International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Fuzzy Logic Rule Based Paradigm for Wine Quality Prediction

Onima Tigga[1] , Jaya Pal[*2] and Debjani Mustafi[3]

Submitted:11/03/2024 **Revised**: 26/04/2024 **Accepted**: 03/05/2024

Abstract: We consistently employ machine learning to uncover fascinating patterns and trends from large and intricate datasets. Utilizing supervised machine learning to classify real-life data is a common practice. This study employs multiple linear regression (MLR), random forest (RF), and fuzzy rule-based systems (FRBS) to classify wine quality and evaluate performance metrics. A FRBS demonstrates high accuracy. The calculations for magnitude of relative error (MRE) and mean magnitude of relative error (MMRE) demonstrate the achievement of model perfection. The MRE and MMRE achieved through a fuzzy rule-based system (FRBS) are lower than those obtained using MLR and Random Forest. The fuzzy rule-based system, multiple linear regression, and fuzzy logic system demonstrate MRE values of 0.1312 and MMRE values of 0.0043, respectively. This investigation utilizes the white wine dataset from the UCI Machine Learning Repository. Fuzzy logic represents a cutting-edge approach to creating wine quality prediction models. This research presents a fuzzy logic system that forecasts wine quality. Additionally, evaluate the system's performance in relation to MLR and random forest models. The results show that the MMRE value derived from fuzzy logic is less than the MMRE value derived from MLR. Furthermore, the values of Pred (0.25) and Pred (0.05) derived from fuzzy logic beat those achieved through multiple linear regression and random forest techniques.

Keyword: *Fuzzy Rule Based System, Multiple Linear Regression, MRE, MMRE, Random Forest*

1. Introduction

Machine Learning (ML) aims to understand data structures and incorporate them into models for specific tasks on novel data. Several areas, including business, medicine, astrophysics, and other scientific challenges, have extensively used machine learning. Motivated by machine learning's achievements across several industries, we use it to forecast wine quality based on multiple characteristics [1]. Classifier analysis, a subset of supervised learning in machine learning, predetermines the class label for each training tuple. Bayesian classification, rule-based classification, decision trees, k-nearest neighbours, and multiple linear regression are many classification methodologies [2]. Pattern recognition, medical diagnosis, and loan approval are just a few of the many uses for the work. The system sorts information based on established standards. But unlike the classification model, it can predict both labels and numerical data values [3]. Additionally, it may use current data to figure out distribution patterns. Classification determines the types of datasets by assigning labels to each class, categorizing determines the types of datasets and allows for the construction of models using test data. The theoretical analysis guides the FRBS-based classifier's implementation. The process involves creating a model, resolving the issue of classifying attributes or classes wi thin a dataset, and finally constructing the model. When testing unknown data, they build a model using training data to identify the class or attribute. They examined the prior model's performance to determine if it met the necessary requirements and then evaluated it using measurable criteria. This paper employs a white wine dataset from the UCI Machine Learning Repository to evaluate the FRBS, RF, and MLR classifiers [4]. It also compares their performance. When it comes to white wine, all seven classes classify the dataset into different classes and make quality predictions (3, 4, 5, 6, 7, 8, 9). Eleven attributes and 4,898 instances make up the dataset. The UCI machine learning library provided the white wine dataset for this study. The collection contains 4897 instances and eleven attributes. There are a total of seven classes in the dataset: 3, 4, 5, 6, 7, 8, and 9. Out of all the categories, 71 fall into Class 2, while 48 fall into Category 3. To build the proposed model, we select attributes using the feature selection approach. In order to discover the connections between characteristics, this research used a genetic algorithm. The analysis has identified three critical characteristics: chloride, citric acid, and fixed acidity.

Standardization follows the preprocessing stage. Equation (1) shows how we standardized the data using Z-score standardization. This method uses a scaling of -1 to 1 for feature values [5]. In this research, the properties of a normal distribution with a mean (μ) and a standard deviation (σ) of 0 and 1, respectively, have been used for each feature.

[1]Birla Institute of Technology, Mesra, Ranchi-834001, INDIA
ORCID ID: 0009-0006-3959-9294
[2]Birla Institute of Technology, Mesra, Ranchi-834001, INDIA
ORCID ID: 0000-0001-7843-283X
[3]Birla Institute of Technology, Mesra, Ranchi-835215, INDIA
ORCID ID: 0000-0002-7055-5031
*Corresponding Author Email: jayapal@bitmesra.ac.in

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

The abstract view of this research paper is depicted in Fig. 1.

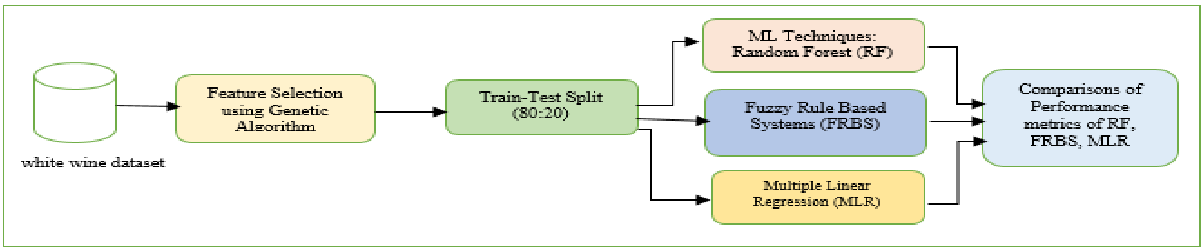


Fig. 1. Abstract view of research paper

The following is a list of the research paper's main contributions:

1. Feature selection on the white wine dataset using a genetic algorithm.
2. Divide the dataset into training and testing subsets.
3. Utilized machine learning techniques, including random forests (RF), multiple linear regressions (MLR), and fuzzy rule-based systems (FRBS).
4. Evaluate the performance metrics of the aforementioned techniques using MRE, MMRE, and prediction level (Pred (l)).

In the referenced research paper [6], the hierarchical fuzzy rule-based system yields a lower MMRE value than both multiple regression and conventional fuzzy rule-based systems. Nevertheless, the hierarchical fuzzy rule-based approach yields higher Pred (25) and Pred (5) values compared to the multiple linear regression and simple fuzzy rule-based system. In the paper [7], the authors utilized the rule-based classifier and decision tree approaches to construct rules and evaluate the accuracy and precision. As a result, the decision tree classifier yielded high accuracy. The authors used the metrics of entropy and information gain to identify the best characteristic for splitting and building a decision tree for the Iris dataset. They also utilize the confusion matrix to assess the accuracy of both methods. The experiment's results indicate that decision trees exhibit high accuracy and low mistake rates. Furthermore, as the number of data points increases, the decision tree classifier's accuracy confidence interval narrows compared to the rule-based classifier.

This research paper is organized as follows: Section 2 shows the related work. Section 3 describes the methodology. Section 4 delves into the discussion of the evaluation criteria. Section 5 presents the experiment's results and comparison analysis, while Section VI concludes this research work.

2. Related Work

In order to predict the quality of wine based on a variety of performance parameters, Dahal et al. (2021) used ridge regression, support vector machines (SVM), gradient-boosting neural networks (XGB), and multi-layer artificial neural networks (ANN). In terms of R-squared error (MSE) and mean absolute percentage error (MAPE), the gradient boosting regression model performs much better than the other models; its MSE results were 0.6057, while its MAPE values were 0.3741 and 0.0873 [1]. Dhaliwal et al. (2022) provided the dataset of red and white wines for preprocessing. Machine learning techniques, such as RF classifiers, decision trees, KNN, and ANN classifiers, have condensed the dataset from thirteen attributes to nine without compromising performance. Regarding accuracy and root-mean-squared error (RMSE) metrics, Random Forest significantly outperforms the other two classifiers in predicting wine quality [8].

Kumar et al. (2020) used RF, SVM, and NB techniques to predict wine quality. The used performance metrics are F1-score, precision, recall, accuracy, specificity, and misclassification error. Support Vector Machines (SVM) demonstrated the greatest performance, with an accuracy of 67.25% [9]. Geetanjali et al. (2021) employed LR, DT, RF, and Extra tree classifier algorithms to identify distinct levels of wine quality, ranging from exceptional to subpar. The study employs a single hot encoder to convert category values into numerical values. The RF and Extra Tree Classifier achieved accuracies of 88.19% and 88.79%, respectively [10].

Khilari et al. (2021) conducted an examination of the red wine dataset. They employ LR, DT, RF, SVM, Adaboost, and Gradient Boosting Machine Learning classifiers. The Random Forest method surpasses previous classifiers and has an accuracy rate of 92% [11]. Gawale (2022) claims ML and hybrid algorithms predict wine quality. The assessment criteria were accuracy, recall, precision, and F1-score. He compared various machine learning methods, including decision trees (DT), random forests

(RF), XGBoost, and a hybrid model. SMOTE enhances model performance. To improve model performance, researchers deleted outliers and null values. The RF approach is 85.57% accurate, the DT algorithm 79.25%, and the high gradient boost 78.07%. A hybrid model using several machine learning methods achieved 77.71% A study by Chaudhari et al. (2023) used a variety of machine learning approaches, such as Decision Tree (DT), K-nearest neighbor classifier (KNN), Random Forest (RF), Support Vector Classifier (SVC), and Logistic Regression (LR). The findings demonstrated that Random Forest (RF) produced the most advantageous result. Furthermore, they used Random Forest (RF) for feature selection during the process and discovered that the alcohol concentration has a more significant influence on the wine's quality. Furthermore, they utilized the SMOTE technique to tackle the problem of imbalanced classes by oversampling the minority class [13]. Moreover, Patkar and Balaganesh (2021) utilize artificial intelligence to predict the quality of wine. Unpredictable acidity is an indication of decay and can lead to an unpleasant smell. They presume that the wine possesses outstanding quality. Furthermore, a wine that possesses a higher liquor concentration typically signifies a higher level of quality [14].

According to Burigo et al., (2023) employed imbalanced data to classify the quality of wine. The researchers examined oversampling and under sampling techniques to enhance the model's operational effectiveness. The researchers found that incorporating oversampling techniques did not improve the performance of Random Forest (RF) in handling multiclass problems with imbalanced data. However, using SMOTE (Synthetic Minority Oversampling Technique) resulted in a substantial improvement in RF performance [15].

Data mining preparation comprises feature selection, according to Rostami et al. Selection removes low-predictive, high-connectivity characteristics. Many meta-heuristics have been developed to eliminate duplicate and superfluous features in high-dimensional datasets. A major difficulty with meta-heuristics is that they ignore variable relationships. The study discusses community detection-based genetic feature selection. The process takes 3 stages. Initial feature similarities are computed. Step two is community discovery-based clustering. A genetic algorithm selects features and implements a unique community-based repair operation in the third step. The method was evaluated on nine benchmark classification problems. The authors compared the proposed method to four feature selection methods. The recommended technique was compared to three innovative feature selection methods employing PSO, ACO, and ABC classifiers. The proposed technique outperforms PSO, ACO, and ABC algorithms by 0.52%,

accuracy. In classification algorithms, balanced data distribution and feature selection improve model performance. Using red and white wine datasets and machine learning classifiers, the research added variation [12].

1.20%, and 1.57% [16]. According to Babatunde Oluleye and colleagues employed the Genetic Algorithm (GA) for the purpose of feature selection. The binary genetic technique effectively reduced dimensionality and enhanced classifier performance. This research extracted 100 attributes from images in the publicly available Flavia dataset. The research presents Zernike Moments (ZM), Fourier Descriptors (FD), Legendre Moments (LM), Hu 7 Moments (Hu7M), Texture Properties (TP), and Geometrical Properties. This study presents two significant contributions. The documentation of the MATLAB GA Toolbox is comprehensive. Secondly, it introduces a GA-based feature selection method that employs a unique fitness function based on KNN classification error. GA identifies the optimal feature combination with the highest accuracy. We compared the outcomes using WEKA feature selection methods. Classification accuracy results exceed those of WEKA feature selectors [17].

According to Jie Lu et al., statistics, cognitive science, and computer science are all necessary for machine learning. Complex conditions and unpredictability continue to be challenges for machine learning, despite its theoretical and practical advancements. Data gaps, erroneous observations, and ambiguous connections are potential stumbling blocks for machine learning systems. Scientists used fuzzy sets, systems, logic, measurements, relations, and other machine learning tools to address these challenges. Every aspect of fuzzy machine learning—its theory, technique, and applications—are discussed in this article. Progress in fuzzy machine learning is summarized. In order to do this, we classified theories and models into five groups: Stream data, classical, reinforcement, transfer, and recommender systems are all examples of fuzzy machine learning methods. Fuzzy machine learning and its applications may be better understood by academics with the help of this material [18]. After evaluating the literature, M. Ivanova et al. conclude that machine learning and fuzzy logic can solve electrical difficulties. Our objective was to incorporate current field results and promising new research paths. A new research idea is proposed. Over the last decade, academics have developed, modelled, and implemented hardware-based intelligent systems using machine learning and fuzzy logic. Researchers search online for material on fuzzy logic and machine learning. The aim is to establish a well-defined conceptual framework and formalize domain knowledge. Thus, new research possibilities open up. They use bibliometrics and article reviews to conduct research. In the past decade, machine learning has become increasingly popular for addressing electrical issues, while fuzzy logic has become less prevalent. Machine learning has improved electronic studies in optimization, energy and control management, feature selection and extraction, defect detection, tracing, safety, identification, diagnostics, treatments, monitoring, and optimization. Many control systems optimize using fuzzy logic [19]. Jayanthi G. tested credit card fraud detection machine learning models. The evaluation uses 284,407 online transactions. Accurate data processing requires purification, scaling, mechanical property analysis, scalability, processing efficiency, and more. The study suggests ways to improve financial security and detect credit card fraud [20].

The study employs multiclass classification models on a well-balanced white wine dataset. The dataset is made up of seven distinct classes and seven extracted features. To

address the issue of an imbalanced multiclass dataset in the majority group, we have utilized 2197 instances of each type of white wine, each containing seven distinct attributes. Furthermore, there are a total of 613 occurrences, each including seven distinct characteristics.

imbalance correction, and transient feature evaluation. After preprocessing, they evaluated ANN, SVM, RF, DT, and NB. SVM, RF, DT, and NB had 95.5%, 94.5%, 92.3%, and 88.9% accuracy. ANN was 97.6% accurate. Each model's confusion matrices have high accuracy, true negatives, false positives, and false negatives for every sample. We found confusion over false negative detection. A fair and effective ANN recognized fraudulent transactions. However, the flaws in the DT-NB model revealed issues. When choosing credit card fraud detection machine learning models, we consider accuracy,

foundation for genetic algorithms. People often use them to provide high-quality solutions for optimization and

search difficulties [21]. This characteristics includes fixed acidity, citric acidity, and chloride.

3. Methodologies

Fig. 2. illustrates the complete workflow diagram for the research work as follows:

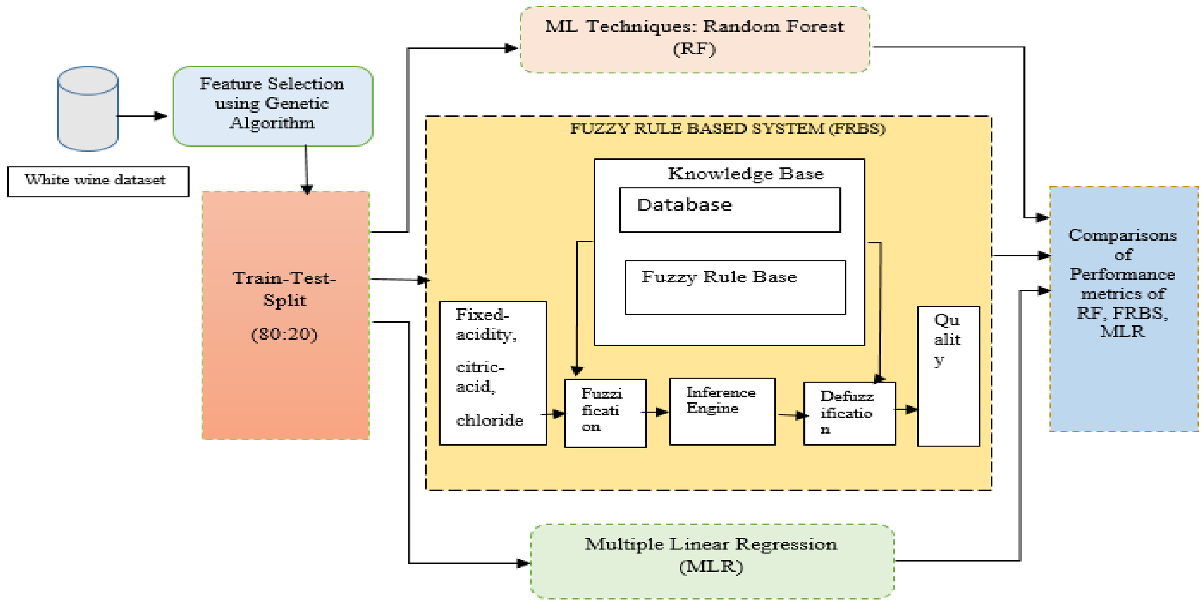


Fig. 2. The complete workflow diagram of the research work

The white wine dataset has eleven characteristics, as seen in this diagram. This strategy is used to extract these qualities. This technique uses a white wine dataset to retrieve three characteristics. This category includes fixed acidity, citric acidity, and chloride. Next, we split the white wine dataset 80:20 into training and testing data. Our technique employed multiple regression (statistical regression), random forest (machine learning), and fuzzy rule-based system (fuzzy logic) methods. We refer to the above procedures as statistical regression. We found that FRBS performed better than RF and MLR

3.1. Feature Selection using Genetic Algorithm

Genetic algorithms (GAs) draw an analogy from the genetic composition and behavior of chromosomes within a population. Genetic algorithms are adaptive heuristic search algorithms that are a subset of evolutionary algorithms. Natural selection and genetics serve as the

The steps of genetic algorithm are as follows:
Step1: Start.

Step2: Create initial population of chromosome or individual.

Step3: Calculate fitness value of each chromosome present in the population.

Step4: Apply selection operation of Genetic Algorithm on population to select best chromosome.

Step5: Apply crossover operation on chromosomes selected in step 4.

Step6: Apply mutation operation on step 5.

Step7: If termination criteria satisfy then Goto step 8 of the algorithm else Goto step 4.

Step8: End.

3.2. Machine Learning Techniques

3.2.1. Multiple Linear Regression (MLR)

The following is a multiple linear equation that represents multiple linear regressions with three independent variables as shown in "(2)":

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \ldots\ldots \quad (2)$$

Here, y is the dependent variable, x_1, x_2, $x_{3,} \ldots$ are the extracted independent variables, and b_0, b_1, $b_{3,} \ldots$ are constants. The coefficients of multiple linear regression are b0, b1, b2, and b3 are obtained by solving the following system of linear equations [22, 23].

$$\Sigma y = nb_0 + b_1(\Sigma x_1) + b_2(\Sigma x_2) + b_3(\Sigma x_3) \quad (3)$$

$$\Sigma x_1 y = b_0(\Sigma x_1) + b_1(\Sigma x_1^2) + b_2(\Sigma x_1x_2) + b_3(\Sigma x_1x_3) \quad (4)$$

$$\Sigma x_2 y = b_0(\Sigma x_2) + b_1(\Sigma x_1x_2) + b_2(\Sigma x_2^2) + b_3(\Sigma x_2x_3) \quad (5)$$

$$\Sigma x_3 y = b_0(\Sigma x_3) + b_1(\Sigma x_1x_3) + b_2(\Sigma x_2x_3) + b_3(\Sigma x_3^2) \quad (6)$$

The dependent variable is obtained by solving the system of linear equations "(3)", "(4)", "(5)", and "(6)" using three extracted independent variables fixed-acidity, citric-acidity, and chloride from the white wine dataset is shown in "(7)" as follows:

$$quality = 7.05358 - 0.12699 * fixed\_acidity + 0.31281 * citric\_acid - 8.93052 * chloride \quad (7)$$

Thus equation "(7)" gives the multiple linear regression model in this research paper.

3.2.2. Random Forest (RF)

The RF ensemble training technique is used to create a classifier model that divides the dataset into several smaller datasets. Bootstrap aggregation is employed in Random Forest (RF) to effectively reduce variance and minimize the risk of over fitting, making it a superior approach in most situations. The ensemble models are constructed using these datasets. RF builds many decision trees without pruning, creating a forest that has a predicted value for a specific data sample. The majority voting of the trees determines the predicted value. A decision tree model serves as the foundation for the RF machine learning model. RF models, derived from the idea of decision trees, generate a large number of trees ('n') that greatly improve prediction accuracy compared to a single tree. We achieve this by randomly selecting a subset of trees from the training set without replacement. Decision trees typically have a tree-like configuration with a primary node, known as the root or decision node, positioned at the top [24, 25].

This is a pseudo code for implementing a Random Forest algorithm.

Step 1: Generates N bootstrap samples from the white wine dataset.

Step 2: Each node (sample) randomly selects a feature of size m,
where m is less than M, where M is the total number of features and m is the sample size.

Step 3: It involves constructing a split for the selected m features from Step 2
and identifying the kth node using the test split point.

Step 4: Continuously divide the tree until only one leaf node is reached and
the tree is fully constructed.
Step 5: We train the algorithm on each bootstrap sample separately.

Step 6: Using tree classification voting, we collect the data predicted by the trained trees (n).

Step 7: We construct the ultimate RF model using the attributes that received
the highest number of votes.

3.3 Fuzzy Rule Based System (FRBS)

A fuzzy rule based system is designed to regulate a process with a specified number of inputs, denoted as I_1, $I_2,\ldots,I_n$ and a function can defined its input-output connection (fuzzy rules) as shown in "(8)".

$$O = f(I_1, I_2, \ldots\ldots, I_n) \quad (8)$$

Let us assume that each variable is denoted by a 'm' number of fuzzy sets. Hence, a total of m^n (where n is the number input variables) fuzzy rules need to be developed in order to create the Fuzzy rule-based system. It follows that when the number of rules, increases exponentially with the number of variables, the algorithm's complexity and the challenges related to process control will also increase. The issue of rule explosion positions a challenge in the development of Fuzzy rule-based systems (FRBS) for attempting complicated real-world situations with numerous variables [26]. The problem can be resolved by employing a Fuzzy Rule Based System. Intelligent Systems offer many approaches that try to make it easier to represent and manipulate data that is unclear, partial, inaccurate, or noisy. Fuzzy Logic provides a straightforward method to create a precise mapping between input and output areas due to the natural formulation of fuzzy rules [27]. This research paper introduces a fuzzy rule-based system that includes three fuzzy inputs: fixed acidity (FA), citric acid (CA), and chloride (CHL), as well as one output, quality (Q), as depicted in Fig. 3.

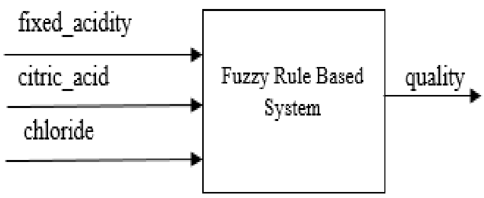


Fig. 3. Fuzzy Rule Based System

A typical fuzzy logic system comprises four primary components, as depicted in the Fig. 4.

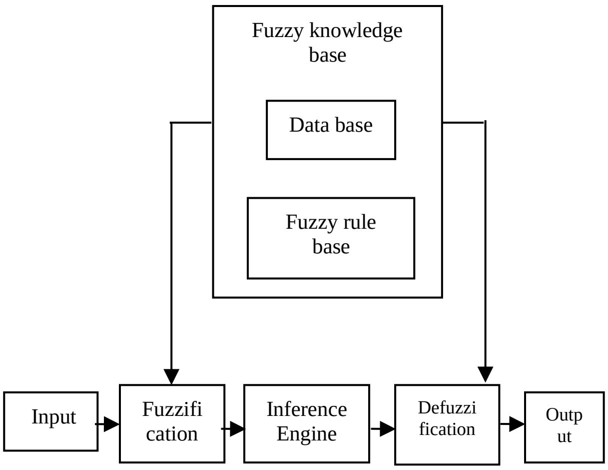


Fig. 4. Fuzzy Logic System

A standard fuzzy logic system comprises four primary components, as depicted in the diagram.

1. Fuzzification involves the utilization of a pre-established collection of linguistic qualities. This procedure's goal is to convert non-fuzzy (deterministic) inputs from a fuzzy system into fuzzy inputs that the inferencing mechanism can use.

2. The knowledge base comprises two primary elements: a database that establishes linguistic variables and conditional statements (fuzzy sets), and a rule base that provides instructions on how to convert fuzzy input sets to fuzzy output sets. We use fuzzy implications to represent rules effectively. We use fuzzy sets and a rule base to demonstrate how to map a fuzzy input set to a fuzzy output set. Rules can be defined as conditional assertions, often known as implications that may have some ambiguity or lack of clarity.

3. Decision logic: a computational model that emulates human decision-making processes by incorporating fuzzy ideas. Logical decision-making determines the conclusion of a specific condition.

4. Defuzzification is the process of transforming fuzzy outputs from a rule base into non-fuzzy numerical values.

The fundamental process of the knowledge base and decision-making logic entails integrating the fuzzy extension of the traditional rule inference concept into fuzzy rule inference. The premises and conclusions of the rules now include vague or ambiguous values. These facts intrinsically represent a continuous and consistent input of attributes. By adopting this approach, a single rule can supplant multiple traditional rules [26, 28].

Fuzzy inferencing rules typically establish connections between m conditional variables (X_1,..., X_m) and n subsequent variables (Y_1,..., Y_n) in the following format:

IF (**X_1** is A_1 and ………..X_m is A_m) **THEN** (**Y_1** is B_1 and ... Y_n is B_n).

Where A_1,..., A_m and B_1,..., B_n are linguistic terms of linguistic variables X_1,...,X_m andY_1,..., Y_n respectively.

The "antecedent" refers to the IF statement, whereas the "consequent" describes the THEN statement. To arrive at a conclusion, a rule-based system follows these steps:

Step1. Fuzzification is a process which is employing the membership functions and truth values that are obtained from the inputs to apply the rules. We execute this procedure to determine the outcome of the previous step.

Step2. After that, the results are converted into a membership function and a truth value, which controls the output variable. This process is called the implication process.

Step3. The process of combining these data is known as aggregation. A popular approach to aggregation is to use the "maximum" of the associated sets.

Step4. Finally, a single value representing of the aggregated fuzzy set is computed using a technique called Defuzzification.

Input and output Membership Functions (MF) are portrayed in Table 1.

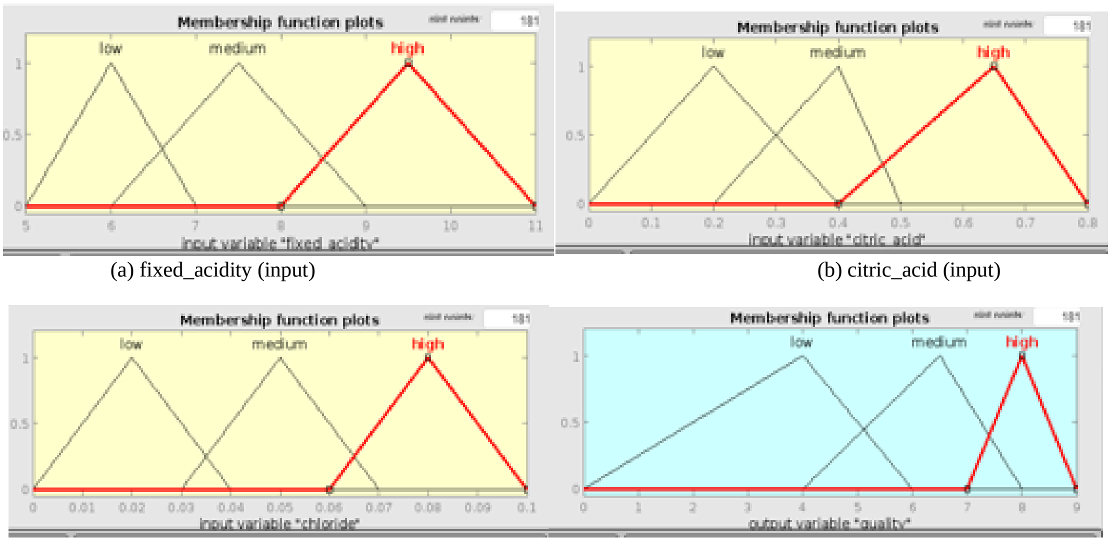
Table 1. Membership Function Characteristics

Inputs

Variable Name	Range	MF	Parameter		
			a	b	c
fixed_acidity	5 - 11	Low	5	6	7
		Medium	6	7.5	9
		High	8	9.5	11
citric_acid	0 – 0.8	Low	0	0.2	0.4
		Medium	0.2	0.4	0.5
		High	0.4	0.6	0.8
chloride	0 - 0.1	Low	0	0.02	0.04
		Medium	0.03	0.05	0.07
		High	0.06	0.08	0.1

Outputs

Variable Name	Range	MF	Parameter		
			a	b	c
quality	0 - 9	Low	0	4	6
		Medium	4	6.5	8
		High	7	8	9



(a) fixed_acidity (input)

(b) citric_acid (input)

(c) chloride (input)

(d) quality (output)

Fig. 5(a), 5(b), 5(c), and 5(d) have been created to illustrate the membership function plots that correlate to Table 2.

The surface area of the above membership plots is shown below:

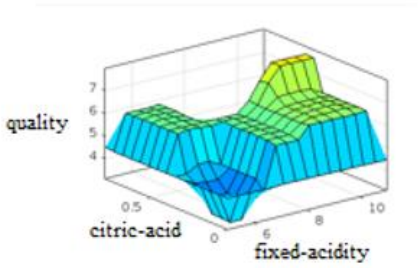


Fig 6 Surface Area of the membership function plots

3.4. Fuzzy Rules

This research paper implements a Mamdani-type fuzzy rule-based system. The general structure of fuzzy rule is as shown below:

R_k: If [x_1 is A_{1k}] & [x_2 is A_{2k}] & …& [x_n is A_{nk}] then [y is B_k]

Where R_k is the k-th rule in the fuzzy rule base (k=1, 2, …, k). Here A_{jk} and B_k are fuzzy sets on appropriate domains (j=1, 2, … , n). The original fuzzy rule base comprises a total of 27 rules, resulting from the combination of 33 different factors.

This paper employs a fuzzy rule-based system, consisting of 27 rules, which are outlined as follows:

Rule1: If fixed_acidity is low, if citric_acid is low, and if chloride is low then quality is low.

Rule2: If fixed_acidity is low, if citric_acid is medium, and if chloride is low then quality is low.

Rule3: If fixed_acidity is low, if citric_acid is high, and if chloride is low then quality is medium.

Rule4: If fixed_acidity is low, if citric_acid is low, and if chloride is medium then quality is low.

……. ……………. ……………… ………………….. ……….

……. ……………… ……………… ………………….. ………..

Rule24: If fixed_acidity is low, if citric_acid is medium, and if chloride is medium then quality is medium.

Rule25: If fixed_acidity is high, if citric_acid is high, and if chloride is low then quality is high.

Rule26: If fixed_acidity is high, if citric_acid is low, and if chloride is medium then quality is medium.

Rule27: If fixed_acidity is medium, if citric_acid is high, and if chloride is high then quality is high.

4. Evaluation criteria

A commonly used metric for assessing wine quality models is the Magnitude of Relative Error (MRE), which is defined in the "(9)" as follows:

$$MRE = \frac{|Actual\_quality - Predicted\_quality|}{Actual\_quality} \quad (9)$$

In order to ascertain the MRE value, it is necessary to do calculations on each observation whose quality is anticipated. It is feasible to obtain the aggregate of MRE over a large number of observations (N) by making use of the Mean MRE (MMRE), which can be found in " (10)" [26, 27].

$$MMRE = \frac{1}{N}\sum_{i}^{N} MRE \quad (10)$$

A complementary criterion is the prediction at level l, which is denoted by "(11)".

$$\text{Pred}(l) = k/N \quad (11)$$

In this equation, k represents the number of observations in which the MRE is equal to or less than l, and N represents the total number of observations. Therefore, Pred (25) and Pred (5) represent the proportion of data that were predicted with a Mean Relative Error (MRE) less than or equal to 0.25 and 0.05, respectively.

5. Experimental results

For this research work, we have taken the white wine dataset, accessible via a web URL, from the UCI machine learning repository [4]. This collection contains 4898 instances of properties such as fixed acidity, citric acid, and chloride. After partitioning the dataset, we allocated 80% of the records to the training data, yielding 3917 records, and the remaining 20% to the testing data, yielding 980 records. Table 2 shows the description of white wine dataset.

Table 2. Datasets Description

Name of Datasets	Instances	Attributes	Classes
White Wine	4898	11	7

The class-wise count of the white wine dataset is 20 for quality 3, 163 for quality 4, 1457 for quality 5, 2197 for quality 6, 880 for quality 7, 175 for quality 8, and 5 for quality 9 respectively. No null values are found in both datasets. MLR, RF, and FRBS use the same data subset. The comparison of actual and predicted values with MRE results using MLR, RF, and FRBS for training and testing data are shown in Table 3 and Table 4 respectively.

Table 3. Comparison of actual and predicted values using MLR, RF, and FRBS for training data

Sl. No.	Actual Value	Techniques					
		Multiple Linear Regression		*Random Forest*		*Fuzzy Rule Based System*	
		Predicted Value	*Error*	*Predicted Value*	*Error*	*Predicted Value*	*Error*
1.	6	5.703	0.297	5	1	6.22	**-0.22**
2.	6	5.721	0.279	5	1	6.13	**-0.13**
3.	6	5.875	0.125	5	1	6.14	**-0.14**
4.	5	5.755	**-0.755**	5	0	4.74	0.26
5.	4	5.785	**-1.785**	5	**-1**	4.34	**-0.34**
6.	5	5.757	**-0.757**	5	0	4.46	0.54
7.	5	5.766	**-0.766**	5	0	4.46	0.54
8.	5	5.836	**-0.836**	5	0	4.5	0.5
9.	5	5.441	**-0.441**	5	0	4.5	0.5
10.	5	5.877	**-0.877**	5	0	4.5	0.5
11.	5	5.795	**-0.795**	5	0	4.5	0.5
12.	6	5.747	0.253	5	**-1**	5.38	0.62
13.	6	5.434	0.566	5	**-1**	6.06	**-0.06**
14.	6	5.743	0.257	4	**--2**	6.13	**-0.13**
15.	5	4.592	0.408	5	0	4.5	0.5
16.	4	5.903	**-1.903**	4	0	4.87	**-0.87**
17.	6	5.924	0.076	5	**-1**	6.09	**-0.09**
18.	4	5.203	**-1.203**	5	**-1**	4.5	**-0.5**
19.	6	6.0600	**-0.06**	5	**-1**	6.14	**-0.14**
20.	5	5.091	**-0.091**	5	0	4.5	0.5
………	……	……..	……	…..	**……**	……	**……**
3913.	6	5.677	0.323	5	**-1**	6.13	**-0.13**
3914.	7	5.659	1.341	5	**-2**	6.67	0.37
3915.	4	5.874	**-1.874**	5	**-1**	3.84	0.12
3916.	4	5.901	**-0.68**	6	**-2**	4.68	**-0.68**
3917.	6	5.869	0.131	5	**-1**	6.11	**-0.11**

Table 4. Comparison of actual and predicted values using MLR, RF, and FRBS for testing data

Sl. No.	Actual Value	Techniques					
		Multiple Linear Regression		*Random Forest*		*Fuzzy Rule Based System*	
		Predicted Value	*Error*	*Predicted Value*	*Error*	*Predicted Value*	*Error*
1.	6	5.767	0.233	5	**-1**	6.13	**-0.13**
2.	5	5.898	**-0.898**	5	0	5.29	**-0.29**
3.	5	4.724	0.276	5	0	4.5	**-0.5**
4.	7	5.880	1.12	5	**-2**	6.45	**-0.55**
5.	6	5.736	0.264	5	**-1**	6.13	**-0.13**
6.	5	5.763	**-0.763**	6	**-1**	4.56	**-0.44**
7.	5	5.892	**-0.892**	4	1	4.87	**-0.13**
8.	5	5.967	**-0.967**	4	1	5.6	**-0.6**
	......	....	....	......	....	....	....
975.	5	5.911	**-0.911**	4	**-1**	5.26	**-0.26**
976.	6	5.848	0.152	6	0	6.12	**-0.12**
977.	6	5.740	0.26	6	0	6.14	**-0.14**
978.	5	5.754	**-0.754**	5	0	4.5	0.5
979.	5	5.844	**-0.844**	5	0	4.5	0.5
980.	4	5.858	**-1.858**	5	**-1**	4.4	**-0.4**

Furthermore, Table 5 presents the Pred (5), Pred (25) and MMRE.

Table 5. Prediction Results

Performance Metrics	Techniques Multiple Linear Regression	Random Forest	Fuzzy Rule Based System
MMRE	0.0041	0.0043	**0.0023**
Pred(5)	0.4	0.33	**0.5**
Pred(25)	0.86	0.9	**1.0**

Table 5. shows that FRBS has lower MMRE values than Random Forest and Multiple Linear Regression. As a result, we classified the white wine dataset using FRBS.

6. Conclusions

This research paper evaluates the efficacy of machine learning approaches and a fuzzy rule-based system by assessing their performance using Pred and MMRE as metrics. The experimental results demonstrate that the Fuzzy Rule Based System (FRBS) can serve as an alternate strategy for forecasting new tuples. This paper demonstrates the application of prediction using Python code and MATLAB. It showcases the results achieved through the use of a fuzzy logic system, as well as ordinary multiple linear regression and machine learning techniques as random forest. The results indicate that fuzzy logic can serve as a viable option for estimating wine quality. The dataset specifically focuses on white wines. The results indicate that the MMRE value is lower when applying fuzzy logic compared to applying multiple linear regression. Additionally, the Pred (25) and Pred (5) values for fuzzy rule based paradigm are greater than MLR and RF as shown in Table 4. A future work needs applying additional datasets to establish a comprehensive fuzzy tool for predicting quality. Our future measurements will incorporate additional elements to improve our predictions' accuracy.

7. Future Work

In conjunction with our current efforts, the further tests are scheduled:

1.Balancing of datasets may be performed, and the improvement in accuracy will be assessed.
2. Ensemble methods may be used, and comparisons will be conducted for specific datasets.

Conflict of Interest There is no conflict of interest declared by the authors.

Data Availability The dataset is sourced from the UCI Machine Repository.

References

[1] K. R. Dahal, J. N. Dahal, H. Banjade, and S.Gaire, “Prediction of Wine Quality Using Machine Learning Algorithms”, Open Journal of Statistics, 11, 278-289, 2021. https://doi.org/10.4236/ojs.2021.112015.

[2] P. N. Tan, M. Steinbach, A. Karpatne, and V. Kumar, "Introduction to Data Mining," 2nd ed., Pearson Publications, 2022.

[3] J. Han, M. Kamber, and J. Pei, Data Mining Concepts & Techniques, 3rd ed., Morgan Kaufmann Publishers, ISBN: 978-93-80931-91-3, 2008.

[4] https://archive.ics.uci.edu/ml/datasets/Wine+Quality

[5] O. Tigga, J. Pal, and D. Mustafi, " Performance Analysis of Machine Learning Algorithms for Data Classification", International Conference on Machine Intelligence with Applications (ICMIA 2023), Birla Institute of Technology, Mesra , Off Campus Lalpur, Ranchi, India,07-08, 2023, https://doi.org/10.1063/5.0214183.

[6] J. Pal, D. Mustafi and O.Tigga, "Prediction of Software Quality using Hierarchical Fuzzy Rule Based System" International Journal of Emerging Technologies and Innovative Research(UGC Approved), ISSN:2349-5162, Vol.11, Issue 5, pp.m630-m636, 2024. http://www.jetir.org/papers/JETIR2405C82.pdf

[7] O. Tigga, J. Pal, D. Mustafi, "A Comparative Study of Rule-Based Classifier and Decision Tree in Machine Learning", In Proceedings of 4th International Journal of Advances in Soft Computing and Intelligent Systems (IJASCIS), Vol 02, Issue 01, 40-47, 2023. https://sciencetransactions.com/ijascis/uploads/2023/02/j23-40-47.pdf

[8] P. Dhaliwal, S. Sharma, and L. Chauhan, "Detailed Study of Wine Dataset and its Optimization", I. J. Intelligent Systems and Applications. 5, 35-46, 2022. doi: https://doi.org/10.5815/ijisa.2022.05.04

[9] S. Kumar, K. Agarwal, and N. Mandan, "Red Wine Quality Prediction Using Machine Learning Techniques", International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India. pp. 1-6. 2020. https://doi.org/10.1109/ICCCI48352.2020.9104095

[10] T. M. Geethanjali, M. Y. Sowjanya, S. N. Rohith, and B. E. Shubashree, "Prediction of Wine Quality using Machine Learning", Journal of Emerging Technologies and Innovative Research (JETIR). volume 8, Issue 11, 2021. http://www.jetir.org/papers/JETIR2111328.pdf

[11] N. Khilari, P. Hadawale, H. Shaikh, and S. Kolase, "Analysis of Machine Learning Algorithm to Predict Wine Quality", International Research Journal of Engineering and Technology (IRJET). volume: 08 Issue 12, 2022. DOI: 10.32628/IJSRSET229235

[12] Gawale, "Wine Quality Prediction using Machine Learning and Hybrid Modeling", School of Computing, National College of Ireland, 2022.

[13] M. S. Chaudhari, A. Kiran , H. Shahare., V. Helwatkar, S. Shinde, D. Janbandhu, and S. Rangari, "VinQCheck: An Intelligent Wine Quality Assessment", International Journal of Innovative Science and Research Technology. Volume 8, Issue 12. 2023. doi : https://doi.org/10.5281/zenodo.10393843

[14] G. S. Patkar, D. Balaganesh, "Smart Agri Wine: An Artificial Intelligence Approach to Predict Wine Quality," Journal of Computer Science 2021. DOI: https://doi.org/10.3844/jcssp.2021.1099.1103

[15] B. Robert, F. Scott, K. Eli, and L. Nibhrat, "Comparison of sampling Methods for Predicting Wine Quality Based on Physicochemical Properties", SMU Data Science Review. Vol. 7: No. 1, Article 8. 2023. https://scholar.smu.edu/datasciencereview/vol7/iss1/

[16] M. Rostami, K. Berahmand, and S. Forouzandeh, "A novel community detection based genetic algorithm for feature selection ", Journal of Big Data, Rostami et al. J Big Data 2021. 8:2 https://doi.org/10.1186/s40537-020-00398-3.

[17] B. Oluleye, A. Leisa, J. Leng, and D. Dean, "A Genetic Algorithm-Based Feature Selection", International Journal of Electronics Communication and Computer Engineering Volume 5, Issue 4, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209, 2014.

[18] L. Jie, M. Guangzhi, and G. Zhang, " Fuzzy Machine Learning: A Comprehensive Framework and Systematic Review", IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 32, No. 7, 2024.

[19] M. Ivanova, P. Petkova, "Machine Learning and Fuzzy Logic in Elecronics: Applying Intelligence in Practice", Electronics 2021. 10, 2878. https://doi.org/10.3390/electronics10222878

[20] G. Jayanthi, P. Deepthi, B. N. Rao, M. Bharathiraja, A. Loga Priya, " A Comparative Study on Machine Learning and Fuzzy Logic-Based Approach for Enhancing Credit card Fraud Detection", International Journal of intelligent systems and applications in engineering, IJISAE, 2024. 12(12s), 192-199.

[21] V. Pandit, "A Study On Multiple Regression Analysis In Era Of Demonetization," , Issue. 3 of the 2017 IJARIIE ,Vol. 2, C-1542, 2017.

[22] M. Gera and S. Goel, "Data Mining- Techniques, Methods and Algorithms: A Review on Tools and their Validity," International Journal of Computer Applications (0975-8887), Volume 113-No. 18, March 2015.

[23] Y. Cao, H. Chen, and B. Lin, " Wine Type Classification Using Random Forest", Highlights in Science, *Engineering, and Technology,* SDPIT2022, Volume 4, 2022.

[24] H. Lingjun, R. A. Levine, J. Fan, J. Beemer, and J. Stronach, "Random Forest as a Predictive Analytics Alternative to Regression", *Institutional Research. ISSN 1531-7714,* Volume 23, Number 1, 2018.

[25] V. Bhattacherjee, and J. Pal, "Fuzzy Rule Based System for Fine Prediction in Accidents" in proceedings of Conference on information Science, Technology and Management, CISTM 2009, MDI Gurgaon, India, July 13-15, 2009.

[26] J. Pal, and V. Bhattacherjee, "Fuzzy Rule Based Classifier for SoftwareQuality Data", Journal of Theoretical and Applied Information Technology(JATIT),Vol 72,No 1, ISSN 1992-8645, pp.58-66, 2015.

[27] N. Viraj, and V. Bhattacherjee, "Software Development Time Estimation Using Fuzzy Logic", presented at TECHTATVA 2009, Manipal Institute of Technology, Karnataka, India, September 8-11, 2009.