Received: January 26, 2024. Revised: February 20, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

A Hybrid Feature Selection Gradient Recurrent Neural Network (HFSGRNN) Model for Rainfall Prediction in India Regions

Nishant N. Pachpor[1]* **B. Suresh Kumar[2]** **Prakash S. Prasad[3]** **Salim G. Shaikh[4]**

[1]*Amity University Jaipur & IIMS, Chinhcwad, Pune, India*
[2]*Amity University Jaipur, India*
[3]*Priyadarshini College of Engineering, Nagpur, India*
[4]*Kalesekar, Panvel, India*
* Corresponding author's Email: nishantpachpor@gmail.com

Abstract: In current studies, India is a farming country, and the accomplishment or disappointment of the crop mainly depends on the country's rainfall design. Generally, India's farming production is primarily based on the nature of the precipitation of the rainy season rainfall. The rainy season is the primary source of water in India. Regular rainfall forecasting is the primary source for crop development. Several analyses have defined the direct effect of rainwater on harvests. The main motive of this research work is proper and early rainfall prediction, which is helpful to people who live in northeast regions inclined to natural disasters like floods, etc. It helps agriculture with decision-making in their crop and water management (WM) using extensive dataset analysis that generates maximum terms of production for farmers and profits. This proposed work introduced an improved rainfall forecasting framework, a hybrid feature selection gradient-based RNN (HFSGRNN) model with an RNN algorithm. The research uses the HFSGRNN model steps, such as initial data preprocessing steps, which are used for forecasting rainfall, handling missing value outliers, and typecasting the rainfall dataset collected from the government site. After that, an HFSGRNN method is implemented to select the valuable using stochastic gradient descent (SGD) and optimal solutions calculated by particle swarm optimization (PSO) from the preprocessed data. The hybrid optimized feature sets are fed to the rainfall forecasting of the RNN classifier. Lastly, the valuable feature sets are forecasted using decision-making, and the simulation outcome shows that the research approach performed better in rainfall forecasting. The simulation results define that the HFSGRNN model delivered the minimum value of Root means square error (RMSE= 0.10) and maximum value of accuracy rate (acc = 98.1%) compared with existing methods, such as logistic regression (LR), Long Short Term –Convolutional Neural Network (LSTM-CNN), etc. The outcomes of the research analysis will help the farmers accept efficient modeling methods for forecasting long-term seasonal rainfall.

Keywords: Rainfall forecasting, Rainfall dataset, Hybrid FSGRNN (Feature Selection Gradient Recurrent Neural Network), Recurrent neural network, Long short term memory- Convolutional neural network.

1. Introduction

Rain has been the most significant variable in water resources and climate domains in recent years. Precise and accurate prediction methods for rainfall forecasting can be efficiently practical to study the further incidence of flood and drought actions, improve irrigation practices, and predict natural disasters [1, 2]. Moreover, precisely indicated rainfall data may be utilized to model further variations in stream flow and accurate flood control construction (FCC) [3, 4]. Rainfall is a crucial component in the agricultural sector. It refers to the amount of rain an area receives over a period, usually measured in millimeters. Rainfall patterns can significantly influence crop growth and yield [5]. In agriculture, rainfall is a crucial determinant of crop productivity. It provides the necessary water for plant growth and development. However, the distribution and variability of rainfall can significantly impact crop production. For instance, if there is a decrease in precipitation or if the rainfall is unevenly distributed, it can lead to adverse effects on crops like rice [6]. Changes in rainfall patterns can also disrupt the cultivation of crops, leading to food insecurity. It can reduce agricultural productivity, leading to food shortages and increased prices. This traditional knowledge has been crucial in adapting to weather and climate variability. Conventional methods of rainfall prediction encompass a wide array of approaches, including indigenous knowledge, such as the Afar in Ethiopia, that rely on observations of natural phenomena like animal behaviour and celestial bodies to forecast rainfall [7].

Existing methods for rainfall prediction are classified into two main classes: arithmetical and machine learning (ML) techniques. Arithmetical techniques encompass a generalized additive model (GAM) and k-nearest neighbor (KNN). On the other hand, ML methods consist of random forest (RF), support vector machine (SVM), artificial neural network (ANN), and extreme gradient boosting (XGBoost) [8, 9] methods. Statistical methods focus on historical patterns and relationships, while ML methods leverage algorithms to learn and predict based on data patterns. These methods offer several advantages, including the ability to provide accurate forecasts. These methods have evolved over the years, with the emergence of intelligent computing techniques like Artificial Neural Networks and machine learning algorithms, which have significantly improved the accuracy of rainfall predictions [10]. The development of rainfall prediction methods has facilitated the determination of growing seasons for crops, contributing to the organization and flexibility of farming systems (FS). This analysis identified [11] several gaps in existing work:

1. *Data Reliability and Integrity:* If the data used for prediction is compromised for any reason, it raises doubts about the reliability of the projections. Additionally, malfunctions in weather sensors can adversely impact the accuracy of the rainfall prediction system [12].

2. *Cost and Complexity of Merit Techniques*: Many techniques with promising results in rainfall prediction are often expensive and challenging. These techniques' high cost and complexity may pose barriers to widespread adoption, limiting their practicality and accessibility for various stakeholders [13].

3. *Dependency on Seasonal Conditions for Agriculture:* Despite advancements in rainfall prediction, farmers rely heavily on traditional seasonal conditions to determine the optimal time for planting crops [14].

S. Samantaray, S. S. Das, A. Sahoo, D. P. Satapathy (2022) [15] proposed a hybrid SVM-SSA method for monthly runoff forecasting in the Baitarani River basin in India. Rainfall, temperature, stage, specific humidity, and relative humidity data from 1991-2020 were collected. Inspired by the salp swarm behaviour, the SSA algorithm optimized the SVM parameters, like kernel width, penalty factor, etc., by minimizing the misclassification rate. J. Zhao, R. Chen and H. Xin (2022) [16] proposed a combined autoregressive intergraded moving average model with a radius bias function (ARIMA-RBF) model for rainfall prediction. They use monthly rainfall information from January 1951 to May 2015 in Nanchang City, China. The collective ARIMA-RBF model shows better forecast performance than just ARIMA or RBF alone. The RMSE reduced from 2.798 for ARIMA and 1.724 for RBF to 0.366 for the combined model. S. Poornima, and M. Pushpalat (2019) [17] implemented an Intensified LSTM-based R-CNN for rainfall forecasting. The proposed Intensified LSTM modifies the input and candidate gates of the standard LSTM unit by multiplying the input with the activations to mitigate vanishing gradients. It allows the network to learn long-term needs in time series data. The Intensified LSTM model achieves lower RMSE (0.33), loss (0.0054), and higher accuracy (88%). G. Meti, and R. K. Gunj, (2023) [18] proposed a new ML approach known as "Analogousness Enhanced Rainfall Predictor." The technique uses the XGBoost algorithm as the backbone. It tunes its parameters for higher accuracy—a k-fold cross-validation method used to search optimal hyperparameter values for the balanced XGBoost model. The resultant model achieved an accuracy of 83%.highlighting its efficiency.

This article defines a new model to improve rainfall forecast prediction to identify these challenges. The input data are initially gathered from the Kaggle and the Government's official sites [19]. Then, preprocessing steps are applied to the gathered rainfall forecast data to find missing values, noise reduction, normalize data, and give better quality without losing data. Next, to improve the normalized data, hybrid feature selection has been proposed using particle swarm optimization (PSO)[20] and stochastic gradient descent (SGD)[21] to choose the feature vectors. Merging two feature selection methods includes advantages like reducing dimensionality, reducing error rate, and enhancing the accuracy rate. The nature-inspired feature selection gradient (FSG) proposed to choose the optimal feature vectors. The FSG approach is proposed based on swarm optimization to solve complex optimization problems, which improves the prediction accuracy rate. The obtained optimized feature vectors are fed to the recurrent neural network (RNN) to classify rainfall forecasting prediction. For the model calculation motive, the performance precision of the hybrid model is compared with other models trained using a hybrid of PSO and SGD methods. The last implementation of the mixed FSGRNN method is analyzed in rainfall predicting issues for the northeast region of India. The experimental analysis, the proposed hybrid FSGRNN model, is explained in terms of root mean square error (RMSE), accuracy (Acc), precision (Pre), recall (Rec), and F1-score. Simulation result analysis defined that the proposed hybrid FSGRNN method attained a 98% accuracy rate and 0.104 RMSE on the rainfall forecasting dataset. The proposed approach outperformed the rainfall forecasting dataset compared to the previous methods, such as long short-term memory (LSTM) with convolutional neural network (CNN) [22] and logistic regression (LR) [23] models.

The rest of the paper is planned as trials: Section 2 offers a literature review of recent research papers on "Rainfall Forecasting Prediction." Section 3 outlines the proposed methods, such as CNN, LSTM, LSTM-CNN, PSO, SGD, and RNN. Section 4 shows the research methodology, the dataset description, the proposed steps, and the implemented hybrid FSGRNN model. Section 5 presents the statistical metrics and simulation outcomes of the proposed hybrid FSGRNN approach, and Section 6 concludes the paper, outlining future scope.

2. Literature work

Several analyzers have detailed the rainfall forecasting methods. Some of the review and research articles elaborated in this section between them. In 2020, I. *Cholissodin, and S. Sutrisno* [24] proposed an innovative approach to rainfall prediction by combining deep learning (DL) with PSO. It transformed rainfall time series data into image matrices, enabling the utilization of a D-CNN extreme learning machine (DLCNN-ELM). It employed PSO to optimize weight selection in the DLCNN-ELM, aiming to enhance accuracy and mitigate the risk of local optima. The study focused on a rainfall dataset from the BMKG Meteorology Station in Malang Regency, East Java. Results demonstrated the method's effectiveness, achieving a remarkable lowest mean absolute deviation (MAD) of 0.3418. In 2019, *S. Poornima, and M. Pushpalat* [17] developed an Intensified LSTM NN model designed for enhanced rainfall prediction. The primary innovations of their model included the multiplication of LSTM input gate activations and candidate vectors by the inputs, strategically implemented to mitigate vanishing gradients. This intensification process facilitates more effective learning from the intricate rainfall time series data patterns. The authors employed the Adam optimization technique during backpropagation training to optimize the network, ensuring efficient weight updates. The model underwent training on a comprehensive 34-year meteorological dataset and was subsequently tested on rainfall data from 2014. Notably, their intensified LSTM model outperformed Baseline LSTM, RNN, ARIMA, and other models, achieving a remarkable 88% accuracy, lower Root Mean Square Error (RMSE) at 0.33, and minimized losses. In 2020, *M. A. Duhayyim, H.G. Mohamed, J. S. Alzahrani, R. Alabdan, M. Mousa, A.S. Zamani, and M. I.Alsaid [25]* developed an FCMM-RPS method to forecast rainfall spontaneously and accurately. They preprocessed the data by cleaning and normalizing it. They used an FCM model to classify and predict rainfall. FCMs can capture complex causalities. They optimized the FCM parameters with an MBOA to enhance prediction accuracy. They tested their technique on a rainfall dataset with 25,919 samples. The FCMM-RPS technique outperformed standalone FCM and other recent machine-learning methods in accuracy (94.22%), precision (94.72%) and recall. They tackled the challenges of rainfall's chaotic, nonlinear behaviour, parameter tuning of classification models, and accuracy limitations of existing techniques. The FCMM-RPS technique demonstrated better rainfall prediction abilities. In 2022, *W. Li, X. Gao, Z. Hao, and Rong Sun* [26] established a CNN-LSTM approach for short-term rainfall forecasting in Lanzhou, China, a semi-arid region. They used meteorological data from 25 stations, with feature engineering and Mutual Information feature selection. They applied ADASYN oversampling to balance the rainfall data. CNN-LSTM network used CNN to learn spatial patterns and LSTM to learn temporal dependencies. They evaluated their model on 2013-2019 data and showed higher accuracy than SVM and Random Forest. They addressed the challenges of sparse and volatile Rainfall, complex and nonlinear atmospheric processes, and sample imbalance. Their CNN-LSTM model showed better rainfall prediction

Table 1. Analysis of the Rainfall forecasting prediction models using ML and DL

Authors Name	Proposed Methods	Features	Drawbacks/Limitations	Dataset/Tool	Metrics	Findings
I. holissodin, and S. Sutrisno, (2020) [24]	Improved PSO-DLCNN	1. Uses feature extraction and data transformation into image form. 2. Applies convolution and pooling layers based on problem complexity. 3. Optimizes deep learning weights using PSO for better accuracy.	1. Deep learning can get stuck in local optima due to random weight initialization. 2. Rainfall prediction is challenging, with an accuracy of around 70% when using previous methods.	Dataset: The dataset was collected from a flat place without any experience. Tool: MATLAB	Mean absolute Deviation (MAD) = 0.3418	The outcomes from the implemented method have successfully defined an enhancement as in the attained test analysis with a minimum avg. value of MAD from PSODLCNNELM – 0.38418
S. Poornima (2019) [17]	Intensified LSTM-based RNN	1. It uses the sigmoid activation function (AF) increased with input for the input gate. 2. Uses tanh AF increased with input for candidate vector 3. Employs Adam optimizer for training 4. Leverages long-term temporal contexts	1. Vanishing gradients in standard LSTM models. 2. Rainfall prediction is complex and highly nonlinear.	Dataset: Standard daatset Tool: Python Keras	Accuracy = 88% RMSE = 0.33 Losses = 0.0054 Learning Rate = 0.025	The proposed model attained better outcomes in minimum epochs and mitigates the running time, which helps process and analyze big data (BD).
M. A. Duhayyim (2023) [25]	FCMM-RPS	1. Pre-processes data via cleaning and normalization 2. Uses Fuzzy Cognitive Map (FCM) model for prediction 3. Tunes FCM parameters with Modified Butterfly Optimization Algorithm (MBOA)	1. Complex and chaotic nature of rainfall patterns. 2. Parameter tuning is challenging 3. Existing models have limitations in accuracy	Dataset: Historical meteorological datasets. Tool: Python	Accuracy = 94.22% Precision = 94.72% NPV = 94.72%	The proposed model improved the maximum accuracy by 94.2% compared with other methods.
W. Li, X. Gao (2021)	CNN-LSTM	1. Uses meteorological data from 25	1. Quantitative precipitation forecasting is difficult in semi-arid	Dataset: Meteorological dataset of 25	Precision = 0.74 Recall =	ADASYN method and mutation

[26]		stations Employs Mutual Information (MI) for feature extraction Handles imbalanced data via Adaptive Synthesis (ADASYN) oversampling CNN captures spatial patterns, LSTM extracts temporal dependencies	regions 2. Rainfall patterns are complex, nonlinear and stochastic 3. Sample imbalance between rain and no-rain events	stations in Lanzhou and the nearest cities.	0.77 F1 = 0.74 CSI = 0.63	information (MI) feature selection approach, the LSTM-CNN model introduced an efficient semi-arid field precipitation forecasting method.
R. Venkata B. Ramana, B. S. R. Krishna (2013) [27]	Wavelet and ANN	1. Uses Discrete Wavelet Transform to decompose rainfall time series 2. Extract details and approximation components 3. Uses components as input to ANN	1. Rainfall has chaotic, random characteristics, making it difficult to predict Complex nonlinear relationships between rainfall and other variables Require large historical datasets for prediction model training.	Dataset: Himalayan foothills (Darjeeling)	RMSE = 226.0 R = 0.65 COE = 37.7	Proposed model metrics were standardized using 44 years of information, and left information was utilized for model validation. The research analysis explored which effective index is more than 94% for WNN methods, and the ANN model achieved 64%.

performance. In 2013, *R. Venkata B. Ramana, B. S. R. Krishna, and N.G. Pandey,* [27] proposed a hybrid wavelet NN (WNN) model for monthly rainfall prediction. The method combines wavelet analysis for decomposing the rainfall time series with an ANN model for forecasting. The wavelet transform decomposes the rainfall data into multi-resolution components called approximations and details. These components better capture the dynamic patterns in the data compared to the raw data. The WNN model was applied to predict monthly Rainfall in Darjeeling, India, using past rainfall, minimum temperature, and maximum temperature data. The results showed that the WNN models significantly outperformed regular ANN models regarding prediction accuracy metrics like RMSE, correlation coefficient, and efficiency. The WNN model could forecast monthly rainfall with over 94% efficiency. The authors demonstrate that decomposing complex rainfall data into sub-components using wavelets captures more helpful information for training machine learning models.

2.1 Problem statement

Rainfall forecasting is significant since unbalanced and heavy rainfall may have several impacts, like damage to farms and crops [28]. Impairment of property, so a better rainfall forecasting method is needed for an early-stage warning, which can be required for risk of life and property and helps achieve farming better. Heavy rainfall is a reason for natural disasters, such as floods and droughts [29]. Several methods have been developed to calculate rainfall and predict heavy rain. These methods depend on the supervised and un-supervised ML methods. Precision is the primary concern in machine and deep learning. In this research analysis, we will understand the information and then train the proposed hybrid

Received: January 26, 2024. Revised: February 20, 2024.

FSGRNN model accordingly to predict whether it rains under defined situations. Therefore, various DL and ML-based approaches have been implemented for the rainfall prediction model; features, challenges, and Outcomes are planned in Table 1.

Abbreviations: LSTM: Long Short Term Memory, FCMM-RPS: Fuzzy Cognitive Maps with Meta-heuristics-based Rainfall Prediction System, PSO: Particle Swarm Optimization, DLCNN: Deep Learning Convolutional Neural Network, CNN-LSTM: Convolutional Neural Network- Long Short Term Memory, ANN: Artificial Neural Network, RMSE: Root Means Square Error, COE: Coefficient of Efficiency, R: Correlation Coefficient, ADYSN: Adaptive Synthesis Oversampling.

3. Proposed methods

The main motive of this research is to create a novel rainfall forecasting model that can be utilized as a decision tool to predict the future. This section elaborates on the proposed methods, such as logistic regression (LR), LSTM-CNN, PSO, SGD, CNN, and LSTM, that are used in comparative analysis.

3.1 Convolutional neural network (CNN)

Fig 1 shows the CNN for monthly rainfall forecasting, a system that forecasts the monthly rainfall for a chosen site [38]. Neural Network (NN) with convolutional operators is known as CNN. This method depends on the DL outline, which takes an input dataset and assigns significance to different attributes in the input data values and is required to be able to differentiate from each other. The data preprocessing is necessary and minimal in comparison with several classification methods. CNN layers can be studied from the filters.

The CNN [30] architecture resembles the connectivity of neuron patterns, like the brain design shown in Fig 2. The visual cortex implements this design, individual neuron designs in the human brain that respond to a stimulus in a restricted region defined as the receptive domain. A group of these domains overlap and encompass the complete environment.

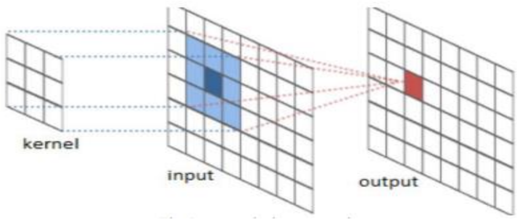

Figure. 1 CNN operations [38]

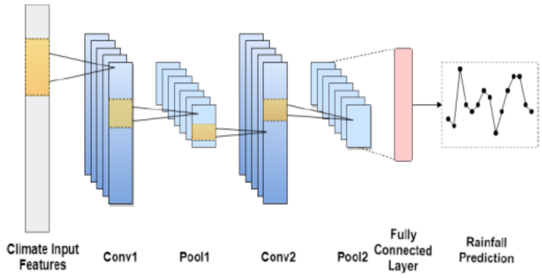

Figure. 2 Design of CNN Model [38] to predict t

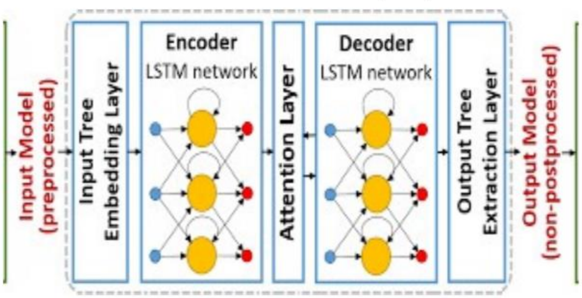

Figure. 3 Architecture of LSTM Model [31, 32]

3.2 Long short-term memory (LSTM)

It is an ANN in a DL framework. It is reliable in execution functions, speech recognition, machine translation, video processing, and the medical sector. Its cells are utilized for classification motives, prediction, and processing in time-series information. LSTM [31] units are proficient in resolving gradient issues. These networks suffer from gradient issues. Fig 3 shows the general architecture of the LSTM model.

3.3Hybrid LSTM-CNN model

This hybrid method, the LSTM method, is utilized to examine short and long-term dependencies. Meanwhile, the CNN method's convolution layers (CLs) are used for feature extraction through input data. This classification method involves a maximum computational analysis (CA) of several train metrics. The framework described below in Fig 4 is to design a feature extraction approach with kernels in the minimum layers. This model optimizes overfitting and reduces the number of metrics utilized for the training process. In the sequence, they are using three distinct LSTM layers [32]. The series of hidden layers meanwhile in the LSTM layer the false network utilized to outcome the hidden layer as the final phase. The no. of layers in the output varies between distinct neurons in the last layer. The input

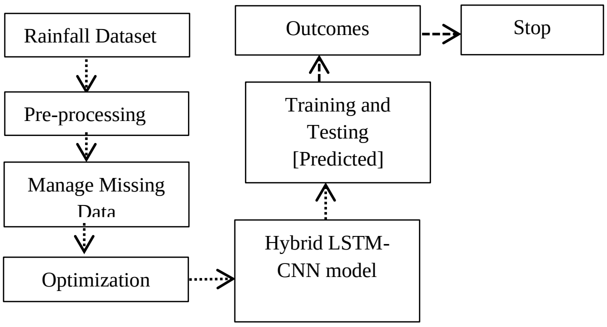

Figure. 4 Architecture of hybrid LSTM-CNN model

data values are divided into a 70:30 ratio as training and testing. It evaluates the RMSE value and other metrics.

3.4Logistic regression (LR)

The machine Learning (ML) method is the most famous LR method under supervised learning (SL) approaches. The LR method predicts the class reliant on a variable from a set of autonomous variables. So, in the LR method, the outcome must be isolated (0,1). The benefits of the LR are that the designed approach is easy to compare to the others, has minimum train time, and extends to various predictions, known as multinomial regression, when the data is linear and separable, giving maximum correctness rate. LR approach has to fit an S-shaped LR function in its place of providing a regression line to predict two maxima values (0,1). This line curve shows whether rainfall will come or not. Regulate the most efficient variables from dissimilar LR feature categories [33]**.** Fig 5 defines the sigmoid function (SF) of the arithmetical method used to predict the results. The LR value must be 1 and 0; thus, the curve represents such an S-shape. It is known as sigmoid or LR.

3.5Particle swarm optimization (PSO)

Depending on the populace, it is a swarm intelligence (SI) optimization approach. Each

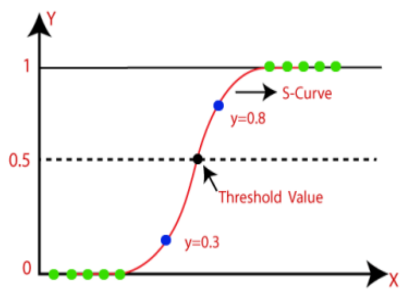

Figure. 5 The curve of LR function [33]

particle of this method has its position and velocity vector (VV). Position and VVs define a reliable solution to an issue. Class determines the rank assigned to train data during standardization. The velocity vector is the running time taken by the train data. Each particle saves its best global position through communication with its nearest particles. It handles and explores by altering the location of particles and VV [34]. The fitness or objective function represents the movement of particles. Particles closest to the optimal solution (OS) have a minimum velocity. Particles distant from the OS have a maximum speed.

3.6 Stochastic gradient descent (SGD)

It is the most famous optimization method in ML and DL fields. A gradient is a *"slope"* method, the degree of alteration of a metric with the amount of alteration in another metric. Arithmetically, it can be defined as the particle calculations of different metrics w.r.t its inputs. Generally, SGD is a convex function (CF). It can be defined as an iterative approach [35] utilized to explore the values of a metric procedure, which reduces the cost function (CF).

3.7 Recurrent neural network (RNN)

DL approaches have been implemented to resolve the existing issues in climate using modelled data. It resembles the workings of the human brain, like transferring and getting data. In the same way, the RNN method's work is that the human brain can also transmit and get information from one neuron to another. Typically, the human brain acts to make decisions, and in the process of making rational decisions, given considerations of the previous. It is essentially an ANN method that utilizes recurrence by using previous information. RNN [36] is used for prediction and classification purposes.

4. Research methodology

The problem elaborated in this paper is forecasting precise seasonal rainfall for the following season for the next year based on the prior weather database by using consistent earlier data. The proposed HFSGRNN model was applied to the northeast dataset. Rainfall prediction, the proposed HFSGRNN method, defines different steps:

- Input data is collected from Kaggle and Govt. sites.
- Preprocessing and typecasting are applied to handle the missing values.

- Hybrid feature selection (FS) using PSO and SGD methods.
- A new feature selection gradient method has been implemented for FS.
- Rainfall forecasting using the RNN classification model.
- Valuable feature sets of rainfall data are attained utilizing the decision-making equation.

The proposed workflow HFSGRNN model is defined in Fig 6.

The HRSGRNN flow shows various steps to process the data and compute the outcome. Data collection includes a source of dataset considered in the research. The govt portal for datasets and the Kaggle data repository are used to collect the datasets. The collected data is processed with the preprocessing phase to see the missing values, data element type conversions, variables selection, etc. This process helps to reduce the error probability during the training and testing of the proposed model. The normalized data was processed with a hybrid optimizer based on the processing architecture of swarm intelligence and SGD optimizer. Optimizer is the process of using the data to build the patterns and reduce the error probability from the outcome. With the help of various steps in algorithm 1, the optimized data is computed and set

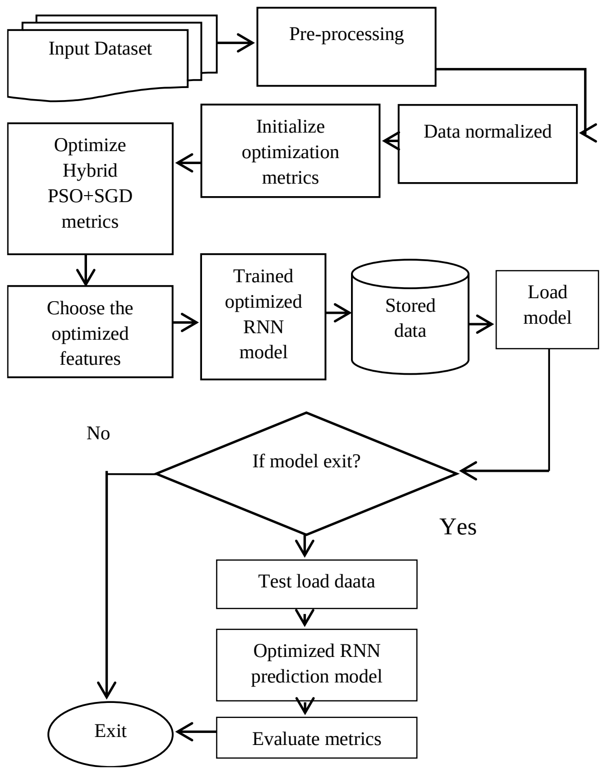

Figure. 6 Proposed methodology of the HFSGRNN method

as a training/ test set for the RNN model. The RNN model is built with various layers, forward epochs, and backpropagation modules to analyze the associated training patterns target sets. The trained model is loaded in the test phase to generate the predicted values for the test set. The test set involves the confusion metrics used to compute the presentation of the trained model.

4.1 Dataset explanation

The Kaggle [19] and data.gov sites [37] were used for the datasets collected in the research analysis. The Indian Government has conducted several research analyses of global warming (GW) and climate alteration on rainfall design in India. The studies utilized observed rainfall information from 3k rain device stations nationwide for one hundred and fifteen years (1901-2015). The main inferences from this analysis depend on the one-hundred-and-fifteen-years of rainfall information are as follows:

The research analysis of one hundred and fifteen years of monsoon rainfall information suggests no longer alteration in the monsoon rainfall avg. over the region. However, there are no variations in India's rainfall; there are essential variations in yearly rainfall in some meteorological sub-sections. Rainfall over Kerala, East MP, JK, AP, NMMT, etc., defines minimizing trends. Rain over coastal Maharashtra, J&K, and Karnataka represents an improving trend, as shown in Fig 7.

The general trend is maximizing the frequency of heavy rainfall events over India, usually over the center sections of India during the southwest monsoon season (June to Sept.) [39]. There is no proof of GW on the observed variations in yearly or seasonal rainfall over India. Moreover, there is developing evidence assuming the increasing frequency of heavy rain is due to GW. The climate variation assessment made by the inter-government

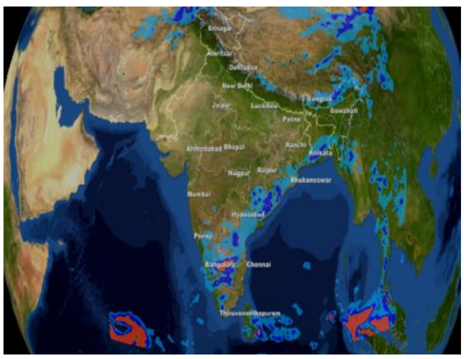

Figure. 7 Region division [19]

panel on climate variation recommends that the frequency of heavy rain may increase in GW even though there are no other long-term rainfall variations over India. That can be attributed to GW. The Indian Monsoon (IM) is explored to be a reliable system. The information defines details on climate average district-wise rainfall (in mm) evaluated with the data from 1951 to 2000 shown in Table 2.

Period – 1951 – 2015

Table 2. Dataset

STATE_UT_NAME	**DISTRICT**	**JAN**	**FEB**	**MAR**	**APR**	**MAY**	**JUN**	**SEP**	**OCT**	**NOV**	**DEC**
ANDAMAN And NICOBAR ISLANDS	NICOBAR	107.3	57.9	65.2	117	358.5	295.5	354.8	326	315.2	250.9
ANDAMAN And NICOBAR ISLANDS	SOUTH ANDAMAN	43.7	26	18.6	90.5	374.4	457.2	455.6	301.2	275.8	128.3
ASSAM	CACHAR	13.3	50.2	168.3	262.5	386.4	532.1	360.8	182.4	34.8	11.4
ASSAM	DARRANG	13.1	21.4	53.5	168.8	320	419.7	221.5	95.4	17.2	9.3
MEGHALAYA	EAST KHASI HI	15.4	24.1	129.7	312.5	733.7	1476.2	607.8	277.9	40.3	10.7
MEGHALAYA	JAINTIA HILLS	33.8	44.1	115.1	282.3	598.8	1316.1	826.3	517.7	110.9	9.7
MANIPUR	BISHNUPUR	54.5	50	112.4	108.1	159.3	435.6	219.4	237	56.9	15
MANIPUR	IMPHAL WEST	22.3	31	63.8	109.1	134.6	337.3	160.5	144.5	34.3	19.5
MANIPUR	CHURACHANDPUR	13	31.1	72.7	187.3	238	422.1	348.3	173.2	50.1	11.4
MIZORAM	AIZAWL	13.8	31.2	107.9	185.8	351.4	467.7	390.9	254.5	65.3	16.5
MIZORAM	CHAMPHAI	13.4	21.8	83	122.7	261.5	350.5	296.1	226.7	64.5	22.5
NAGALAND	MON	13.1	33.5	48.2	137.9	193.9	271.1	101.8	62.1	19.8	7.5
TRIPURA	WEST TRIPURA	9.6	27.9	72.1	194.2	359.9	426.6	250.6	174.9	41.5	10.3
TRIPURA	DHALAI	13.6	38.9	105.5	246.3	431.3	482.7	255.9	182.8	42.1	11.4
WEST BENGAL	COOCH BEHAR	8.9	16	32.2	138.9	345.4	668.8	470.9	141.3	15.1	8.3
SIKKIM	NORTH SIKKIM	61.6	98.5	199.5	238.3	355.4	503	389.7	265	43.5	22.4
WEST BENGAL	PURULIA	14.3	20.7	24.6	36.1	57.3	222.1	266.7	91.5	16.7	7.6
ORISSA	NAYAGARH	13.6	22.5	26.2	42	47.7	213.4	242.2	152.2	47.9	2.8
ORISSA	KHURDA	10.1	25.9	26.9	30.9	68	186.4	251.5	167.7	45.5	6.5
JHARKHAND	EAST SINGHBHU	14.7	18.4	20.7	31.9	63.3	225.5	225.7	68.4	12.5	5.2
BIHAR	KHAGARIA	9.5	4.2	7.7	15.4	46.3	184	265	82.2	7	3.9
BIHAR	SAHARSA	6.1	10.9	12.8	39.6	107.1	249.4	290.8	94.6	3.7	10
UTTAR PRADESH	ALLAHABAD	17.5	10	7.6	3.6	6.6	82.1	182.3	34.6	9.4	4.6

Received: January 26, 2024. Revised: February 20, 2024.

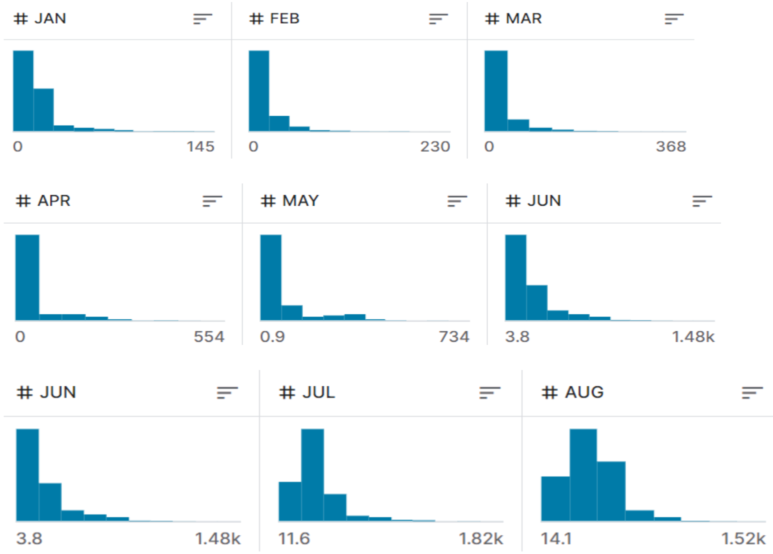

Figure. 10 Rainfall for each month: histogram plotting

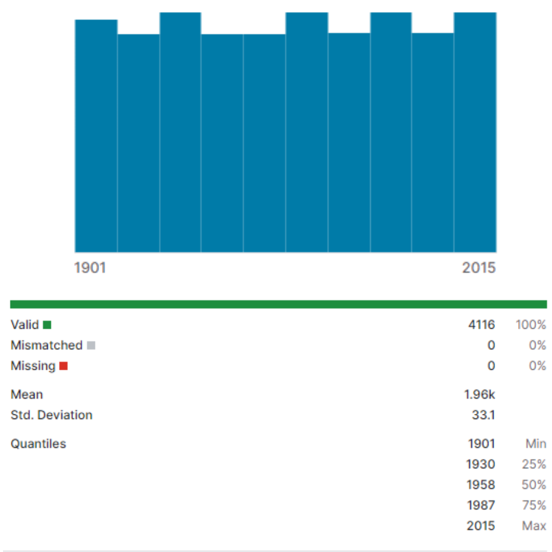

Figure. 8 Yearly rainfall in the north-east from 1951-2015

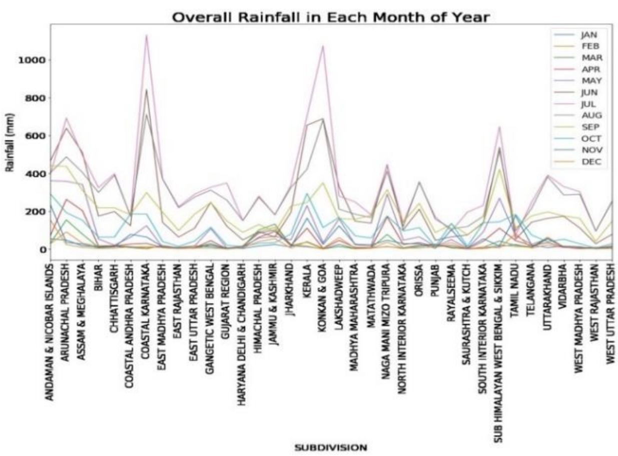

Figure. 9 Division of rainfall month-wise

Granularity – Monthly
Location – North - East
Unit of Rainfall – in mm

Initially, the statistics exposed the division of rainfall in yearly and monthly times as defined in Figs 8 and 9

4.2 Data preprocessing

The input dataset is applied to the preprocessing steps to remove unnecessary data from the collected data. Preprocessing is the most famous and significant step When preparing a dataset for a training model. Data is given unpredictably, incomplete, and comprises distorted data. After the preprocessing, maxima quality data is attained by managing missing values, outliers, scaling, type casting, etc. Using the dataset, we considered the division of rainfall from Jan to Aug using the histogram in Fig 10.

4.3 Feature selection using hybrid optimization algorithm

This research article implements a hybrid feature selection gradient (HFSG) based on PSO and SGD approaches to select the optimized feature sets from the preprocessed feature sets. In the end, the RNN method is implemented to classify the rainfall. Hybrid soft computing (SC) uses PSO and SGD methods to improve the balanced features.

The PSO method aims to optimize the model performance and network complexity—the association weights (wts) of the simulations are managed by objective function (OF) or fitness calculation Eq. (1) during standardization.

$$OF = \frac{1}{2}\sum_{P}\sum_{I=1}^{X}(DI - YI)^2 \qquad (1)$$

Here, YI = Output attained from the implemented model, DI = Target Value, Q = no. of patterns, and P = no. of nodes in the last layer.

To attain the fitness value (FV), PSO searches for the probable result of an issue and calculates its quality using RNN, the typical network. PSO explores reliable results. The reliable results rate is calculated using RNN to conclude the regular. The subsequent phases are used in the PSO-based preparation method:

- Initially, choose the architecture of the network and metrics of PSO.
- Starting locations and velocities of a populace. Everyone's positions comprise network connections.
- Depending on the grouped data, evaluate the FV of each particle using eq (1). Supposing the individual vest and worldwide best location.
- Inform position and velocity for the complete particle swarm (PS).
- If the exit situation of the method is not fulfilled, move step 3; then eliminate the epochs and acquire the best optimal solution wts. from the worldwide BS.

The combined PSO-SGD methods are due to the following steps:

- PSO mainly focuses on global search, trying to discover promising areas of the search space.
- SGD achieves local manipulation and fine-tuning. It converts speedily to local optima utilizing gradient data of 1st and 2nd order.

PSO is viewed as reliable for optimizing multimodal unconstrained nonlinear methods. They identify the merits and demerits of SGD-based techniques and swarm intelligence (SI))-based approaches. The main motive behind hybridizing PSO and SGD optimization methods is to control both ways while reducing their faults. By merging the two optimizers, we can merit global explore capabilities to escape local optima and utilize gradient-based methods to refilter the solution once a promising area is verified. Hybridization can speed up convergence, optimizing the time required to explore a high-quality solution. SGD is highly effectiveness for filtered and well-managed methods, generally in high-dimensional (HD) spaces. It trails a deterministic and clear path toward the OS, which can be beneficial in particular cases. SGD can perform unwell when dealing with distorted or non-filter methods. PSO is better at searching the complete search space and can give search global optima in complex, non-convex issues.

Algorithm 1. Hybrid optimization process for HFSGRNN

Input: Dataset as D, Dependent variable Dv, Independent variable IDv

output: Optimized set OD

1. Load data as D
   :map data elements{D}
2. Build Matrix of Dependent vs Independent variables
   1. Dv ≧ {D[:,1:X]}
      :X as number of columns
   2. Di ≧ {D[:,X+1]}
3. Initialize population Pi where I is max generation as Dv
4. For fp=1:max_gen
5. For p=1: pop_size
6. build positions Dp=(X,Y)
7. endFor
8. Compute F over Dp;
9. build Matrix of F as Ffp;
10. endFor
11. For epochs,e=1: N
    :E[1,..........N] total iterations
12. Process preduction y[e] :=P(F[e])
13. Process Loss as L for predicted elements.
14. Process gredient value Gw
15. Set parameters as final weight W
16. build set of best fitness evaluations as FV
17. return FV

Above, algorithm 1 has various execution steps that follow in the execution and processing of datasets. It includes the loading of data and the separation of data elements that belong to dependent and independent variables. These variables help prioritize the impact of existence during the execution of training and testing within the HDSGRNN model. Dv and Di are divisions of training and target sets that are considered to process the data separately for the next step of the execution. The hybrid optimizer initializes the population with PSO architecture and builds a deployed swam matrix using the size of maximum generations. It is a dynamic part of a deployment that belongs to the size of database elements considered swarms. The hybrid optimizer runs the fitness calculations as F over deployed particles Dp with customized iterations and builds an Ffp as a fitness matrix. The fitness matrix helps the SGD optimizer construct a priority vector and provide an optimized set. This reduced error probability-based set provides a maximum suitable set of values for a particular case. It includes the loss calculated as L, the gradient as Gw, and the weight at various iterations to find the best fitness value for the uploaded samples. The returned Optimized set is processed with the help of the training model module in the HRSGRNN.

4.4 Classification using HFSGRNN algorithm

Algorithm 2 is a combination of optimization and classification part of the proposed architecture. The processed data in algorithm 1 is collected at step 5 and loaded with various weight segments in the proposed classification architecture. It helps to understand the impact of calculated solutions in different conditions so that the model can understand the effects of particular elements on the target set. The recursive layer in the module reduces the error probability and enhances the training process of RNN. RNN runs the training part with given epochs and computes loss. If the loss is more significant than a threshold, it backpropagates to restore the training model and reduce the error probability. Once all the Epochs are completed, the model is stored and prepared for the test stage. The test stage includes the confusion matrix and calculates the performance metrics to analyze the accuracy of the trained model.

Algorithm 2. HFSGRNN processing architecture

Input: Dataset as D,
output: Predicted set PV

1. Load data as D
   :map data elements{D}
2. Preprocess D with
   1. missing value filtration
   2. format conversions etc.
3. Build Matrix of training vs target variables
   1. Ds ≥ {D[:,1,X]}
      :X as number of columns, Ds as data
   2. Dts ≥ {D[:,X,X+1]}
      :Dts as target variable
4. Optimized Ds with with Hybrid optimizer
5. Compute FV over Ds
   :FV as optimized dataset
6. init X as FV
7. Add weight segments for X as wX
8. set recursive weight Segments as rwX
9. Train RNN with [X,wX,rwX etc.]
10. Model.Train
    1. [
    2. Forward propagate as per epoch K
    3. computation of loss L
    4. Backpropagate to reduce L
    5. Restore the trained model
    6.]
11. Model.test(testSet)
12. performance evaluations as accuracy, RMSE etc. :compute over Dts
13. stop

5. Experiment result analysis

The proposed HFSGRNN model was implemented in MATLAB 2021a, and software with the system needs, operating system (OS) window 10, RAM 128GB, Intel core i7 processor, and hard disk 4TB. To evaluate the effective HFSGRNN model, the currently implemented model is compared with existing approaches, such as LR and LSTM-CNN, on the Indian rainfall prediction [19,37] dataset. In this proposed analysis, the performance of the HFSGRNN algorithm is implemented in accuracy (Acc), Precision (Pre), Recall (Rec), Root means square error (RMSE), and F1-score. The calculation performance metrics are represented in eqs (2), (3), (4), (5), and (6).

RMSE is the number of squared alterations between the outcome and explanations, and it is calculated by Eq. (2)

$$RMSE = \sqrt{\frac{\sum_{k=1}^{l}(|Avg - Forecast_{Val}|)^2}{L}} \quad (2)$$

Accuracy is a spontaneous performance metric, the ratio of precisely classified observations between the total observations by Eq. (3).

$$Accuracy = \frac{(truePos + trueNeg)}{(truePos + trueNeg + falsePos + falseNeg)} * 100 \quad (3)$$

Precision calculates how many positive classifications made are precisely (truePos). The formula is shown in Eq. (4);

$$Pre = \frac{truePos}{truePos + falsePos} \quad (4)$$

Recall calculates how many positive cases the prediction accurately predicted and the overall number of positive cases in the data shown in Eq. (5);

$$Rec = \frac{true\ Pos}{truePos + falseNeg} \quad (5)$$

F1-score calculates combining both pre and rec. It is typically defined as the harmonic mean (HM) of two. HM calculates average values, generally defined as more reliable for rations than the traditional mathematical mean shown in Eq. (6);

$$F1 - score = 2 * \frac{Pre * Rec}{Pre + Rec} \quad (6)$$

5.1 Result Analysis

The result analysis, the dataset (https://data.gov.in/catalog/district-rainfall-normal-mm-monthly-seasonal-and-annual-data-period-1951-2000 (Accessed: 02 January 2024) is assumed to analyze the calculation of the Hybrid FSGRNN algorithm. The proposed HFSGRNN method is commonly compared with existing prediction methods. In this research analysis, we measure different features in the rainfall dataset, in which 80% of the data is used for training and 20% for testing. In Table III, the HFSGRNN method performed better in the rainfall prediction than other methods using RMSE, F1-score, Pre, Rec, etc.

The implemented HFSGRNN method achieved 98.1% accuracy rate, 95.07% precision, 99.1% recall, 97.04 % of F1-score, and 0.10 RMSE value compared to the existing techniques. The implemented HFSGRNN method Rainfall dataset showed a maximum of 22.7% and a minimum 10.9% improvement in classification accuracy rate compared with other methods. The graphical representation comparison analysis of the HFSGRNN method on the Rainfall dataset (Kaggle and Govt.) is shown in Figs 10 and 11.

Table 3. Performance Evaluation of HFSGRNN Method on Rainfall Dataset

Methods	Accuracy (%)	Precision (%)	Recall (%)	RMSE	F1-score (%)
HFSGRNN	98.1	95.07	99.1	0.10	97.04
LR [33]	75.4	79.5	77.4	0.46	78.4
LSTM-CNN [32]	87.28	90.3	96.9	0.21	93.5

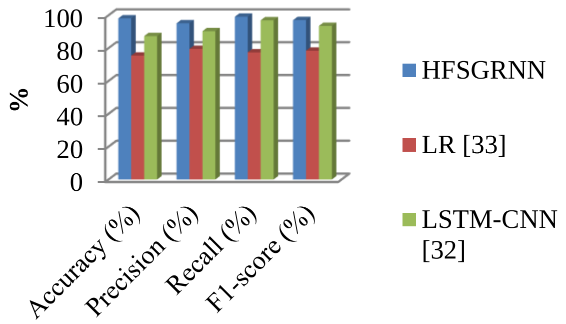

Figure. 10 Prediction Outcomes with different metrics and methods (HFSGRNN, LR, and LSTM-CNN)

Table 4. Classification Model with Different Data Dimensions and Metrics

Training Set(%)	Test set(%)	Accuracy	RMSE	Precision	Recall	F1-score
60	40	90.83	0.2104	91.45	91.99	93.22
80	20	98.10	0,10413	95.07	99.105	97.04
30	70	78.02	0.4001	78.02	82.88	81.2
70	100	99.26	0.0902	98.99	95.7	98.49
80	50	93.4	0.1291	95.2	96.78	95.99

Table 4 shows various executions that were performed with the hybrid classifier. It executes with different training and testing ratios to see the final prediction variations. It helps to determine the classifier performance and quality of features that the hybrid optimizer processes with the help of PSO and SGD modules. The highest accuracy achieved within these test scenarios is 70% training and 100

Table 5. Comparison Analysis with Accuracy Rate (%)

Methods	Accuracy (%)
HFSGRNN	98.1
LR	75.4
LSTM-CNN	87.2
LSTM-RNN [17]	88
FCMMRPS[25]	94.2
Wavelet ANN [27]	80.06

Table 6. Comparison Analysis with Precision, Recall, and F1-score

Methods	Precision	Recall
HFSGRNN	95.0	99
LR	79.5	77.4
LSTM-CNN	90.3	96.9
CNN-LSTM [26]	74	79.5
Weighted SOM-DNN [40]	83.19	82.75

Table 7. Comparison Analysis with RMSE

Methods	RMSE
HFSGRNN	0.1
LR	0.4
LSTM-CNN	0.21
LSTM-CNN [17]	0.33
WA-SVM[41]	0.126
SVR[42]	0.59
ARIMA[43]	0.4
LRA [44]	0.4
LORA[44]	0.38
RBF+HGWPSO [45]	0.32
EMLRM [46]	0.27

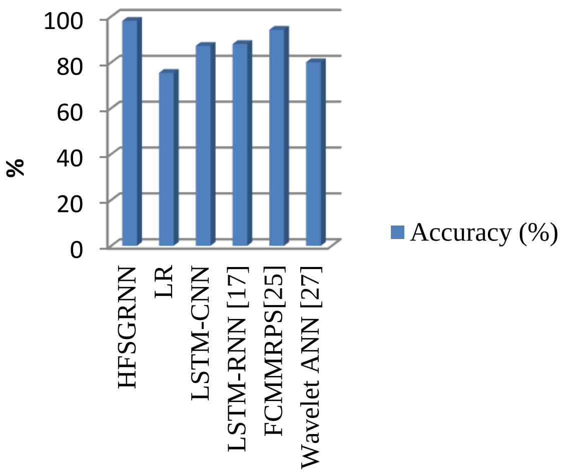

Figure. 12 Comparative Analysis – Accuracy (%)

usages as testing data. It shows more than 99% accuracy during the test stage. Along with accuracy, the other parameters like precision, recall, and F1-score also offer high performance with the give dataset ratio. This training and testing approach shows the proposed architecture's processing quality.The efficiency of the researched model is evaluated, and the hybrid method is compared with several methods for prediction purposes, such as FCMMRPS[25], LSTM-CNN, LR, Wavelete CNN, and LSTM-RNN[17]. In the above figs 12, 13, and 14, the efficiency of the overhead system is evaluated using the accuracy, precision, and recall of the HFSGRNN model. It is compared with that of other models.

6. Conclusion and future scope

Using the Indian rainfall prediction dataset, we implemented a novel method for predicting the rainfall in this analysis. In this research work, the Hybrid HFSGRNN method is implemented to predict the rain from the annual and monthly. This research article aims to execute a complete hybrid feature selection (FS) and classification method for rainfall datasets in different months. After enhancing the rainfall data using preprocessing, feature selection (FS) is done using PSO and SGD to select the feature sets that are usually high-dimensional (HD). So, the FS method is called a hybrid feature selection gradient-based method to choose the reliable feature from the preprocessed data, where the selected features are fed to RNN for classification. In this research study, the implemented HFSGRNN approach defined the maximum accuracy of 98.10 percent and reduced the RSME by 0.10. The research approach has been enhanced in rainfall forecasting and classification, using the accuracy rate on the Kaggle dataset when comparison is done with different methods, such as LR, LSTM-CNN, etc. The future enhancement will enhance the hybrid FSGRNN model with the DL method to manage the accuracy rate and computation time in real-world monsoon rainfall prediction.

Conflicts of Interest

Any of the authors has declared no conflicts of interest.

Author Contributions

For research articles with several authors, a short paragraph specifying their contributions must be provided. The following statements should be used as follows: "Conceptualization, Nishant Pachpor and Dr. B. Sushrsh Kumar; methodology, Nishant Pachpor; software, Nishant Pachpor; validation, Nishant Pachpor, Dr. Prakash Prasad, and Prof. Salim Shaikh; formal analysis, Nishant Pachpor; investigation, Nishant Pachpor; resources, Nishant Pachpor; data curation, Nishant Pachporand Prof. Salim Shaikh; writing—original draft preparation, Nishant Pachpor; writing—review and editing, Nishant Pachpor; visualization, Nishant Pachpor; supervision, Nishant Pachpor; project administration, Nishant Pachpor; funding acquisition, Dr. B. Suresh Kumar, etc. Authorship must be limited to those

who have contributed substantially to the work reported.

References

[1] Y. Hong, K. L. Hsu, S. Sorooshian, and X. Gao "Self-organizing nonlinear output (SONO): A neural network suitable for cloud patch–based rainfall estimation at small scales", *Water Resources Research*, Vol. 41, No. 33, pp. 1-15, 2005.

[2] J. Abbot, and J. Marohasy, “Application of artificial neural networks to rainfall forecasting in Queensland, Australia”, *Advances in Atmospheric Sciences*, Vol.29, No. 4, pp. 717-730,2012.

[3] A. Altunkaynak, and T. A. Nigussie, “Prediction of daily Rainfall by a hybrid wavelet-season-neuro technique”, *Journal of Hydrology*, Vol. 529, pp. 287-301, 2015.

[4] R. Prasad, R. C. Deo, Y. Li and T. Maraseni, "Input selection and performance optimization of ANN-based streamflow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm", *Atmospheric Research*, Vol. 197, pp. 42-63, 2023.

[5] M. Shah, and B. M. Steinberg, "Drought of opportunities: Contemporaneous and long-term impacts of rainfall shocks on human capital", *Journal of Political Economy*, Vol. 125, No.2, 2017.

[6] K. Ogundari, A. A. Ademuwagun, and O. Appah, "A note on rainfall variability and trends in nigeria: implications for agricultural production", *Studies in Microeconomics*, Vol. 11, No. 2, pp-184-194, 2021.

[7] M. Balehegn, S. Balehey, F. Chao and L. Wu, "Indigenous weather and climate forecasting knowledge among afar pastoralists of north eastern ethiopia: role in adaptation to weather and climate variability", *Pastoralism*, Vol. 9, No. 1, pp. 2-14, 2019.

[8] A. D. Dubay, "Artificial neural network models for rainfall prediction in Pondicherry", *International Journal of Computer Applications*, Vol. 120, No. 3, pp. 30-35, 2015.

[9] Q. B. Pham, T. C. Yang, H. W. Tseng and P. S. Tseng, "Combing random forest and least square support vector regression for improving extreme rainfall downscaling", *Water*, Vol. 11, No. 3, pp. 11-17, 2019.

[10] Z. Zhou, J. Ren, and X. He, "A comparative study of extensive machine learning models for predicting long-term monthly rainfall with an ensemble of climatic and meteorological predictors", *Hydrological Processes*, Vol. 35, No. 11, pp.1-14, 2021.

[11] I. Wahyuni, P.F.E. Adipraja, and W. F. Mahmudy. "Determining growing season of potatoes based on rainfall prediction result using system dynamics", *Indonesian Journal of Electrical Engineering and Informatics*, Vol. 6, No. 2, pp. 210-216, 2018.

[12] A. Rahman, S. Abbas, M. Gollapalli, R. Ahmed, S. Aftab, M. Ahmad, M. A. Khan and A. Mosavi, "Rainfall Prediction System using machine Learning fusion for smart cities", *Sensors*, Vol. 22, No. 9, pp. 2-15, 2022.

[13] Y. Luo, J. Sun, Y. Li, R. Xia, Y. Du, S. Yang, and M. Li. "Science and prediction of heavy Rainfall over China: Research progress since the reform and opening-up of new China", *Journal of Meteorological Research*, Vol. 34, pp. 427-459, 2020.

[14] B. A. Ari, L. O. Oyedele, M. Bilal, T. D. Akinosho, J. M.D. Delgado, L. A. Akanbi, "Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting", *Machine Learning with Applications*, Vol. 7, No. 100204, pp. 1-20, 2022.

[15] S. Samantaray, S. S. Das, A. Sahoo, D. P. Satapathy. "Monthly runoff prediction at Baitarani river basin by support vector machine based on Salp swarm algorithm", *Ain Shams Engineering Journal*, Vol. 13, No. 5, 2022.

[16] J. Zhao, R. Chen and H. Xin, "Rainfall study based on ARIMA-RBF combined model", In: *Journal of Physics: Conference Series*, Vol. 2294, No. 1, pp. 1-9, 2022.

[17] S. Poornima, and M. Pushpalat, "Rainfall Prediction System using machine Learning fusion for smart cities" *International Journal of Intelligent Systems and Applications in Engineering*, Vol. 11, No. 2, pp. 329-335, 2023.

[18] G. Meti, and R. K. Gunj, "Analogousness Enhanced Rainfall Predictor using XGB at: https://www.kaggle.com/code/anbarivan/indian-rainfall-analysis-and-prediction/notebook (Accessed: 02 January 2024).

[19] Anbarivan, "Indian rainfall analysis and prediction, Kaggle", Available at: https://www.kaggle.com/code/anbarivan/indian-rainfall-analysis-and-prediction/notebook (Accessed: 02 January 2024).

[20] R. E. Caraka, R.C. Chen, T. Toharudin, M. Tahmid, M. B. Pardamean, and R. M. Putra. "Evaluation performance of SVR genetic algorithm and hybrid PSO in rainfall forecasting", *ICIC Express Letter Part B Application*, Vol. 11, No. 7, pp. 631-639, 2020.

[21] P. Zhang, Y. Jia, J. Gao, J., W. Song, and H. Leung. "Short-term rainfall forecasting using multi-layer perceptron", *IEEE Transactions on Big Data*, Vol. 6, No. 1, pp 93-106, 2018.

[22] Adaryani, F. Rezaei, J. Mousavi, S. Jafari, and Fatemeh, "Short-term rainfall forecasting using machine learning-based approaches of PSO-SVR, LSTM and CNN", *Journal of Hydrology*, Vol. 614, No. 128463, (2022).

[23] C. W. W. Yang, Z. Q. B. Liu, J. S. H. Kwan, L. Chen and N.G Lai. "Spatiotemporal modelling of rainfall-induced landslides using machine learning", *Landslides*, Vol. 18, pp. 2499-2514. 2021.

[24] I. Cholissodin, and S. Sutrisno, "Prediction of Rainfall using improved deep learning with particle swarm optimization", *Telecommunication Computing Electronics and Control*, Vol. 18, No. 5, pp. 2498-2504, 2020.

[25] M. A. Duhayyim, H.G. Mohamed, J. S. Alzahrani, R. Alabdan, M. Mousa, A.S. Zamani, and M. I. Alsaid, "Modeling of Fuzzy Cognitive Maps with a Metaheuristics-Based Rainfall Prediction System", *Sustainability*, Vol. 15, No. 1, pp. 1-6. 2023.

[26] W. Li, X. Gao, Z. Hao, and Rong Sun, "Using deep learning for precipitation forecasting based on spatio-temporal information: a case study", *Climate Dynamics*, Vol. 58, No. 12, pp. 443-457. 2021.

[27] R. Venkata B. Ramana, B. S. R. Krishna, and N.G. Pandey, " Monthly rainfall prediction using wavelet neural network analysis", *Water Resources Management*, Vol. 27, pp. 3697-3711, 2013.

[28] K. Sarvani, Y.S. Priya, C. Teja, T. Lokesh, and E. B. B. Rao, " Rainfall analysis and prediction using machine learning techniques", *J Eng Sci*, Vol. 12, No. 0377-9254, pp. 531-537, 2021.

[29] W. M. Ridwan, M. Sapitang, A. Aziz, K. F. Kushiar, A.N. Ahmed, and A. N. A El-Shafie, "Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia", *Ain Shams Engineering Journal*, Vol. 12, No. 2, pp. 1651-1663, 2021.

[30] L. Alzubaidi, J. Zhang, A. J., Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, and L. Farhan, "Review of deep learning: Concepts, CNN architectures, challenges, applications", *Journal of Big Data*, Vol. 8, pp. 1-74, 2021.

[31] L. Burgueno, J. S. Cabot, and S. Gérard, "A generic LSTM neural network architecture to infer heterogeneous model transformations", *Software and Systems Modeling*, Vol. 21, No.1, pp. 139-156, 2022.

[32] S. Salma, A. Ashwitha, A. "Hybrid Cnn-Lstm Model: "Rainfall Analysis And Prediction For Karnataka Region", *Journal of Theoretical And Applied Information Technology*, Vol. 100, No. 22, pp. 6715-6727, 2022.

[33] R. Praveena, T. G. Babu, M. Birunda, G. Sudha, P. Sukumar, and J. Gnanasoundharam, "Prediction of Rainfall Analysis Using Logistic Regression and Support Vector Machine", *Journal of Physics: Conference Series*, Vol. 2466, No. 1, 2022.

[34] K. Keshav, V Singh, and T. Roshni,. "Application of the PSO–neural network in rainfall–runoff modeling", *Water Practice & Technology*, Vol. 18, No. 1, 2023.

[35] D. P. Hapsari, M .I. Utoyo, and S.W. Purnami, "A Prediction of Rainfall Data Based On Support Vector Machine With Stochastic Gradient Descent", *IICST*, Vol. 10. No. 2, pp. 70-75, 2022.

[36] M. A. Priatna, and E.C. Djamal, "Precipitation prediction using recurrent neural networks and long short-term memory", *Telecommunication Computing Electronics and Control*, Vol. 18, No. 5, pp. 2525-2532, 2022.

[37] Home "District Rainfall Normal (in mm) Monthly, Seasonal And Annual : Data Period 1951-2000 | Open Government Data (OGD) Platform India", Available at: https://data.gov.in/catalog/district-rainfall-normal-mm-monthly-seasonal-and-annual-data-period-1951-2000 (Accessed: 02 January 2024).

[38] O. O. S. Lekshmi. "A pinpoint model for rainfall prediction using CNN model", *Journal of Emerging Technologies and Innovative Research*, Vol. 6, No. 6, pp. 518-524, 2019.

[39] India Meteorological Department (no date) *Home*. Available at: https://mausam.imd.gov.in/imd_latest/contents/rainfall_statistics_3.php (Accessed: 29 January 2024).

[40] P. Mohan, and K. K. Patil, "Deep learning based weighted SOM to forecast weather and crop prediction for agriculture application", *International Journal of Intelligent Engineering and Systems*, Vol.1, No. 4, pp. 167-176, 2018.

[41] Q. Feng, Q, Wen, and X. Li, "Wavelet analysis-support vector machine coupled

models for monthly rainfall forecasting in arid regions", *Water Resources Management*, Vol. 29, No. 4, 2015.

[42] Q. Ouyang, and W. Lu, "Monthly rainfall forecasting using echo state networks coupled with data preprocessing methods", *Water Resources Management*, Vol. 32, pp. 659-674. 2018.

[43] A. Geetha, and G. M. Nasira, "Time-series modelling and forecasting: Modelling of rainfall prediction using ARIMA model", *International Journal of Society Systems Science*, Vol. 8, No. 4, pp. 361-372. 2017.

[44] T. S. Manepalli, and C. Subramanian, "Time Series Analysis of Large Scale Rainfall Data Using Regression Automata Models", *International Journal of Intelligent Engineering and Systems*, Vol.11, No. 6, pp. 118-128, 2018.