A MLP Prediction Model with Reduced Feature Complexity for Effective Rainfall and Crop Water Needs

Rupinder Kaur
Research Scholar
Guru Gobind Singh College of Modern Technology, Kharar
Punjab
itsrupinderkaur@gmail.com

Harpreet Kaur
Assistant Professor, CSE Department
Guru Gobind Singh College of Modern Technology, Kharar
Punjab
Billing02harpreet@gmail.com

Abstract— **The agriculture plays an important role to enhance economy of a country. In country like India, the half population survives on the agriculture as a profession or a source of income. In order to enhance the growth of crops, the effective rainfall and suitable climate is mandatory. Thus a large number of researches has been conducted till now to develop the effective recommendation system to predict the rainfall in the region to enhance the growth of the crops.**

This study implements the crop water needs and rainfall prediction system on the basis of the dataset that has been gathered online. The dataset is collected for Indian states and for foreign state as well. The FCM is used for cluster formation and MLP classifier is used for classification of the clustered dataset. The simulation is done in MATLAB and a comparison analysis is done among proposed and traditional work i.e. J48, Naive Bayes and Regression. The results prove the FCM-MLP as the most suitable and efficient technique for data mining based recommendation system for crop water needs and rainfall prediction.

Keywords—Data Mining, Recommendation System, Fuzzy C Mean Clustering, Multilayered Perceptron Classification.

I. INTRODUCTION

The soil has a specific water ingestion rate and moisture holding ability. The efficient fraction of rainfall obtained, enhancing run-off and decreasing infiltration are decreased generally through the higher quantities and rainfall intensities. Equivalently, the amount of efficient rainfall achieved reduced through the rough distribution whereas this is improved by an even distribution. Comparative to the heavy downpours to the crop development the well distributed rainfall in frequent light showers is much favorable. Therefore, In order to establish the efficient rainfall the quantity of frequency distribution on a region with time of rainfall is necessary. In the year 2016 for specific days the efficient rainfall has been calculated and is utilized in order to establish the crop water requirements in the data set for the initial classifier. Under the terms of abundant water supply the Potential evapotranspiration is principally managed through the evaporative demand. From the mutual effect of four parameters that are temperature, radiation, wind velocity and humidity an estimation of the evaporative demand can be achieved. The evaporation would be improved with the enhancement in the first third and reduction in the fourth parameter. That circumstance can direct to superior shortages of moisture content in the soil and then the ratio of efficient rainfall enhanced as a total. Comparative to the entire rainfall the mean monthly significances of temperature, radiation, wind velocity and humidity illustrate minimal variation from year to year. Through the Blaney Criddle mechanism the evapotranspiration is computed and in order to discover an uneven approximation of the efficient rainfall this is surrogated in the formula.

II. PROBLEM FORMULATION

The existing method was created with the aim to overcome the obstacles such as the first one is to find the effective amount of rainfall, the second being to use the effective rainfall to find the irrigation water required and third to suggest suitable irrigation systems that should be implemented by the farmers to increase crop productivity. The problem in the exiting technique is that the classifier which is used is less effective and tends to enhance the mean square error. Additionally, no reduction technique has been introduced which can increase the size of data. The problems in the existing technique insisted to propose a new technique which can overcome the issues present in the current technique.

III. PROPOSED WORK

Considering the issues of the traditional work, a new approach is proposed in this work. Initially, clustering algorithm to reduce the data size is introduced which can enhance the prediction accuracy rate. The Fuzzy c-mean clustering approach is used for the purpose of clustering. Additionally, MLP classifier is used for the classification purpose. The MLP classifier is an effective classifier than other classifiers. Thus, the proposed technique can be proven to be an effective and efficient approach in terms of prediction accuracy rate and less error in identification.

The flow or process of proposed work is as follows:

1. **Start**
2. **Data Preprocessing:** The first steps in the proposed work are to gather the dataset from online resources. The parameters in the dataset define the weather climate of the region. The dataset for the implementation of proposed work is gathered from the following links:

   https://www.kaggle.com/rtatman/did-it-rain-in-seattle-19482017/home
   https://github.com/sayalidzambre/CaseStudy

   After creating the dataset, the next step is to remove the ambiguities from the dataset. The ambiguities like miss-matching of values within the same parameter, to reset the null values etc is removed to make it more sensible and useful for deriving the efficient results.

3. **Cluster Formation:** After pre-processing the data, the next step is to create the clusters by using the available dataset. In proposed work, the Fuzzy C-Mean technique is used for cluster formation. This mechanism is used for cluster formation due to its various advantages such as it can handle the large dataset with less complexity by dividing the dataset to the multiple clusters. Fuzzy C-Mean is a clustering technique in which one dataset can relate to more than one clusters. This technique is efficient technique foe pattern recognition. This technique was developed by Dunn in 1973 and enhanced by Bezdek in 1981. The technique has to initialize the objective function firstly by using the following equation:

$$J_m = \sum_{i-1}^{N}\sum_{j-1}^{C} u_{ij}^{m} \left\|x_i - c_j\right\|^2 \qquad 1 \le n < \infty \ldots..(1)$$

   In this equation m is a real number greater than 1;

   u_{ij} Defines the degree of membership of x_i in cluster which is defined by j_i.

Feature Reduction: Next step is to reduce the features by using the FCM. To reduce the features the dataset is divided to the 2 clusters and 2 columns. After this the training and testing of the dataset is performed. For the purpose of training, the training of dataset is done and then this trained dataset is passed to the MLP (Multi Layered Perceptron) system to train the system for classification purpose. The MLP is used for classifying the data. The MLP is a system is a type of feed forward artificial neural network. It utilizes the back propagation mechanism for training purpose. MLP is the simplest and easy to implement type of Artificial Neural Network (ANN). In MLP, the flow of information is in single direction i.e. the data is forwarded to the hidden nodes and then output nodes from input nodes. It did not comprise of loops or iterations. The advantage of MLP is that it can resolve the non-linear and stochastic problems effectively.

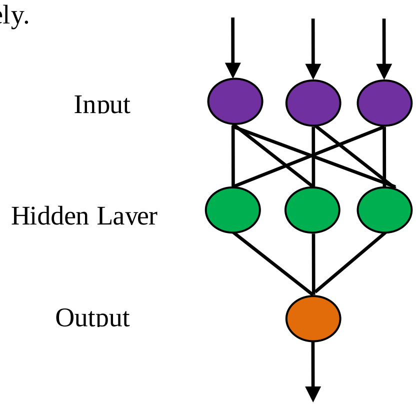

Figure 1 Proposed MLP Architecture

4. **Classification of dataset:** After training the MLP system, the classification of the trained dataset is done to classify the category or decision that is driven at last.
5. **Performance Evaluation:** After classification, the performance of the system is evaluated in the terms of following result matrices:
   a. **RMSE:** The RMSE refers to the Root Mean Square Error. The following formulation is used for evaluating the RMSE. It refers to the larger absolute error thus it is mandatory to have lower RMSE.

$$RMSE = \sqrt{\frac{1}{N}\sum_{u,i} |p_{u,i} - r_{u,i}\,|} \ldots \ldots (2)$$

   Where $p_{u,i}$ denotes the predicted rating for user, $r_{u,i}$ depicts the actual rating and N refers to the total number of rating on the available dataset.

   b. **MSE:** MSE is mean square error that is a performance matrix which is used to measure the mean square error in the observed output. The MSE should be always low to assure the quality of the observed results or output. The following is the formulation of MSE:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(AC - PR)^2 \ \ldots \ldots (4.3)$$

   Where AC refers to the Actual Results and PR refers to the Predicted Results. The N is the total number of samples.

   c. **F-Measure:** F-Measure is a performance matrix that is used to evaluate the harmonic mean of precision and recall. The formulation is as follows:

$$FMeasure = 2 * \frac{Precision * Recall}{(Precision + Recall)} \ldots \ldots (4.4)$$

The figure 2 depicts the flowchart for the proposed work.

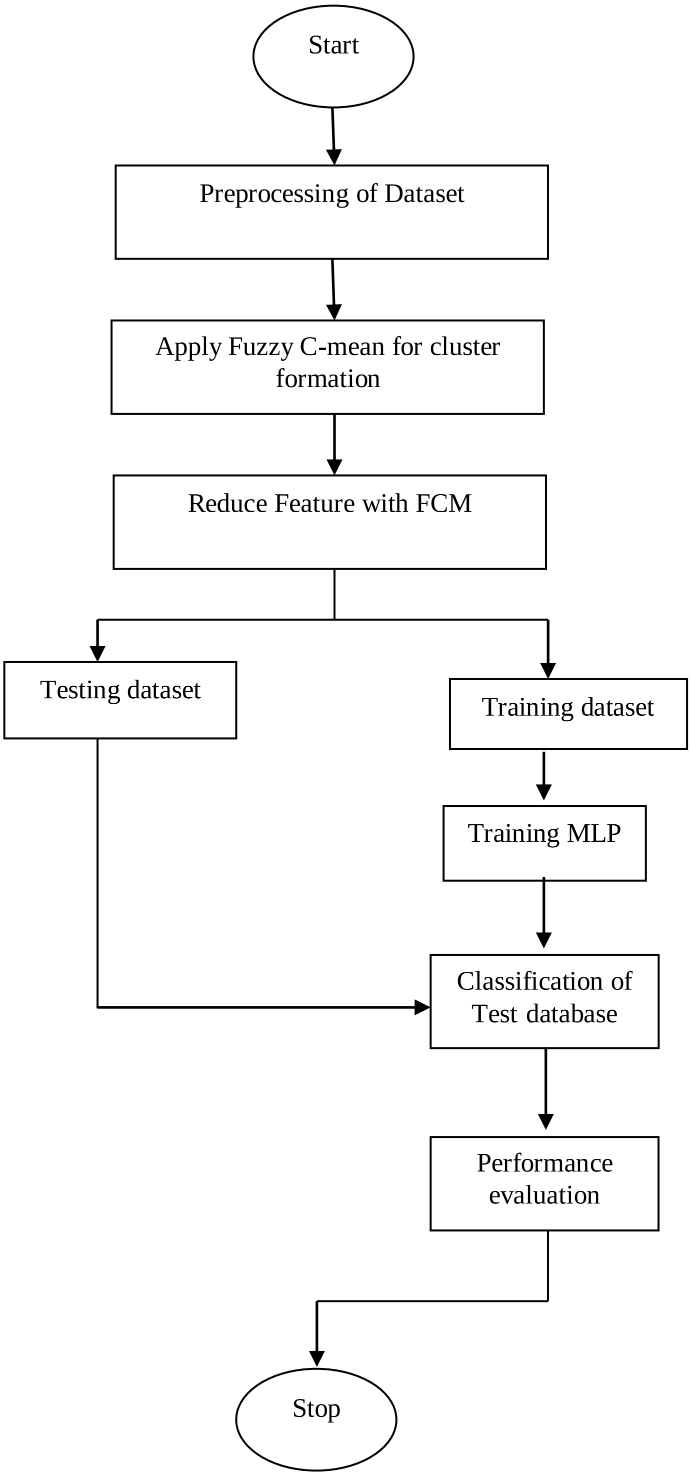

Figure 2 Framework of Proposed Work

IV. RESULTS

In this study a novel approach is developed to predict the effective rainfall for Rain Irrigation and Crop water by using the fuzzy C-mean for clustering and MLP for classification of the dataset. This section is organized to have a review to the results that are observed after implementing the proposed work in MATLAB. The result analysis is done by considering three different scenarios on the basis of various climatic conditions with respect to the dataset of different states of India. Along with this the analysis is performed in two different scenarios on these collected dataset.

5.1 Effective Rain Prediction Analysis

In present work, the Effective rainfall prediction analysis is done by considering the rainfall based dataset of Himachal Pradesh and Andaman Nicobar island of India. Along with this the dataset of foreign climatic condition in terms of rainfall of previous years is also considered. This section represents the results for rainfall prediction done in present work.

Case Studies

The first case study is derived for the Seattle dataset for rainfall prediction. The graph in figure 3 and 4 depicts the clusters formed by the proposed fuzzy C-mean clustering technique. In figure 3 the graph states the cluster 1 and in graph of figure 4 states the cluster 2.

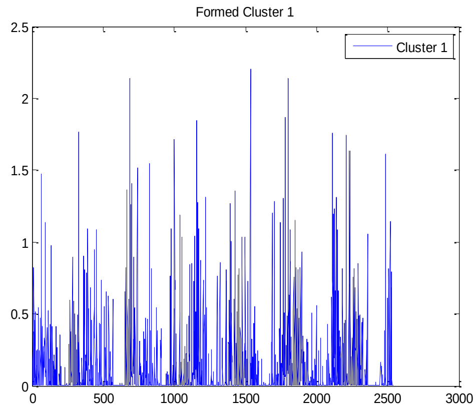

Figure 3 Cluster1 formed by FCM

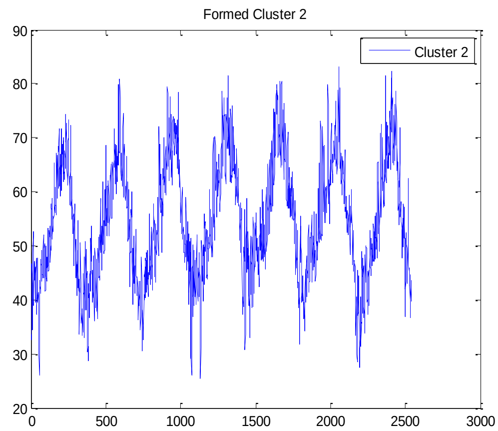

Figure 4 Cluster1 formed by FCM

The graph in figure 5 shows the combined cluster values on the basis of the observations of the graph of figure 3 and figure 4. The lines in red color depict the cluster 2 and the lines in blue color shows the cluster1.

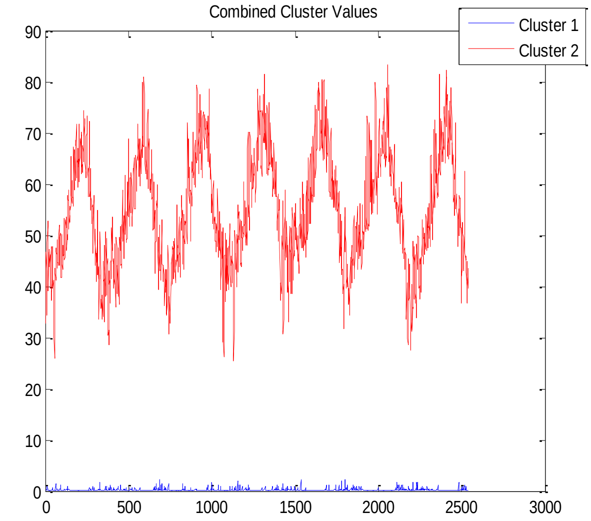

Figure 5 Combined Cluster Representation of FCM

The graph in figure 6 shows the MSE of proposed work. The MSE is evaluated on the basis of the training epochs that is represented by axis x and ranges from 0 to 2000. The y axis in the graph calibrates the data in the form of MSE.

As discussed in above section, it is essential to have lower MSE for an ideal system, the graph explains that initially the MSE of the proposed work is little higher but then with the increment in the number of training epochs the MSE started falling and till reaching to the 2000 epochs the observed MSE is quite lower.

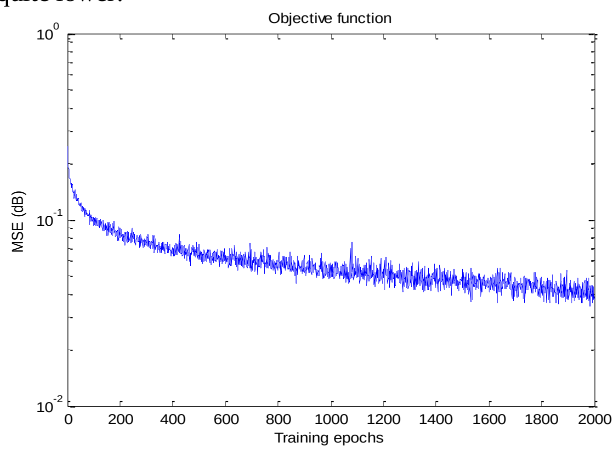

Figure 6 MSE of proposed work (Case Study 1)

The graph in figure 7 represents the classification performance of the proposed work with respect to the trained dataset. The classification performance is measured in the form of accuracy. The accuracy defines the exactness of the observed results. The highly accurate result proves the efficiency of the system. The accuracy is evaluated on the basis of the training epochs i.e. 0 to 2000. The y axis defines the data in the form of classification accuracy and the value ranges from 0.5% to 1%. The graph proves that initially the accuracy was low but at the end of the epochs the accuracy is evaluated nearby 0.95%.

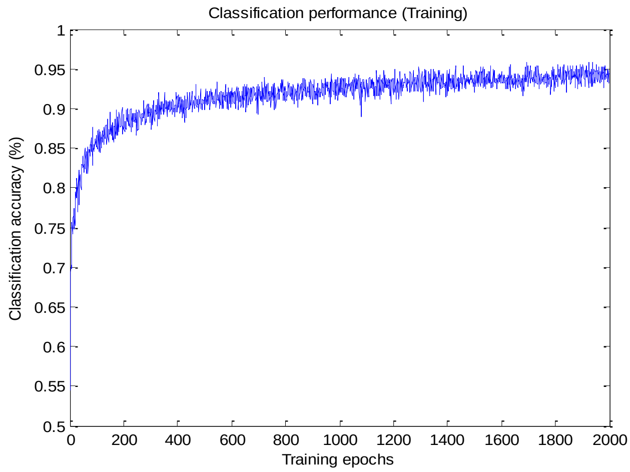

Figure 7 Classification Accuracy of proposed work (Case Study 1)

The graph in figure 8 delineates the comparison of various classification techniques. The comparison is done in the terms of Root Mean Square Error (RMSE) for rain irrigation. The comparison is done among various classification algorithms i.e.j458, Naive Bayes, Regression and proposed FCM-MLP mechanism. The bar in red color depicts the RMSE of J48 technique, the bar in blue refers to the RMSE of Naive bayes, and the bar in black refers to the RMSE of regression and the bar in green color shows the RMSE of proposed work. The RMSE of proposed work is quite effective and lower than the RMSE of traditional techniques.

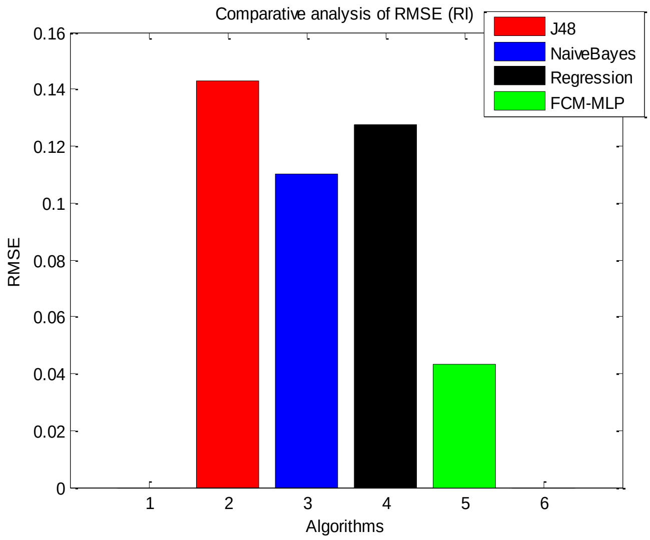

Figure 8 RMSE Analysis for RI(Case Study 1)

The graph in figure 9 depicts the comparison analysis of proposed and other classification techniques in the terms of accuracy. The analysis is done for RI. The accuracy is parameters that defend the efficiency of the work. If the accuracy is high the work is quite efficient and if the accuracy is low the work is not effective and is not capable to generate the exact or reliable results. The accuracy of the proposed work is observed to be higher in contrast to other classification techniques.

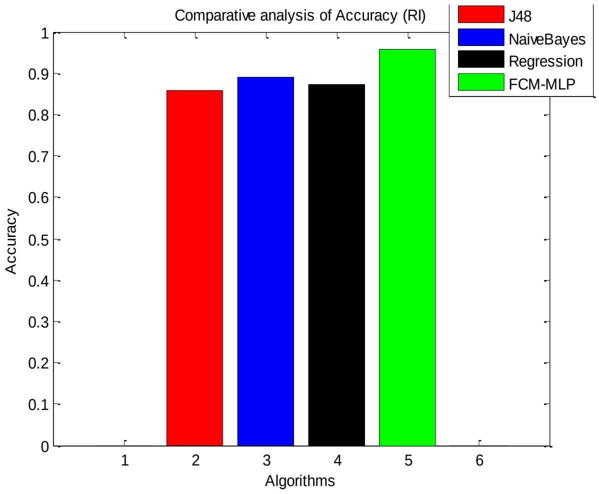

Figure 9 Accuracy Analysis for RI (Case Study 1)

The graph in figure 10 shows the F-Measure of the various classification technique and proposed technique. The comparison is done for RI. The x axis in the graph shows the algorithms that are considered for comparison and the y axis in the graph gathers the value of F-Measure that varies from 0 to 1. The F-measure of the proposed work is better than the F-measure of the other classification techniques.

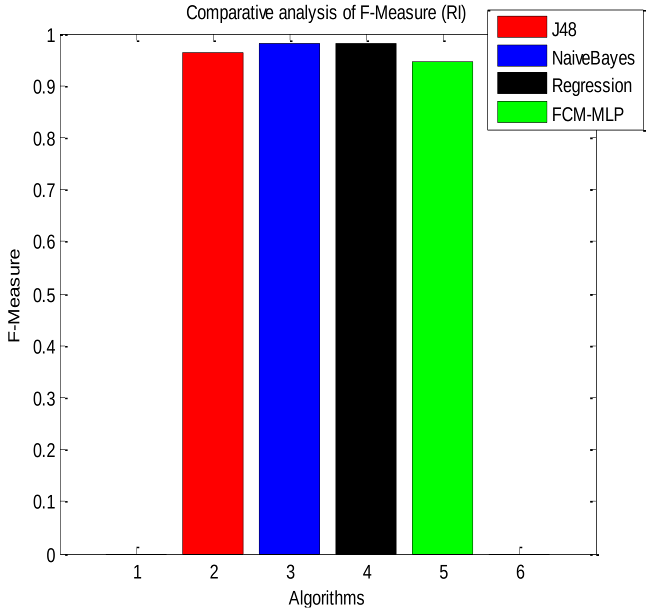

Figure 10 F-Measure Analysis for RI

Table 1 Performance analysis for Case Study 1

Techniques	RMSE	Accuracy	F-Measure
J48	0.1429	0.8571	0.9640
Naive Bayes	0.1103	0.8897	0.9820
Regression	0.1276	0.8724	0.9820
FCM-MLP	0.0432	0.9568	0.9469

Table 1 shows the facts and figures that are observed from the above comparison graphs. Similarly the analysis has been done for rest of the two case studies. The observations these case studies has been shown in table 2 and 3.

Table 2 Performance analysis for Case Study 2

Techniques	RMSE	Accuracy	F-Measure
J48	0.1429	0.8571	0.9640
Naive Bayes	0.1103	0.8897	0.9820
Regression	0.1276	0.8724	0.9820
FCM-MLP	0.0909	0.9091	0.9524

Table 3 Comparison analysis of Case Study 3

Techniques	RMSE	Accuracy	F-Measure
J48	0.1429	0.8571	0.9640
Naive Bayes	0.1103	0.8897	0.9820
Regression	0.1276	0.8724	0.9820
FCM-MLP	0.1000	0.9000	0.9474

5.2 Crop Water Analysis

This section depicts the results that are obtained after implementing the proposed work for Crop water requirement over collected dataset. The graph in figure 11and 12 depicts the graph for cluster 1 and cluster 2 that are formed by proposed work for the analysis of crop water required for irrigation. The x axis in the graph shows the values from 0 to 60 and y axis shows the values from 0 to 2000. Similarly, the graph in figure 13 shows the combination of both clusters.

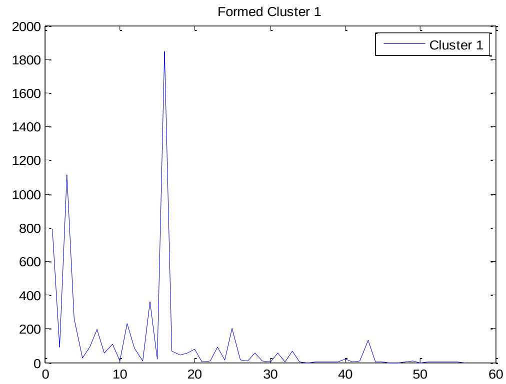

Figure 11 Cluster1 formed by FCM (Crop Water Analysis)

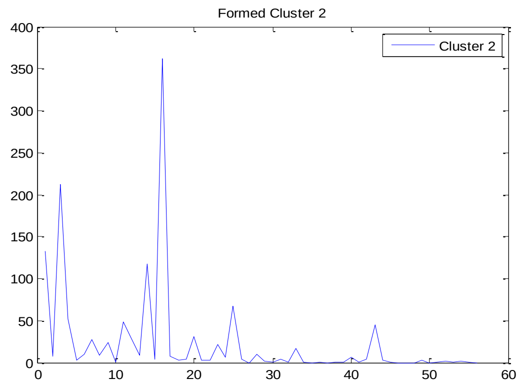

Figure 12 Cluster 2 formed by proposed work (Crop Water Analysis)

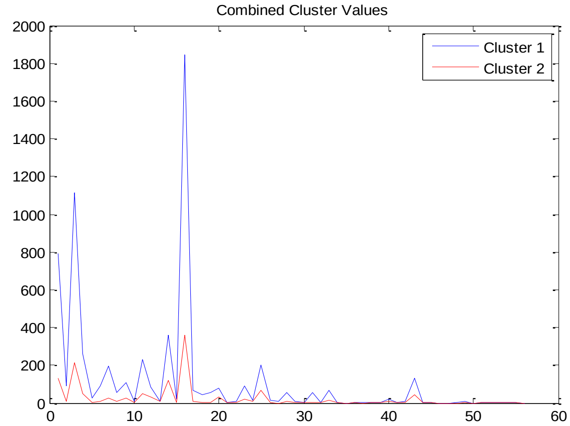

Figure 13 Combined clusters formed by propose work (Crop Water Analysis)

The graph in figure 14 depicts the MSE evaluation for crop water analysis. The MSE is evaluated on the basis of the number of training epochs. The evaluated MSE is quite lower that ensures the proficiency of proposed work.

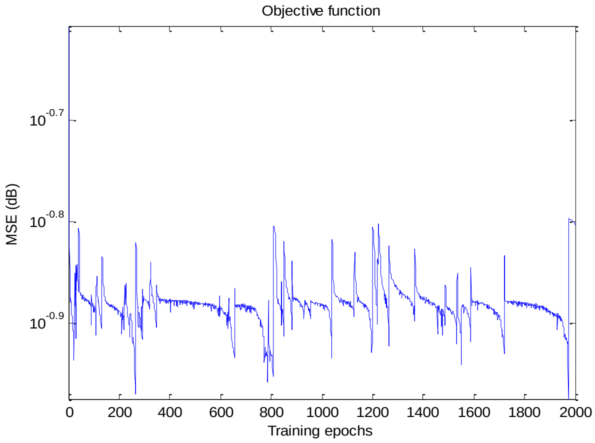

Figure 14 MSE of propose work (Crop Water Analysis)

The graph in figure 15 depicts the classification accuracy of proposed work. It is also evaluated to be higher.

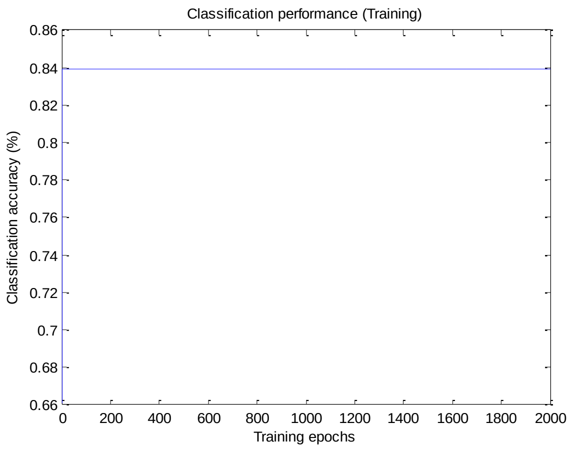

Figure 15 Classification Accuracy of propose work (Crop Water Analysis)

The graph in figure 16 shows that the RMSE of proposed work is higher in comparison to the traditional techniques. The bar in red color depicts the results of J48, the bar in blue depicts the RMSE of Naive Bayes, the bar in black depicts the RMSE of regression and the bar in green depicts the RMSE of proposed work i.e. FCM-MLP. The graph shows that the regression mechanism has the higher RMSE which reduces the performance quality. Consequently, the RMSE of proposed work is found to be quite lower in comparison to all the traditional techniques that are considered for the analysis purpose.

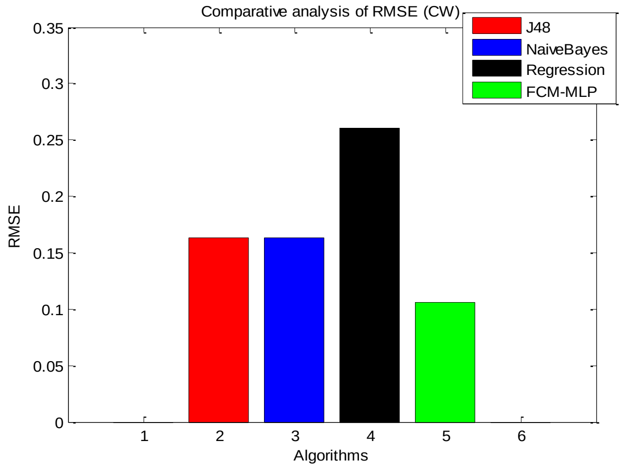

Figure 16 Comparison of RMSE for Crop Water

The graph in figure 17 draws the comparison of proposed and traditional techniques in the terms of accuracy. The comparison analysis is done among J48, Naive Bayes, Regression and FCM-MLP. The observations from the graph prove that the accuracy of proposed work is higher than the accuracy of the traditional works.

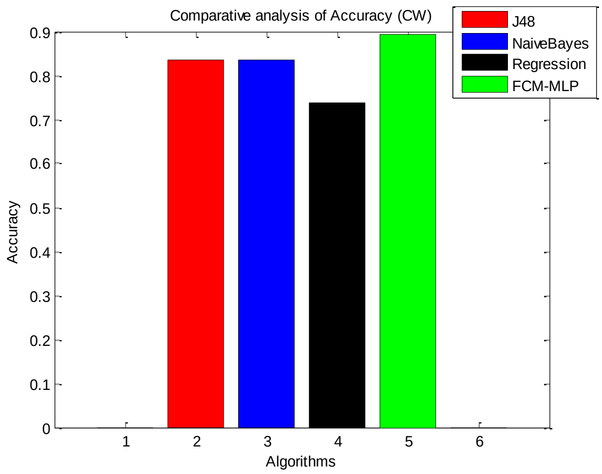

Figure 17 Comparison of Accuracy for Crop Water

Table 4 Comparison Analysis for Crop Water

Techniques	RMSE	Accuracy
J48	0.1633	0.8367
Naive Bayes	0.1635	0.8395
Regression	0.2608	0.7392
FCM-MLP	0.1064	0.8936

V. CONCLUSION

To sum up, this study develops a novel recommendation system for rainfall and crop water needs prediction on the basis of the climatic conditions of the respective area. For the purpose of gathering the historic data related to the climatic whether, the online dataset is gathered and further processing is performed over it. The clustering is done by applying the FCM cluster formation technique. The clusters are formed by using the parameters that are considered in dataset. The MLP classifier is applied to train the classification system for the trained dataset. The results that are observed after implementing the FCM-MLP proves that the proposed technique outperforms the traditional Naive Bayes, J48 and regression mechanism in the terms of RMSE, F-Measure and Accuracy.

As the results observed from the simulation, the FCM-MLP is found to be quite effective and efficient but still more amendments are possible in this work to make it more reliable and efficient. In order to make the amendments the work can be done on multiple features and hybrid classification as well.

REFERENCES

[1] B. Abishek; R. Priyatharshini; M. Akash Eswar; P. Deepika,"Prediction of effective rainfall and crop water n eeds using data mining techniques",TIAR, Pp 231 – 235, 2017.

[2] C. Piruthevi; C. S. Kanimozhi Selvi, "Filtering of anomalous weather events over the region of Tamil Nadu", INCOS, Pp 1-7, 2017.

[3] K. L. Ponce-Guevara; J. A. Palacios-Echeverría; E. Maya-Olalla; H. M. Domínguez-Limaico; L. E. Suárez-Zambrano; P. D. Rosero-Montalvo; D. H. Peluffo-Ordóñez; J. C. Alvarado-Pérez, "GreenFarm-DM: A tool for analyzing vegetable crops data from a greenhouse using data mining techniques (First trial)", ETCM, Pp 1-6, 2017.

[4] Kuljit Kaur; Kanwalpreet Singh Attwal, "Effect of temperature and rainfall on paddy yield using data mining", ICCCDSEC, Pp 506-511, 2017.

[5] Sandeep Kumar Mohapatra; Anamika Upadhyay; Channabasava Gola, "Rainfall prediction based on 100 years of meteorological data", ICCCTSN, Pp 162-166, 2017.

[6] Shuangxi Miao; Qing Zhu; Linlin Zhang; Chang Liu; Bo Zhang; Min Chen, "A knowledge-guided landslide deformation prediction approach based on SVR", ICG, Pp 1-5, 2017.

[7] R.Senthil Kumar; C. Ramesh, "A study on prediction of rainfall using datamining technique", ICICT, Volume 3, Pp 1-9, 2016.

[8] R.Beulah; M. Punithavalli, "Prediction of sugarcane diseases using data mining techniques", ICACA, Pp 393-396, 2016.

[9] R.Senthil Kumar; C. Ramesh," A study on prediction of rainfall using data mining technique", ICICT, Volume 3, Pp 1-6, 2016.

[10] R.Sujatha; P.Isakki, "A study on crop yield forecasting using classification techniques", ICCTIDE, Pp 1-4, 2016.

[11] Rupanjali D. Baruah; Sudipta Roy; R. M. Bhagat; L. N. Sethi, "Use of Data Mining Technique for Prediction of Tea Yield in the Face of Climate Change of Assam, India", ICIT, Pp 265-269, 2016.

[12] A. Geetha; G. M. Nasira, "Data mining for meteorological applications: Decision trees for modelingrainfall prediction", CCICR, Pp 1-4, 2014.

[13] Andrew Kusiak; Xiupeng Wei; Anoop Prakash Verma; Evan Roz, "Modeling and Prediction of Rainfall Using Radar Reflectivity Data: A Data-Mining Approach", GRS, Volume 51, Issue 4, Pp 2337-2342, 2013.

[14] Valmik B Nikam; B. B. Meshram, "Modeling Rainfall Prediction Using Data Mining Metho d: A Bayesian Approach", ICCIMS, Pp 132-136, 2013.

[15] Dingsheng Wan; Yaming Wang; Nan Gu; Yufeng Yu, "A novel approach to extreme rainfall prediction based on data mining", ICCSNT, Pp 1873-1877, 2012.