Webology (ISSN: 1735-188X)
Volume 18, Number 6, 2021

Gain Ratio With Optimization Based Feature Selection Method

A. TAMILMANI[1] , M. SUGHASINY[2]

[1]Research Scholar, Department of Computer Science, Srimad Andavan Arts and Science College (Autonomous) (Affiliated to Bharathidasan University, Tiruchirappalli), Tiruchirappalli, Tamilnadu, India.

[2]Assistant Professor, Department of Computer Science, Srimad Andavan Arts and Science College (Autonomous) (Affiliated to Bharathidasan University, Tiruchirappalli), Tiruchirappalli, Tamilnadu, India.

ABSTRACT

Precipitation in any form—such as rain, snow, and hail—can affect day-to-day outdoor activities. Rainfall prediction is one of the challenging tasks in weather forecasting process. Accurate rainfall prediction is now more difficult than before due to the extreme climate variations. Machine learning techniques can predict rainfall by extracting hidden patterns from historical weather data. Selection of an appropriate classification technique for prediction is a difficult job. In this research work, Gain Ratio and Differential Evolution hybridized for choosing the most relevant features. Once, the suitable features subset is obtained, the classification algorithm called Artificial Neural Network (ANN), and other classifiers are adopted, that could classify the info in the manner that is effective the selected features.

KEYWORDS: Feature Selection, Classification, Information Gain, Gain Ratio, Optimization, Decision Tree, Artificial Neural Network

1. INTRODUCTION

In the current scenario, rainfall is a significant factor for most essential things happening throughout the world. The farming sector is regarded as one of the most critical factors determining the country's economy, and farming relies entirely on rainfall. This research uses machine learning techniques for rainfall prediction and conducts the comparative analysis of two machine learning techniques, respectively, depicting an efficient rainfall prediction method [1] [2]. Rainfall prediction facilitates water resources management, flood alerts, flight operations management, limiting transportation, construction activities, and other factors that are most important to humankind. Rainfall data for forecasting is collected using weather satellites, wired and wireless instruments, and high-speed computers are used. Rainfall prediction has been a fascinating and captivating sector since the dawn of civilization, and it remains one of the most complex and enticing domains. Scientists use various methods and techniques to predict rainfall, some of which are more precise than others. Weather forecasting gathers atmospheric conditions such as humidity, temperature, pressure, rainfall, wind direction &speed, evaporation, etc.

Presently, Rainfall prediction is the most crucial factor for most water storage schemes worldwide. The uncertainty of rainfall data is one of the most complex problems. Today, most rainfall forecasting methods are incapable of detecting hidden patterns or non-linear trends in rainfall data. This research would help discover all hidden patterns and non-linear trends, which would be necessary for predicting accurate rainfall [3]. Due to the presence of complex issues in existing methods that cannot find the hidden patterns and non-linear trends efficiently the majority of the time, the forecast predictions were incorrect, resulting in massive losses. Thus, this research aims to find a rainfall prediction system that can solve all issues, find complexity and hidden patterns present, and provide proper and reliable predictions, therefore assisting the country in developing agriculture and the economy [4] [5].

2. RELATED WORKS

Grace, R. Kingsy, and B. Suganya [6] For the Indian dataset, a rainfall prediction model based on Multiple Linear Regression (MLR) was presented. Multiple meteorological characteristics are included in the input data, allowing for a more precise prediction of rainfall. The metrics utilised in this paper are MSE, RMSE, and Correlation.

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

Ahmed, Kamal, et al [8] Multi-Model Ensembles (MMEs) were used to minimise the uncertainty in GCM simulations and projections. The goal of this research was to assess the performance of MMEs created with machine learning (ML) algorithms using various combinations of GCMs that were ranked based on their performance, and to find the optimal amount of GCMs to include in an MME.

Le, Vuong Minh, et al [9] In order to estimate daily rainfall in Hoa Binh City, Vietnam, a prediction model based on Nonlinear Autoregressive Neural was proposed. To do this, an eight-year time series of meteorological data was collected, including temperature, wind speed, relative humidity, solar radiation as input variables and daily rainfall as an output variable (NARX). For evaluation, measures such as Correlation Coefficient, MAE, RSME, Error Mean, Median, and STD are used.

Sardeshpande, Kaushik D., and Vijaya R. Thool [10] presented a case study on the use of neural networks to forecast time series. In India, a case study was conducted to estimate rainfall using a local database. The findings came from a comparison of neural network topologies such as back propagation (BPNN), generalised regression (GRNN), and radial basis function (RBF) (RBFNN).

Singh, Nitin, Saurabh Chaturvedi, and Shamim Akhter [11] The major goal of this project is to create a weather forecasting system that can be used in remote places. To forecast weather conditions, data analytics and machine learning methods such as random forest classification are used. This study proposes a low-cost and portable weather forecast system.

Balan, M. Selva, et al [12] To predict rainfall, a few statistical analytic techniques including the use of Artificial Neural Networks were proposed. The influence on any system's prediction is defined by the correlation between the attributes. Because they do not contribute to the activation of the neuron, attributes with no correlation can be eliminated. The accuracy of the system is also affected by loss functions, activation functions, the number of neurons, and the number of hidden layers.

Parashar, Anubha [13] The major goal of this paper is to track and report weather conditions so that one can be informed ahead of time and take the required precautions to minimise the damage caused by any tragedy by forecasting it. The authors employ a variety of sensors to collect data, and previous data is used to train the system, while current data collection is used to make predictions. The evaluation measures in this work are Explained Variance, Mean Absolute Error, and Median Absolute Error.

Kalteh, Aman Mohammad [14] developed an extreme learning machine-based rainfall forecasting system based on combining wavelet analysis and a revolutionary artificial neural network technique (ELM). The specific properties of each technique are integrated in this way to capture various patterns in the data. To forecast rainfall, wavelet analysis is first used to decompose rainfall time series into wavelet coefficients, which are then used as inputs into the ELM model. The performance measurements are the correlation coefficient (r), root mean square errors (RMSE), and Nash–Sutcliffe efficiency coefficient (NS).

Lathifah, Siti Nur, et al [15] The rainfall in Bandung Regency was forecasted using the Classification and Regression Tree (CART) algorithm. In addition, an Adaptive Synthetic Sampling (ADASYN) approach was used to improve the model created as a result of a data class imbalance. In this paper, performance measurements like as Precision, Recall, Accuracy, and F1-Score are used.

3. PROPOSED INFORMATION GAIN WITH OPTIMIZATION BASED FEATURE SELECTION METHOD

In this proposed technique, the detailed description of the filter-based feature selection techniques and Differential Evolution (DE) optimization algorithm are given. In this proposed approach, the best and worst solution are obtained by converting the real code into binary values string to speed up the process, for reducing the computation time.

3.1 Gain Ratio Feature Selection

The Gain Ratio [17] is the non-symmetrical measure that is presented to pay back on the bias of the Information Gain (IG). GR is given by Equation (1):

$$\mathrm{GR} = \frac{\text{Information Gain(IG)}}{\mathrm{H(X)}} \quad (1)$$

Information Gain (IG) is a symmetrical measure.

$$\mathrm{IG} = \mathrm{H(Y)} - \mathrm{H(Y|X)} = \mathrm{H(X)} - \mathrm{H(X|Y)} \quad (2)$$

The information gained about Y after observing X is alike to the information gained about X after observing Y in the Equation (2). There, a weakness of the GR criterion is that it is biased in favor of features with more values even when they are not more informative.

As in the above Equation (2) presents, when the variable Y has to be predicted, then regularize the GR by distributing the entropy of X, and vice versa. Owing to this normalization, the GR values constantly fall in the range [0, 1]. A value of GR = 1 specifies that the knowledge of X totally forecasts Y, and GR = 0 means that there is no relation between Y and X. In opposition to IG, the GR favor variables with lesser values.

3.2 Differential Evolution Optimization

Differential evolution (DE) is merely one of several approaches through evolutionary algorithm where in actuality the features are search and centred on ant colony. An easy and yet effective, DE give you the benefits usually requires like many optimization methods [18][19]. There are several actions from DE such; 1) ability to handle non-differentiable, nonlinear and value this is certainly multimodal, 2) parallelizability to cope with computation cost that is intensive, 3) simplicity of good use, 4) good convergence properties.

Like GA, DE employ factors which can be same of mutation, selection and crossover. The efficiency of DE depends on the handling of target vector and difference in order to acquire a task vector in exploring procedure. Every real-value this is certainly d-Dimensional, a population of NP members is provided. NP will be the population size and D will be the true range that is wide of to be fine-tuned. Among the members of two population like $\mathrm{y_{s2}}$ and $\mathrm{y_{s3}}$ added the vector of weight difference to the $\mathrm{y_{s1}}$ which is third member for creating a trial vector. This action is termed as mutation. A mutant vector is generating relating to for every target vector $\mathrm{y_{(I,G)}}, \mathrm{j} = 1,2,3, \dots, \mathrm{M}$ a mutant vector using the given equation:

$$\mathrm{w_{j,G+1}} = \mathrm{y_{s1,H}} + \mathrm{G}\left(\mathrm{y_{s2,H}} - \mathrm{y_{s2,H}}\right) \qquad (3)$$

Where $\mathrm{s_1, s_2, s_3} \in \{1,2,\dots,\mathrm{NP}\}$ are integers that are chosen randomly, should be specific from 1 another plus unique through the operating index j. The control rate of Scaling factor F(0,1) that your particular population comprises. In order to improve the variety in connection with perturbed factor vectors, introduction of crossover is takes place. The trial vector:

$$\mathrm{v_{j,H+1}} = \left(\mathrm{v_{1,j,H+1}}, \mathrm{v_{2,j,H+1}}, \dots, \mathrm{v_{E,j,H+!}}\right) \qquad (4)$$

Is from where;

$$\mathrm{v_{kj,H+1}} = \begin{cases} \mathrm{w_{kj,H+1}} \text{ if } \mathrm{rand}(0,1) \le \mathrm{d_s} \\ \quad \mathrm{y_{kj,H+1}} \text{ otherwise} \end{cases} \qquad (5)$$

Where the H is the current population and the trial vector $\mathrm{k^{th}}$ $\mathrm{j^{th}}$ for the dimension of $\mathrm{v_{(kj,H)}}$. The probability of crossover $\mathrm{d_s}(0,1)$ is a person described value that operates the portion in connection with parameter values which are often and that can be replicated through the mutant. Selection will be the stage to get the vector among the target vector as well as trial vector making use of the aim of generating an individual in terms of generation this is certainly next. Then your causing vector substitutes the vector with which it absolutely was compared [20] if the recently created vector leads to a lower objective feature value (better fitness) as compared to population member that is predetermined. But, many factors from DE are instantly transformative without needed user to see by learning from your own error's strategy. In this work that is ongoing size of generation and population are adaptively identifying predicated on a total of features remained from relief-f. Hence, the buyer doesn't always have to initialize those factor values manually.

3.3 Proposed Optimal Feature Selection Method

In this proposed OFS method, instead of using Crossover and Mutation operator of DF optimization, encoding of solution (converting real code to binary string) is introduced to consume the computation time for weather datasets during the classification of diseases. The following are stages involved in this proposed OFS method is given below:

Stage 1: Encoding of Solution

Each individual solution is expressed in this work as a binary string in the population. The length of each solution (binary string) is equal to that of different features in the weather data sets. The solution's binary code 1 indicates the feature selection and the solution's binary code 0 is the feature not selected. The $S = [F_1, F_2, F_3, .. F_m]$ is the solution where m is the different dataset features. For example, a solution described as a [1, 1, 0 , 0, 1 , 1, 1, 0 , 0 , 0, 0 , 1] feature with index 1, 2 , 5, 6 and 10 is selected while the other features are not selected. Each location in the solution is binary value, q as a solution. The generation i of p^{th} solution is represented with $S_p^{(i)}$ and $S_{p,q}^{(i)}$ is represented by q^{th}position.

Stage 2: Initial population

Set the population size for this project to be 50. We produce 50 random solutions with real values randomly varying from 0 to 1. After that, each individual solution is used to convert real values to binary values based on the following equation: The digitization process:

$$S_{p,q}^{(i)} = \begin{cases} 1 & S_{p,q}^{(i)} > \text{rand} \\ 0 & \text{Otherwise} \end{cases}$$

Rand is a random number distributed uniformly from 0 to 1.

Stage 3: Fitness function

In optimization problems, fitness plays a key role. A fitness function measures the output of a single positive integer. With the aid of classifier error rate formulated in the following equations, the fitness for each solution in the population is calculated:

$$\text{fitness}\left(S_p^{(i)}\right) = \text{ClassifierErrorRate}\left(S_p^{(i)}\right)$$

The classifier error rate (solution) is the testing error rate:

$$\text{ClassifierErrorRate}\left(S_p^{(i)}\right) = \frac{\text{Number of misclassifiered records}}{\text{Total number of records}} \text{ X } 100$$

Stage 4: Finding new solutions

To produce the new solution the best and worst solutions in the output t are used. The lowest fitness value of the best solution (error rate) and the lowest fitness value of the generation i. $S_{bt}^{(i)}$ is represents an i iteration best solution and the $S_{wt}^{(i)}$ is used to represent the worst solution. Taking into account the best and worst generation solutions ‘i’, the q^{th} position of old solution $S_{p,q}^{(i)}$ as formulated by the below equation:

$$S_{p,q}^{(i)} = S_{p,q}^{(i)} + A\left|S_{bt,q}^{(i)} - S_{p,q}^{(i)}\right| + B\left|S_{wt,q}^{(i)} - S_{p,q}^{(i)}\right|$$

In case A, B is between 0 and 1 random numbers. Afterwards, the digitalization process is used in the next generation 'i+1' to transform real values into binary values based on the following equation for each position of the candidate:

$$S_{p,q}^{(i)} = \begin{cases} 1, & S_{p,q}^{(i+1)} > \text{rand} \\ 0, & \text{Otherwise} \end{cases}$$

Here rand is a uniformly distributed random number between 0 and 1.

Stage 5: Termination criteria

An iterative process is the proposed work. By the following considerations the termination conditions for the iterative process can be determined:

- Total number of iterations (T_{max})
- Fitness convergence rate and
- Threshold for iterative process running time. In this proposed work, the maximum number of iterations (T_{max}) is used as the termination criterion.

Algorithm: GRDE Feature Selection Method

Input: Rainfall Datasets (RD)
Output: Optimal Feature Subset (selected best features)
Step 1: Splitting of the Dataset into Training and Test (MD = $MD_{tr} + MD_{ts}$)
Step 2: Applying T = GR ← MD
Step 3: $m_f = |T|$
Step 4: Constructing Initial Population Table
  Step 4.1: foreach solution S_p p= 1 to M do
    Step 4.1.1: foreach position q of solution S_p; $q = 1$ to m_f do
    Step 4.1.2: $S_{p,q} = rand(0,1)$
    Step 4.1.3: $S_{p,q} = Digitization(S_{p,q})$
    Step 4.1.4: end
  Step 4.2: $S_p^{fitness} = computeFitness(S_p, C, f, MD_{tr}, MD_{ts})$
  Step 4.3: end
Step 5: $S_{bt} = findBestSolution()$
Step 6: $S_{wt} = findWorstSolution()$
Step 7: Iterative Process
  Step 7.1: foreach iteration i 1 to T_{max} do
    Step 7.1.1: foreach position q of solution S_p $p = 1$ *to* M do
      Step 7.1.1.1: foreach position q of solution S_p; $q = 1$ *to* m_f do
      Step 7.1.1.2: $A = rand(0,1)$; $B = rand(0,1)$
      Step 7.1.1.3: $F_{p,q} = S_{p,q}^{(i)} + A\left|S_{bt,q}^{(i)} - S_{p,q}^{(i)}\right| + B\left|S_{wt,q}^{(i)} - S_{p,q}^{(i)}\right|$
      Step 7.1.1.4: $F_{p,q} = Digitization(F_{p,q})$
      Step 7.1.1.5: end
    Step 7.1.2: $F_p^{fitness} = computeFitness(F_p, C, f, MD_{tr}, MD_{ts})$
    Step 7.1.3: if $F_p^{fitness} < S_p^{fitness}$ then
      Step 7.1.3.1: $S_p = F_p$
    Step 7.1.4: end
  Step 7.2: end
  Step 7.3: $S_{bt} = findBestSolution()$
  Step 7.4: $S_{wt} = findWorstSolution()$
Step 8: end
Step 9: Extracting the best optimal feature subset from the Ψ_{best}
Step 10: foreach position j of solution S_{bt} j=1 to m_f do
  Step 10.1: if isPositionSelected($S_{bt,q}$) then
    Step 10.1.1: $OF_{bt} = OF_{bt} \cup T[q]$
  Step 10.2: end
Step 11: end
Return OF_{bt}

4. RESULT AND DISCUSSION

4.1 Dataset Description

The weather dataset is taken from the Kaggle Repository. Table 1 depicts the features involved in the weather rainfall dataset [22].

Table 1: Description of the Dataset

Feature Number	Feature Name
1	meantempm
2	maxtempm
3	mintempm
4	meantempm_1
5	meantempm_2
6	meantempm_3
7	meandewptm_1
8	meandewptm_2
9	meandewptm_3
10	meanpressurem_1
11	meanpressurem_2
12	meanpressurem_3
13	maxhumidity_1
14	maxhumidity_2
15	maxhumidity_3
16	minhumidity_1
17	minhumidity_2
18	minhumidity_3
19	maxtempm_1
20	maxtempm_2
21	maxtempm_3
22	mintempm_1
23	mintempm_2
24	mintempm_3
25	maxdewptm_1
26	maxdewptm_2
27	maxdewptm_3
28	mindewptm_1
29	mindewptm_2
30	mindewptm_3
31	maxpressurem_1
32	maxpressurem_2
33	maxpressurem_3
34	minpressurem_1
35	minpressurem_2
36	minpressurem_3
37	precipm_1
38	precipm_2
39	precipm_3

40	Weather_Class (rain, no rain)

4.2 Number of Features Obtained

Table 2 depicts the number of features obtained by the Proposed Optimization based Feature Selection Algorithm, Information Gain, Differential Evolution algorithm. From the table 2, it is clear that the proposed optimization-based feature selection method gives a smaller number of features than the existing optimization algorithm like GR and DE algorithm. In the table 2, Proposed optimization-based feature selection algorithm gives only 27 features, GR gives 34 features whereas DE algorithm gives 31 features.

Table 2: Number of Features obtained by the Proposed Optimization based Feature Selection Method, Information Gain and Differential Evolution algorithm

S.No	Feature Selection Techniques		
	Proposed Optimization based Feature Selection	**Differential Evolution algorithm**	**Gain Ratio**
1	meantempm	meantempm	meantempm
2	maxtempm	maxtempm	maxtempm
3	mintempm	mintempm	mintempm
4	meantempm_2	meantempm_1	meantempm_1
5	meantempm_3	meantempm_2	meantempm_2
6	meandewptm_3	meantempm_3	meantempm_3
7	meandewptm_1	meandewptm_1	meandewptm_1
8	meanpressurem_2	meandewptm_2	meandewptm_2
9	meanpressurem_3	meandewptm_3	meandewptm_3
10	maxhumidity_1	meanpressurem_1	meanpressurem_1
11	maxhumidity_3	meanpressurem_2	meanpressurem_2
12	minhumidity_1	meanpressurem_3	meanpressurem_3
13	minhumidity_2	maxhumidity_1	maxhumidity_1
14	maxtempm_2	maxhumidity_2	maxhumidity_2
15	maxtempm_3	maxhumidity_3	maxhumidity_3
16	mintempm_2	minhumidity_1	minhumidity_1
17	mintempm_3	minhumidity_2	minhumidity_2
18	maxdewptm_3	minhumidity_3	minhumidity_3
19	mindewptm_3	maxtempm_1	maxtempm_1
20	maxpressurem_2	maxtempm_2	maxtempm_2
21	maxpressurem_3	maxtempm_3	maxtempm_3
22	minpressurem_1	mintempm_1	mintempm_1
22	minpressurem_2	mintempm_2	mintempm_2
23	minpressurem_3	mintempm_3	mintempm_3
24	precipm_1	maxdewptm_1	maxdewptm_1

25	precipm_2	precipm_1	maxdewptm_2
26	precipm_3	precipm_2	maxdewptm_3
27	mindewptm_2	mindewptm_1	mindewptm_1
28		mindewptm_2	minpressurem_3
29		mindewptm_3	mindewptm_3
30		maxpressurem_1	maxpressurem_1
31		minpressurem_3	maxpressurem_2
32			precipm_2
33			minpressurem_1
34			precipm_3

4.3 Performance analysis of the proposed Optimization based Feature Selection Method

The classification techniques like Artificial Neural Network, Random Forest and Naïve Bayes are utilized in this research work to analyze the proposed Optimization based Feature Selection, Gain Ratio and Differential Evolution algorithms. The metrics like Accuracy, True Positive Rate (TPR or Recall), False Positive, Precision, Miss Rate and Specificity. Table 3 depicts the performance metrics for the evaluation.

Table 3: Performance Metrics

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 4 depicts the performance analysis of the Original Dataset, GR processed dataset, DE processed dataset and proposed GRDEFS method processed dataset using ANN classification technique.

Table 4: Performance analysis of the original dataset, GR dataset, DE dataset and proposed GRDEFS dataset using ANN classification

Performance Metrics (in %)	Feature Selection Techniques with ANN Classification			
	Original Dataset	GR Dataset	DE Dataset	Proposed GRDEFS Dataset
Accuracy	45.65	57.05	59.9	89

TPR	54.62	60.77	63.21	93.62
FPR	67.62	51.47	46.38	16.2
Precision	63.13	71.13	73.21	86.78
Miss Rate	26.79	39.23	36.8	6.38
Specificity	32.38	48.53	53.62	83.8

Table 5 depicts the performance analysis of the Original Dataset, GR processed dataset, DE processed dataset and proposed GRDEFS method processed dataset using RF classification technique.

Table 5: Performance analysis of the original dataset, GR dataset, DE dataset and proposed GRDEFS dataset using RF classification

Performance Metrics (in %)	Feature Selection Techniques with RF Classification			
	Original Dataset	GR Dataset	DE Dataset	Proposed GRDEFS Dataset
Accuracy	43.1	53.25	56.64	67.95
TPR	50.46	57.75	60.311	69.83
FPR	70.11	56.16	51.68	35.42
Precision	56.34	66.34	70.69	77.91
Miss Rate	49.54	42.25	39.69	30.17
Specificity	29.89	43.84	48.32	64.55

Table 6 depicts the performance analysis of the Original Dataset, GR processed dataset, DE processed dataset and proposed GRDEFS method processed dataset using Naïve Bayes classification technique.

Table 6: Performance analysis of the original dataset, GR dataset, DE dataset and proposed GRDEFS dataset using Naïve Bayes classification

Performance Metrics (in %)	Feature Selection Techniques with Naïve Bayes Classification			
	Original Dataset	GR Dataset	DE Dataset	Proposed GRDEFS Dataset
Accuracy	41.85	51.2	54.85	65.25
TPR	49.51	56.78	59.18	67.81
FPR	73.82	58.77	54.04	39.28
Precision	57.82	63.30	69.21	75.30
Miss Rate	50.49	43.22	4082	32.19
Specificity	26.18	41.23	45.94	60.72

5. CONCLUSION

The dimensions regarding the weather data were high, and therefore the choice regarding the optimal features was a task that is complex that the proposed model used the suitable feature selection technology referred as GRDEFS method to pick the suitable features. The investigation was dedicated to the aim of diminishing the correlation amongst the selected features, while they were pertaining to the generation of diverse information which was pertaining to different classes of information. Further, the classification regarding the feature was carried out following the variety of the features that are optimal. The classification regarding the selected features was through with NN, RF and NB which had the capacity to classify the info in a manner that is effective the selected features. Thus, the complete analysis that is experimental the effective performance of proposed GRDEFS method way for the weather rainfall dataset classification method.

REFERENCES

[1] Datta, Anisha, Shukrity Si, and Sanket Biswas. "Complete Statistical Analysis to Weather Forecasting." Computational Intelligence in Pattern Recognition. Springer, Singapore, 2020. 751-763.

[2] Wang, Bin, et al. "Deep uncertainty quantification: A machine learning approach for weather forecasting." Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019.

[3] Refonaa, J., et al. "Machine Learning Techniques for Rainfall Prediction Using Neural Network." Journal of Computational and Theoretical Nanoscience 16.8 (2019): 3319-3323.

[4] Kachwala, Sakina, et al. "Predicting Rainfall from Historical Data Trends." Available at SSRN 3571738 (2020).

[5] Rasp, Stephan, et al. "Weather Bench: A benchmark dataset for data-driven weather forecasting." arXiv preprint arXiv:2002.00469 (2020).

[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. 213-22.

[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 (ICo ICT). IEEE, 2019.

[16] Sudeeptha, J., and C. Nalini. "Hybrid Optimization of Cuckoo Search and Differential Evolution Algorithm for Privacy-Preserving Data Mining." International Conference on Artificial Intelligence, Smart Grid and Smart City Applications. Springer, Cham, 2019.

[17] Liang, Jing, et al. "Multimodal multi objective optimization with differential evolution." Swarm and evolutionary computation 44 (2019): 1028-1059.

[18] Brezočnik, Lucija, Iztok Fister, and Grega Vrbančič. "Applying Differential Evolution with Threshold Mechanism for Feature Selection on a Phishing Websites Classification." European Conference on Advances in Databases and Information Systems. Springer, Cham, 2019.

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

[20] 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.

[21] 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.

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