaussian. The Gaussian membership function was chosen for this study because it fits the situation and is simple to comprehend. A fuzzy set B's membership function on the realm of discourse Y is written as "B: Y" [0, 1], which gives each element in Y a value between 0 and 1. How closely a given element in Y resembles a member of the fuzzy set B is represented by this value, which is also known as the membership value. The x-axis represents the realm of conversation, and the degrees of membership within the [0, 1] range are displayed on the y-axis., in a visual representation of membership functions.

$$\mu B\ (K, R, W, F) = \exp\ [\ -½|\ (K-R)/W\ |^{F}] \qquad \text{[exp. 1]}$$

R: Centre
W: Width
F: Fuzzification factor (e.g., F=2)
R = 5, W = 2, F =2

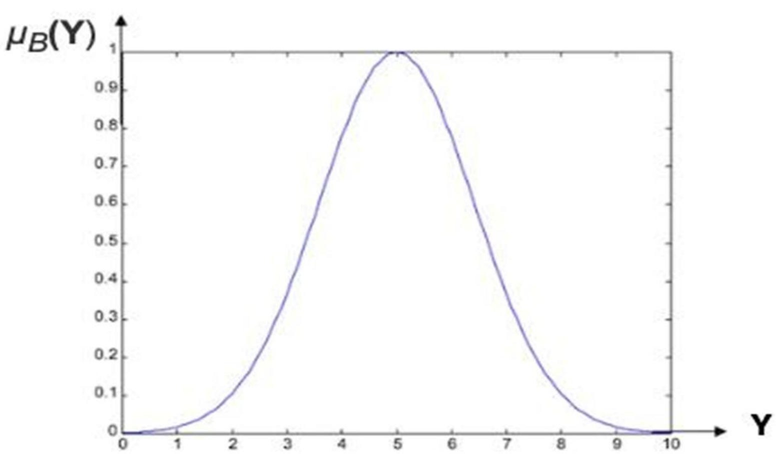

Fig 2: Function for Gaussian Membership

B. System of Fuzzy Inference

A system of fuzzy inference, based on specified fuzzy rules within a knowledge domain, is a helpful tool for mapping inputs to their corresponding outputs. Fuzzification, a technique that transforms inputs into fuzzy values, is necessary before this system can be used effectively. As a result, it is possible to apply the fuzzy rules that link the input and output fuzzy sets. The knowledge domain and the stated if-then rules must be carefully evaluated in order to construct the fuzzy rules. After applying a membership function to the input values, to determine the rule strength, the fuzzy rules and fuzzy input are combined. The output distribution is then produced by combining rule strength and the output membership function. Finally, the defuzzification procedure is used to transform the output from its fuzzy values into a crisp value in order to provide the output in a more intelligible style. A fuzzy inference system can offer beneficial insights and answers to challenging issues in a range of disciplines by following these steps.

C. Defuzzification

A crucial step in symbolic logic is the process of defuzzification, which entails producing a measurable outcome from fuzzy sets and the related membership degrees. In fuzzy control systems, when decisions must be made using clear values, it is essential. The Centre of Area (COA), commonly referred to as the centroid approach, is the defuzzification technique that is most frequently used. With this technique, the crisp value corresponding to the fuzzy set's area's centre is determined. The bisector, centroid, Middle Of Maximum (MOM), Largest Of Maximum (LOM) and Smallest Of Maximum (SOM) are the five defuzzification techniques that are most often utilized. The centroid approach was selected for the present project since it is shown in Figure 4 and offers the location of the area beneath the curve's center. This qualifies it as a technique that can be used to extract a precise value from the fuzzy set.

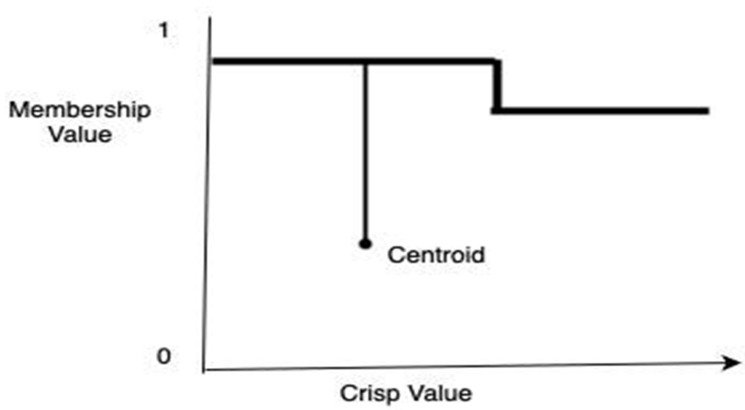

Fig 3: Defuzzification Method

IV. CONCLUSION

Heart disease is a major global public health concern, and early diagnosis is essential for successful treatment and avoiding unfavourable consequences. Fuzzy rule-based classification, which can handle ambiguous and imprecise data and is the right tool for this task, has demonstrated tremendous potential for heart condition identification. The diagnosis of cardiac disease is proposed using a classifying fuzzy rules model in this research, and evaluates performance of the model using real-world datasets. The model's performance will be evaluated against that of other models in order to determine its advantages and disadvantages.

In order for healthcare professionals to understand and accept the model's conclusions, the study also aims to provide light on the model's interpretability. The predicted outcomes of this study will further our knowledge of how to detect heart illness in people and provide healthcare professionals with a practical strategy for doing so. This discovery might have a huge influence on the healthcare sector, ultimately improving patient outcomes and quality of life.

REFERENCES

[1] The World Health Organisation. CVDs (cardiovascular illnesses). https://www.who.int/news-room/fact- sheets/detail/cardiovascular-diseases-(cvds)
[2] Mendis, S.; Wong, W. H. (2019). A practical introduction to machine learning in medicine. 19(1), 64; BMC Medical Research Methodology.
[3] (2018) Yaseen, F., Khan, M. A., and Amin, M. B. Systems for classifying data using fuzzy rules: a thorough analysis of the models, algorithms, and methods. 29(2), 371-394, Neural Computing and Applications.
[4] In 2020, Alomari, R., AlRyalat, S. A., and Jararweh, Y. A mixed fuzzy-neural network approach is used to categorize cardiac disease. 88, 106016, Applied Soft Computing.

[5] In 2019, Huang, Y., Li, Q., Li, X., Li, D., and Liang, Y. Based on enhanced fuzzy c-means clustering and SVM, the diagnosis of heart disease. Medical Imaging and Health Informatics Journal, Volume 9(9), 1917–1922.

[6] A. Haidar, A. Khawaja, and S. Lutfi. Using big data analytics and fuzzy logic, a prediction method for heart disease. 43(10), 312 in Journal of Medical Systems.

[7] In 2019, Khan, S. S., and Hayat, M. a method based on fuzzy rules for predicting heart disease. 43(5), 106; Journal of Medical Systems.

[8] Goyal, M., Das, S. K., and (2018). Performance analysis of several machine learning methods for predicting heart disease. 42(7), 129, Journal of Medical Systems. UCI Database for Machine Learning. (n.d.). Data Set on Heart Disease. the heart disease dataset was retrieved from https://archive.ics.uci.edu/ml/datasets/

[9] Kaggle. (n.d.). UCI Heart Disease. The information was collected from https://www.kaggle.com/ronitf/heart-disease-uci

[10] N.S. Altman, an introduction to kernel and nearest-neighbor nonparametric regression, Am Stat 46 (3) (1992) 175–185.

[11] I. Kononenko, "Inductive and Bayesian learning in medical diagnosis," Appl. Artif. Intell. Int. J., vol. 7, no. 4, 1993, pp. 317–337.

[12] The Nature of Statistical Learning Theory by V. Vapnik. the year 2013; Springer Science & Business Media.

[13] A. Jantan, A.E. Omolara, O.I. Abiodun, K.V. Dada, N.A. Mohamed, H. Arshad, Current uses of artificial neural networks: a survey, Heliyon 4 (11) (2018) e00938.

[14] Heart Disease Data Set by D. Robert, 1988. Scientific African 14 (2021) e01019 by K. Bahani, M. Moujabber, and M. Ramdani