Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 5 (2023)

# Performance Analysis of Machine Learning Techniques and Fuzzy Rule Based Systems on Classification Problems

**A.K. Agrawal[1], Amresh Kumar[2], Piyush Kumar Tripathi[3],**
[1,3]Amity School of Applied Science, Amity University Uttar Pradesh Lucknow UP India,
[2]BBDNIIT, Lucknow UP India
Email: akagrawal@lko.amity.edu, pktripathi@lko.amity.edu, amreshsrivastava17@gmail.com

**Abstract.** Machine learning techniques and fuzzy rule-based systems (FRBS) are important tools to solve classification problems. In this paper, Machine learning and FRBS are used to design and implement systems for classification problems. The system is modelled using logistic regression (LR), random forest classifier (RF), gradient boosting classifier, gaussian naive bayes (NB), decision tree classifier (DT), K-nearest neighbour classifier and support vector machine (SVM). The prediction accuracy of these machine learning models is approximately 77%. Also, the same classification problem was modeled in FRBS and a big improvement was achieved in the accuracy, which was 99.7% with fuzzy partition Count 5. Hence, in this work it is concluded that FRBS are more accurate than machine learning and prediction Models.
**1991 Mathematics Subject Classification:** 03B52.
**Corresponding Author:** Piyush Kumar Tripathi

**Keywords and phrases:** Fuzzy Rule Based Systems (FRBS), Logistic Regression (LR), Random Forest Classifier (RF), Support Vector Machine (SVM), K-nearest neighbour, Decision Tree Classifier(DT), Fuzzy Partition, Membership Function.

## 1. Introduction

Machine learning is the process to learn and train computers by analyzing data using statistical tools [1]. Several techniques are used to model prediction systems like logistic regression [2], random forest classifier [3], gradient boosting classifier [4], gaussian naive bayes [4], decision tree classifier [5], K-nearest neighbour classifier [6], support vector machine [7]. Machine learning techniques are utilised in different areas like health & care, insurance, farming, retail, banking and finance, real estate etc.
On the other hand, fuzzy systems are used to design and implement classification systems [8]. Fuzzy systems are based on strong mathematical framework to deal with impression and uncertainty existing in the real-world systems [9]. Mamdani type fuzzy rule-based systems are used commonly to implement classification systems. The components of Mamdani FRBS are as follows:

- Fuzzification interface: To convert crisp information into fuzzy.
- Inference Engine: To take decision on the input using available knowledge in if then rule format.
- Knowledge base: It has two components i.e. rule base and database. Fuzzy if then rules are stored in the rule base and membership functions are stored in the database.
- Defuzzification interface: Converts the fuzzy output into crisp output which will be the desired output [10, 11].

This paper is divided into four sections. section 1 is the introduction. section 2 is the proposed model and implementation. Discussion and result analysis is presented in section 3. section 4 is the conclusion and future scope of the work done in this paper.

## 2. Proposed model and implementation

**2.1. Machine learning implementation.** Machine learning models are used to implement the classifier which are logistic regression (LR), random forest classifier (RF), gradient boosting classifier, gaussian naive bayes (NB), decision tree classifier (DT), K-nearest neighbour classifier and support vector machine (SVM). PIMA dataset [12] is used for the classifier generation for machine learning and FRBS. This dataset is concerned with the prediction of diabetes in the patients. This dataset has 9 variables which are as follows;

- Number of times pregnent
- Concentration of plasama glucose measured in oral glucose tolerence test.
- Diastalic Blood Pressure
- Triceps skinfold thickness
- Serum insuline
- BMI (Body Mass Index)
- Diabetes pedigree function
- Age in years
- Decision Variable (0 or 1)

The statistical analysis has been done.

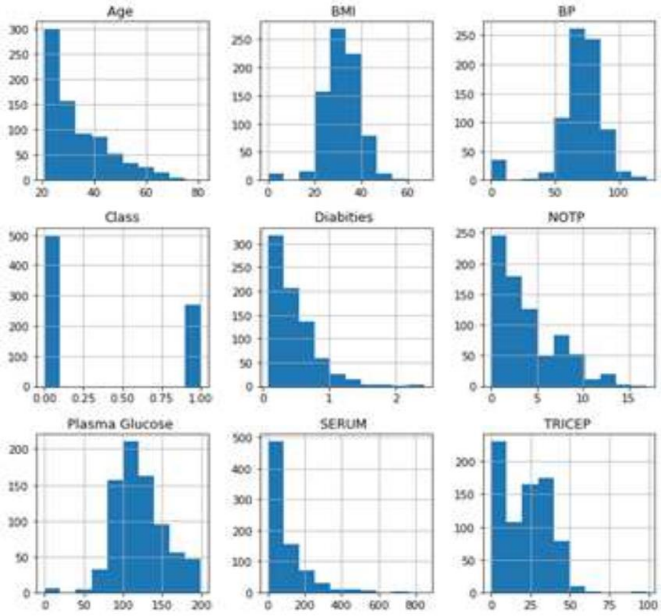


**Figure 1. Histrograms on the input data set**

The null values are given Table-(1) which shows no null values. Missing values are tabulated in Table- (2) which shows that there is no null values. After excuting all machine learning models accuracy and prediction are given in Table- (3).

TABLE 1. Null Values

| Name | Scores |
|---|---|
| NOTP | 0 |
| Plasma Glucose | 0 |
| BP | 0 |
| TRICEP | 0 |
| SERUM | 0 |
| BMI | 0 |
| Diabities | 0 |
| Age | 0 |
| Class | 0 |

TABLE 2. Missing Values

| Name | Scores |
|---|---|
| NOTP | 0 |
| Plasma Glucose | 0 |
| BP | 0 |
| TRICEP | 0 |
| SERUM | 0 |
| BMI | 0 |
| Diabities | 0 |
| Age | 0 |
| Class | 0 |

TABLE 3. Accuracy values of the models

| Sl.No. | Name | Scores |
|---|---|---|
| 0 | LR | 0.77921 |
| 1 | RF | 0.77921 |
| 2 | GB | 0.787879 |
| 3 | GN | 0.761905 |
| 4 | DT | 0.748918 |
| 5 | KN | 0.748918 |
| 6 | SV | 0.753247 |

TABLE 4. Accuracy values of the models (with cross validation)

| Sl.No. | Name | Scores |
|---|---|---|
| 0 | LR | 0.773479 |
| 1 | RF | 0.757809 |
| 2 | GB | 0.761705 |
| 3 | GN | 0.756494 |
| 4 | DT | 0.709604 |
| 5 | KN | 0.721377 |
| 6 | SV | 0.757861 |

In this experiment 70% data is used for training and 30% is used for testing.
Now, the experiment is done with K-fold cross validation.

The accuracy results are given in Table- (4) and comparative chart is shown in Fig-(2).

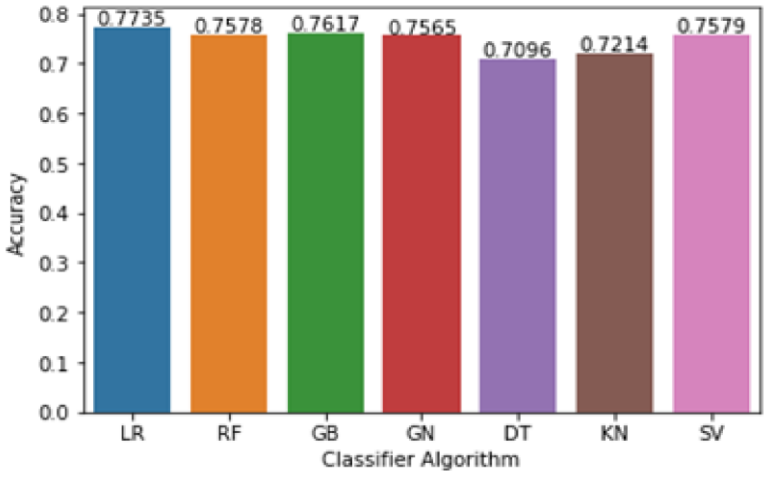


**Figure 2. Bar chart showing accuracy values with cross validation**

**2.2. Classification using Fuzzy Rule Based Systems.** Mamdani Fuzzy Rule Based Systems are used to design and implement classifier. The experiment is done using Guaje open source software.

- Rule generation technique: Wang and Mendel
- Partition: K-Means

The experiments are done with different values of number of Fuzzy Partitions (NOP).

- Experiment 1 - NOP = 3
- Experiment 2 - NOP = 5
- Experiment 3 - NOP = 7

The comparative results are shown in Table-(5).

TABLE 5. Accuracy & Interpotability Results

| Parameters | Experiment-1 | Experiment-2 | Experiment-3 |
|---|---|---|---|
| Number of partitions | 3 | 5 | 7 |
| Accuracy(%) | 77.7 | 99.2 | 99.7 |
| MSE | 0.143 | 0.091 | 0.089 |
| NOR | 116 | 745 | 763 |
| TRL | 928 | 5960 | 6104 |
| ARL | 8 | 8 | 8 |

The comparative results charts are given in Fig.- [3,4,5,6,7].

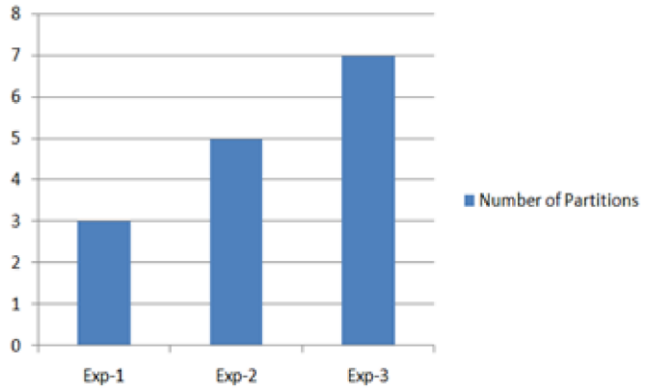


**Figure 3**

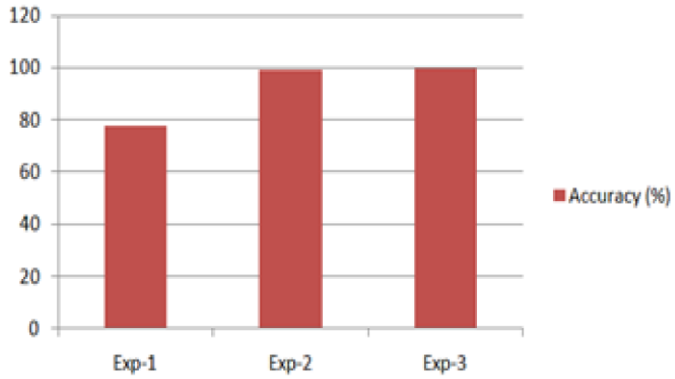


**Figure 4**

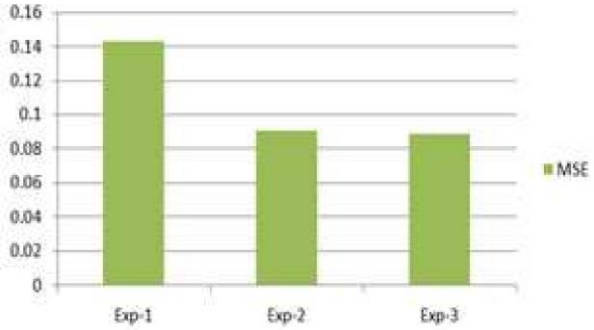


**Figure 5**

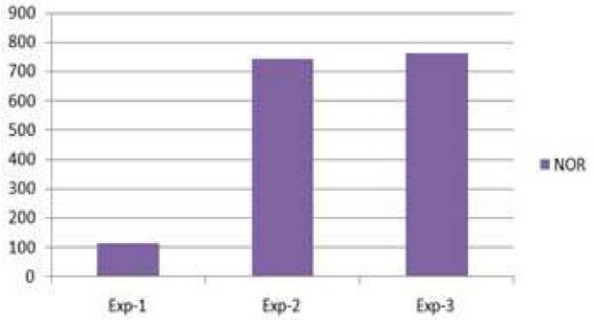


**Figure 6**

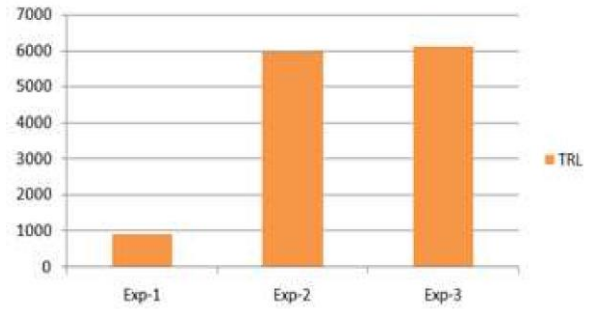


**Figure 7**

3. **Result Analysis and Discussions**

After the experiments done as discussed in previous sections, following findings are drawn:

- Gradient Boosting Classifier is better in terms of accuracy without K-fold cross validation.
- Logistic Regression is better in accuracy while using K-Fold Cross Validation.
- In fuzzy implementation, higher number of fuzzy partition results into improved accuracy with a higher difference.
- Fuzzy Rule Based Systems with NOP = 7 is found most accurate with accuracy of 99.7% which is an excellent improvement compairing to other models.

4. **Conclusion and Future Scope**

Machine learning tools are extremely utilized in desiring classifier systems. On the other hand, Fuzzy Rule Based Systems are also applicable to solve classification problems. In this paper, Fuzzy Rule Based Systems are found more competive compared to Machine learning models. Cross Validation is also used to doing more reliable systems.

In future the authors, are interest to doing more accurate classifier systems using type-2 and interval type-2 fuzzy systems.

**References**

[1] C. Molnar, *Interpretable Machine learning*, Leanpub, (2020).

[2] T. Rymarczyk, E. Kozlowski, G. Klosowski, and K. Niderla *Logistic regression for machine learning in process tomography*, Sensors, Vol. **19**, No.15. (2019).

[3] A.R.Chowdhury, T. Chatterjee and S. Banerjee, *Random forest classifier based approach in the detection of abnormalities in the retina*, Medical & Biological Engineering and computing ,Vol. **57**, No.15.(2019) 193-203.

[4] A. A. Bataineh, *A comparative analysis of nonlnear machine learning algorithms for breast cancer detection*, International Journal of Machine Learning and Computing, Vol. **9**, No.3 (2019).

[5] F. Jauhari and A. A. Supianto, *Building Student's performance decision tree classifier using boosting algorithms*, Vol. **14**, No.3. (2019) 1298-1304.

[6] T. Sathish, S. Rangarajan, A. Muthuram and R. Praveen Kumar, *Analysis and modelling of dissimilar materials needing based on K-nearest neighbour predictor*, Materials today Proceedings, Vol.**21**, Part 1, (2020), 108-112.

[7] G. Battineni, N. Chintalapudi and F. Amenta *Machine learning in medicine: performance calculation of dementia prediction by Support Vector Machine (SVM)*, Informatics in Medicin unlocked, Vol. **16**(2019).

[8] P. K. Shukla and S. P. Tripathi, *A Reviewed on the interpretability accuracy trade off in evolutionary multi objective fuzzy systems (EMOFS)*, Information, Vol. **3**, Issue 3(2012) 256-277.

[9] P. K. Shukla and S. P. Tripathi, *Interpretability issues in evolutionary multi objective fuzzy knowledge base systems*, Proceedings of Seventh International Conference on Bio Inspired computing: Theories and Applications (BIC-TA2012), 473-484.

[10] P. K. Shukla and S. P. Tripathi, *Handling high dimensionality and interpretability accuracy trade off issues in evolating multi objective fuzzy classifiers,*, International Journal of Scientific and Engineering Research, Vol.**5**,No.6 (2014) 665-671.

[11] P. K. Shukla and S. P. Tripathi, *Interpretability and accuracy issuing in evolating multi objective fuzzy classifiers*, International Journal of Soft Computing and Networking, Vol.**1**,No.1 (2014) 55-69.

[12] https://www.Kaggle.com/uciml/pima - indians - diabetes - databse, accessed on 30.06.2020