ISSN: 0974-8571 Vol.13 No. 2 December, 2021

International Journal of Computational Intelligence in Control

Enhanced Self-Organizing Map Clustering Method based Feature Selection for Rainfall Prediction

A. MARY SUBASHINI

Research Scholar, PG and Research Department of Computer Science, Jamal Mohamed College (Autonomous) (Affiliated to Bharathidasan University), Tiruchirappalli, Tamilnadu, India.

Dr. M. RAJAKUMAR

Assistant Professor, PG and Research Department of Computer Science, Jamal Mohamed College (Autonomous) (Affiliated to Bharathidasan University), Tiruchirappalli, Tamilnadu, India.

Date of Submission: 15th August 2021 Revised: 27th October 2021 Accepted: 18th December 2021

How to Cite: **A. MARY SUBASHINI et.al.(2020). Enhanced Self-Organizing Map Clustering Method based Feature Selection for Rainfall Prediction. International Journal of Computational Intelligence in Control 12(2),**

Abstract - **Heavy rainfall prediction is a major problem for meteorological department as it is closely associated with the economy and life of human. It is a cause for natural disasters like flood and drought which are encountered by people across the globe every year. Accuracy of rainfall forecasting has great importance for countries like India whose economy is largely dependent on agriculture. Prediction of rainfall gives awareness to people and know in advance about rainfall to take certain precautions to protect their crop from rainfall. Many techniques came into existence to predict rainfall. Machine Learning algorithms are mostly useful in predicting rainfall. In this research work, an enhanced clustering method is proposed to find the most pre-dominant feature subset for the efficient prediction of the rainfall. In this work, Self- Organizing Map (SOM) based Neural Network and Rice Distribution is combined to find the most pre-dominant features in the rainfall prediction dataset. The feature set obtained by the proposed ECM method is evaluated with the existing feature selection techniques using various evaluation metrics like Accuracy, True Positive Rate, False Positive Rate, Precision, Miss Rate and Specificity using three different classifiers.**

Index Terms - Rainfall prediction, classification, clustering based feature selection, Neural Network.

INTRODUCTION

India's welfare is agriculture. The achievement of agriculture is dependent on rainfall. It also helps with water resources [1]. Rainfall information in the past helps farmers better manage their crops, leading to economic growth in the country. Prediction of precipitation is beneficial to prevent flooding that saves people's lives and property. Fluctuation in the timing of precipitation and its amount makes forecasting of rainfall a problem for meteorological scientists. Forecasting is one of the utmost challenges for researchers from a variety of fields, such as weather data mining, environmental machine learning, functional hydrology, and numerical forecasting, to create a predictive model for accurate rainfall [2]. In these problems, a common question is how to infer the past predictions and make use of future predictions. A variety of sub-processes are typically composed of the substantial process in rainfall. It is at times not promising to predict the precipitation correctly by on its global system. Climate forecasting stands out for all countries around the globe in all the benefits and services provided by the meteorological department [3]. The job is very complicated because it needs specific numbers and all signals are intimated without any assurance. Accurate precipitation forecasting has been an important issue in hydrological science as early notice of stern weather can help avoid natural disaster injuries and damage if prompt and accurate forecasts are made. The theory of the modular model and the integration of different models has recently gained more interest in rainfall forecasting to address this challenge. A huge range of rainfall prediction methodologies is available in India [4][5]. In India, there are two primary methods of forecasting rainfall. Regression, Artificial Neural Network (ANN), Decision Tree algorithm, Fuzzy logic and team process of data handling are the majority frequently used computational methods used for weather forecasting. The basic goal is to follow information rules and relationships while gaining intangible and potentially expensive knowledge. Artificial NN is a promising part of this wide field. Through this work, an efficient neural network clustering-based feature selection method is proposed to enhance the prediction of the rainfall.

RELATED WORKS

Grace, R. Kingsy, and B. Suganya [6] proposed a rainfall prediction model using Multiple Linear Regression (MLR) for Indian dataset. The input data is having multiple meteorological parameters and to predict the rainfall in more precise. MSE, RMSE, and Correlation are the metrics used in this paper.

Pham, Binh Thai, et al [7] studied the main objective is to develop and compare several advanced Artificial Intelligence (AI) models namely Adaptive Network based Fuzzy Inference System with Particle Swarm Optimization (PSOANFIS), Artificial Neural Network (ANN) and Support Vector Machine (SVM) for the prediction of daily rainfall in Hoa Binh Province, Vietnam. Correlation Coefficient (R) and Mean Absolute Error (MAE), Score Skill (SS), Probability of Detection (POD), Critical Success Index (CSI) and False Alarm Ratio (FAR) are used as the performance evaluation metrics.

Ahmed, Kamal, et al [8] employed Multi-Model Ensembles (MMEs) to reduce the uncertainties related to GCM simulations/projections. The objective of this study was to evaluate the performance of MMEs developed using machine learning (ML) algorithms with different combinations of GCMs ranked based on their performance and determine the optimum number of GCMs to be included in an MME.

Le, Vuong Minh, et al [9] proposed a prediction model using Nonlinear Autoregressive Neural in order to forecast daily rainfall at Hoa Binh city, Vietnam. For this aim, eight-year time series of meteorological data were first collected, involving temperature, wind speed, relative humidity, solar radiation as input variables and daily rainfall as output variable Networks with external variables (NARX). The metrics like Correlation Coefficient, MAE, RSME, Error Mean, Median, STD are considered for evaluation.

Sardeshpande, Kaushik D., and Vijaya R. Thool [10] presented a case study on time series prediction as an application of neural networks. The case study was done for the rainfall prediction using the local database in India. The results were obtained by the comparative study of neural network architectures like back propagation (BPNN), generalized regression (GRNN), and radial basis function (RBFNN).

Singh, Nitin, Saurabh Chaturvedi, and Shamim Akhter [11] developed a weather forecasting system that can be used in remote areas is the main motivation of this work. The data analytics and machine learning algorithms, such as random forest classification, are used to predict weather conditions. In this paper, a low-cost and portable solution for weather prediction is devised.

Balan, M. Selva, et al [12] proposed a few statistical analysis techniques and the use of Artificial Neural Network to predict the rainfall. The correlation between the attributes defines the influence on the prediction of any system. Attributes with no correlation can be removed as they do not contribute to the activation of the neuron. The loss functions, activation functions, number of neurons and the number of hidden layers also affect the accuracy of the system.

Parashar, Anubha [13] The main aim of this paper is to monitor and report weather conditions so that one is informed beforehand and necessary actions can be taken to reduce the damage by any calamity by forecasting it. Here the authors are using various sensors in order to collect the data and previous data is used in order to train the system and with current data collection we do the prediction. Explained Variance, Mean Absolute Error, and Median Absolute Error are used as the evaluation metrics in this paper.

Kalteh, Aman Mohammad [14] proposed a rainfall forecasting method based on coupling wavelet analysis and a novel artificial neural network technique called extreme learning machine (ELM). In this way, the unique characteristics of each technique are combined to capture different patterns in the data. At first, wavelet analysis is used to decompose rainfall time series into wavelet coefficients, and then the wavelet coefficients are used as inputs into the ELM model to forecast rainfall. The correlation coefficient (r), root mean square errors (RMSE) and Nash–Sutcliffe efficiency coefficient (NS) statistics are used as the performance metrics.

Lathifah, Siti Nur, et al [15] used Classification and Regression Tree (CART) algorithm to forecast the rainfall in Bandung Regency. Furthermore, an Adaptive Synthetic Sampling (ADASYN) algorithm was added to optimize the model produced due to a class imbalance in the data. The performance metrics like Precision, Recall, Accuracy and F1-Score are used in this paper.

PROPOSED ENHANCED CLUSTERING METHOD (ECM) FOR FINDING PREDOMINANT FEATURE SET

Rice clustering with Self Organizing Map (SOM) is named in this work as ECM for rainfall dataset is introduced by evaluating the most pre-dominant features by find the relative strengths among the feature subset. Figure 1 depicts the flowchart of the proposed ECM for rainfall dataset.

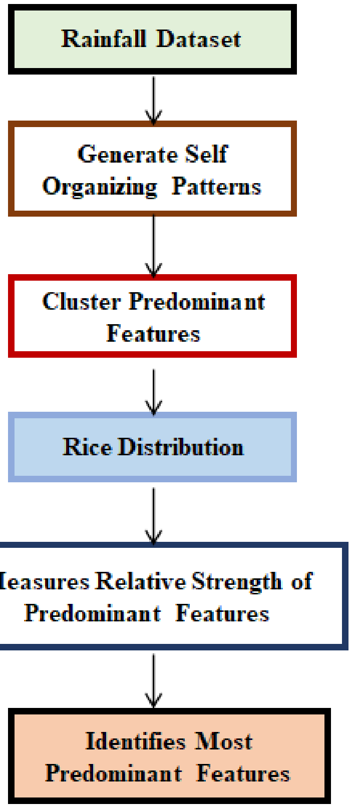

FIGURE 1
Proposed Enhanced Clustering Method (ECM) for rainfall dataset

I. Self-Organizing Map

The self-organizing map (SOM) in data mining that performs important steps in developing knowledge discovery in a huge textual as well as rainfall dataset. SOM is described like normally unsupervised neural-network process which develops a graph of similarities of data.

By using the large rainfall dataset, evaluating predominant patterns is one of the major topics being analyzed in dataset area. Though, the self-organizing map (SOM) was engaged for dataset-feature projection and dimensionality reduction, conventional SOM cannot represent similarity-based relativity because the SOM effort is only in a type of a flat vector.

The problem studied in this paper now is as follows. Let us assume a collection '$D = \{d_1, d_2, \ldots., d_n\}$' of '$n$' features from the rainfall dataset 'DS'. Each data 'D' comprises of a set of patterns '$p = p_1, p_2, .., p_n$'. It is the goal of this paper to study how to measure the relative strength of the predominant patterns of features and improve the execution time and computation space using Rice clustering.

II. Feature Pattern Generation based on SOM

Self-Organizing Pattern Generation derived by the Self Organizing Map (SOM) identifies a predominant pattern in the corresponding dataset through efficient mapping. The SOM comprises of an input vector and an output vector. The keywords in the input data set are presumed to be in a vector format. The input vector 'P_i' is linked with a weight 'w_{ij}' to obtain an output vector 'Q_i' and are formalized as shown below.

$$Q_i = w_{i1}P_1, w_{i2}P_2, \ldots, w_{in}P_n \quad (1)$$

Where 'i *and* j' represent the dimension of input vector ($j = 1,2, ..n$). Let us consider that the input space possesses 'n' dimension, then every feature on the map holds an n-dimensional vector of weights. The '$2 - n$' map with weight vectors is as given below.

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ \ldots & \ldots \\ w_{n1} & w_{nn} \end{bmatrix} \quad (2)$$

Self-Organizing Pattern Generation based on Self Organizing Map works as follows. When the training vectors (i.e. dataset) are provided as input, the input and output vectors are linked with a weight to obtain the output vector. Next, the Euclidean distance involving the weight vector 'w_{ij}' and the input training vectors 'P_i' are used to identify the patterns possessing the minimum distance and is formalized as given below.

$$R = MIN\,(P_i - w_{ij}) \quad (3)$$

Followed by this, the weights of the resultant vector and its neighboring neurons (i.e. patterns) are adjusted, to make them similar to the input dataset. As the volume of neurons generated is smaller than the number of items in the dataset, important patterns have to be preserved. So, the objective behind the design of SOM is to obtain a correlation between the similarities of items in the dataset and therefore identify the most feature predominant pattern in the corresponding rainfall dataset features and is mathematically formalized as given below.

$$w_{ij}^{(m+1)} = w_{ij}^{m} + \eta\,(m)\left(P - w_{ij}^{m}\right) \quad (4)$$

Here '$\eta\,(m)$' is the learning rate parameter for iteration 'm'. The neurons 'ij' are connected to each other by Euclidean distance relation [18]. This Euclidean distance relation determines the feature predominant pattern in the corresponding dataset.

III. Rice Distribution Method

As explained in the previous section, SOM model cannot represent similarity-based relativity, which may increase the execution time for clustering. Thus, improves relative strength. Then, performs semantic-based relativity and employ a Rice distribution model is to incorporate efficient rainfall data clustering.

The Rice distribution model is applied to the clustered similar feature pattern dataset to list out the relative strength of predominant pattern for each dataset data. If 'P' and 'Q' are the two features of dataset 'P' and 'Q', then the similarity is evaluated as given below.

$$Sim\ (P, Q) = \frac{w(p_i,P)*w(p_i,Q)}{\sqrt{\Sigma(wp_i,P)^2}*\sqrt{(wp_i,Q)^2}} \quad (5)$$

Where '$w(p_i, P)$' represent the weight of pattern 'p_i' from pattern set 'P'. The distance measure of feature pattern similarity to relative strength of the predominant features patterns of dataset is calculated using the Rice Distribution model. With relative distance measure values, Rice clustering process is formulated as given below.

$$w(p_i, P) = w_i\ (p_i^P, P) * Rel\ (p_i^P, p_i) \quad (6)$$

From (6), by identifying the relative distance values '$Rel\ (p_i^P, p_i)$', the efficiency of the dataset cluster is improved notably.

Algorithm: Identifying Feature Predominant Patterns from the rainfall dataset

Input: Dataset 'DS', Data '$D = \{d_1, d_2, \ldots, d_n\}$', Pattern '$K = k_1, k_2, .., k_n$', weight vector '$w_{ij}$', input training vector 'P_i', output vector 'Q_i'

Step 1: Begin

Step 2: For each 'DS' and Data 'D'

Step 3: Obtain the weights of the patterns 'K' using (1)

Step 4: Calculate the Euclidean distance between weight vector and input training vectors using (3)

Step 5: Adjust the weights (obtained using 4) to determine the feature predominant pattern

Step 6: For each feature predominant pattern 'p_i' from the Pattern set 'P'

Step 7: Search for 'p_i' from dataset 'DS'

Step 8: If 'p_i' if found then

Step 9: Identify strength of the pattern of 'p_i' and similarity using (5)&(6)

Step 10: End for

Step 11: End for

Step 12: End.

Output: Feature Predominant Pattern

RESULT AND DISCUSSION

The Indian rainfall dataset is considered for this research work from Kaggle repository [19]. The performance of the proposed ECM is evaluated with the following existing feature selection methods like Differential Evolution, and Information Gain using the performance metrics considered in the table 1. The three different classifications like Random Forest (RF), Artificial Neural Network (ANN) and Naïve Bayes (NB) [16] [7] [18] are used to evaluate the performance of the proposed ECM and existing feature selection methods.

TABLE I

Performance Metrics used in this research work

Performance Metrics	**Equation**
Accuracy (in %)	$\frac{TP+TN}{TP+FN+TN+FP}$
True Positive Rate (in %)	$\frac{TP}{TP+FN}$
False Positive Rate (in %)	$\frac{FP}{FP+TN}$
Precision (in %)	$\frac{TP}{TP+FP}$
Miss Rate (in %)	1-TPR
Specificity (in %)	1-FPR

Table 2 depicts the classification accuracy (in %) obtained by the proposed ECM, DE and IG based feature selection methods using RF, ANN and NB classification methods. Figure 2 gives the graphical representation of the classification accuracy (in %) obtained by the proposed ECM, DE and IG based feature selection methods using RF, ANN and NB classification methods. From the table 2 and figure 2, it is clear that the proposed ECM gives better accuracy with ANN classifier than other feature selection methods and classifiers.

TABLE 2: Classification Accuracy (in %) obtained by Proposed ECM, IG, DE based feature selection methods using ANN, RF and NB classification methods

Feature Selection Methods	**Classification Accuracy (in %) obtained by Classification Methods**		
	ANN	**RF**	**NB**
Original dataset	45.65	43.1	41.85
Information Gain	57.05	53.25	51.2
Differential Evolution	59.9	56.64	54.85
Proposed ECM	75.25	65.87	63.36

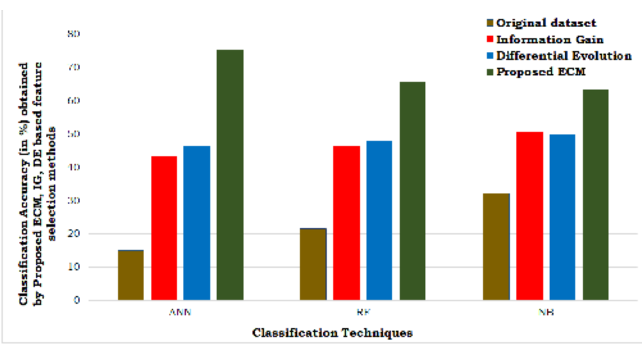

FIGURE 2: GRAPHICAL REPRESENTATION OF THE CLASSIFICATION ACCURACY (IN %) OBTAINED BY PROPOSED ECM, IG, DE BASED FEATURE SELECTION METHODS USING ANN, RF AND NB CLASSIFICATION METHODS

Table 3 depicts the True Positive Rate (in %) obtained the proposed ECM, DE and IG based feature selection methods using RF, ANN and NB classification methods. Figure 3 gives the graphical representation of the True Positive Rate (in %) obtained the proposed ECM, DE and IG based feature selection methods using RF, ANN and NB classification methods. From the table 3 and figure 3, it is clear that the proposed ECM gives better TPR with ANN classifier than other feature selection methods and classifiers.

TABLE 3: True Positive Rate (in %) obtained by Proposed ECM, IG, DE based feature selection methods using ANN, RF and NB classification methods

Feature Selection Methods	True Positive Rate (in %) obtained by Classification Methods		
	ANN	**RF**	**NB**
Original dataset	54.62	50.46	49.51
Information Gain	60.77	57.75	56.78
Differential Evolution	63.21	60.311	59.18
Proposed ECM	89.53	67.75	64.92

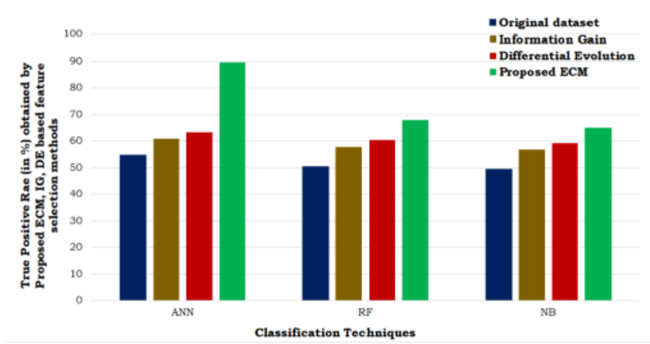

FIGURE 3: GRAPHICAL REPRESENTATION OF THE TRUE POSITIVE RATE (IN %) OBTAINED BY PROPOSED ECM, IG, DE BASED FEATURE SELECTION METHODS USING ANN, RF AND NB CLASSIFICATION METHODS

Table 4 depicts the False Positive Rate (in %) obtained the proposed ECM, DE and IG based feature selection methods using RF, ANN and NB classification methods. Figure 4 gives the graphical representation of the False Positive Rate (in %) obtained the proposed ECM, DE and IG based feature selection methods using RF, ANN and NB classification methods. From the table 4 and figure 4, it is clear that the proposed ECM gives reduced FPR with ANN classifier than other feature selection methods and classifiers.

TABLE 4: False Positive Rate (in %) obtained by Proposed ECM, IG, DE based feature selection methods using ANN, RF and NB classification methods

Feature Selection Methods	False Positive Rate (in %) obtained by Classification Methods		
	ANN	**RF**	**NB**
Original dataset	67.62	70.11	73.82
Information Gain	51.47	56.16	58.77
Differential Evolution	46.38	51.68	54.04
Proposed ECM	20.41	37.61	41.39

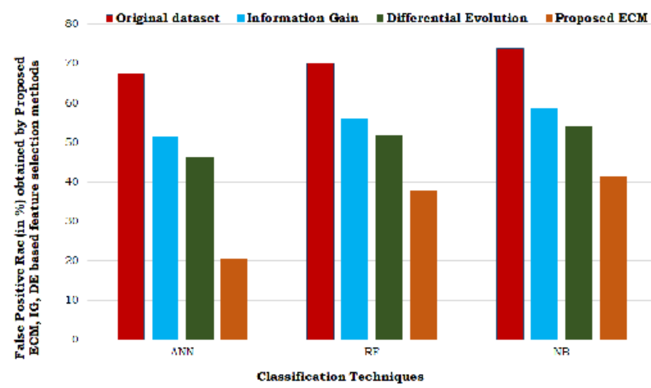

FIGURE 4: GRAPHICAL REPRESENTATION OF THE FALSE POSITIVE RATE (IN %) OBTAINED BY PROPOSED ECM, IG, DE BASED FEATURE SELECTION METHODS USING ANN, RF AND NB CLASSIFICATION METHODS

Table 5 depicts the Precision (in %) obtained the proposed ECM, DE and IG based feature selection methods using RF, ANN and NB classification methods. Figure 5 gives the graphical representation of the Precision (in %) obtained the proposed ECM, DE and IG based feature selection methods using RF, ANN and NB classification methods. From the table 5 and figure 5, it is clear that the proposed ECM gives better precision rate with ANN classifier than other feature selection methods and classifiers.

TABLE 5: Precision (in %) obtained by Proposed ECM, IG, DE based feature selection methods using ANN, RF and NB classification methods

Feature Selection Methods	Precision (in %) obtained by Classification Methods		
	ANN	**RF**	**NB**
Original dataset	63.13	56.34	57.82
Information Gain	71.13	66.34	63.30
Differential Evolution	71.13	70.69	69.21
Proposed ECM	83.59	74.72	72.41

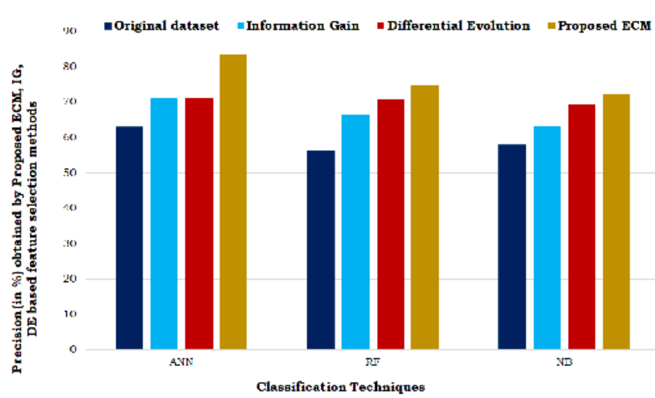

FIGURE 5: GRAPHICAL REPRESENTATION OF THE PRECISION (IN %) OBTAINED BY PROPOSED ECM, IG, DE BASED FEATURE SELECTION METHODS USING ANN, RF AND NB CLASSIFICATION METHODS

Table 6 depicts the Miss Rate (in %) obtained the proposed ECM, DE and IG based feature selection methods using RF, ANN and NB classification methods. Figure 6 gives the graphical representation of the Miss Rate (in %) obtained the proposed ECM, DE and IG based feature selection methods using RF, ANN and NB classification methods. From the table 3 and figure 3, it is clear that the proposed ECM gives reduced miss rate with ANN classifier than other feature selection methods and classifiers.

TABLE 6: Miss Rate (in %) obtained by Proposed ECM, IG, DE based feature selection methods using ANN, RF and NB classification methods

Feature Selection Methods	Miss Rate (in %) obtained by Classification Methods		
	ANN	RF	NB
Original dataset	45.38	49.54	50.49
Information Gain	39.23	42.25	43.22
Differential Evolution	36.79	39.689	40.82
Proposed ECM	10.47	32.25	32.25

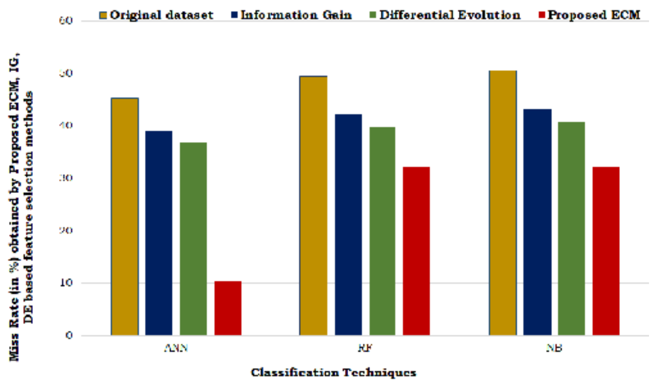

FIGURE 6: GRAPHICAL REPRESENTATION OF THE MISS RATE (IN %) OBTAINED BY PROPOSED ECM, IG, DE BASED FEATURE SELECTION METHODS USING ANN, RF AND NB CLASSIFICATION METHODS

Table 7 depicts the Specificity (in %) obtained the proposed ECM, DE and IG based feature selection methods using RF, ANN and NB classification methods. Figure 7 gives the graphical representation of the Specificity (in %) obtained the proposed ECM, DE and IG based feature selection methods using RF, ANN and NB classification methods. From the table 7 and figure 7, it is clear that the proposed ECM gives better specificity rate with ANN classifier than other feature selection methods and classifiers.

TABLE 7: Specificity (in %) obtained by Proposed ECM, IG, DE based feature selection methods using ANN, RF and NB classification methods

Feature Selection Methods	Specificity (in %) obtained by Classification Methods		
	ANN	RF	NB
Original dataset	32.38	29.89	26.18
Information Gain	48.53	43.84	41.23
Differential Evolution	53.62	48.32	45.96
Proposed ECM	79.59	62.39	58.61

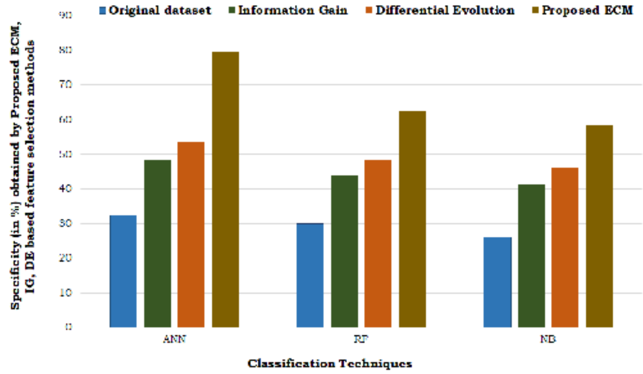

FIGURE 7: GRAPHICAL REPRESENTATION OF THE SPECIFICITY (IN %) OBTAINED BY PROPOSED ECM, IG, DE BASED FEATURE SELECTION METHODS USING ANN, RF AND NB CLASSIFICATION METHODS

CONCLUSION

Rain fall prediction plays the major role in agriculture production. The growth of the agricultural products is based on the rainfall amount. So it is necessary to predict the rainfall of a season to assist farmers in agriculture. In this paper, the rainfall dataset is considered to improve the efficiency of the rainfall prediction using proposed Enhanced Clustering Method (ECM). This proposed ECM utilizes the SOM and Rice Distribution (used to calculate the relative strength of the feature pre-dominant pattern) to improve the rainfall accuracy by getting the most pre-dominant feature subset. The performance of the proposed ECM is evaluated with existing feature selection methods using three classifiers. From the results obtained in this research work, the proposed ECM with ANN classifier performs better in terms of accuracy, TPR, FPR, precision, miss rate and specificity than other feature selection methods with other classifiers as well.

REFERENCES

[1] Cramer, Sam, et al. "An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives." *Expert Systems with Applications* 85 (2017): 169-181.

[2] Shah, Urmay, et al. "Rainfall prediction: Accuracy enhancement using machine learning and forecasting techniques." *2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC)*. IEEE, 2018.

[3] Thirumalai, Chandrasegar, et al. "Heuristic prediction of rainfall using machine learning techniques." *2017 International Conference on Trends in Electronics and Informatics (ICEI)*. IEEE, 2017.

[4] Dash, Yajnaseni, Saroj K. Mishra, and Bijaya K. Panigrahi. "Rainfall prediction for the Kerala state of India using artificial intelligence approaches." *Computers & Electrical Engineering* 70 (2018): 66-73.

[5] Namitha, K., A. Jayapriya, and G. Santhosh Kumar. "Rainfall prediction using artificial neural network on map-reduce framework." *Proceedings of the Third International Symposium on Women in Computing and Informatics*. 2015.

[6] Grace, R. Kingsy, and B. Suganya. "Machine Learning based Rainfall Prediction." *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*. IEEE, 2020.

[7] Pham, Binh Thai, et al. "Development of advanced artificial intelligence models for daily rainfall prediction." *Atmospheric Research* 237 (2020): 104845.

[8] Ahmed, Kamal, et al. "Multi-model ensemble predictions of precipitation and temperature using machine learning algorithms." *Atmospheric Research* 236 (2020): 104806.

[9] Le, Vuong Minh, et al. "Daily Rainfall Prediction Using Nonlinear Autoregressive Neural Network." *Micro-Electronics and Telecommunication Engineering*. Springer, Singapore, 2020. 213-221.

[10] Sardeshpande, Kaushik D., and Vijaya R. Thool. "Rainfall Prediction: A Comparative Study of Neural Network Architectures." *Emerging Technologies in Data Mining and Information Security*. Springer, Singapore, 2019. 19-28.

[11] Singh, Nitin, Saurabh Chaturvedi, and Shamim Akhter. "Weather Forecasting Using Machine Learning Algorithm." *2019 International Conference on Signal Processing and Communication (ICSC)*. IEEE, 2019.

[12] Balan, M. Selva, et al. "Rainfall Prediction using Deep Learning on Highly Non-Linear Data.", International Journal of Research in Engineering, Science and Management, Volume-2, Issue-3, March-2019, pp. 590-592.

[13] Parashar, Anubha. "IoT Based Automated Weather Report Generation and Prediction Using Machine Learning." *2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT)*. IEEE, 2019.

[14] Kalteh, Aman Mohammad. "Modular wavelet–extreme learning machine: a new approach for forecasting daily rainfall." *Water Resources Management* 33.11 (2019): 3831-3849.

[15] Lathifah, Siti Nur, et al. "Rainfall Forecasting using the Classification and Regression Tree (CART) Algorithm and Adaptive Synthetic Sampling (Study Case: Bandung Regency)." *2019 7th International Conference on Information and Communication Technology (ICoICT)*. IEEE, 2019.

[16] Poornappriya, T. S., and M. Durairaj. "High relevancy low redundancy vague set- based feature selection method for telecom dataset." *Journal of Intelligent & Fuzzy Systems* Preprint: 1-18.

[17] Durairaj, M., and T. S. Poornappriya. "Choosing a spectacular Feature Selection technique for telecommunication industry using fuzzy TOPSIS MCDM." *International Journal of Engineering & Technology* 7.4 (2018): 5856-5861.

[18] Durairaj, M., and T. S. Poornappriya. "Why Feature Selection in Data Mining Is Prominent? A Survey." *International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*. Springer, Cham, 2019.

[19] https://www.kaggle.com/datasets?search=weather+dataset