Journal of Theoretical and Applied Information Technology
15th July 2021. Vol.99. No 13
© 2021 Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A HYBRID ACO-CS BASED OPTIMIZED KNN CLASSIFIER ALGORITHM FOR RAINFALL DETECTION & PREDICTION

K. VARADA RAJKUMAR [1], Dr. K. SUBRAHMANYAM[2]

[1]Research Scholar, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India.

[2]ProfessorDepartment of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India.
E-mail: [1]varadarajkumar18@gmail.com, [2]smkodukula@kluniversity.in

ABSTRACT

The detection and prediction of rainfall is an important task in recent years. The usage of machine learning approach in agriculture has enhanced the efficiency of farming. The rainfall detection and prediction will be helpful to farmers to take appropriate actions on sowing, irrigation etc. In this paper, the rainfall detection and classification are done, the investigation of various machine learning approaches like Support Vector Machine, K-Nearest Neighbor (KNN), Decision Tree Neural Networks were done using rainfall data. In this paper we propose a new approach for the optimization of the nearest neighbor numbers in KNN algorithm using a hybrid Ant colony optimization and Cuckoo search algorithm for efficient rainfall detection. The experiments were performed in MATLAB platforms using monthly rainfall data sets that are downloaded from Indian meteorological Department (IMD). Monthly rainfall for years 1901 to 2019 are taken for analysis. The performance of various classification algorithm for rainfall data using the parameters like precision, sensitivity, specificity, and accuracy has been done.

Keywords: *ACO, Cuckoo search, SVM, neural networks, KNN.*

1. INTRODUCTION

Rainfall is the very essential for all the atmospheric activities and it not only aids for the society and environment but also for every living being living on the earth. It is the most important and necessary natural phenomenon and so it affects everything directly or indirectly. At the same time, it is necessary for humans to examine on the changes in precipitation with respect to the climatic changes [6]. The rainfall has an important cause on the universal gauge of atmospheric circulation, and it also have effect on the local weather conditions. It also aids to balance the increasing weather temperatures and for the existence of the humans [9]. The increase in temperature is directly related to global warming and it is the fact that water is the scare and the most wanted resources that results in the increasing weather conditions that leads to the evaporation of water from the reserves. The compensation of these reserves is given by rainfall. It is also essential for agricultural production. The occurrence of rainfall varies with the variation in the latitude and longitude. Different regions, planes, mountains, and plateaus also affects rainfall.

Rainfall prediction is a great challenge for climatologists. Rainfall is one of the most essential elements of our climate system. Many disasters such as global warming, floods, drought, heat waves, soil erosion and many other climatic issues are caused because of rainfall. Also, agriculture is related to rainfall [18] and it is the important element of economic activities in most of the countries in the world. Hence for increasing the production of crop and to protect crop, human life, and ecosystem the demand for prediction of rainfall is growing in rate day-by-day for reliable prediction from policy makers. This explains the need for accurate prediction of rainfall. Many methods are followed for prediction of rainfall, but the accuracy is the necessary factor that must be tested in all the methods. The world is affected by many disasters caused by rainfall and those includes Drought, Flood and intense summer heat etc., and this also have effect on water resources around the world. Fig 1 shows the downfall of yearly rainfall in millimetre.

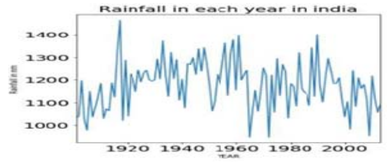

Figure .1: Rainfall in particular year

2. LITERATURE SURVEY

Here artificial neural network (ANN) was employed with general circulation models (GCM) to get information about the future precipitation and mean temperature in Tabriz synoptic station at north-west of Iran by Nourani et.al., (2019) [1]. Three different GCM are used such as Can-ESM2 and BNU-ESM from IPCC AR5 models and CGCM3 from IPCC AR4 models. Future projection and mean temperature for 2020-2060 was predicted with the use of ANN. Documentation of acid rain is essential for taking control measures. Zhang et.al., (2012) [2] focuses in spatial and temporal variations that exists in acid rains in north-eastern China. Prediction of acid rain is done by geographic position, terrain characteristics, routinely monitored meteorological factors and column concentrations of atmospheric SO2 and NO2. Decision tree approach is applied on the observed data that yield an accuracy of 98.04%. Here prediction performance of rule-based decision tree (DT) and combination of frequency ratio (FR) and logistic regression (LR) statistical methods are carried out for Kelantan, Malaysia proposed by Tehrany et.al., (2013) [3]. Flood inventory map for 115 flood locations is extracted for Kelantan. 70% of the dataset is used for testing and the remaining 30% is used for validation purposes. Validation results for DT and integrated FR and LR is 87% and 90% for success rate and 82% and 83% for prediction rate.

Here Hasan et.al., (2015) [4] used Support Vector Regression (SVR) technique for prediction of rainfall. Recent rainfall data of Bangladesh is preprocessed and fed as the input to the algorithm. The proposed method yields a highest prediction accuracy of 99.92% and the experimentation also prove that SVR is better in terms of process running time. The monthly rainfall forecasting is done with the help of particle swarm optimization algorithm that uses SVR's optimal parameters to construct SVR model by Zhao and Wang (2010) [5]. The rainfall data of Guangxi, China during 1985–2001 are applied as the dataset. The results prove that SVR performs much better than BPNN and ARIMA. Gupta and U. Ghose (2015) [7] done rainfall forecasting with the help of 2245 data samples collected form Weather Underground website. Training is done with the help of Pattern Recognition Neural Network with 82.1% accuracy, K nearest Neigh bour with 80.7% accuracy, classification, and regression tree algorithm with the accuracy of 80.3% and Naive Bayes approach with 78.9% accuracy.

M. Huang et al (2017) [8] used K-nearest neighbor (KNN) algorithm. The KNN algorithm was one among the machine learning techniques. It provides robustness ability against various choices of neighborhood size k, especially in an irregular class distribution of precipitation dataset. Jan et al (2008) [9] innovated a Seasonal to Inter-annual Climate Prediction with the help of Data Mining KNN Technique. This Data Mining KNN Technique used historical weather data of a region and classifies historical data to a particular span of time. The k nearest neigh bours is used to analyze and predict the weather before within reasonable time for months.

Lathifah et al (2019) [10] coined Classification and Regression Tree (CART) Algorithm and Adaptive Synthetic Sampling. The CART Algorithm forecasts rainfall in Bandung Regency, and it also used an Adaptive Synthetic Sampling (ADASYN) algorithm. The Experiment used CART and ADASYN algorithm resulted with rainfall prediction accuracy of 93.94% and 1.38 s running time but alone obtained rainfall prediction accuracy of 98.18% and 1.48 s running time. Devak et al (2015) [11] implemented a Dynamic coupling of support vector machine and K-nearest neighbor for the purpose of downscaling the daily rainfall data. The Dynamic coupling of support vector machine and K-nearest neighbor for downscaling daily rainfall was comparatively analyzed with simple KNN and SVM models based on their performance parameters. M. kbari et al (2011) [12] constrained Clustered K Nearest Neighbor Algorithm for Daily Inflow Forecasting. This Clustered K nearest neighbor algorithm traps inconsistent data points and performs robust against noisy data. This Clustered K Nearest Neighbor Algorithm is suited for synthetic linear data set manipulated by the existence of noise and it was also illustrated at Karoon1 reservoir located in Iran. Yu et al (2017) [13] coined a Comparison of random forests (FA) and support vector machine (SVM) for real-time radar-derived rainfall forecasting. It also constructed single-mode forecasting model (SMFM) and multiple-mode forecasting model

(MMFM) in terms of RF and SVM. The Experiment resulted that SMFMs obtained better results than MMFMs for both SVM-based and RF-based SMFMs gives good performances for 1-h ahead forecasting but SVM-based SMFM was higher than RF-based SMFM at 2- and 3-h ahead forecasting. Lin et al (2013) [14] invoked an integrated two-stage Support Vector Machine (SVM). In the initial stage, monitored rainfall and typhoon characteristics are applied for the production of rainfall forecast and then predicted rainfall and monitored runoff are constructed for a runoff forecast at second state. Two stage SVM model are evaluated with the help of 16 typhoon storm dataset from Taiwan. The Experiment Resulted with a runoff forecasts of 1–6 h lead time and substantial performance improvement of flood forecast was obtained for 4- to 6 h lead time.

D C R Novitasari et al (2019) [15] used ANFIS method and Support Vector Regression (SVR). The ANFIS used MSE and RMSE for the accurate prediction of rainfall Predictions of parameters that affected rainfall. The ANFIS method resulted with RMSE of 3.871590 for predictions of relative humidity, RMSE of 1.975004 for wind speed predictions and RMSE of 0.742332 for temperature predictions. S. Georganos et al (2017) [16] coined a Geographically Weighted Regression (GWR). A Normalized Vegetation Difference Index and rainfall was used for applying GWR. The operating scale of the Sahelian NDVI–rainfall relationship was stabilized to about 160 km. The NDVI-rainfall relationship are expressed in terms of spatial pattern and appropriate scale selection by the GWR than the conventional Ordinary Least Squares (OLS) regression.

S. Cramer et al (2017) [17] constrained an Extensive Evaluation of Seven Machine Learning methods. This Experiment was processed by the rainfall time series of 42 cities and compared to current state of the art predictive performance and also with six various machine learning algorithms including: Radial Basis Neural Networks, Genetic Programming, Support Vector Regression, k-Nearest Neighbors, M5 Rules, M5 Model trees.

R.C Deo et al (2017) [20] stimulated Drought Modelling using multivariate adaptive regression splines (MARS), least square support vector machine (LSSVM) and M5Tree models. The Periodicity presence reduces RMSE value 3.0–178.5% and increased the r2 value 0.5–8.1%. The MARS dominated other counterparts of three out of five stations with lower MAE 7.3–42.2% and 15.0–73.9%. The Results showed M5Tree perform well when compared with MARS/LSSVM with decreased MAE by 25.7–52.2% and 13.8–13.4% and droughts are identified by SPI ≤ − 0.5. Kusiak et al (2013) [21] used Radar Reflectivity Data. It is a data mining technique based on Tipping-Bucket (TB) and Radar Reflectivity data. It has three models and five algorithms namely Random Forest, Support Vector Machine, classification and regression tree and Neural Network algorithms and K-nearest neighbor for predicting the rainfall in a watershed basin at Oxford.

S. Bhomia et al (2016) [22] coined a Dynamical-Model-Selection-based on Multimodal Ensemble (DMS-MME) technique. It was compared with individual models and regression-based MME model. The DMS-MME forecasts had excellent skills based on verification scores up to 120 h compared with MME forecasts. F.S. Marzano et al (2007) [23] used Neural Network. It also uses a microwave (MW), infrared (IR) passive sensor image and Neural Combined Algorithm for Storm Tracking (NeuCAST). The NeuCAST had two steps namely Measuring IR radiance field from a geostationary satellite radiometer and combining MW-IR and rain retrieval algorithm exploiting GEO-LEO observations. J. Pucheta et al (2013) [24] innovated a Sub sampling Nonparametric Methods. The Non-parametric methods are subdivided into stage of smoothing. It also uses techniques for forecasting the high roughness time series and also for generating a smooth time series. The Results of this experiment evaluated in terms of Mackey Glass Equation MG and generate cumulative month-wise historical data of rainfall. Haidar and B. Verma (2018) [25] constrained a One-Dimensional Deep Convolutional Neural Network. These Experiments results are compared with Australian Community Climate and Earth System Simulator Seasonal Prediction System (ACCESS) and conventional Multi-Layered Perceptron (MLP) and results are evaluated based on RMSE with difference of 37.006 mm and compared with ACCESS and RMSE with difference of 15.941 comparatively analyzed with respect to the conventional MLP.

3. BACKGROUND METHODOLOGY

The advancement in the field of Artificial Intelligence has given rise to number of newer technologies and they include Machine Learning Approaches and Deep Learning Techniques which uses numerous features of artificial intelligence .The application areas of machine learning techniques are wide and some of the important fields of machine learning are medical image analysis, robot path planning , flood detection in a particular city or area and land cover classification .Machine learning is a process in which machines learn a particular task without any human influence by a continuous learning process by means of improving the performance by gaining experience with the previous outcome .The learning process in machine learning consists of two main types supervised learning and unsupervised learning. Supervised learning is one type of learning process that involves labelling the given set of features in the training dataset and Unsupervised learning is a type of learning process that contains no labels and the system itself groups and labels the features of the dataset. The vital step in machine learning approach is feature extraction. The extracted features from the target can be classified by using various other approaches such a classification or regression. There are wide variety of classifiers in machine learning techniques, some of the commonly used classifiers include SVM classifiers, Decision Trees, Naive Bayes and Linear Regression and Random Forest Trees. Here various classifiers are used for classifying the input data and to predict the rainfall occurrence and its performance was compared. Time series data-based forecasting models are developed for predicting different variables. Regression is a based on statistical empirical technique and it is widely used in business, social, behavior and biological science, climate prediction and many other fields. Also, linear, and non-linear multiple regression models of different orders can also be used for predicting based on time series data. These models can make use of one or more predictor for rainfall prediction. Fig 2 shows the various stages involved in rainfall prediction.

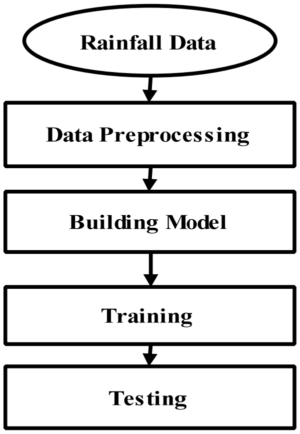

Figure 2: Rainfall prediction stages

3.1. Ant colony optimization metaheuristic

Optimized ant colony algorithm into metaheuristic algorithm for solving combination problem. ACO is a methodology by which the ant traverses the graph G(V, E) that are connected together for finding optimum solutions. V represents the set of vertices and E is the set of edges. The ant gives solutions by travelling from one vertex to the other along the edge. This ACO metaheuristic algorithm is depicted in Algorithm 1

```
-------------------------------------------------------------
Algorithm 1 The Ant Colony Optimization Meta heuristics
Step1: Initialization of parameters, pheromone, pheromone trials.
Step2: Terminate if the conditions does not satisfy.
Step3: Construction of ant solutions.
 Step4: Local search.
Step5: Update the Pheromones
Step6: End
-------------------------------------------------------------
```

3.2. Cuckoo search algorithm

Cuckoo search algorithm is the inspiration of its egg laying process in host bird nest. If the host bird found the alien egg it throws away the egg or it destructs the nest. In some cases, the color and size of the egg are found to be similar to the host. Hence it is used for various optimization problems. It is described as follows

```
Objective function $f(x), x=(x_1, x_2,….x_d)^T$
    Generate initial population of n host nests
        $xi(i=1,2,….n)$
While (t<Max Generation) or (stop criteria)
Get a cuckoo (say i) randomly ny Levy distribution;
Evaluate its quality/fitness $F_i$;
Choose a nest among n (say j) randomly;
Evaluate its quality/fitness $F_j$;
If ($F_i > F_j$)
Replace j by the new solution;
End
A fraction of ($p_a$) of worse nests are abandoned and
New ones are built at new locations via Levy flights;
Keep the best solutions (or nests with quality solutions);
Rank the solutions and find the current best;
End while
Post processing
```

4. RAINFALL PREDICTION MODEL

4.1. Support Vector Machine Classification

Binary pattern recognition includes a process of building a decision rule for the classification of vectors into any of the two classes on the basis of the training set of vectors classification also called as priori. The use of support vector machine can execute it simply by mapping the training data in a higher dimensional feature space. Then the construction of hyperplane is done for the feature space which bisects both of the categories and increases the separation margin between themselves and the points nearby them (known as support vectors). For unknown vector classification this decision surface is used as a base. For the binary classification setting, consider,

((x1, y1) • • •(xn, yn)) - training dataset,

Xi - feature vectors that represents the instances (i.e., observations) and

yi ∈ {−1, +1} are the labels of the instances.

Support vector learning is a process of determining the separated hyperplane that separates the positives labels (+1) and then negative labels (-1) having the maximum margin. The margin in the hyperplane is described as the minimum distance between the positive and negative labels which are nearer to the hyperplane. The main reason behind keeping the maximum margin in the hyperplane is that, the one with the maximum margin is highly resistive to noise than the hyperplane with a minimum margin.

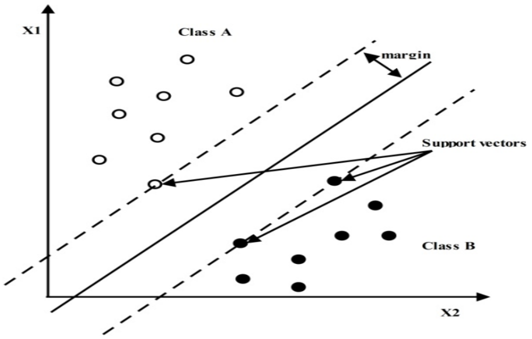

Figure3: Largest-margin hyperplane for a SVM classifier. Samples within the margin are known as support vectors.

Usually, if every data satisfies the above conditions

$$w \cdot x_i + b \geq +1 \quad y_i = +1 \qquad (1)$$

$$w \cdot x_i + b \leq -1 \quad y_i = -1 \qquad (2)$$

Here,

w -the normal to the hyperplane,

|b|/||w|| - the perpendicular distance between the hyperplane and the origin

||w|| - the Euclidean norm of w.

The above two conditions can be easily joined together as follows,

$$y_i(w \cdot x_i + b) \geq 1 \ \forall_i \qquad (3)$$

The training examples of the (3) lies in the canonical hyperplanes (Figure 3). Then the margin ρ will be calculated as the distance between H1 and H2.

$$\rho = \frac{|1-b|}{\|w\|} - \frac{|-1-b|}{\|w\|} = \frac{2}{\|w\|} \qquad (4)$$

Thus, the largest margin that separates the hyperplane can be builded by giving solution to the above prima optimization problem.

$$\min_{w \in H} \tau(w) = \frac{1}{2}\|w\|^2 \quad \textit{subject to} \quad y_i(w \cdot x_i + b) \geq 1 \ \forall_i \qquad (5)$$

4.2. K-Nearest Neighbor (KNN)

K-nearest neighbor (KNN) classifier is the basic and easier technique that works based on similarity measure by storing all the obtainable examples and by classifying the newest cases of the example language. It makes use of lazy learning method. For every test image that are needed to be predicted, this method locates the k closest members (the K nearest neighbors) of the training data set. For the calculation of the closeness of every member of the training set is to the test class is measured by Euclidean Distance measure. Class labels can be found by the use of K nearest neighbor and then majority voting is done is establish the class label of test image. The value k is depended on the data we use, the larger the value of k the greater the effect of noise is reduced in the classification, but the boundaries made with the class are little varying.

Euclidean distance is measured as:

$$d_x(x,w_k = \sqrt{(x-w_k)^T(x-w_k)} \qquad (6)$$

4.3. Decision Tree

Decision trees are known to be one of the best-known techniques for the classifier representation. From the available data a decision tree is generated by this technique. It looks like a structure of tree that gives prediction model in which every internal node represents a test on an attribute and every outgoing branch denotes the test outcome and similarly every leaf node is named with a class of the image. For classification purpose and prediction, decision tree is commonly used. This method looks simply but is the most influential method to represent knowledge. The decision tree models are generally depicted in the form of tree like structure. Decision tree learning includes the determination of where to make split at every node and the approximate the depth of the tree structure by properly analyzing it. The class of the data is denoted by the leaf node. By sorting the decision tree from root node to leaf node the instances of the tree can be classified. Mostly decision trees are known as noise resistant due to the pruning strategies that eliminates over-fitting the data in common and Gaussian noisy data.

Decision trees looks like as a flowchart tree structure, in which every internal node denotes a test on attribute, every branch represents a test outcome, every leaf node/terminal node represents a class label. For a tuple X given, the values of attributes are tested over the decision tree. From the root to leaf node a path is traced that handles the class prediction for the tuple. The conversion of decision tree into classification rules is easier. Learning a decision tree makes use of the decision tree as a predictive model that plots the items observation to the items target value conclusions. It is used in many fields that includes statistics, data mining, machine learning etc., and it is considered to be one of the predictive modeling approaches. The construction of decision tree is comparatively fast when studied with rest of the methods of classification. For accessing the database effectively, decision tree aids in the construction of SQL statements. They accuracy of decision tree is calculated to be similar or in some cases it is better in performance when studied comparatively with the rest of the classification methods.

From a given dataset, decision trees can be builded automatically by use of decision tree inducer algorithm. The main aim here is to discover ideal decision tree by decreasing the generalization error. But, the rest of the target functions also can be explained, for example, decreasing the number of nodes or minimizing the average depth.

Decision tree can be classified into the following two types.

•Classification decision trees- Is a kind of decision tree in which the decision variable is categorical.

•Regression decision trees – Is a kind of decision tree in which the decision variable is continuous.

4.4. Neural Networks

Artificial neural networks (ANNs) are the tools used for modelling an unlimited complex and intelligent task. It is the popular machines in the rising technology with parallel and distributed processing system. This Processing system executes complex tasks such as recognition, prediction and detection. It applies various models for data-processing such as feed-forward back propagation, NARX model with various functionalities for every model. It has one input layer, and one output layer and hidden layers is between them, which process and forwards the result data to the next hidden layer and the end result data are collected at output layer. These Dynamic machines have capability solve to unlimited complex and intelligent tasks for humans.

4.4.1 Neurons

The Artificial Neural Network performance is similar to that of Neurons performance in human brain. The Artificial Neural Network executes an unlimited difficult and intelligent process with the help of neurons that are fed into the layers and do processing with the data. The Neural Network is a non-linear function in which neurons are initially trained with the old data in order to get the new and predicted data.

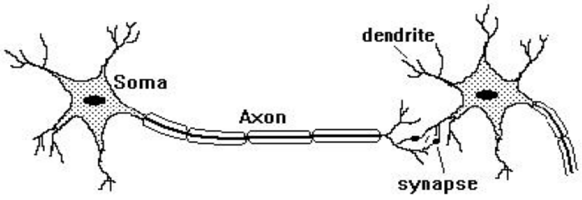

Figure 4: Neuron scheme

After completing training, it was applied to test checkup that has different results with different data and a comparison was made by feeding the system with a different number of neurons that varies and depends upon data and processing complexity. Hence, these architectures get differs with each other in terms of layers and input/output complexity in the system.

4.4.2. Structure of ANN

The ANN Architecture has three main layers with a huge number of neurons. These Neurons are also known as units which are ordered in a sequential layer.

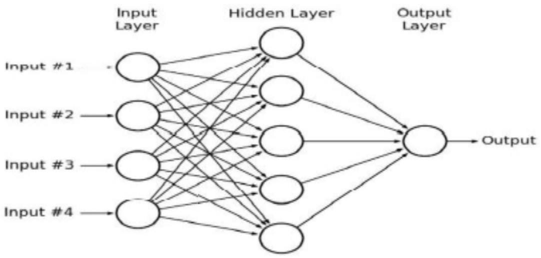

Figure 5: Structure of ANN

INPUT LAYER is the initial layer of ANN structure. The Input layer collects the input for processing.

HIDDEN LAYER is the next layer of ANN structure. This Hidden Layer is used for the processing of the data gathered from the input layer for processing via neurons and updating are made endlessly in terms weights for accuracy and correctness of the output.

OUTPUT LAYER is the last and foremost layer of ANN structure. This Output Layer shows the obtained results, as shown in figure 4.2 above.

4.4.3 Weights

The ANN architecture Weights are nothing but the memory storage. This Weights are used for the storage of information and data to get the expected results. During the training, testing and validation the weights are varied at each step such output acquires and stores accuracy of data further for future usages.

4.4.4. Feedforward neural network

The Feedforward neural network consists of many layers for processing the elements. In Feed forward neural network every layer has to processes the input data that are received by it and it also forwards the results obtained to next layer and these processes are made independently by every layer to generate result which is to be transferred to the next layer. The output layer obtains processing results of every layer. An Artificial neuron is an element similar to human brain between every input and output layer there exists hidden layers. The neurons send messages or information to the nearby neurons via a channel called as connections.

5. PROPOSED METHODOLOGY

There are numerous methods for the carrying out rainfall prediction. For performing this neural network is preferred since it is believed that when compared to SVM, KNN and tree classifier the results of neural networks are accurate. Therefore, we presented a novel modified KNN algorithm for enhancing the accuracy. The main issue with KNN is the approximation and estimation of hyperparameter k. If k is small, then the algorithm will be outlier sensitive. If K is of a larger value, then the neighborhood may include too many points from other classes. Hence a proper selection of K is crucial. Here we have used an efficient method for selecting the k value by the use of ACO algorithm. Additionally, cuckoo search algorithm is employed for aboding the worst k. this decreases the allocation time. The logarithmic scale of evenly spaced is used for the estimation of number of neighbors. The classification error is eliminated, and the k value is tuned to properly. Then the classifier is designed with the estimated optimal nearest neighbor, and it is noted that the performance of the classifier is enhanced.

The above steps describe the proposed algorithm. The nearest neighbor is calculated by using evenly spaced log scale. Once the nearest neighbor is calculated our next step is to find the classification error by employing ACO with CS scheme.

Proposed Algorithm

```
Step 1: Select Number of neighbors by approximately evenly spaced on a logarithmic scale.
Step 2: select k randomly by cuckoo search algorithm
Step 3: for selected ant (k)
            Find classification error
Step 4: update pheromone
Step 5: update global best
Step 4: design KNN classifier with global best (k)
Step 5: Train the model
Step 6: Test the model
Step 7: Validate and compare the proposed model with existing model.
```

The optimal nearest neighbor is chosen by this process. The testing and evaluation of KNN design is performed once the design of KNN is completed.

Finally, a comparative analysis is done between the results of the proposed methodology and the conventional KNN classifier.

6. DATA SET TAKEN

The meteorological data that used in this research has been brought from Indian meteorological Department, based on previous 120 years data set calculation of Monthly Rainfall prediction made in Andhra Pradesh, India.

7. IMPLEMENTATION RESULTS AND DISCUSSIONS

The experiments were performed in MATLAB platforms using monthly rainfall data sets that are downloaded from Indian meteorological Department (IMD). Monthly rainfall for years 1901 to 2019 are taken for analysis. Data were preprocessed by filling missing data and normalized by min-max normalization. Here we categorize the rainfall data into moderate rainfall and heavy rainfall. Then the processed data is given to various classifiers for evaluating its performance.

Performance Analysis

The presented algorithm is observed and analyzed by statistical measures of sensitivity, specificity, and accuracy.

$$Sensitivity = \frac{TP}{TP+FN} \quad (7)$$

$$Specificit\ y = \frac{TN}{FP+TN} \quad (8)$$

$$Accuracy = \frac{TP+TN}{Total\ \ Frames} \quad (9)$$

True positive (TP) is the condition in which the classifier predict the heavy rainfall input correctly as heavy rainfall. False positive (FP) is a condition in which the classifier mistakenly predict the moderate rainfall as heavy rainfall. True negative (TN) is the phenomenon by which the classifier predicts the moderate rainfall input correctly as moderate rainfall. False negative (FN): It is the condition where the classifier mistakenly predict the heavy rainfall input as moderate rainfall

The table 1 and table 2 give the performance of various classification algorithm for rainfall data. In table 1 Out of 22 true positive data, SVM classify 21 data correctly (TP) and one wrong (FP) and out of 13 true negative, SVM correctly predict 11 true negative (TN) and 2 False Negative (FN). In KNN classification, out of 22 true positive data, it classify 20 data correctly (TP) and two wrong (FP) and out of 13 true negative, it correctly predict 12 true negative (TN) and 1 False Negative (FN). In decision tree classification, Out of 22 true positive data, it classify 21 data correctly (TP) and one wrong (FP) and out of 13 true negative, it correctly predict all 13 true negative (TN) and 0 False Negative (FN) .

In table 2 Feed –Forward Neural Network Model classification, Out of 22 true positive data, it classify 20 data correctly(TP) and 2 wrong(FP) and out of 13 true negative, it correctly predict all 13 true negative(TN) and 0 False Negative(FN) .In Cascade-Forward Neural Network Model classification, Out of 22 true positive data, it classify 21 data correctly(TP) and 1 wrong(FP) and out of 13 true negative, it correctly predict all 13 true negative(TN) and 0 False Negative(FN) .

In Pattern Recognition Neural Network Model classification, out of 22 true positive data, it classify 21 data correctly (TP) and 1 wrong (FP) and out of 13 true negative, it correctly predict all 13 true negative (TN) and 0 False Negative (FN) which is similar to the previous one.

From the table, it shows that decision tree gives better accuracy than SVM and KNN. cascade-Forward and Pattern Recognition NN Classifier accuracy is 97.14, which means it gives better result than all other classification. In future hybrid methods may implement to further enhance its performance.

The table 1 shows that the decision tree has 100% precision while SVM has higher sensitivity. When specificity and accuracy is considered once again decision tree is dominant in performance when compared to other methods. Fig 6 shows the comparison chart of various classifier performance.

Table 1: Performance Comparison of SVM, KNN and Decision Tree Classifier

Parameters	Feed - Forward NN	Cascade-Forward NN	Pattern Recognition NN
Total (T)	35	35	35
True Positive (TP)	20	21	21
False Negative (FN)	2	1	1
False Positive (FP)	0	0	0
True Negative (TN)	13	13	13
Precision (%)	100	100	100
Sensitivity (%)	90.91	95.45	95.45
Specificity (%)	100	100	100
Accuracy (%)	94.29	97.14	97.14

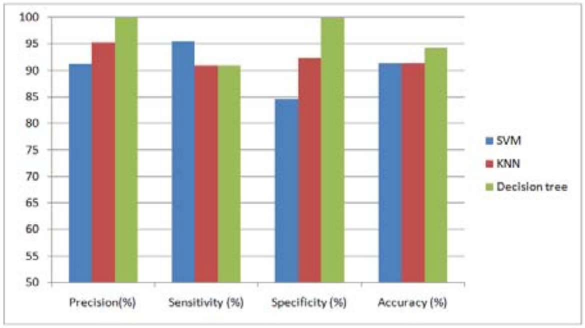

Figure 6: Comparison chart of performance of SVM, KNN and Decision Tree classification.

The table 2 shows the performance of different neural network architectures. The methods show 100% precision and specificity. The sensitivity and accuracy are better in Cascade forward NN and pattern recognition NN.

Table 2: Performance Comparison of Feed-Forward, cascade-Forward and Pattern Recognition NN Classifier

Parameters	SVM	KNN	Decision tree
Total (T)	35	35	35
True Positive (TP)	21	20	20
False Negative (FN)	1	2	2
False Positive (FP)	2	1	0
True Negative (TN)	11	12	13
Precision (%)	91.30	95.24	100
Sensitivity (%)	95.45	90.91	90.91
Specificity (%)	84.62	92.31	100
Accuracy (%)	91.43	91.43	94.29

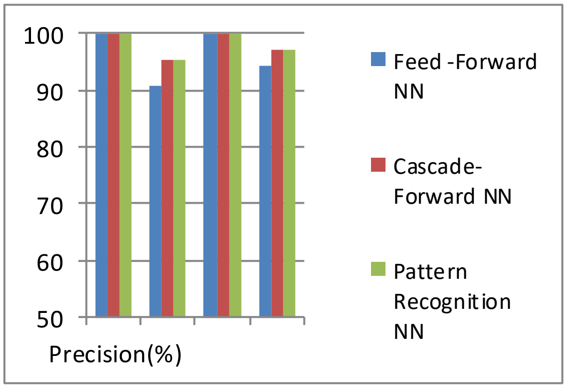

Figure 7: Comparison chart of performance of various Neural Network model

Table 3: Performance Comparison of KNN and Proposed modified KNN.

Parameters	KNN	Proposed modified KNN
Total (T)	35	35
True Positive (TP)	20	20
False Negative (FN)	2	2
False Positive (FP)	1	0
True Negative (TN)	12	13
Precision (%)	95.24	100
Sensitivity (%)	90.91	90.91
Specificity (%)	92.31	100
Accuracy (%)	91.43	94.29

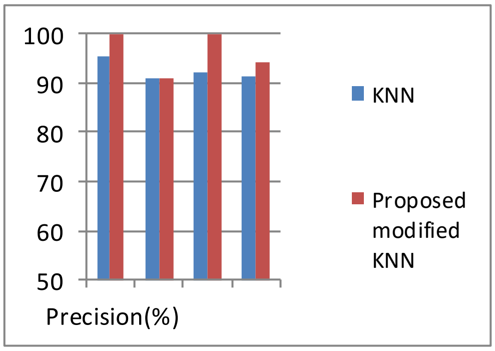

Figure 9: Comparison chart of Existing and proposed KNN model

8. CONCLUSION

The paper presents the detail investigation of the detection and prediction of rainfall using machine learning approach. The results show the efficiency of different methods on precision, sensitivity, specificity, and accuracy. The work will enhance the way machine learning approach be used in agriculture. The Support Vector Machine, K-Nearest Neighbor (KNN), Decision Tree Neural Networks were implemented in MATLAB and tested with rainfall data sets available at Indian meteorological Department (IMD). Cascade and pattern recognition Neural network are compared with the new approach where the nearest neighbor numbers in KNN algorithm is optimized using the ACO and CS optimization algorithm. The precision, sensitivity, specificity, and accuracy parameters were investigated for different methods. In future hybrid NN models will be implemented with optimized hidden nodes. The future work more focus on increased optimization level in rainfall prediction .

REFRENCES:

[1]. V. Nourani, Z. Razzaghzadeh, A.H. Baghanam, et al, “ANN-based statistical downscaling of climatic parameters using decision tree predictor screening method,” Theor Appl Climatol, vol. 137, pp. 1729–1746, 2019, doi: 10.1007/s00704-018-2686-z.

[2]. X. Zhang, H. Jiang, J. Jin, X. Xu and Q, “Zhang Analysis of acid rain patterns in northeastern China using a decision tree method,” Atmospheric Environment, vol. 46, pp. 590–596, 2012, doi: 10.1016/j.atmosenv.2011.03.004.

[3]. M. S. Tehrany, B. Pradhan and M. N. Jebur, “Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS,” Journal of Hydrology, vol. 504, pp. 69–79, 2013, doi: 10.1016/j.jhydrol.2013.09.034.

[4]. N. Hasan, N. C. Nath and R. I. Rasel, "A support vector regression model for forecasting rainfall," 2015 2nd International Conference on Electrical Information and Communication Technologies (EICT), Khulna, pp. 554-559, 2015, doi: 10.1109/EICT.2015.7392014.

[5]. S. Zhao and L. Wang,” The Model of Rainfall Forecasting by Support Vector Regression Based on Particle Swarm Optimization Algorithms,” In: Li K., Fei M., Jia L., Irwin G.W. (eds) Life System Modeling and Intelligent Computing, ICSEE 2010, LSMS 2010, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, vol. 6329, 2010, doi: 10.1007/978-3-642-15597-0_13.

[6]. J. N. K. Liu, B. N. L. Li and T. S. Dillon, "An improved naive Bayesian classifier technique coupled with a novel input solution method [rainfall prediction]," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 31, no. 2, pp. 249-256, May 2001, doi: 10.1109/5326.941848.

[7]. D. Gupta and U. Ghose, "A comparative study of classification algorithms for forecasting rainfall," 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), Noida, pp. 1-6, 2015, doi: 10.1109/ICRITO.2015.7359273.

[8]. M .Huang, R. Lin, S. Huang and T. Xing, ‘‘A novel approach for precipitation forecast via improved K -nearest neighbor algorithm,’’ Advanced Engineering Informatics, vol.33, pp. 89-95, 2017, doi:10.1016/j.aei.2017.05.003.

[9]. 2Z. Jan, M. Abrar , S. Bashir , A.M Mirza, ‘‘Seasonal to Inter-annual Climate Prediction using Data Mining KNN Technique,” 2008, In: D.M.A Hussain, A.Q.K. Rajput , B.S Chowdhry B.S., Gee Q. (eds) Wireless Networks, Information Processing and Systems. IMTIC 2008.Communications in Computer and Information Science, vol. 20, Springer, Berlin, Heidelberg, 2008,doi:10.1007/978-3-540-89853-5_7.

[10]. S. N. Lathifah, F. Nhita, A. Aditsania and D. Saepudin, "Rainfall Forecasting using the Classification and Regression Tree (CART) Algorithm and Adaptive Synthetic Sampling (Study Case: Bandung Regency)," 7th International Conference on Information and Communication Technology (ICoICT), Kuala Lumpur, Malaysia, 2019, pp.1-5,2019, doi: 10.1109/ICoICT.2019.8835308.

[11]. M. Devak, C.T Dhanya and A.KGosain, ‘‘Dynamic coupling of support vector machine and K-nearest neighbour for downscaling daily rainfall,’’ Journal of Hydrology, vol.525, pp. 286–30, 2015, doi:10.1016/j.jhydrol.2015.03.051.

[12]. M. kbari, P.J.V Overloop and A. Afshar, ‘‘Clustered K Nearest Neighbor Algorithm for Daily Inflow Forecasting,’’Water Resources Management, vol.25, pp.1341–1357, 2011, doi.org/10.1007/s11269-010-9748-z.

[13]. P.S Yu, T.C Yang, S.YChen, C.M Kuo and H.W Tseng, ‘‘Comparison of random forests and support vector machine for real time radar derived rainfall forecasting,” Journal of Hydrology, vol. 552, pp. 92–104, doi:10.1016/j.jhydrol.2017.06.020.

[14]. G.F. Lin, Y.C Chou and M.C Wu, ‘‘Typhoon flood forecasting using integrated two-stage Support Vector Machine approach,” Journal of Hydrology, vol.486, pp. 334–342, 2013, doi: 10.1016/j.jhydrol.2013.02.012.

[15]. D.C. R. Novitasari, H. Rohayani, Suwanto, Arnita, Rico, R. Junaidi, Rr. D. N. Setyowati, R. Pramulya and F. Setiawan, “Weather Parameters Forecasting as Variables for Rainfall Prediction using Adaptive Neuro Fuzzy Inference System (ANFIS) and Support Vector Regression (SVR),” Journal of Physics, Conference Series, International Conference on Science & Technology, Yogyakarta, Indonesia, vol. 1501, pp. 1-11, 2019, issue. 2-3, doi: 10.1088/1742-6596/1501/1/012012.

[16]. S. Georganos, A. M. Abdi, D. E. Tenenbaum and S. Kalogirou, “Examining the NDVI-rainfall relationship in the semi-arid Sahel using geographically weighted regression,” Journal of Arid Environments, vol. 146, pp. 64–74, 2017, doi: 10.1016/j.jaridenv.2017.06.004.

[17]. S. Cramer, M. Kampouridis, A.A. Freitas and A.K. Alexandridis, “An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives,” Expert Systems with Applications, vol. 85, pp. 169–181, 2017, doi: 10.1016/j.eswa.2017.05.029.

[18]. J. George, L. Janaki and J. Parameswaran Gomathy, “Statistical Downscaling Using Local Polynomial Regression for Rainfall Predictions – A Case Study,” Water Resour Manage, vol. 30, pp. 183–193, 2016, doi: 10.1007/s11269-015-1154-0.

[19]. X. Zhang, S. N. Mohanty, A. K. Parida, S. K. Pani, B. Dong and X. Cheng, "Annual and Non-Monsoon Rainfall Prediction Modelling Using SVR-MLP: An Empirical Study From Odisha," in IEEE Access, vol. 8, pp. 30223-30233, 2020, doi: 10.1109/ACCESS.2020.2972435.

[20]. R. C. Deo, O. Kisi and V. P. Singh, “Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model,” Atmospheric Research, vol. 184, pp. 149–175, 2017, doi: 10.1016/j.atmosres.2016.10.004.

[21]. A.Kusiak, X. Wei, A. P. Verma and E. Roz, "Modeling and Prediction of Rainfall Using

Radar Reflectivity Data: A Data-Mining Approach," in IEEE Transactions on Geoscience and Remote Sensing, vol. 51, no. 4, pp. 2337-2342, April 2013, doi: 10.1109/TGRS.2012.2210429.

[22]. S. Bhomia, N. Jaiswal, C. M. Kishtawal and R. Kumar, "Multimodel Prediction of Monsoon Rain Using Dynamical Model Selection," in IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 5, pp. 2911-2917, May 2016, doi: 10.1109/TGRS.2015.2507779.

[23]. F.S. Marzano, G. Rivolta, E. Coppola, B. Tomassetti and M. Verdecchia, "Rainfall Nowcasting From Multisatellite Passive-Sensor Images Using a Recurrent Neural Network," in IEEE Transactions on Geoscience and Remote Sensing, vol. 45, no. 11, pp. 3800-3812, Nov. 2007, doi: 10.1109/TGRS.2007.903685.

[24]. J. Pucheta, C. Rodriguez Rivero, M. Herrera, C. Salas and V. Sauchelli, "Rainfall Forecasting Using Sub sampling Nonparametric Methods," in IEEE Latin America Transactions, vol. 11, no. 1, pp. 646-650, Feb. 2013, doi: 10.1109/TLA.2013.6502878.

[25]. A.Haidar and B. Verma, "Monthly Rainfall Forecasting Using One-Dimensional Deep Convolutional Neural Network," in IEEE Access, vol. 6, pp. 69053-69063, 2018, doi: 10.1109/ACCESS.2018.2880044.

[26]. P. Zhang, Y. Jia, J. Gao, W. Song and H. Leung, "Short-Term Rainfall Forecasting Using Multi-Layer Perceptron," in IEEE Transactions on Big Data, vol. 6, no. 1, pp. 93-106, 1 March 2020, doi: 10.1109/TBDATA.2018.2871151.

[27]. C. Rodriguez Rivero, J. Pucheta, M. Herrera, V. Sauchelli and S. Laboret, "Time Series Forecasting Using Bayesian Method: Application to Cumulative Rainfall," in IEEE Latin America Transactions, vol. 11, no. 1, pp. 359-364, Feb. 2013, doi: 10.1109/TLA.2013.6502830.

[28]. S. M. Hosseini and N. Mahjouri, “Integrating Support Vector Regression and a geomorphologic Artificial Neural Network for daily rainfall-runoff modeling,” Applied Soft Computing, vol. 38, pp. 329–345, 2016, doi:10.1016/j.asoc.2015.09.049.

[29]. Q. Ouyang and W. Lu, “Monthly Rainfall Forecasting Using Echo State Networks Coupled with Data Preprocessing Methods,” Water Resour Manage, vol. 32, pp. 659–674, 2018, doi: 10.1007/s11269-017-1832-1.

[30]. Wen-jingNiu, Zhong-kai Feng, Wen-fa Yang and Jun Zhang, “Short-term streamflow time series prediction model by machine learning tool based on a data preprocessing technique and swarm intelligence algorithm,” Hydrological Sciences Journal, 2020, doi: 10.1080/02626667.2020.1828889.