Minimax Theory and its Applications

Volume 11 (2026), No. 2s, 596 to 604

Fuzzy Rule-Based Systems For Uncertainty Handling In Big Data

J.Biju[1], Gowtham Gopal[2], Dr. Hrushikesh Jaiwant Joshi[3], Drishya Darwin [4], Dr. C J Raman [5], Sridevi. T[6], Manna Sheela Rani Chetty[7]

[1]Assistant Professor , Division of Data Science and Cyber Security Karunya Institute of Technology and Sciences ,Coimbatore.Tamilnadu, India. Email: jbijuinfo@gmail.com, ORCID ID:0009-0009-1628-7871
[2]Assistant Professor , Kalaignar Karunanidhi Institute of Technology, Coimbatore, Email: gowthamgopal16@gmail.com, Orcid Id:0009-0009-9326-8965
[3]Assistant Professor, Savitribai Phule Pune University (SPPU), Pune, Email: hrushikesh.joshi@vit.edu,Orcid id . 0000000181322746
[4]Assistant Professor of Information Technology , Dr.M.G.R.Educational and Research Institute, Chennai, Tamilnadu. Email: drishya86@gmail.com
[5]Professor, Department of Artificial Intelligence and Machine Learning, St. Joseph's College of Engineering, Old Mamallapuram Road, Chennai - 600119, Tamilnadu, India, Email: cjramancj@gmail.com, Orchid id: 0000-0001-9376-6424
[6]SRM Institute of Science and Technology Ramapuram Campus, Chennai, India, Email: sridevit1@srmist.edu.in, Orchid ID: 0000-0002-7512-429X
[7]Professor of Computer Science Koneru Lakshmaiah education Foundation, Green Fields, Vaddeswaram, Guntur District- 522502, Email: sheelarani_cse@kluniversity.in, ORCID ID:0000-0002-1378-6917

Abstract
Uncertainty is a major difficulty for big data analytics, which can result in poor decision-making and erroneous forecasts. Inadequate or imprecise knowledge is the cause of uncertainty in medical data. When it comes to handling cognitive uncertainties like ambiguity and vagueness, data mining techniques are ineffective. The inability to draw clear or exact distinctions in the world is linked to vagueness. When there are two or more options, ambiguity arises because it is unclear which option to choose. A further victim of the curse of dimensionality is medical data. Repetitive or unnecessary attributes may be included in high dimensionality (number of attributes), which could cause noise in the data mining process. Systems based on fuzzy rules are a mathematical technique for dealing with data that is ambiguous. A thorough framework is created to combine big data analytics and fuzzy logic, allowing for the efficient handling of ambiguity and uncertainty. Real-world big data sets are used to assess the effectiveness of fuzzy rule-based systems, showing that they can yield more accurate and trustworthy insights. By helping to create strong and durable big data analytics systems, the research helps businesses make more informed choices when faced with uncertainty.

Keywords: uncertainty, Bigdata, decision-making, fuzzy rule-based systems.

1. Introduction

Rapid technological advancements and the exponential growth of data and information from many sources are producing massive data heaps that are impossible for traditional tools and systems to efficiently store, handle, and analyze. Big data is the term used to describe such enormous data and information packages [1]. "Roger Magoulas" was the first to use the term "Big Data." The fact that big data is a "relative" term—that is, its definition varies from organization to organization and from time to time—is an essential consideration. Big data's explosive expansion has changed how businesses function, make decisions, and engage with their clientele. But managing uncertainty is also made more difficult by the growing amount, speed, and diversity of big data. Big data uncertainty can come from a number of sources, such as missing or incomplete data, noisy or inaccurate data, and ambiguous or imprecise data [2]. Global technology advancements have resulted in the rapid accumulation of vast amounts of data that are beyond the capacity of traditional tools, processes, systems, and sub-systems to handle. Big data has undoubtedly brought about a paradigm shift and changed how we work, live, and perceive the world. An ocean of information is concealed within these enormous data mounds. We can gain cutting-edge insights about the data and the system overall if we can uncover those hidden patterns and information. These insights can then be utilized to create new tools, techniques, technologies, methodologies, forecasting, and predictions, among other things, for the benefit and well-being of people and organizations. Conventional big data analytics techniques frequently depend on exact and certain data and have trouble

handling uncertainty, which can result in poor decision-making, erroneous forecasts, and a decline in confidence in the outcomes. A viable method for handling uncertainty in large data is fuzzy logic, which can deal with ambiguous and uncertain data. Particularly, fuzzy rule-based systems have demonstrated significant promise in managing ambiguity and uncertainty in large systems. [3].

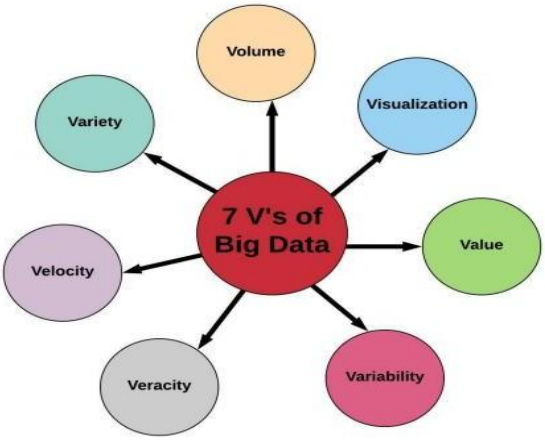


Figure 1: V's of Big Data

By using fuzzy logic and fuzzy rule-based systems, organizations can develop more robust and resilient big data analytics systems that can effectively handle uncertainty and provide more accurate and reliable insights. This research aims to investigate the application of fuzzy rule-based systems in handling uncertainty in big data analytics. The study will explore the current state-of-the-art in fuzzy logic and big data analytics, and develop a comprehensive framework for integrating fuzzy rule-based systems with big data analytics. The research will also evaluate the performance of fuzzy rule-based systems in handling uncertainty in big data, and provide recommendations for organizations seeking to develop more robust and resilient big data analytics systems.

2. Literature review

Davoudian, Ali & Chen (2018) [4] Ongoing requests for putting away and questioning big data have uncovered different weaknesses of conventional social database frameworks. This, thusly, has prompted the development of another sort of corresponding nonrelational data store, named as NoSQL. Ghorbanian, Maedeh & Hacopian Dolatabadi (2019) [5] The main force structures are focused on creative data and correspondence that lead to a data tempest starting from various sources. Large data processing on shrewd force frames is unavoidable in order to overcome these difficulties with collected voluminous data. Gan, Wensheng& Lin (2020) [6] Modern IoT applications generate huge measurements of data, some of which are readings, occasions and log sections. In particular, the vast majority (i.e. recurrence and uncertainty) of these items contain rich inserted details (such as the unit hazard and utility of items, the interesting nature of products, cost or weight) and variable significance. Andrienko, Gennady & Andrienko, et al. (2020) [7] In terms of data representation and analysis, this report illustrates part of the problems and applications that occur in the large data era. In particular, 14 renowned researchers in the academic and industry worlds and various networks, such as visualisation of information, computer-human interaction, machine learning, board data and mining, were welcome to express their opinions. Martin Macak (2020) [8] Digitalization of our general public brings different new computerized environments.

Traditional big data analytics methods often struggle to effectively manage uncertainty, leading to inaccurate predictions, poor decision-making, and decreased confidence in results [9]. Therefore, there is a need for innovative approaches that can effectively handle uncertainty in big data, and fuzzy rule-based systems have shown promise in this regard. However, the application of fuzzy rule-based systems in big data analytics is still in its infancy, and there is a lack of comprehensive frameworks, methodologies, and tools for effectively integrating fuzzy logic with big data analytics.".

- To investigate the application of fuzzy rule-based systems in handling uncertainty in big data analytics.
- To develop a fuzzy logic-based framework for managing uncertainty in big data.
- To evaluate the performance of fuzzy rule-based systems in handling uncertainty in big data compared to traditional methods.

The paper is arranged as follows: The object detection and its back ground of the study is briefly described in section 1 and 2. The proposed research project's system architecture is shown in Section 3. The simulation result is depicted in Section 4. And finally, section 5 concludes the research projects.

3. Proposed Method

Big health data, as controlled by its immense data qualities (i.e.,volume, speed, variety, veracity, regard, change, multifaceted nature), needs an advantageous profitable requesting procedure for the effective recovery of data with these attributes Indexing is one of the worries that have been accepting a great deal of consideration from the information research local area. The layout of the file stores the value of a certain field or field set, as requested from the field value. The association of registration divisions arranges successful search operation for parallels. Indexing approaches are striking techniques for optimising query handling in database systems, in particular related databases [16]. Whatever the case, it is not clear if the traditional indexing methods and architectures can be changed to big data due to the fascinating characteristics of both data and the simple infrastructure of the big data handling. Or possibly, the information is obtained and stored in large groups of files at the stage. The data size often exceeds the connection database systems will normally handle the machine time to retrieve the data [10].

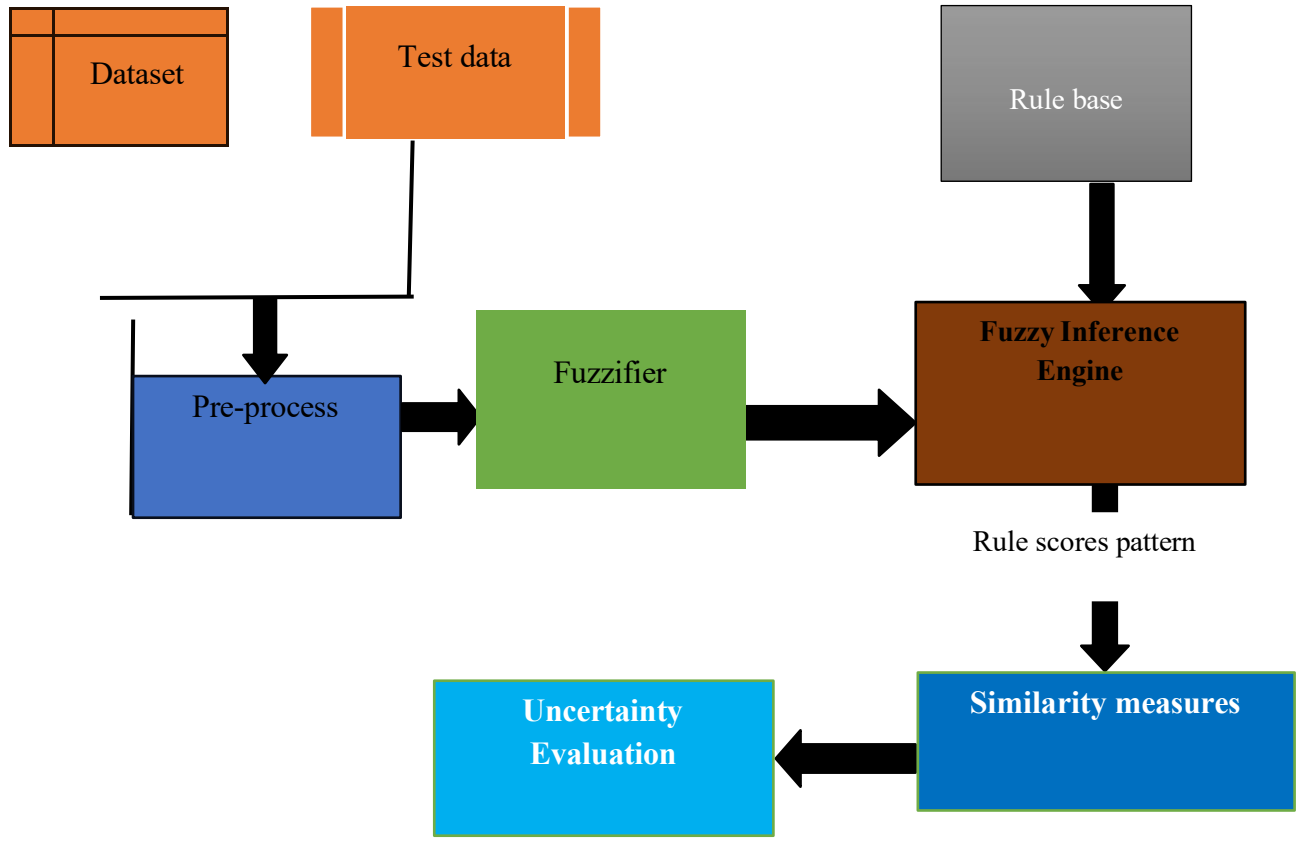


Figure 2: Proposed flow

The suggested methodology for evaluating uncertainty in big data using a fuzzy logic framework is shown in Fig. 1. This method compares and suggests pertinent findings for the data being considered by utilizing a database of previously examined data. Configuring the model, which centers on the use of fuzzy rule-based classifier systems (FRBCS) for uncertainty evaluation, is one of the processing processes [11]. This method makes it possible to assess uncertainty in big data, which promotes better data-driven insights and well-informed decision-making.

3.1. Data Collection and Preprocessing

Data Collection:

One of the most crucial phases of research is data collection. It is sense to take into account the fact that a medical dataset contains uncertainty in order to address the study of evaluating uncertainty in medical datasets and to find answers to the previously mentioned research topics. The degree of membership and indiscernibility can be used to identify datasets that exhibit uncertainty in the form of vagueness and ambiguity. The Center of Machine Learning and Intelligent Systems hosts the UCI Machine Learning Repository. [12].

Data Preprocessing:

The UCI Machine Learning Repository dataset is preprocessed by using train_test_split to separate the data into training and testing sets after addressing missing values with SimpleImputer's mean technique. The data is then scaled to have a unit variance and zero mean using StandardScaler. After preprocessing, the data is transformed into a Pandas DataFrame for additional modeling and analysis. The data is cleaned, standardized, and prepared for usage in machine learning models thanks to this preparation procedure.

3.2. Fuzzy Rule-Based System Development: Design Frame work

By completing the above steps, the relevant dataset ready for further analysis and processing using fuzzy rule-based systems. A dataset can be represented by a table known as information system where each row represents an instance and column represents an attribute. Formally, it is a pairLet X = {x1, x2, x3,…….xn}, and let X be the discourse universe.

The membership function is given by,

$$f_{\tilde{A}}(X) = \begin{cases} 0, x < a1 \\ \dfrac{(x-a1)}{a2-a1}, a1 \le x \le a2 \\ \dfrac{(a3-x)}{a3-a2}, a2 \le x \le a3 \\ 0, x > a3 \end{cases}$$

The operational laws are given as follows:

$$\tilde{A}(+)\tilde{B} = (a_1, a_2, a_3)(+)(b_1, b_2, b_3) = (a_1 + b_1, a_2 + b_2, a_3 + b_3)$$
$$\tilde{A}(-)\tilde{B} = (a_1, a_2, a_3)(-)(b_1, b_2, b_3) = (a_1 - b_1, a_2 - b_2, a_3 - b_3)$$
$$\tilde{A}(\times)\tilde{B} = (a_1, a_2, a_3)(\times)(b_1, b_2, b_3) = (a_1 b_1, a_2 b_2, a_3 b_3)$$
$$\tilde{A}(\div)\tilde{B} = (a_1, a_2, a_3)(\div)(b_1, b_2, b_3) = (a_1/b_1, a_2/b_2, a_3/b_3)$$

Where a_1, a_2, and a_3 are real numbers and $a_1 \le a_2 \le a_3$.

The decision-making process used by the group is crucial to coming up with a good answer. A valid judgement may be reached since the process of reaching a consensus is based on the response of numerous people. The Normalization is expressed as

$$xa^k_{1ij} = (a^k_{1ij} - \min a^k_{1ij})/\Delta^{\max}_{\min}$$
$$xa^k_{2ij} = (a^k_{2ij} - \min a^k_{2ij})\Delta^{\max}_{\min}$$
$$xa^k_{3ij} = (a^k_{3ij} - \min a^k_{3ij})/\Delta^{\max}_{\min}$$
$$\text{Where } \Delta^{\max}_{\min} = \max r^n_{ij} - \min l^n_{ij}$$

Calculate the normalized values for the right (rs) and left (ls):

$$xs^k_{ij} = xa^k_{2ij}/(1 + xa^k_{2ij} - xa^k_{1ij})$$
$$xrs^k_{ij} = xa^k_{3ij}/(1 + xa^n_{3ij} - xa^k_{2ij})$$

Total normalized crisp values should be calculated:

$$x^k_{ij} = [xls^k_{ij}(1 - xls^k_{ij}) + xrs^k_{ij} \times xrs^n_{ij}]/(1 - xls^k_{ij} + xrs^k_{ij})$$

Compute total normalized crisp values:

$$\omega_{ij}^{k} = \min a_{ij}^{n} + x_{ij}^{n} \Delta_{\min}^{\max}$$

Integrate crisp values from different opinions of k respondents:

$$\omega_{ij} = 1/k(\omega_{ij}^{1} + \omega_{ij}^{2} + \cdots + \omega_{ij})$$

Fuzzy logic takes into account incomplete information and the development of available knowledge (Kannan, De Sousa, et al. 2014). It permits ambiguous entry. It permits some regulations to incorporate difficulties with complicated issues. Fuzzy models are difficult to expand for hazards. Before they can be applied to actual circumstances, they frequently need to conduct numerous simulations.

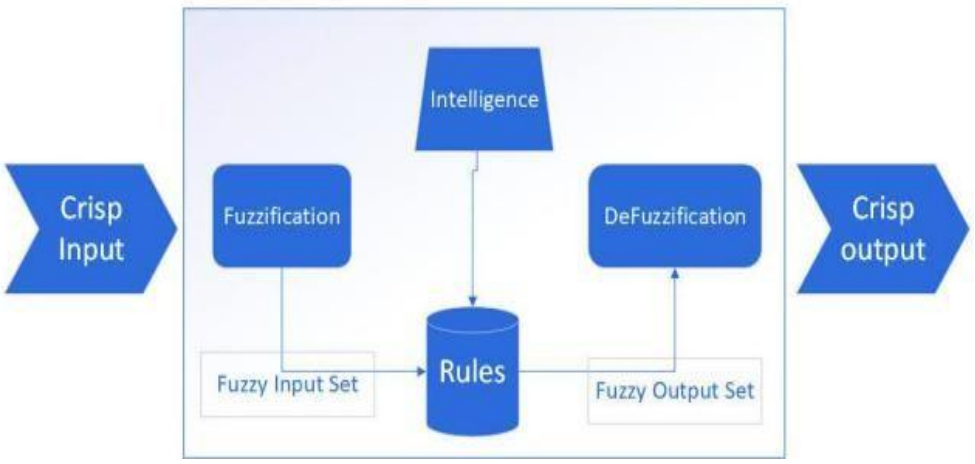


Figure 3: Architecture of the Fuzzy Rule based model

Another characteristic of decision systems is decision. Decision tables are used to generate concept approximations in supervised learning, a process that can be recognized with training samples that are well-known in machine learning. Because things are seen based on the values of available qualities, they are indistinguishable. As a result, items with identical attribute values cannot be distinguished:

Step 1. Determine the positive ideal solution $(x_j)^{+}$

$$(x_j)^{+} = \max_{i=1}^{m} x_{ij}$$

and the negative ideal solution $(x_j)^{-}$

$$(x_j)^{-} = \min_{i=1}^{m} x_{ij}$$

Step 2. Calculate the value S_i

$$s_i = \sum_{j=1}^{n} w_j \frac{(x_j)^{+} - x_{ij}}{(x_j)^{+} - (x_j)^{-}}$$

and the value R_i :

$$R_i = \max_{j} \frac{(x_j)^{+} - x_{ij}}{(x_j)^{+} - (x_j)^{-}}$$

Step 3. Determine the value Q_i :

$$Q_i = \eta \frac{S_i - \min_i S_i}{\max_i S_i - \min_i S_i} + (1-\eta) \frac{R_i - \min_i R_i}{\max_i R_i - \min_i R_i}$$

Ensemble learning is a machine learning framework to solve classification/prediction problems where multiple learners are trained to construct a set of hypotheses unlike other approaches which learn one hypothesis from training data [13-15]. Ensemble methods combine these hypotheses to obtain the classification/prediction for new instances. In general, an ensemble construction comprises of producing multiple homogenous base learners. This work has employed the proposed fuzzy-rough rule induction as a single base learning algorithm for the ensemble construction. Ensembles compensate such issues and achieve better performance by combining the opinions of multiple learners. Base learners can be influenced by small changes in the training samples during training phase. Subsequently, classifiers will overfit their training data and do not generalize well to other data. This is a major concern when using ensemble.

4. Results and discussion

This research work is focused on analysing uncertainty in medical datasets so that efficient diagnosis of disease can be made by the medical practitioners. Benchmark datasets are considered in order to assess the relative performance of a proposed model on par with the existing ones. The results obtained in each phase are showcased subsequently. Also, performances of the proposed methods are evaluated by comparing the results obtained with that of the state-of-the-art methods. The proposed framework is developed under Windows 7 platform on Intel (R) Core (TM) i5 (3.30 GHz) processor. The proposed techniques were implemented using MATLAB.

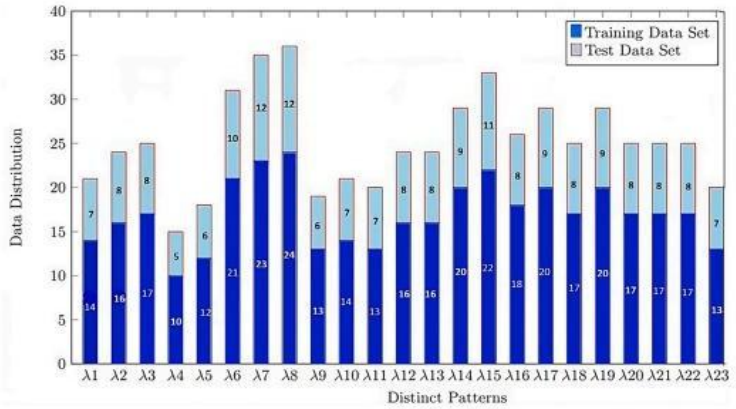


Figure 4: Dataset distribution.

Because any additional processing or transformation applied to the data after the addition of the aforementioned noise will likewise alter the matching noises. If we don't know what kind of noise it is, we will have trouble de-noising there. Figure 4 illustrates how the most valuable features are recovered using convolutional and max-pooling.

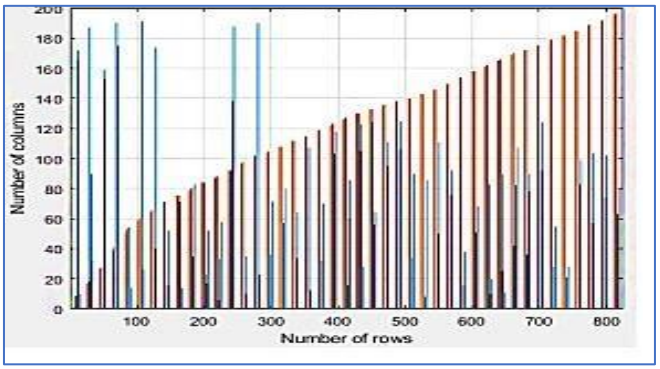
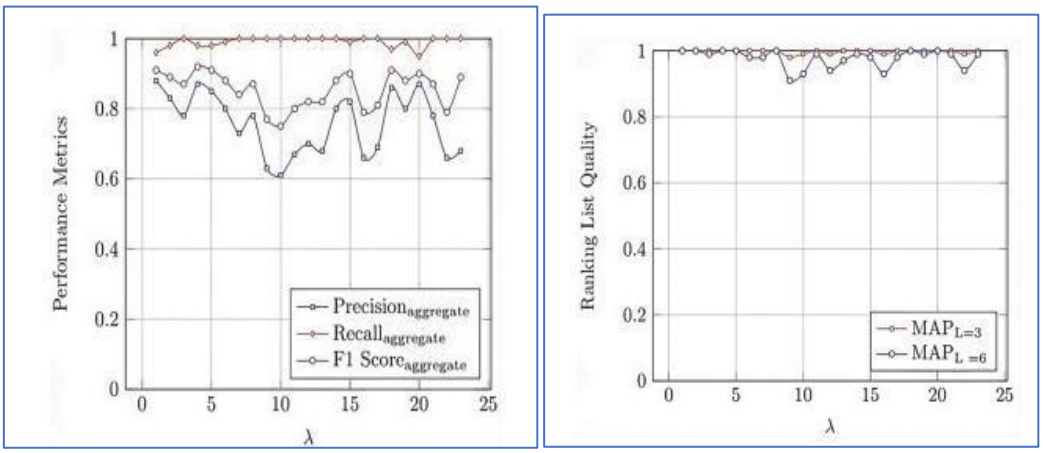


Figure 5: Feature vectors output

It is advised that the fuzzy rule feature extractor do two feature extraction operations, followed by object identification, in order to produce a feature vector, based on Figure 5. The feature vector in this extraction technique denotes two indications: the interest point and the non-interest point. In this case, the interest point is shown by the blue color bar line, and the remaining red lines represent non-interest locations. To categorize randomized feature vectors are extracted from each pixel and compared with the trained dataset's corresponding

features in test images. Many methods are used to analyze the interest and non-interest point. All that the feature extractor can do is locate the point of interest and cut out the unnecessary parts. Figure 6 displays the performance evaluation using several methods.



(a) **(b)**

Figure 6: Performance evaluation

Finally, validatestheproposedmodel. Anensemble of fuzzy-rule based classifiers is built for the purpose. This model is trained by 70% of the reduced dataset. Remaining 30% of dataset is used for testing the ensemble's performance. The training dataset is further distributed as approximatelyequal size samples to the classifiers.

Table 1: Performance evaluation

Rule Based Classifiers	Precision	Recall	Accuracy	Error Rate
Fuzzy Bayesian Network	0.795	0.833	0.837	0.163
Fuzzy Decision Tree	0.872	0.874	0.873	0.127
Fuzzy Random Forest	0.658	0.713	0.804	0.196
Fuzzy Support Vector Machine	0.825	0.834	0.861	0.139
Proposed Fuzzy rule-based model	0.951	0.972	0.98	0.02

From Table 1, It is observed that the fuzzy-rule based classifier achieved a classification accuracy of 98%. It is observed that incorporation of fuzzy-rule based theory in data preprocessing has reduced the noise and inconsistency of data. Further, rule-based classifier produced efficient classification results when tested for benchmark medical datasets. Performance evaluation of the proposed model is performed using above mentioned datasets.

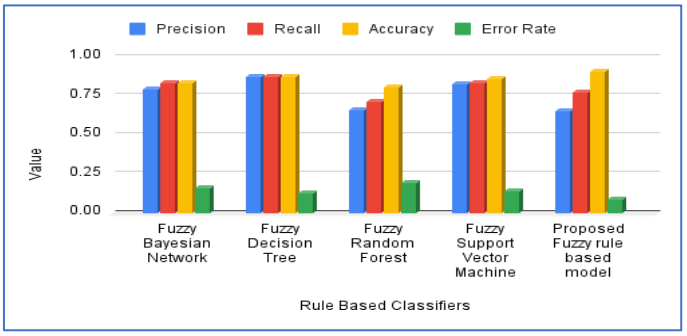


Figure 7: Classification Performance Fuzzy Rule based Classifiers

Medical datasets pose critical challenge for soft computing techniques in extracting accurate comprehensive knowledge for further medical diagnosis. This work emphasized on classification accuracy as a major criterion. Therefore, employed ensemble of proposed classifiers instead of single classifier. Consequently, the cost of computation (time and space complexity) may increase during learning and classification by the proposed

model. However, the current computational environment could easily facilitate working of such hybrid models. Results of all the phases are presented in tabular and graphical formats for better visualization and understanding.

5. Conclusions

Medical diagnosis plays a vital role to put forward the accurate treatment for proper curing of diseases in health care. Computerized information systems for healthcare have gained popularity over the last decade to enhance the safety of the patients by providing better accuracy in diagnosis. However, a major issue in modelling such information systems is to address uncertainty. Uncertainty is spread across in every part of healthcare. Physicians cannot prevent uncertainty; it is inevitable when making decision about diagnosis. The sources that contribute to uncertainty are patient's biological variability, patient and physician bias and also medical errors. Although uncertainty in practice cannot be completely eliminated, numerous strategies can be adopted to reduce uncertainty for better diagnosis. Modern researchers have introduced hybridized computational intelligence techniques to solve complex problems in various fields like medical diagnosis, aviation, banking, e-commerce, home appliances etc. Hybridized techniques have proved to be the best way to address problems in present scenario. Inspired with the outcome of such investigations, this research work is focused on developing a fuzzy rule-based systems model to aid the decision making in medical diagnosis by using hybridized computational intelligence.

References

1. Mendel, Jerry M. "Uncertain rule-based fuzzy systems." *Introduction and new directions* 684 (2017).

2. Fernández, Alberto, Sara del Río, Abdullah Bawakid, and Francisco Herrera. "Fuzzy rule based classification systems for big data with MapReduce: granularity analysis." *Advances in Data Analysis and Classification* 11 (2017): 711-730.

3. Lopez, Victoria, Sara del Rio, Jose Manuel Benitez, and Francisco Herrera. "On the use of MapReduce to build linguistic fuzzy rule based classification systems for big data." In *2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1905-1912. IEEE, 2014.

4. Wang, Hai, Zeshui Xu, and Witold Pedrycz. "An overview on the roles of fuzzy set techniques in big data processing: Trends, challenges and opportunities." *Knowledge-Based Systems* 118 (2017): 15-30.

5. Elkano, Mikel, Mikel Galar, Jose Sanz, and Humberto Bustince. "CHI-BD: A fuzzy rule-based classification system for Big Data classification problems." *Fuzzy Sets and Systems* 348 (2018): 75-101.

6. Bobek, Szymon, and Grzegorz J. Nalepa. "Uncertainty handling in rule-based mobile context-aware systems." *Pervasive and Mobile Computing* 39 (2017): 159-179.

7. Adriaenssens, Veronique, Bernard De Baets, Peter LM Goethals, and Niels De Pauw. "Fuzzy rule-based models for decision support in ecosystem management." *Science of the Total Environment* 319, no. 1-3 (2004): 1-12.

8. Yang, Long-Hao, Jun Liu, Ying-Ming Wang, and Luis Martínez. "A micro-extended belief rule-based system for big data multiclass classification problems." *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 51, no. 1 (2018): 420-440.

9. del Rio, Sara, Victoria López, Jose Manuel Benitez, and Francisco Herrera. "A mapreduce approach to address big data classification problems based on the fusion of linguistic fuzzy rules." *International Journal of Computational Intelligence Systems* 8, no. 3 (2015): 422-437.

10. Rajan.C. (2026). Thermal-Aware Co-Design of GaN-Based Power Amplifiers for High-Linearity 6G Sub-THz Transceivers. *National Journal of RF Circuits and Wireless Systems* , 3(3), 1-8.

11. Awan, Malik Shahzad Kaleem, and Mian Muhammad Awais. "Predicting weather events using fuzzy rule based system." *Applied Soft Computing* 11, no. 1 (2011): 56-63.

12. Segatori, Armando, Francesco Marcelloni, and Witold Pedrycz. "On distributed fuzzy decision trees for big data." *IEEE Transactions on Fuzzy Systems* 26, no. 1 (2017): 174-192.

13. Fernandez, Alberto, Cristobal Jose Carmona, Maria Jose del Jesus, and Francisco Herrera. "A view on fuzzy systems for big data: progress and opportunities." *International Journal of Computational Intelligence Systems* 9, no. Suppl 1 (2016): 69-80.

14. Ahlawat, Khyati, and Amit Prakash Singh. "Human activity recognition in imbalanced big data using fuzzy rule-based classification system." In *Soft Computing and Signal Processing: Proceedings of 2nd ICSCSP 2019 2*, pp. 31-38. Springer Singapore, 2020.Liu, Han, and Li Zhang. "Fuzzy rule-based systems for recognition-intensive classification in granular computing context." *Granular Computing* 3, no. 4 (2018): 355-365.

15. Riza, Lala Septem, Christoph Bergmeir, Francisco Herrera, and José M. Benítez. "frbs: Fuzzy rule-based systems for classification and regression in R." *Journal of statistical software* 65 (2015): 1-30.