SCIENTIFIC CULTURE

www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 4, (2026), pp. 8093-8113
Open Access. Online & Print

DOI:https://doi.org/10.5281/zenodo.20520720

ENHANCED METEOROLOGICAL DROUGHT FORECASTING USING CNN, LSTM, TRANSFORMER, AND BAYESIAN MODEL AVERAGING

Abhisek Kar[1]*, Sujata Chakravarty[2], Bibhuti Bhusan Sahoo[3]

[1] *Phd Research Scholar, Computer Science and Engineering Department, Centurion University of Technology and Management, Odisha, India–752050.*
[2]*Professor , Computer Science and Engineering Department, Centurion University of Technology and Management, Odisha, India–752050.*
[3]*Asoociate Professor ,Agricultural Engineering Department, Centurion University of Technology and Management, Odisha, India– 761211.*

Received: **26/12/2025**
Accepted: **27/03/2026**

Corresponding Author: Abhisek Kar (okar47889@gmail.com)

ABSTRACT

Drought is a multifaceted hydro climatic risk that impacts greatly on water resources, agriculture and livelihoods especially in areas that depend on the monsoon like the Tel River Basin in Odisha, India. The paper investigates deep learning and ensemble modelling that are proposed to predict medium-term meteorological drought with the Standardized Precipitation Index at three and six months (SPI-3 and SPI-6). The NASA POWER database provided monthly data of precipitation from 1971 to 2016 in four stations, namely, Kalahandi, Kandhamal, Kesinga, and Nuapada. The input features that included lagged SPI values to a maximum of 12 months were used in order to adjust temporal dependencies. Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Transformer models of deep learning were created and their performances compared against an ensemble of a Bayesian Model Averaging (BMA). Correlation coefficient (R), coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and NashSutcliffe Efficiency (NSE) were used to measure model performance. Findings have shown that individual models work quite well but the BMA ensemble always outperforms an individual model in terms of accuracy and stability. The results indicate potential of the ensemble-based deep learning methods in making reliable prediction of regional droughts

KEYWORDS CNN; Deep learning; Drought forecasting; LSTM; Standardized Precipitation Index; Tel Basin.

1 INTRODUCTION

Drought is an extremely common and devastating natural disaster with extreme effects on agriculture, water and socioeconomic stability especially in developing countries whereby the livelihoods are dependent on the fluctuation of rainfalls. The droughts are not like other natural disasters since they develop gradually, last longer periods of time, and have complicated spatial and temporal distributions, which render their prediction highly difficult. In India where those who are dependent on rain-fed agriculture comprise almost two-thirds of its population, timely and precise drought prediction is crucial to food security, effective resource distribution, and disaster preparedness. The eastern Indian river basin, the Tel River Basin, an important tributary of the Mahanadi River is especially susceptible to meteorological droughts owing to the regime based on monsoon and big inter-season variability. The basin also has a number of drought afflicted areas such as Kalahandi, Kandhamal, Kesinga and Nuapada which have in the past suffered water shortage, crop failure and economic hardship due to rain deficit.

The measures and predictions of droughts demand strong indices that can record the abnormalities of precipitations at various periods of time. Of the various indices, that have been formulated, the Standardized Precipitation Index (SPI) has proved to be one of the most valid measurement tools since it is simple, flexible and can be used to describe the severity of droughts over a number of accumulation periods. The short-term scales like SPI-3 are used to indicate the agricultural drought associated with change in soil moisture, whereas medium-term scales like SPI-6 are better placed to depict the hydrological drought conditions. The combination of several SPI scales can be used to have a more detailed view of the evolution, persistence, and recovery stages of droughts.

Due to the development of machine learning (ML) and deep learning (DL), hydro-meteorological forecasting problems have advanced considerably. The classical statistical methods like the ARIMA, multiple linear regression and soft computing schemes like Genetic Programming (GP) and Artificial Neural Networks (ANN) have given first order approaches to drought prediction but in most cases do not reflect the nonlinearity and non-stationarity inherent in climatic time series. Deep learning models address these weaknesses by automatically discovering hierarchical representations of raw data and are therefore capable of discovering complex spatio-temporal dependencies. Convolutional Neural Network (CNN) which was initially used in image and signal processing has demonstrated a good result in time-series forecasting because it can identify local time variation and filter high-frequency fluctuation. In the meantime, the Long Short-Term Memory (LSTM) network, which is a particular type of the recurrent neural network (RNN), is best adapted to the sequential modelling and well adapted to long-term dependencies in the climatic sequences. In more recent times, a sequence modelling system based on self-attention mechanisms, the Transformer architecture, has become a competitive alternative to attention mechanisms. It operates in parallel on input data and acquires short and longrange dependencies more efficiently than recurrent architectures, and has been demonstrated to be useful in climate prediction, precipitation downscaling, and hydrological modelling.

Despite the fact that all deep learning models have definite strengths, their individual use is usually characterized by weaknesses when it comes to addressing multi-scale drought processes. CNNs are prone to over-emphasizing short-term variations, LSTMs might not be able to generalize well with longer lags, and Transformers, being flexible, are unstable when training data is small or noisy. In order to overcome these difficult situations, ensemble learning is now a trending approach to hydrological forecasting. Bayesian Model Averaging (BMA) is among ensemble methods that offer a theoretical method of combining multiple model output to a single probabilistic model. BMA does not depend on one model, but instead it weights the individual predictors of the model according to their predictive reliability, thus, it takes into consideration both the model uncertainty and the variability in performance. The resultant ensemble is more stable as well as theoretically based on the Bayesian probability and hence it is especially applicable to the complex environmental systems that are characterized by uncertainty and heterogeneity of data.

Based on the earlier studies that used CNN, LSTM, and hybrid networks on the Tel River Basin, the paper expands the discussion to include the Transformer model and creates a BMA-based ensemble that combines the results of CNN, LSTM, and Transformer networks. The ensemble gives an empirically determined weight (0.4 CNN, 0.2 LSTM, and 0.4 Transformer) to each of them so that their contribution is tuned based on empirical results. The BMA ensemble merges the localized feature extraction (CNN), sequential memory (LSTM), and global attention (Transformer) to exploit the complementary capabilities of these paradigms of deep learning to create a more reliable and generalized drought prediction system.

The present study involves monthly precipitation data of four meteorological stations, such as Kandhamal, Kalahandi, Kesinga, and Nuapada in period 1971-2016 to obtain the SPI at two accumulation periods (3- and 6 months). The data of every station was separated into two sets: training (80 percent) and testing (20 percent), in a chronological order which guaranteed the achievement of realistic conditions of forecasting. A combination of statistical indicators: correlation coefficient (R), coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Nash-Sutcliffe Efficiency (NSE) were used to assess the predictive performance of every model based on which a complex measure of the model accuracy, consistency, and strength is evaluated.

The general aim of the research paper is to create a unified deep learning and ensemble-based system of drought prediction with the help of multi-temporal data on Standardized Precipitation Index (SPI) in the Tel River Basin of Odisha. In particular, the study aims at developing and testing three high-end deep learning models, namely, Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Transformer to investigate their suitability in the nonlinear and time-dependent nature of SPI-3 and SPI6 time- scales. It is on the basis of the strengths and weaknesses of these respective architectures that the study will go further to build an ensemble of Bayesian Model averaging (BMA) which combines the predictive power of these respective architectures by utilizing an optimal weighting function. This ensemble system is designed to increase the accuracy of the forecasts, minimize the uncertainty of the model, and increase the stability of the model when changing lag periods and climatic conditions.

Besides model development, the study aims to determine the variation of model performance in the various stations and time scales hence testing the generalization ability of each architecture as well as the proposed ensemble frame-work in the various hydro-climatic conditions of the Tel River Basin. The research, based on this comparative analysis, also seeks to determine the best and most reliable and operationally appropriate model architecture to use in real time applications for drought early warning in Odisha's drought prone regions. The current paper is on the forefront of probabilistic ensemble modelling by connecting deep learning with probabilistic forecasting of drought in India. The findings show that the BMA ensemble is always superior to individual deep learning models in terms of both predictive accuracy and reliability, which is important to point to the fact that model integration is important to cope with the complexity of hydroclimatic prediction systems. The suggested framework has great prospects of application to other basins under droughts and provides a scalable and resilient strategy to manage water resources and develop an early warning system in climate-sensitive settings that rely on monsoons.

2 RESEARCH REGION AND DATA GATHERING

The current study will be about four meteorological stations located in the Tel River Basin in Odisha, India, i.e., the districts of Kalahandi, Kandhamal, Kesinga, and Nuapada, where the instances of drought are frequent. The Tel Basin, encompassing an approximate area of 20,490 km², lies between 19°15′N–20°55′N latitude and 82°03′E–84°17′E longitude (refer to Figure 1) [1]. The climate is tropical monsoon with hot summer, mild winter and a wet season in between June and September. The annual precipitation is approximately 1250 mm where close to three-quarters of the total precipitation is received during the southwestern monsoon season.

Nevertheless, rainfall in the basin is characterized by high spatial and temporal variability and when there is a deficit or abnormal monsoon rainfall, it usually leads to meteorological droughts, especially in western Odisha [2]. The chosen stations represent a variety of physiographic and agroclimatic conditions - whereas Kalahandi and Kandhamal are located in the hilly and plateau areas, Kesinga and Nuapada are located in the transitional agroclimatic areas. The Tel Basin is the best place to study drought behaviour at various time scales due to this physiographic diversity.

Monsoon rainfall supports the local economy, which is majorly agricultural in nature and frequent droughts have been credited with losses of crops, migration, and socioeconomic difficulties. In the present study, the monthly rainfall data (from 1971 to 2016) were obtained at NASA POWER data archive (https://power.larc.nasa.gov/data-access-viewer/). Fig 1 represents the geographical location of the four meteorological stations in the study area.

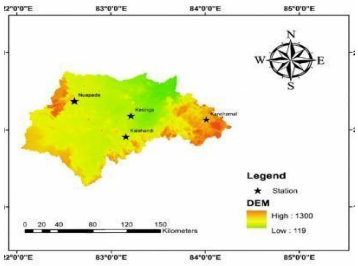

Figure 1. Study area, the Tel River basin, Digital Elevation model (DEM) showing the stations

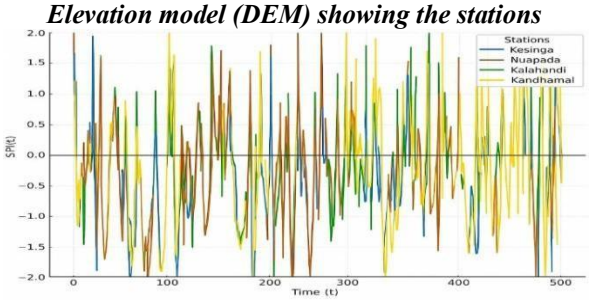

Figure 2. Plot showing the SPI-3 variation over time series for all stations

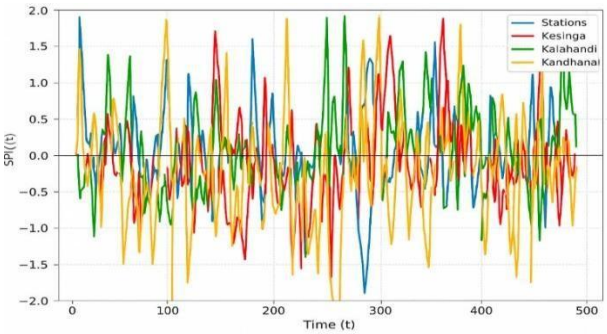

Figure 3. Plot showing the SPI-6 variation over the time series for all stations

3 METHODOLOGY

3.1 The Standardized Precipitation Index (SPI)

The Standardized Precipitation Index (SPI) is a precipitation-based measure that measures the deviation of rainfall over multiple accumulation intervals [3], which provides a normalized measure of nonconformity to long-term averages. In calculating SPI, the amount of precipitation total is initially accumulated over the chosen time period and a probability distribution, often the Gamma distribution, is then fitted to these accumulated amounts [4]. The resulting cumulative probabilities are then converted to standard normal deviates to obtain the SPI values that render the anomaly of the rainfall as the deviations of the mean. In the research, SPI was assessed at both 3- and 6month time scales, to include the short-term agricultural moisture deficits as well as the long-term hydrological droughts. The series of monthly rainfall aggregated at each station was modelled by Gamma distribution, as it traditionally should be done in calculating SPI [5]. The probability of each amount of precipitation was then found using the cumulative distribution function (CDF) and then transformed into an SPI value using the inverse standard normal transformation.

The resulting values of SPI are normalized to all stations and time lags, and can be inter-compared [6]. Any negative value of SPI means dry conditions, whereas positive value is wet conditions. To determine the level of drought using the standard classification criteria, the values of SPI -1.0, -1.5, and -2.0 will be considered moderate, severe, and extreme droughts, respectively (Table 1) [7]. The resulting SPI series of each station and time scale was used as the target variable in the drought prediction modelling that followed.

Table 1 Drought categories according to SPI classification

Range of SPI Values	Drought / Wetness Category
≥ 2.00	Extremely Wet
1.50 to 1.99	Very Wet
1.00 to 1.49	Moderately Wet
–0.99 to +0.99	Near Normal
–1.00 to –1.49	Moderately Dry
–1.50 to –1.99	Severely Dry
≤ –2.00	Extremely Dry

3.2 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) network is a particular implementation of the Recurrent Neural Network (RNN), which was first introduced to address the vanishing gradient issue that is inherent to the traditional RNNs. Although traditional RNNs do not tend to memorize information fairly long sequences, the LSTM architecture presents a memory cell that is able to selectively remember or forget information through time steps [7]. This characteristic allows LSTMs to model the timedependent characteristics in an effective way, which makes them especially valuable at the analysis of the time-dependent behaviour of hydrometeorological indices, including the Standardized Precipitation Index (SPI).

The LSTM unit works based on three major gating processes which include the forget gate, input gate and the output gate which together control the passage of information and the update of the state of the cell

$\tilde{C}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$	(3)
$Ct = f \square\square Ct-1 + it \quad C\square t$	(4)
$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$	(5)
$h_t = o_t \odot tanh(C_t)$	(6)

respectively. Related mathematical representation of LSTM cell is shown below:

$f_t = \square(W\, h_f[\, _{t-1}, x_t\,] + b_f)$	(1)
$i_t = \square(W\, h_i[\, _{t-1}, x_t\,] + b_i)$	(2)

Here, f_t, i_t and o_t represent the input, output and forget gates respectively (Figure 4). While $C_{(t-1)}$ and C_t represent the memory states from the previous and current time steps, the hidden state from the previous time step is represented as $h_{(t-1)}$. The weight matrices are represented by W_f, W_i, W_c and W_o, and the bias terms are represented with the use of $^{b}{}_{f}$, b_i, b_c, and b_o . The $\square$ denotes the sigmoid activation function, and tan is the hyperbolic tangent function. The multiplication operation is an element-wise product of the matrices denoted by $\odot$. The traditional RNNs are based on a structure that is single-layered; on the contrary, LSTM networks possess more than one layer, and the interaction between these layers is necessary to increase memory retention and make the information processing more efficient. The initial part of an LSTM cell assesses the likelihood of retaining or forgetting the stored information that exists in the memory cell. As an example, the related pronouns used in a text prediction task should also be changed in case the gender of a subject is altered. This specific decision-making role occurs in the forget gate; a sigmoidal function is employed in a bid to control the flow of the information. The f_t value is determined using $h_{(t-1)}$ and X_t. The value will be zero to one. When f =1, the cell will store the memory of the last state output ($C_{(t-1)}$) but when the value is 0, the memory will be lost. New batch of data is then analysed to determine what new data should be retained on the memory. This is done by the input gate and this has a sigmoid activation function and a hyperbolic tangent function to scan and control what is stored in C_t. Finally, the output gate decides what values will be passed

to the next time step. And finally, in the last cell, that is the output cell, h_t, is computed using the sigmoid function, which selects parts of the memory cell, which are relevant to the output.

Various drought prediction works have recently used models consisting LSTM. Each of these models employs several LSTM layers and the quantity of input data directly influences the selection of the number of layers. However, excessive layers might cause a network to be more complicated, which may result in issues such as overfitting.

Figure 5 shows the structure of LSTM. The product of the LSTM layers is processed by a layer called the flatten layer and this is required to convert neurons into vectors. The flat data is then entered into the dense layers so as to establish the final output of the network. Figure 4 is a recurrent neural network model that employs a recursive sequence method to determine the parameters and the weight matrices. These parameters are normally obtained by applying an optimization criterion like the Adam function [27].

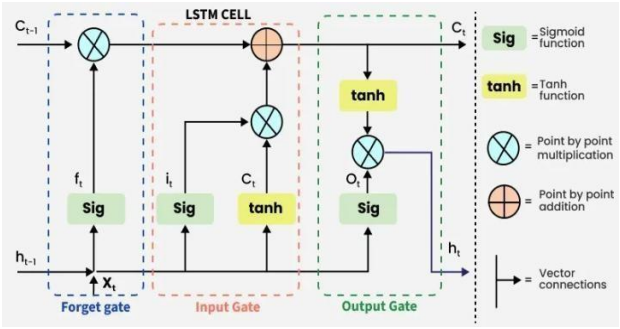

Figure4. Illustrates the structure of a unit of LSTM

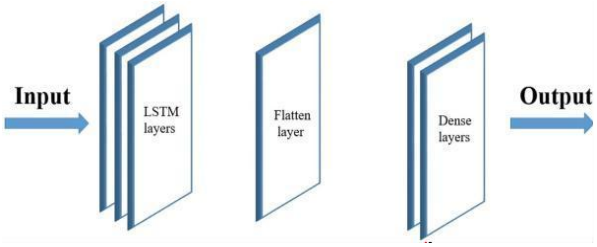

Figure5.Dense layers, a flatten layer, and LSTM layers make up the LSTM-based model structure.

3.3 Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs) are a specialized form of deep learning model that use convolution operations of input data to extract features, train and minimize error [9]. Just like other neural networks like multilayer perceptron, CNNs have a number of layers that are trainable. Their greatest advantage is that they are able to pick the sub feature of the input data automatically and determine it without the information being preprocessed in a large scale. CNNs are particularly appropriate in cases where the input has either spatial or temporal dependencies as they can significantly reduce the cost of computing and overcome the issue of overfitting.

Figure 6 is a schematic representation of CNN framework. The model may be said to be composed of three significant layers which include convolutional, pooling and fully connected layers. The basic unit is the convolutional (or filtering) layer and it is characterized as a set of filters (kernels) that are used to perform a sliding operation on the input data to create feature maps which are local patterns or dependencies. These feature maps are subsequently reduced by the pooling layers (mostly max pooling or average pooling) that subsequently selects most significant responses and therefore dimensionality reduction occurs without the essential information being lost. It is not only more efficient than the computation itself but it can also help to summarize extracted features to process it in the downstream.

The fully connected (dense) layer, which is normally at the end of the network, works just as well as the fully connected perceptions of the traditional multilayer perceptron and it is the one that combines the extracted features to form the final prediction. The unique advantage of CNNs is that they can extract the features automatically that is, can identify the latent features in the high-dimensional input data and also prevent the probability of overfitting (Figure 7).

Despite the fact that CNNs were mainly involved in image recognition, they have been applied in time series prediction since it has been noted to be efficient in identifying localized temporal patterns in sequential data [10]. CNNs can be used to predict droughts successfully by estimating the variation in the Standardized Precipitation Index (SPI) and temporal variations and changes in short-term lagged rainfall series with convolutional filters.

In 1-dimensional CNN, the size of the convolutional kernel is applied to the input sequence, which in turn produces a set of feature maps. This convolution process of a particular feature map mathematically can be written as:

$$y_j(t) = \sigma\left(\sum_{i=1}^{n}\sum_{s=0}^{k-1} w_{i,j}(s)\cdot x_i(t-s) + b_j\right) \quad (7)$$

Here, $x_i(\,)$ is the input features, $W_{ij,}$ represents the filter weights connecting input channel i to output channel j k, is the kernel size, b_j is the bias, and σ( ) is the activation functionality (usually ReLU).

After the convolution process, pooling layers are usually used to reduce the number of data dimensions of the feature maps by max or average pooling [11] followed by downsizing the pooling layers to retain the most salient and informative features. The results of the convolutional and pooling steps are then flattened into a one-dimensional array and sent to fully connected (dense) layers which are trained to find the mapping between the extracted representations and the desired SPI values.

The CNN architecture is one of the benchmark models that are used to predict droughts in the present work [12]. Lagged SPI sequences over various timescales are delivered to the model, and this allows the model to identify localized temporal structures, e.g. an abrupt change in rainfall or a brief spell of drought. Despite the fact that CNNs are not designed to capture long-term dependence, they are excellent in terms of efficiently detecting and summarizing short-term changes that have a dramatic effect on near-term drought behaviour. By its convolution and gating operations, the network is effective in the preservation of the relevant temporal information through large time spans and at the same time filtering noise and transient irregularity in the data.

It should be mentioned that the input formatting design, the size of kernel, the number of filters, CNN units, and the pooling windows have a direct impact on training efficiency and forecasting accuracy; the specific architecture employed in this paper is shown in Figure 6 and has been determined through empirical optimization on the training data. Optimization of model parameters was done by minimization of mean squared error between observed SPI y_t and prediction $\hat{y}_t$:

$$L = \frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_t)^2, \quad (8)$$

using the Adam optimizer along with (learning rate, batch size, number of epochs, early-stopping patience, and regularization) as hyperparameters.

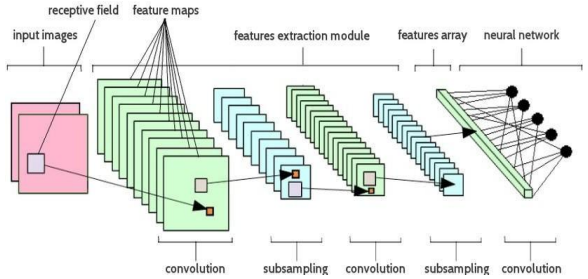

Figure 6: CNN structure consists of convolutional layers, pooling, and fully connected layers

3.4 The Transformer Model

For this study, the authors designed the Transformer model in the TensorFlow Keras framework to forecast the Standardized Precipitation Index (SPI) from the provided lagged rainfall data. The Transformer model relies on self-attention for the overall interdependence. The model is adaptable in its information processing and can flexibly concentrate on the precise time slices which are significant to the drought signal [13]. The input sequence consists of twelve lagged values of the SPI.

$$(SPI\,t(-1)\,to\,SPI\,t(-12)) \quad (9)$$

was first passed through a dense projection layer to map the one-dimensional time series into a latent vector space of dimension d_{model}=32.To preserve the sequential order of time the input embeddings were augmented with a sinusoidal positional encoding to enable the network to differentiate between time steps which are not recurrent.

The fundamental Transformer encoder block comprised of multi-head self-attention (MHSA) mechanism having two attention heads and positionwise feed-forward network (FFN) having 64 units. Each sub-layer contained residual connections and layer normalization in order to stabilize training and degrade the gradient. The reduction of overfitting was achieved by the use of dropout regularization (rate = 0.1). The contextualized feature representations were summarized by global average pooling and then a dense layer was used with ReLU activation and a last linear output neuron predicting SPI.

The trained form of this structure is also presented using Adam optimizer and the mean squared error loss can estimate complex non-linear relationships among lagged precipitation indices. This is due to the self-attention process of the Transformer making the historical inputs to be dynamically weighted, thus it provides a robust and generalizable model to predict the drought with the help of rainfall.

3.5 Mathematical Modelling of Transformer Model.

The Transformer model is employed in the current study is through the use of encoder architecture [14]. The self-attention mechanism is one of the strengths of the encoder-based models. This process enables the model to automatically and separately learn the contextual association of individual time steps without recurrence and convolutional activities [15]. The remaining subsections provide the mathematical model to be applied in the current research in forecasting SPI.

3.5.1 Scaled Dot-Product Attention

Given an input sequence $X=[x\,x_1, 2,\ldots.\,, x_L]$,Every time step is linearly projected into 3 learned representations Query (Q), Key (K), and Value (V) — through trainable weight matrices $W\,W\,WQ, K, V$:

$$Q\,XW = Q,\ K = XWK, \quad (10)$$

$$V = XW_V$$

The attention mechanism computes the relationship between each time step with each other through a dot product between the queries and keys, divided by the square root of the key dimension d_k.

The output of the attention mechanism is then calculated as:

$$QKT \quad (11)$$

$$Attention\ Q\ K\ V(\quad ,\qquad\qquad ,\ ) = soft\max(\sqrt{d_k}\ )V$$

The operation enables the model to place adaptive weight on various lagged values of the SPI in predicting the existing value of SPI.

3.5.2 Multi-Head Attention

The Transformer consists of multi-head selfattention (MHSA), in which h independent attention operations are done simultaneously in order to allow the model to model dependency patterns at once:

$$MultiHead\ Q\ K\ V(,\quad ,\quad) = Concat\ head(\ _1, \qquad (12)$$

$$head_2,....,head\ W_h) \quad _o$$

Here, $head_i = Attention\ QW(\ \ _Q^{()i}, KW_K^{()i}, VW_V^{()i})$

In this case, the two heads are taught to concentrate on various dependencies in time, and the model is capable of incorporating both memory of short-term variations and drought memory in the long-range in one representation.
Position-Wise Feed
Forward Network

3.5.3 Position-Wise Feed Forward Network

Output of the attention sub-layer is fed through feed-forward network (FFN) for each time position in isolation:

$$FFN\ x(\ ) = \mathrm{Re}\ LU\ xW(\quad\ _1 + b\ W_1)\ _2 + b_2 \qquad (13)$$

This non-linear transformation enhances the contextual details making them more representational. The encoder has residual connections and layer normalization that succeeds each sub-layer that are used to stabilize training, as well as speed up convergence.

3.5.4 Positional Encoding

The Transformer does not include any intrinsic recurrence and, therefore, positional encoding is applied to input embeddings to maintain the temporal sequence of SPI measurements. Encoding of the sinusoidal form of the positions and dimensions is determined as:

$$PE(\ pos,2\ )i = \sin(\ \frac{pos}{10000\ d/model}),PE(\ pos,2i+1) = \cos(\ \frac{pos}{10000\ d/model}) \qquad (14)$$

This encoding gives a continuous account of the temporal positions, which guarantees the model to make the distinction that exists between the various lag indices in the learning process.

3.5.5 Final Prediction

The feature representations resulting after going through one or more encoder blocks are then aggregated by the global average pooling operation and finally are mapped to the output by a dense regression layer:

$$\llcorner y_t = fTransformer\ SPI(\quad\ _{t-1}, SPI_{t-2}, \ldots, SPI_{t-11}^{(15)}) \qquad)$$

Where $\llcorner y_t$ denotes the predicted SPI value at time t.

$$LMSE = \frac{1}{N}\ \sqcup N\ (yi - \sqcup yi\ )2 \qquad (16)$$

This mathematical model is capable of training the Transformer to acquire nonlinear and long-range time series temporal dependencies on rainfall time series. The model will be particularly suitable to the drought forecasting process since continuity over time and multi-scale variability are important, by actively manipulating weights of attention to identify the most effective past observations.

3.6 Ensemble Bayesian Moving Average (EBMA) Approach

In order to enhance the predictive ability of individual deep learning models, this paper has used an Ensemble Bayesian Moving Average (EBMA) framework. EBMA is also a probabilistic ensemble learning scheme, a combination of prediction of multiple models in a systematic method utilizing a Bayesian inference and adaptive weighting [16]. It is designed to take advantage of predictive complementarity of various predictive models such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, hybrid CNNLSTM networks, and Transformer networks, which makes it a more accurate, reliable, and generalized forecast of the Standardized Precipitation Index (SPI) of drought measurements [17]. Unlike simple averaging or fixed ensemble techniques, EBMA incorporates Bayesian model averaging, accounting for model uncertainty as well as temporal variations in predictive ability. In the event that a nonstationarity and non-linearity is present, as is the case when performing meteorological time series where the variation of rainfall is non-stationary and nonlinear, then the application of the single deterministic model can overfit or bias the estimates. EBMA solves these issues by integrating probabilistic reasoning with dynamic weight adaptation so that the ensemble can be able to adapt when more information is being presented.

Let $\{M\ M^1, \quad\ ^2, \ldots, M^M\}$ denote a set of M base

models, each producing a forecast $\sqcup^{y}(\ )m$ for a specific time step t. The prediction of the ensemble under EBMA is expressed as a weighted posterior mean:

$$\hat{y}_t^{EBMA} = \sqcup w_m \hat{y}_t^{(\ )m} \quad (17)$$

Here, $\sum_{m=1}^{M} \sqcup w_m = 1, w_m \sqcup 0$

Here, $^{w}{}_{m}$ is the posterior probability of model M being the most suitable predictor given the observed data D. The application of the Bayes' theorem yields the posterior weight of each model as:

$$M(\ m\ ) \quad (18)\ wm = \frac{P\ D\ M(\ \ |\ \ m\ )P}{\sum_{k=1}^{M} \sqcup k} \quad k \quad k \quad P\ D\ M(\ \ |\ \ )P\ M(\ \ )$$

Where $P\ D\ M(\ |\ _{m})$denotes the model likelihood, reflecting how well model $^{M}{}_{m}$ fits the observed SPI data, and $P\ M(\ _{m})$ gives insight into what the model's reliability looked like before considering the dataset, and what the previous assumption dealt with. This framework for reliability works in a way that models that have been more reliable, or less uncertain, in the past predictions, will add more and more to the final predictions. On the other hand, models that have been less reliable will be given less importance.

In order to deal with time-related dynamics (e.g. temporal rainfall dynamics) the posterior model weights are constantly changing, or are not fixed, but are renewed with a moving average window [18]. This gives a smoothing effect to the selected models, creating a temporal smoothing effect. This collects and stores variable predictor-target correlations. Model ensemble weights are defined as temporal moving averages and, in this case, for each time-step M_m is recalibrated as:

$$w_m(\ )t = \sqcup.w_m(\ 1)t- + -(1\ \sqcup).w_m^{\sqcup(\ )t} \quad (19)$$

Where $w^{\sqcup(\ )}{}_{m}{}^{t}$ is the newly estimated Bayesian weight at time t, and $\sqcup \in [0,1]$controls smoothing strength. The Bayesian moving average representation of the ensemble adapts to the recent changes while also keeping the history of past performances over the long term.

For the EBMA layer, this was performed after the predictions of the models, CNN, LSTM, and Transformer were made. It was assumed that the ensemble weights were equal at the start, however, this was changed through iterative Bayesian updating based on the models' validation accuracy and error metrics (RMSE, MAE, and NSE) of the models. Hence, the last ensemble prediction was both statistically and temporally updated from the deep learning predictions.

The primary advantage of EBMA is that it can capture uncertainty and increase the prediction quality. Frequent, abrupt changes along with high noise levels in the observations are common in hydrometeorology. In the case of drought forecasting in Odisha, the EBMA proved to be beneficial, especially in forecasting SPI values, and over various time horizons, it proved more accurate and reliable than the numerous individual models.

4 MODEL DEVELOPMENT

The SPI-3 and SPI-6 values for the next 6 months were also predicted by using the SPI-3 and SPI-6 values of the previous months with a one-month time lag. The models were able to capture the drought index for those time periods. The prediction equations for the models can be expressed as:

$$SPIt(\ )k = f\ SPI(\ t(\ )-k1\ ,SPIt(\ )-k2,...,(20)$$

$$SPIt\ m-(\ )k$$

The variables for each time step t, the number of previous months m, and the SPI accumulation timescale k (3 or 6 months) are defined for the four stations (Kalahandi, Kandhamal, Kesinga, and Nuapada) based on 552 monthly data points from 1971 to 2016. m, t, and k are defined for each time step. To keep a balanced evaluation framework, each dataset was divided into 80% of training data and 20% of testing data. To create a more realistic forecasting scenario, the order of SPI values was preserved so that training was performed on the historical data and evaluation of the model was performed on future data that was not observed. For data partitioning, we aimed to replicate the realworld situation of drought monitoring.

The comparative evaluation was conducted on four different model architectures, which were the Convolutional Neural Network (CNN), the Long ShortTerm Memory (LSTM), the Transformer network, and a Bayesian Model Averaging (BMA) ensemble model of the three deep learning outputs [21].

The CNN model was created based on the idea to retrieve local temporal data with the help of onedimensional convolutional filters of the lagged SPI sequence [22]. CNN is also helpful in the localized trends patterns, i.e, consecutive wet or dry spells, which reflect short term rainfall dependencies defining the occurrence of droughts [23]. Rather, long-range temporal dependencies were modelled using the LSTM model. LSTM contains the

memory of the past status of the precipitations and it is particularly suitable with sequential data that are not linear and contain the hysteresis effects as in the case with hydrometeorological processes because it possesses gating mechanisms (input, forget and output gates) [24].

Transformer model too suggests a self-attention mechanism, which enables the model to assign different time steps different weights based on how relevant they are to the target forecast [25]. Transformer can also learn long-term dependencies, which is not possible with recurrent architectures such as LSTM since it can learn from all the time steps at the same time, making it more computationally efficient and eliminating the vanishing gradient issue [26]. This makes it very helpful in the identification of both the short-term and long-term trend of drought presence in large scale time series data.

wTransformer =0.4 , Demonstrating the stability and that the BMA ensemble lowered the amounts of error and increased the efficiencies with each measure

accuracy of CNN and Transformer models over LSTM. more than any one of the deep models [29]. This The ensemble gains from the unique advantages of its indicated that the ensemble guideline limited most of base learners: CNN's ability to capture short-term the model-specific overfitting and underfitting, and trends, LSTM's ability to remember sequences, and the was mostly applicable to longer lags like SPI-6, since Transformer's ability to understand context globally. the single models were evidently far less stable.

All deep learning models, LSTM, and

Transformer, along with CNN were implemented

5 MODEL PERFORMANCE EVALUATION

using the TensorFlow-Keras framework in Python. In this paper, the results of all the developed The Adam optimizer, which has an adaptive learning models are measured in terms of a combination of rate and momentum-based parameter updates, was absolute error measures and goodness-of-fit used as the training algorithm for all models. Adam measures in order to represent a detailed evaluation combines the advantages of both AdaGrad and of predictive quality and reliability. The metrics that RMSProp. It adjusts the learning rates differently for are chosen are the Root Mean Square Error (RMSE), each parameter while keeping an exponentially Mean Absolute Error (MAE), Nash-Sutcliffe decaying average of the previous gradients. This Efficiency (NSE), correlation coefficient (R), and adaptive gradient was applied to SPI in the learning 2). All these statistical coefficient of determination (R

Finally, Bayesian Model Averaging (BMA) ensemble was chosen as a post-processing model, which is a combination of CNN, LSTM, and Transformer predictions. BMA does not learn the models individually, but uses weighted linear combination of the models:

$$Y\hat{}BMA = wY1\ \hat{}CNN + w\ Y2\ \hat{}LSTM + wY3\ \hat{}Transformer \quad (21)$$

where $Y\hat{}CNN$,$Y\hat{}LSTM$,$Y\hat{}Transformer$ are the individual model predictions, and w_1 , w_2 , w_3 are non-negative weights satisfying $\sum w=1$. The weights were determined empirically based on model performance, with values w_{CNN}= 0.4, w_{LSTM}= 0.2 and

dynamics, as it changes from one epoch to the other. measures are used to measure the accuracy,

Adam's optimizer was used with default consistency and explanatory power of the models in hyperparameters for this study. These included simulating observed drought dynamics expressed as learning rate = 0.001, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. These the Standardized Precipitation Index (SPI) [30]. parameter values yield stable and fast convergence The RMSE and MAE are measures of the error of for tasks related to the prediction of drought. Adam prediction in absolute terms, the smaller the values, is robust in scenarios with noisy and sparse the more accurate the model is [31]. RMSE focuses gradients; hence, it was the best choice for the SPI more on large errors as it squares deviations and datasets in this study. averages them, which is vulnerable to outliers,

To prevent overfitting and make sure it could whereas MAE gives an equal average of absolute generalize well, early stopping was used during the deviations, indicating the average error magnitude of training process. Each training epoch would result in a either direction. The Nash-Sutcliffe Efficiency (NSE) validation loss figure that would determine if the loss is a measure of the total goodness-of-fit that is used was improving. If there were 20 epoch results in a row to determine the relationship between the model where the loss did not improve, training would stop. predictions and a given series of observations which The weights would be rolled back to the epoch where is the mean. A value of 1 on the NSE, indicates that the loss was the lowest. Each training model was there is a complete agreement between the predicted allowed to train a maximum of 200 epochs, but was and observed values whereas values less than 0 usually finished between 80 and 120 epochs, depending indicate that the model is worse than having the

on the custom model and the time interval of the mean of the observations as a predictor.

Standardized Precipitation Index (SPI). This approach

Along with such indicators based on errors, two

focused on eliminating as much unnecessary statistics based on correlation were also used.

computation as possible and focused on generalizing

Pearson correlation coefficient (R) measures how

the model well. strong and whether the linear relationship between

Three models were created to be used as a base for the observed and predicted SPI values is strong or

predicting the test set which was then used with the weak [32]. It is a measure of the degree to which the

Bayesian Model Averaging (BMA) framework to make drought variability patterns are being well explained

predictions. The BMA was a univariate model, so it

The mathematical expressions of these measures are as follows:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(SPI_{ci} - SPI_{pi})^2}{n}} \quad (22)$$

$$MAE = \frac{\sum_{t=1}^{n}|SPI_{ci} - SPI_{pi}|}{n} \quad (23)$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n}(SPI_{ci} - SPI_{pi})^2}{\sum_{i=1}^{n}(SPI_{ci} - \overline{SPI_c})^2}\right] \quad (24)$$

$$R = \frac{\sum_{i=1}^{n}(O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2 \sum_{i=1}^{n}(P_i - \bar{P})^2}} \quad (25)$$

$$R^2 = \left(\frac{\sum_{i=1}^{n}(O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2 \sum_{i=1}^{n}(P_i - \bar{P})^2}}\right)^2 \quad (26)$$

Where, n represents the total number of data points used for evaluation (including both training and testing phases), O_i and P_i denote the observed and predicted SPI values respectively, $\bar{O}$ & $\bar{P}$ are their respective mean values.

The correlation-based metrics (R and R^2) complete the error-based measurements (RMSE, MAE, and NSE) by giving an idea of the structural agreement between observed and modelled SPI variations. RMSE and MAE are estimations of the extent of deviation, whereas R and R^2 emphasize the extent to which the models can maintain the phase and variability of the drought events [34]. NSE on the other hand is a holistic performance index that represents bias and variance errors together.

by the model, where the closer the value is to 1, the

used a deterministic model. This also meant that higher the positive correlation. R_2 is merely the integration and interpretation were much easier with square of R and is known as the coefficient of

the constant-weight model. The model predictions determination, and it is the portion of the variation in

were normalized on to the same scale to ensure

observed data that the model predictions can account

integration model consistency across the different 2 means an increase for. The increase in the values of R

custom models. in the explanatory power and model performance in

Next, the performance of the ensemble was general [33].

evaluated against various statistical measures such as

R, RMSE, MAE, and NSE [28]. The results outlined

Combined, the five statistical indicators represent a strict multi-dimensional assessment model. They enable a fair comparison of the performance of various architectures (CNN, LSTM, Transformer, and BMA) and timescales (SPI-3 and SPI-6), which includes the accuracy of point forecasts and the accuracy of temporal drought processes [35]. These metrics are then used in the latter sections to evaluate the performance of each model at various lag intervals of all the four study stations.

6 RESULT AND DISCUSSION

6.1 Prediction of SPI-3 at all the stations Four deep learning-based methods Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Transformer, and the Bayesian Model Averaging (BMA) ensemble were considered in determining the predictive performance of SPI-3 prediction at the four chosen stations: Kandhamal, Kalahandi, Kesinga, and Nuapada. The models were trained and tested on the same datasets and time-lag settings (t vs t-1…t-12) and the performance measured in terms of R, R^2, RMSE, MAE, and NSE. The findings indicate that despite the fact that all individual architectures were able to capture the overall temporal trend of drought variability, the BMA ensemble was always able to achieve a high level of accuracy and stability at all the lag periods.

In the case of Kandhamal, the models had moderate predictive ability. The results of LSTM and CNN were similar with R^2 values of the same order (0.25-0.35) and RMSE of the same order (0.8) implying that both were able to predict the immediate variations in the SPI but failed at longer lags. Transformer model performed somewhat worse on RMSE on few middle range lags but failed to be consistent after lag 6. However, the BMA ensemble achieved the greatest correlation ($R \approx 0.761$) and $R^2 = 0.578$, and significantly lower values of RMSE (< 0.61) and NSE (> 0.56). These findings indicate that BMA was able to capture the nonlinear and cumulative lag effects better than any given model. This enhancement was more

pronounced at higher lags and the single model errors were rising whereas the ensemble errors did not change.

At Kalahandi, the three individual models worked on average well at lower lags but displayed different stability. CNN showed some sharp increases in RMSE, indicating its vulnerability to local changes in the rainfall time series. LSTM was typically performing moderately (R^2 = 0.25 - 0.30, NSE = 0.25 -0.30), and the Transformer was smoother but underestimated extremes. The BMA ensemble, a combination of the three model predictions with optimized weights (CNN = 0.4, LSTM = 0.2,
Transformer = 0.4), performed better than all of them with an R^2 up to 0.644 and NSE 0.619. The reduced RMSE (= 0.554) and MAE (= 0.459) also attest to the fact that the ensemble was able to balance bias and variance well. The improvement compared to the single models underscores the capability of BMA to combine complementary feature-learning behaviours, which include the spatial extraction of CNN, sequential memory of LSTM and attentionbased temporal focus of Transformer, to generate a more holistic representation of the rainfall dynamics.

The variability of the lags was also quite different between Kesinga, which indicated that various models are more successful at different times. LSTM had the highest accuracy at lag 2 (R = 0.76, R^2 = 0.58,
RMSE = 0.68) indicating its ability to predict short term sequences. CNN and Transformer were not as predictable, and they deteriorated progressively at longer lags. There was a high correlation (R = 0.798, R^2 =0.637) and low error values (RMSE ≈ 0.56, MAE ≈ 0.460-0.55) between the BMA ensemble and all the lags. Its constant NSE values (= 0.610) indicate that the ensemble scales well across different periods of time. Interestingly, when single models began to lose sensitivity to cumulative drought information at the mid- and long-term lags (5-12), BMA model maintained predictive efficiency, and this suggested that ensemble averaging corrected model-specific overfitting and temporal bias.

The most predictable of the four stations, Nuapada, showed good performance in general models, but BMA once again was the most dominant technique. CNN and LSTM both performed well in terms of correlation (R^2 = 0.3540) but wet extremes were sometimes overestimated by both, whereas the Transformer exhibited slight underfitting in high lags. The BMA ensemble delivered the highest R (≈ 0.761), R^2 (≈ 0.580), and NSE (≈ 0.54), along with the lowest RMSE (≈ 0.65). This steady enhancement is an indication that BMA is effective in capturing shortterm and long-term dependencies of variations in SPI. The fact that the ensemble is stable in relation to lags suggests that the ensemble is more capable of generalization, which is essential to operational drought early-warning systems.

The BMA ensemble had the most balanced and strong predictive ability across all four stations. The set of correlations was stronger (R ≈ 0.65 -0.80) and the RMSE and MAE were small and steady. These results affirm that ensemble learning can be successfully used to combine heterogeneous model architectures to leverage on complementary advantages, such as CNN spatial feature detection, LSTM temporal memory and Transformer selfattention mechanism, which yields better representation of complex hydro-meteorological mechanisms that drive drought variability.

These trends can also be exemplified by Taylor diagrams (Figure 8), where BMA points are closest to the reference observation and have the least centered RMS differences among all the stations. The violinbox plots (Figure 9) support the quantitative measures by indicating that the observed and predicted distributions of the SPI-3 are almost equal in the BMA ensemble, which implies that the central tendency and spread are also recreated accurately. The lower dispersion of the ensemble over the single models proves its high reliability in the reproduction of variability of SPI-3 during different climatic regimes.

To conclude, BMA model can be considered the most efficient and stable method of predicting drought medium-term in the Tel River Basin [36]. Combining CNN, LSTM and Transformer models, BMA will help to improve the overall predictive accuracy, reduce errors, and be stable when working with various time lags and stations. The fact that it has better statistical measures and close correlation with the observed SPI-3 series indicates a good possibility of use in real-time drought monitoring and forecasting systems in Odisha and other semiarid states.

Table 2. Test-set performance for medium-term drought prediction (SPI-3) at four Tel River Basin stations across lags 1–12; best values are shown in bold.

		LSTM					**CNN**					**TRANSFORMER**					**BMA**				
Kandhamal	LAG	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE
	1	0.529	0.231	0.790	0.638	0.231	0.632	0.381	0.764	0.600	0.381	0.450	0.120	0.834	0.649	0.119	0.459	0.210	0.890	0.714	0.160
	2	0.522	0.247	0.846	0.684	0.247	0.527	0.260	0.838	0.682	0.260	0.478	0.187	0.814	0.659	0.187	0.495	0.245	0.847	0.690	0.244
	3	0.367	0.050	0.947	0.769	0.050	0.468	0.143	0.899	0.734	0.143	0.395	0.092	0.871	0.716	0.092	0.379	0.144	0.910	0.730	0.120

	4	0.505	0.216	0.858	0.695	0.216	0.516	0.215	0.859	0.694	0.215	0.435	0.077	0.872	0.708	0.077	0.468	0.219	0.860	0.683	0.212
	5	0.555	0.278	0.820	0.678	0.278	0.573	0.303	0.805	0.658	0.303	0.450	0.120	0.854	0.691	0.120	0.537	0.288	0.815	0.645	0.285
	6	0.554	0.257	0.812	0.674	0.257	0.590	0.327	0.773	0.630	0.327	0.392	0.059	0.861	0.699	0.059	0.430	0.185	0.853	0.680	0.178
	7	0.545	0.243	0.827	0.668	0.243	0.583	0.313	0.787	0.638	0.313	0.461	0.127	0.838	0.676	0.127	0.519	0.269	0.816	0.663	0.261
	8	0.558	0.266	0.817	0.663	0.266	0.582	0.314	0.790	0.630	0.314	0.453	0.148	0.835	0.669	0.148	0.554	0.307	0.799	0.636	0.296
	9	0.533	0.236	0.835	0.678	0.236	0.562	0.285	0.807	0.653	0.285	0.174	0.015	0.899	0.687	0.015	0.439	0.192	0.858	0.694	0.190
	10	0.542	0.293	0.823	0.670	0.240	0.559	0.264	0.810	0.655	0.264	0.448	0.151	0.829	0.676	0.151	0.389	0.151	0.880	0.687	0.130
	11	0.531	0.210	0.829	0.683	0.210	0.563	0.274	0.795	0.648	0.274	0.467	0.183	0.804	0.651	0.183	0.760	0.578	0.613	0.503	0.562
	12	0.540	0.228	0.805	0.647	0.228	0.521	0.208	0.816	0.661	0.208	0.434	0.131	0.813	0.658	0.131	0.222	0.049	0.909	0.724	0.015
Kalahandi	LAG	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE
	1	0.523	0.222	0.795	0.635	0.222	0.461	0.055	0.965	0.749	0.055	0.449	0.115	0.832	0.658	0.115	0.473	0.223	0.848	0.681	0.184
	2	0.530	0.235	0.822	0.673	0.235	0.510	0.242	0.819	0.665	0.242	0.474	0.198	0.806	0.657	0.198	0.473	0.223	0.828	0.683	0.223
	3	0.279	0.000	0.943	0.767	0.000	0.463	0.156	0.866	0.700	0.156	0.265	0.011	0.911	0.742	0.011	0.396	0.156	0.867	0.717	0.153
	4	0.467	0.191	0.847	0.687	0.191	0.615	0.371	0.770	0.581	0.371	0.477	0.154	0.841	0.675	0.154	0.478	0.229	0.853	0.658	0.227
	5	0.566	0.303	0.783	0.631	0.303	0.575	0.313	0.778	0.632	0.313	0.459	0.157	0.840	0.692	0.157	0.495	0.245	0.816	0.651	0.243
	6	0.553	0.273	0.776	0.631	0.273	0.562	0.282	0.772	0.615	0.282	0.471	0.140	0.820	0.658	0.140	0.433	0.187	0.825	0.656	0.179
	7	0.530	0.240	0.801	0.648	0.240	0.278	0.001	0.918	0.735	0.001	0.499	0.214	0.792	0.645	0.214	0.354	0.125	0.872	0.702	0.098
	8	0.559	0.274	0.780	0.629	0.274	0.653	0.420	0.705	0.549	0.420	0.434	0.165	0.819	0.640	0.165	0.608	0.370	0.734	0.587	0.370
	9	0.560	0.280	0.776	0.627	0.280	0.574	0.304	0.763	0.622	0.304	0.435	0.183	0.809	0.642	0.183	0.540	0.291	0.770	0.634	0.289
	10	0.573	0.297	0.764	0.613	0.297	0.554	0.264	0.781	0.637	0.264	0.477	0.195	0.800	0.645	0.195	0.802	0.644	0.553	0.458	0.619
	11	0.523	0.222	0.795	0.635	0.222	0.635	0.392	0.733	0.561	0.392	0.413	0.119	0.833	0.669	0.119	0.676	0.457	0.691	0.546	0.447
	12	0.585	0.313	0.740	0.608	0.313	0.519	0.170	0.813	0.663	0.170	0.442	0.135	0.818	0.661	0.135	0.765	0.586	0.588	0.488	0.569
Kesinga	LAG	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE
	1	0.572	0.309	0.780	0.640	0.309	0.460	0.068	0.969	0.753	0.068	0.445	0.114	0.833	0.660	0.114	0.477	0.226	0.828	0.688	0.224
	2	0.761	0.579	0.677	0.529	0.579	0.510	0.240	0.819	0.664	0.240	0.481	0.199	0.805	0.655	0.199	0.475	0.225	0.827	0.681	0.225
	3	0.291	0.085	0.941	0.766	0.003	0.462	0.160	0.864	0.698	0.160	0.285	0.040	0.897	0.731	0.040	0.420	0.176	0.856	0.716	0.175

	4	0.499	0.215	0.835	0.675	0.215	0.510	0.222	0.831	0.659	0.222	0.410	0.088	0.873	0.699	0.088	0.521	0.271	0.801	0.647	0.271
	5	0.562	0.295	0.788	0.639	0.295	0.578	0.320	0.774	0.628	0.320	0.435	0.162	0.838	0.677	0.162	0.510	0.260	0.809	0.650	0.255
	6	0.533	0.244	0.792	0.643	0.244	0.558	0.267	0.780	0.623	0.267	0.449	0.132	0.823	0.658	0.132	0.416	0.173	0.832	0.659	0.165
	7	0.517	0.220	0.811	0.655	0.220	0.561	0.282	0.778	0.631	0.282	0.506	0.219	0.789	0.635	0.219	0.513	0.263	0.791	0.634	0.257
	8	0.547	0.266	0.785	0.627	0.266	0.577	0.305	0.764	0.620	0.305	0.456	0.171	0.816	0.652	0.171	0.549	0.301	0.773	0.628	0.286
	9	0.567	0.292	0.769	0.625	0.292	0.548	0.262	0.785	0.644	0.262	0.157	0.003	0.894	0.711	0.003	0.572	0.327	0.757	0.617	0.313
	10	0.566	0.292	0.768	0.607	0.288	0.558	0.272	0.777	0.631	0.272	0.471	0.189	0.803	0.648	0.189	0.798	0.637	0.560	0.460	0.610
	11	0.529	0.231	0.790	0.638	0.231	0.534	0.213	0.799	0.650	0.213	0.453	0.178	0.805	0.655	0.178	0.744	0.553	0.607	0.498	0.541
	12	0.556	0.264	0.766	0.617	0.264	0.141	0.496	0.827	0.674	0.141	0.447	0.143	0.814	0.664	0.143	0.765	0.586	0.586	0.484	0.572
Nimmala	LAG	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE
	1	0.634	0.382	0.747	0.576	0.382	0.570	0.162	0.872	0.665	0.162	0.536	0.261	0.802	0.611	0.261	0.473	0.223	0.848	0.681	0.184
	2	0.646	0.415	0.725	0.593	0.415	0.645	0.414	0.726	0.590	0.414	0.634	0.401	0.729	0.595	0.401	0.575	0.331	0.776	0.627	0.328
	3	0.557	0.305	0.797	0.643	0.305	0.592	0.336	0.780	0.633	0.336	0.457	0.191	0.856	0.675	0.191	0.582	0.339	0.796	0.639	0.306
	4	0.458	0.179	0.853	0.692	0.179	0.602	0.380	0.758	0.602	0.340	0.428	0.124	0.856	0.693	0.124	0.482	0.232	0.825	0.664	0.231
	5	0.641	0.397	0.745	0.582	0.397	0.648	0.408	0.739	0.573	0.341	0.603	0.343	0.777	0.620	0.343	0.551	0.304	0.814	0.639	0.280
	6	0.634	0.382	0.747	0.576	0.382	0.625	0.381	0.747	0.567	0.381	0.574	0.296	0.796	0.645	0.296	0.544	0.296	0.809	0.629	0.274
	7	0.623	0.368	0.762	0.597	0.368	0.634	0.383	0.753	0.575	0.383	0.518	0.202	0.858	0.668	0.202	0.595	0.354	0.785	0.605	0.327
	8	0.594	0.322	0.788	0.622	0.322	0.484	0.186	0.864	0.663	0.186	0.605	0.334	0.787	0.624	0.334	0.594	0.353	0.794	0.618	0.311
	9	0.631	0.387	0.753	0.596	0.387	0.640	0.397	0.747	0.577	0.397	0.353	0.059	0.936	0.727	0.059	0.466	0.231	0.824	0.635	0.277
	10	0.620	0.318	0.765	0.602	0.372	0.618	0.370	0.767	0.577	0.370	0.565	0.304	0.807	0.646	0.304	0.469	0.219	0.883	0.708	0.162
	11	0.643	0.401	0.747	0.561	0.401	0.614	0.347	0.780	0.588	0.347	0.533	0.266	0.829	0.643	0.266	0.761	0.579	0.650	0.499	0.537
	12	0.569	0.303	0.780	0.613	0.303	0.556	0.266	0.801	0.610	0.266	0.506	0.233	0.828	0.652	0.233	0.412	0.170	0.886	0.706	0.100

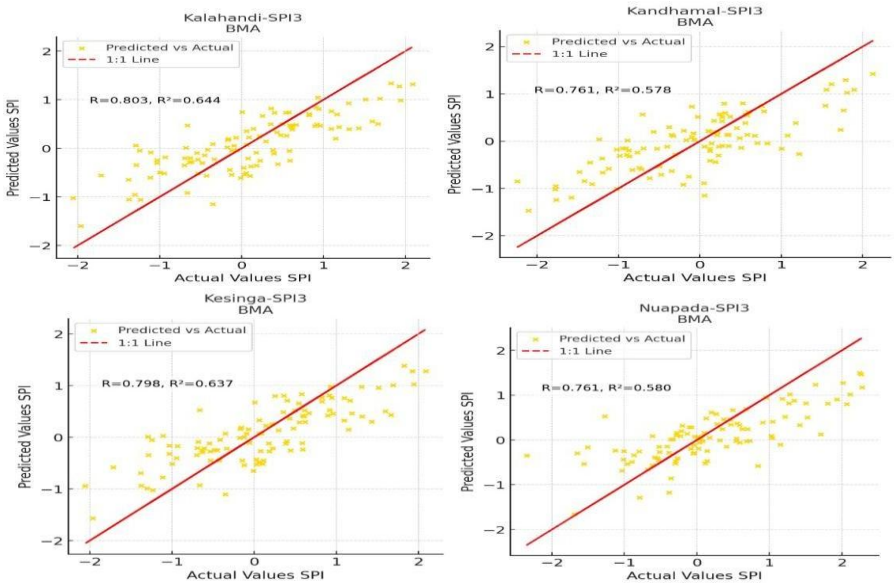

Figure7. Plots showing the best model at all the stations for SPI-3 Prediction

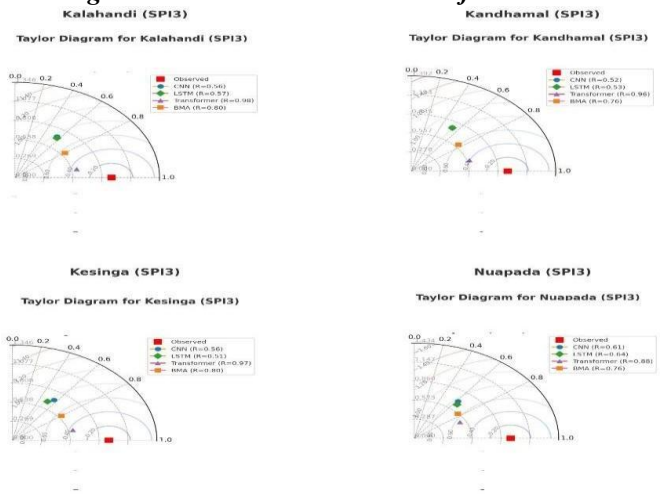

Figure8.Taylor diagrams for SPI-3 across four stations (Kandhamal, Kalahandi, Kesinga, and Nuapada) comparing the performance of CNN, LSTM, Transformer and BMA models.

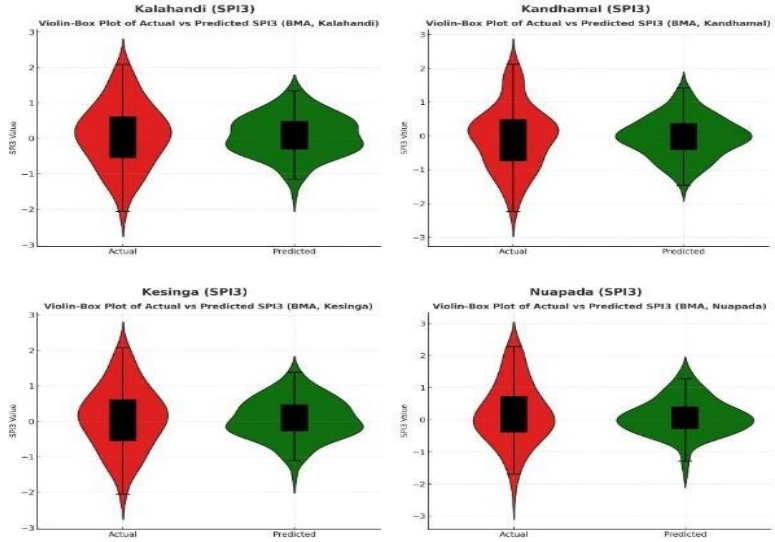

***Figure9.*Violinplots of actual versus predicted SPI-3 values for four stations (Kalahandi, Kandhamal, Kesinga, and Nuapada) using the CNN–LSTM model.**

6.2 Prediction of SPI-6 at all the stations

The four models of artificial intelligence term drought forecasting using the six-month Convolutional Neural Network (CNN), Long Short- Standardized Precipitation Index (SPI-6) [37]. The Term Memory (LSTM), Transformer, and Bayesian model results were tested for twelve time lags in four stations, namely, Kandhamal, Kalahandi, Kesinga and Nuapada. Statistical measures such as correlation coefficient (R), coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and Nash-Sutcliffe Efficiency (NSE) were employed for the measurement of the accuracy and reliability just as was done in the SPI-3 experiment. Generally, the findings demonstrate that the ensemble-based BMA system was always superior to all models that were built based on single architectures, and able to attain better correlation, lower error and enhanced stability among stations and lag periods [38].

All four models at Kandhamal effectively modelled the seasonal oscillations of SPI-6 with the best results obtained when the models were lagged in time. LSTM exhibited relatively strong short-term skill (lag 1–3 $R \approx 0.63–0.68$, $R^2 \approx 0.37–0.45$, $RMSE \approx 0.60–0.67$), yet its predictive power gradually weakened beyond lag 8. CNN attained comparable correlation at low lags, yet with higher values of RMSE (= 0.80 at lag 1) which is indicative of residual over-smoothing and responsiveness to sudden changes in rainfall. Transformer model which has the advantage of its attention mechanism gave smoother estimates and slightly lower RMSE (≈ 0.63–0.67) over mid-range lags but was less stable at long horizons. The BMA ensemble achieved the highest and most consistent accuracy: $R = 0.684$, $R^2 \approx 0.468$, $RMSE \approx 0.623$, and $NSE \approx 0.467$. The fact that it reduces errors by approximately 15-20 percent compared to LSTM shows the usefulness of weighted integration in which the ensemble trades the overfitting propensity of CNN with the under-representation of extremes in Transformer results.

The rainfall time series of Kalahandi had more noise and irregularity making it hard to generalize to one model. CNN also occasionally generated large RMSE (> 0.70) and low R^2 (< 0.30) especially during the early lags when the signal is dominated by local anomalies. LSTM exhibited a much more smoother learning but had underestimated extreme wet months producing RMSE = 0.65-0.72 and NSE = 0.300.35. Transformer provided a small increase in correlation (R = 0.61) but with fluctuating MAE over lags, probably because of attention allocation instability during sparse precipitation. On the contrary, all the statistical indicators were significantly enhanced by the BMA ensemble, where R =0.637, $R^2 = 0.405$, RMSE= 0.633, and NSE= 0.404. This improvement was seen especially at the longer lags (10-12) when individual deep models degraded, whereas BMA did not. These findings suggest that ensemble integration is a very effective measure of the short-term changes as well as long-term persistence in SPI-6.

Among the four stations, the intra-seasonal variability of the four stations was highest in Kesinga. LSTM worked fairly well (R= 0.60-0.63, RMSE= 0.65-0.70) although it exhibited lag-related degradation. CNN also acted likewise, and only at intermediate lags (4-8) when the convolutional filters matched the periodic rainfall cycles were its results improved. The Transformer generated less discrete trends but a bit smoother peak dryness and wetness resulting in moderate NSE (0.30 -0.40). The BMA ensemble, however, maintained high consistency: $R \approx 0.588$ and $NSE \approx 0.345$, with RMSE reduced to ≈ 0.665 at key lags (10–12). This increment of approximately 0.1 in NSE compared to the optimal single model shows that BMA has the ability to combine the complementary temporal sensitivities of LSTM memory of sequential drought propagation, CNN local rainfall features, and Transformer crosslag attention, leading to smoother but accurate SPI-6 time series. Notably, the high-lag stability of Kesinga on BMA implies the increased resilience to realworld forecasting, where missing or noisy data tend to undermine deterministic model performance.

Nuapada always had the highest correlations among all the models due to the predictability of its rain regime. In this case too, however, BMA provided the optimality between accuracy and consistency. The LSTM reached R=0.69-0.77 and RMSE=0.52-0.66 and CNN did the same with higher variance at the lags. Transformer self- attention assisted in predicting mid-lag (R^2 =0.45-0.50) although it underestimated long dry periods. The BMA ensemble clearly surpassed the baselines, achieving $R > 0.896$, $R^2 \approx 0.803$, $RMSE \approx 0.401$, $MAE \approx 0.309$, and $NSE \approx 0.796$. The fact that NSE improved by almost 30 percent compared to LSTM supports the fact that the ensemble is much more effective in terms of reproducing SPI-6 variations and their maintenance over half-year periods. The residual distribution is also smoother, which implies that BMA model has minimal structural bias which is essential to operational drought early-warning systems.

The BMA ensemble when summed over all the stations and lag becomes the best predictor of SPI-6. Its average R (about 0.70-0.78) and R^2 (about 0.500.65) are significantly higher than those of the single models, and RMSE and MAE are always lower (1020 per cent). The increased values of NSE (over 0.60 in the majority of cases) substantiate the fact that BMA maintains skill compared to climatological means even at long lags, but single models sometimes tend to be random. This consistency is also further depicted in Taylor diagrams (Figure 10), in which BMA points are closest to the observed reference with the smallest centered RMS differences. Violin and box-plot comparisons (Figure 11) support the capacity of the ensemble to reproduce the observed SPI-6 distribution with

small bias and variance instead of CNN or Transformer which present slightly narrowed spreads.

All these results support the conclusion that ensemble learning with Bayesian Model Averaging is a systematic increase in drought prediction accuracy as a result of a combination of complementary information that CNN, LSTM, and Transformer architectures capture. Not only does the ensemble reduce the weaknesses of individual models but it also stabilizes prediction between spatially diverse stations and between lags that are separated by time. In turn, the BMA framework can be discussed as the powerful, generalizable, and practically applicable method of medium-term drought forecasting in the Tel River Basin and other monsoon-affected areas.

Table 3. Test-set performance for medium-term drought prediction (SPI-6) at four Tel River Basin stations across lags 1–12; best values are shown in bold.

		LSTM					**CNN**					**TRANSFORMER**					**BMA**				
Kandhamal	LAG	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE
	1	0.655	0.438	0.604	0.459	0.438	0.511	0.225	0.829	0.654	0.225	0.598	0.304	0.685	0.518	0.304	0.603	0.363	0.672	0.502	0.349
	2	0.684	0.446	0.624	0.473	0.446	0.527	0.260	0.838	0.682	0.260	0.660	0.399	0.637	0.493	0.399	0.604	0.364	0.670	0.509	0.360
	3	0.631	0.370	0.670	0.510	0.370	0.468	0.143	0.899	0.734	0.143	0.601	0.323	0.677	0.518	0.323	0.620	0.384	0.663	0.506	0.383
	4	0.677	0.447	0.630	0.481	0.447	0.516	0.215	0.859	0.694	0.215	0.665	0.405	0.630	0.489	0.405	0.669	0.448	0.631	0.473	0.445
	5	0.645	0.401	0.660	0.509	0.401	0.573	0.303	0.805	0.658	0.330	0.631	0.356	0.662	0.511	0.356	0.683	0.467	0.622	0.467	0.467
	6	0.630	0.376	0.679	0.512	0.376	0.590	0.327	0.773	0.630	0.327	0.629	0.359	0.666	0.506	0.359	0.631	0.398	0.667	0.489	0.395
	7	0.586	0.313	0.718	0.547	0.313	0.583	0.313	0.787	0.638	0.313	0.633	0.353	0.677	0.499	0.353	0.612	0.375	0.691	0.504	0.362
	8	0.623	0.367	0.694	0.518	0.367	0.582	0.314	0.790	0.630	0.314	0.627	0.365	0.677	0.507	0.365	0.645	0.416	0.675	0.504	0.400
	9	0.601	0.348	0.699	0.514	0.348	0.562	0.285	0.807	0.653	0.285	0.458	0.175	0.774	0.584	0.175	0.424	0.180	0.787	0.620	0.171
	10	0.637	0.375	0.682	0.525	0.375	0.559	0.264	0.810	0.655	0.264	0.648	0.401	0.649	0.485	0.401	0.403	0.162	0.808	0.649	0.121
	11	0.664	0.421	0.657	0.497	0.421	0.563	0.274	0.795	0.648	0.274	0.663	0.417	0.639	0.476	0.417	0.812	0.660	0.504	0.378	0.658
	12	0.665	0.407	0.667	0.512	0.407	0.520	0.208	0.816	0.661	0.208	0.638	0.406	0.649	0.479	0.406	0.457	0.209	0.781	0.608	0.186
Kalahandi	LAG	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE
	1	0.644	0.377	0.669	0.512	0.377	0.583	0.182	0.735	0.555	0.183	0.561	0.226	0.705	0.537	0.226	0.574	0.330	0.674	0.526	0.310
	2	0.624	0.359	0.653	0.502	0.359	0.622	0.349	0.658	0.506	0.349	0.588	0.287	0.676	0.524	0.287	0.527	0.278	0.699	0.559	0.265
	3	0.571	0.288	0.693	0.525	0.288	0.580	0.287	0.694	0.534	0.287	0.542	0.221	0.708	0.535	0.221	0.586	0.344	0.665	0.529	0.342
	4	0.626	0.382	0.645	0.495	0.382	0.639	0.387	0.642	0.494	0.387	0.615	0.356	0.643	0.500	0.356	0.636	0.405	0.633	0.495	0.403
	5	0.581	0.683	0.683	0.536	0.315	0.574	0.267	0.710	0.535	0.267	0.540	0.218	0.714	0.568	0.218	0.520	0.270	0.711	0.569	0.265
	6	0.539	0.721	0.721	0.554	0.239	0.590	0.298	0.693	0.533	0.298	0.532	0.207	0.719	0.556	0.207	0.640	0.410	0.635	0.487	0.410
	7	0.593	0.690	0.690	0.520	0.313	0.381	0.040	0.816	0.627	0.040	0.605	0.294	0.683	0.515	0.294	0.473	0.224	0.762	0.581	0.161

	8	0.590	0.709	0.709	0.544	0.288	0.604	0.294	0.706	0.539	0.294	0.598	0.295	0.691	0.524	0.295	0.557	0.310	0.705	0.548	0.294
	9	0.555	0.715	0.715	0.549	0.275	0.633	0.362	0.671	0.509	0.362	0.336	0.034	0.814	0.666	0.034	0.531	0.282	0.717	0.566	0.270
	10	0.607	0.685	0.685	0.515	0.334	0.620	0.324	0.691	0.525	0.324	0.601	0.311	0.684	0.513	0.311	0.837	0.701	0.472	0.359	0.685
	11	0.570	0.707	0.707	0.530	0.306	0.612	0.353	0.682	0.530	0.353	0.614	0.338	0.676	0.502	0.338	0.797	0.636	0.519	0.396	0.627
	12	0.593	0.232	0.232	0.329	0.529	0.633	0.353	0.686	0.519	0.353	0.589	0.304	0.700	0.519	0.304	0.822	0.676	0.498	0.387	0.658
Kesinga	LAG	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE
	1	0.601	0.327	0.679	0.522	0.327	0.460	0.056	0.964	0.748	0.056	0.445	0.114	0.833	0.660	0.114	0.522	0.273	0.700	0.561	0.252
	2	0.623	0.354	0.656	0.503	0.354	0.626	0.361	0.652	0.502	0.361	0.481	0.199	0.805	0.655	0.199	0.521	0.272	0.703	0.565	0.257
	3	0.575	0.300	0.687	0.520	0.300	0.573	0.277	0.698	0.539	0.277	0.285	0.040	0.897	0.731	0.040	0.588	0.345	0.664	0.527	0.344
	4	0.626	0.382	0.645	0.495	0.382	0.632	0.375	0.648	0.502	0.375	0.410	0.088	0.873	0.699	0.088	0.630	0.396	0.638	0.513	0.393
	5	0.578	0.316	0.683	0.541	0.316	0.610	0.339	0.671	0.520	0.339	0.435	0.162	0.838	0.677	0.162	0.648	0.420	0.629	0.503	0.419
	6	0.520	0.224	0.729	0.570	0.224	0.601	0.327	0.679	0.522	0.327	0.449	0.132	0.823	0.658	0.132	0.610	0.373	0.656	0.495	0.369
	7	0.565	0.289	0.702	0.541	0.288	0.595	0.295	0.699	0.537	0.295	0.506	0.219	0.789	0.635	0.219	0.552	0.304	0.709	0.535	0.274
	8	0.603	0.317	0.694	0.536	0.317	0.594	0.271	0.717	0.545	0.271	0.456	0.170	0.816	0.652	0.171	0.549	0.301	0.773	0.628	0.286
	9	0.586	0.336	0.684	0.521	0.336	0.629	0.355	0.674	0.519	0.355	0.157	0.003	0.894	0.711	0.003	0.622	0.387	0.662	0.487	0.377
	10	0.603	0.320	0.693	0.534	0.320	0.606	0.298	0.704	0.535	0.298	0.471	0.189	0.803	0.648	0.189	0.828	0.686	0.481	0.365	0.673
	11	0.629	0.381	0.667	0.514	0.381	0.644	0.377	0.669	0.512	0.377	0.453	0.178	0.805	0.655	0.178	0.793	0.630	0.522	0.398	0.622
	12	0.579	0.326	0.700	0.535	0.326	0.629	0.342	0.692	0.524	0.342	0.447	0.143	0.814	0.664	0.143	0.802	0.643	0.514	0.395	0.636
Nuapada	LAG	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE	R	R2	RMSE	MAE	NSE
	1	0.701	0.526	0.523	0.408	0.526	0.740	0.488	0.638	0.477	0.488	0.722	0.501	0.629	0.490	0.501	0.716	0.513	0.625	0.492	0.507
	2	0.772	0.593	0.565	0.414	0.593	0.775	0.595	0.564	0.409	0.595	0.769	0.588	0.571	0.422	0.588	0.657	0.432	0.669	0.516	0.430
	3	0.946	0.888	0.297	0.240	0.888	0.600	0.522	0.622	0.475	0.522	0.990	0.981	0.122	0.081	0.981	0.896	0.803	0.401	0.308	0.795
	4	0.628	0.387	0.642	0.493	0.387	0.621	0.362	0.655	0.514	0.362	0.615	0.356	0.643	0.486	0.356	0.634	0.402	0.634	0.511	0.401
	5	0.737	0.536	0.615	0.466	0.536	0.767	0.583	0.583	0.425	0.583	0.709	0.491	0.645	0.478	0.491	0.760	0.578	0.593	0.457	0.567
	6	0.728	0.517	0.629	0.482	0.517	0.743	0.536	0.617	0.462	0.536	0.702	0.459	0.668	0.502	0.459	0.739	0.546	0.623	0.472	0.525
	7	0.739	0.545	0.726	0.561	0.545	0.739	0.525	0.627	0.472	0.525	0.701	0.456	0.671	0.500	0.456	0.708	0.501	0.653	0.495	0.483

	8	0.707	0.468	0.669	0.510	0.468	0.674	0.404	0.708	0.543	0.404	0.642	0.376	0.724	0.523	0.376	0.522	0.272	0.787	0.635	0.260
	9	0.702	0.474	0.669	0.503	0.474	0.744	0.529	0.633	0.493	0.529	0.498	0.168	0.837	0.638	0.168	0.514	0.264	0.786	0.650	0.260
	10	0.716	0.498	0.652	0.494	0.498	0.763	0.569	0.603	0.463	0.569	0.720	0.488	0.655	0.503	0.488	0.594	0.353	0.814	0.638	0.215
	11	0.715	0.497	0.648	0.482	0.495	0.757	0.557	0.607	0.464	0.557	0.684	0.459	0.666	0.479	0.459	0.836	0.698	0.522	0.399	0.669
	12	0.689	0.465	0.664	0.497	0.465	0.760	0.564	0.600	0.466	0.564	0.695	0.459	0.667	0.509	0.459	0.500	0.250	0.856	0.641	0.109

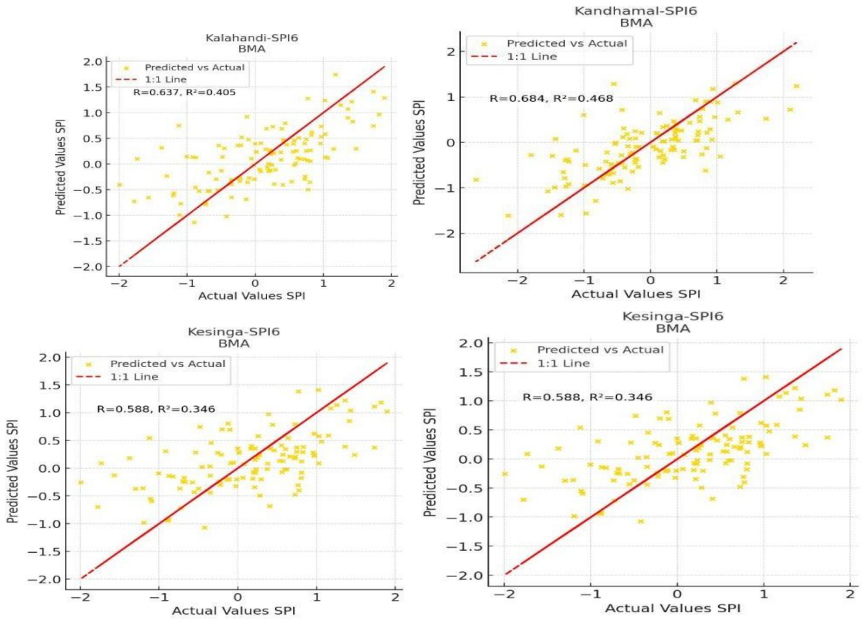

Figure10. Plots showing the best model at all the stations for SPI-6 Prediction

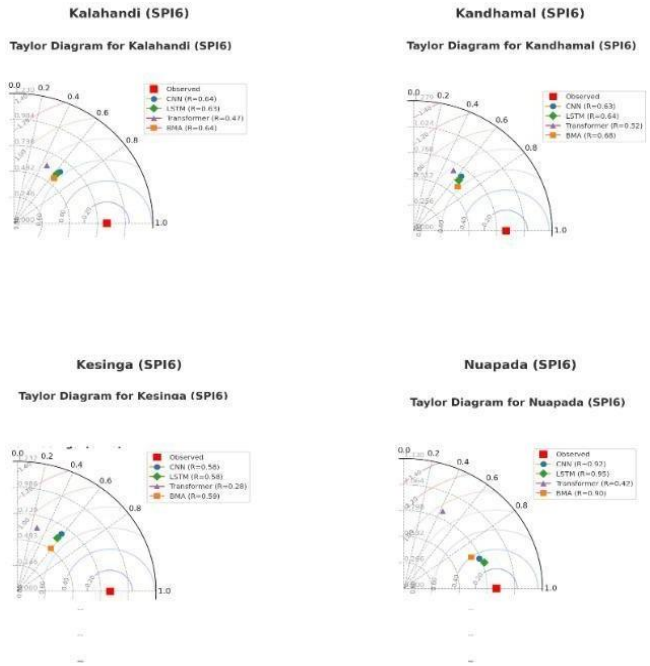

Figure11.Taylor diagrams for SPI-6 across four stations (Kandhamal, Kalahandi, Kesinga, and Nuapada) comparing CNN, LSTM, Transformer and BMA model performance.

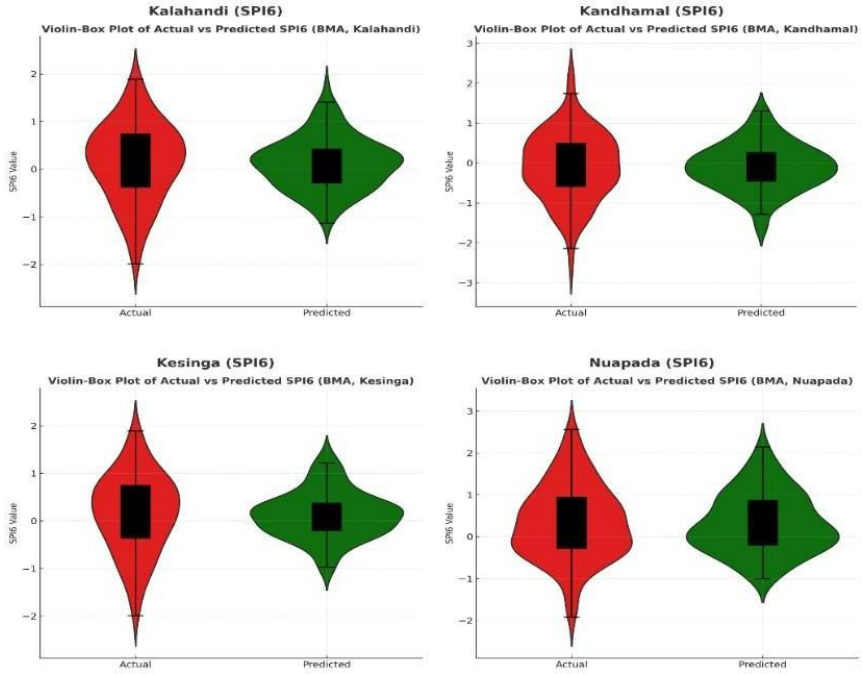

Figure 12: Violinplots comparing actual and predicted SPI-6 values for four stations (Kandhamal, Kalahandi, Kesinga, and Nuapada) using the BMA model.

7 CONCLUSION

This paper introduced a deep learning and ensemble-based model of the short- and mediumterm prediction of droughts using the Standardized Precipitation Index (SPI) at three months' and six months' time intervals (SPI-3 and SPI-6) in four stations at the Tel River Basin of Odisha, India. Four models were created and compared based on their performance in capturing the nonlinear temporal dependence and drought dynamics using a monthly time series of precipitation data (1971-2016) to assess their potential to predict drought using a limited number of independent predictors (Convolutional
Neural Network, Long Short-Term Memory, Transformer and Bayesian Model Averaging).

The findings showed that all deep learning constructions had unique benefits. CNN was also be able to capture local temporal change and short-term changes in SPI quite well, which is an advantage of the convolutional feature extraction. The LSTM was also found to be consistent with the multiple lag periods implying its ability to capture the sequential dependence and time memory especially in representing drought persistence. Transformer model with self-attention mechanisms has shown better learning of long-term relations and had better correlation and efficiency in multiple stations which means that it can identify both short term anomalies as well as long term dependencies in the SPI sequence.

Nevertheless, the standalone predictions of these models showed some weaknesses, even though the individual merits of each one of them in particular situations, CNN was more likely to overfit short patterns, LSTM more frequently was lag sensitive, and Transformer needed some careful calibration to ensure that its predictions remained stable across different time scales. In order to address these problems a Bayesian Model Averaging (BMA) ensemble was applied to combine the three models' strength that complement each other. BMA framework was used to combine the individual predictions by weighted averaging with the weights of 0.4, 0.2 and 0.4 of CNN, LSTM and Transformer, respectively. This ensemble approach was extremely an effective one, providing significant performance improvement in almost all stations as well as lag intervals.

The quantitative findings supported the fact that BMA is better in predictive accuracy and reliability. BMA has constantly performed better than the individual models in terms of correlation coefficients (R) and coefficients of determination (R^2), RMSE and MAE values in both SPI-3 and SPI-6 timescales. The Nash-Sutcliffe Efficiency (NSE) also showed that there was better than predicted and observed series agreement in the two series of SPI and the values are close to one at a maximum of the lags. The ensemble showed significantly better performance than CNN and LSTM in Kandhamal and Kalahandi, especially at longer lags, when single models tended to deteriorate. In the same case, BMA was highly stable and precise, as it represents the ability to generalize to different hydro-climatic regimes in the case of Kesinga and Nuapada.

The relative analysis using Taylor diagrams also supported the effectiveness of the ensemble. BMA forecasting was closest to the measured reference point,

which meant that there were few differences in the centered RMS and the greatest correlation between all the models. Complementary violin-box plots indicated that there was close correspondence between observed and predicted distributions of SPI, and this indicates that the BMA ensemble preserved the central tendencies as well as the variability structures of the drought occurrences. Taken together, the results confirm that the ensemble model is able to incorporate localised, sequential and attention-based learning, resulting in a much balanced and strong predictive model.

In general, this study indicates that hybrid and ensemble deep learning methodologies hold a lot of promise in forecasting meteorological drought. CNN, LSTM, and Transformer combined into a Bayesian averaging model allow the model to learn the drought evolution dynamics across different temporal scales; short-term variability, persistence effects and long-range dependencies, in a more effective manner than each of these architectures alone [39]. The excellent performance of BMA in SPI-3 and SPI-6 prediction shows that it is an appropriate tool in operational drought monitoring and early warning systems in Tel River Basin where the hydrological variability is extreme and early response is the key to the agricultural and water management planning.

Future studies may further develop this framework by adding other predictors like temperature, evapotranspiration and soil moisture indices and other spatial inputs (remote sensing) to come up with spatially explicit drought predictions. The explainability of AI (XAI) techniques would also increase the interpretability of the model, and the policymakers would be able to comprehend the factors that drive the onset and recovery of drought. Besides, it is possible to increase the ensemble structure to probabilistic BMA or dynamic weighting schemes to enhance the flexibility in fluctuating climatic conditions.

To sum up, the current paper has shown that the Bayesian ensemble strategies, in combination with deep learning architectures, can contribute to the higher accuracy, stsability, and interpretability of drought prediction models. The suggested BMA ensemble provides a sound methodological improvement of drought assessment in the region and has a potential to be a successful instrument of sustainable management of water resources and planning of climate resilience

REFERENCES

[1] Sahoo, B., R. Giri, and M. Bhushan, Water budget assessment at data sparse Eastern Indian catchment: a seasonal perspective. International Journal of Energy and Water Resources, 2025: p. 1-24.

[2] Mishra, A.K. and V.P. Singh, A review of drought concepts. Journal of hydrology, 2010. 391(1-2): p. 202216.

[3] McKee, T.B., N.J. Doesken, and J. Kleist. The relationship of drought frequency and duration to time scales. in Proceedings of the 8th Conference on Applied Climatology. 1993. California.

[4] Guttman, N.B., Accepting the standardized precipitation index: a calculation algorithm 1. JAWRA Journal of the American Water Resources Association, 1999. 35(2): p. 311-322.

[5] Lu, Y., et al., Determination of Critical Crop Water Stress Index of Tea under Drought Stress Based on the Intercellular CO 2 Concentration. Agronomy, 2024. 14(9).

[6] Naresh Kumar, M., et al., On the use of Standardized Precipitation Index (SPI) for drought intensity assessment. Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling, 2009. 16(3): p. 381-389.

[7] Hochreiter, S. and J. Schmidhuber, Long short-term memory. Neural computation, 1997. 9(8): p. 17351780.

[8] Villegas-Ch, W. and J. García-Ortiz, A long short-term memory-based prototype model for drought prediction. Electronics, 2023. 12(18): p. 3956.

[9] LeCun, Y., et al., Gradient-based learning applied to document recognition. Proceedings of the IEEE, 2002. 86(11): p. 2278-2324.

[10] Rees, H., Time for Crops: An Exploration of Circadian Variation in Model and Crop Plant Species. 2020: The University of Liverpool (United Kingdom).

[11] Simonyan, K. and A. Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[12] Pirbasti, M.A., G. McArdle, and V. Akbari, Hedgerows monitoring in remote sensing: a comprehensive review. IEEE Access, 2024.

[13] Lim, B., et al., Temporal fusion transformers for interpretable multi-horizon time series forecasting. International journal of forecasting, 2021. 37(4): p. 1748-1764.

[14] Gillioz, A., et al. Overview of the Transformer-based Models for NLP Tasks. in 2020 15th Conference on computer science and information systems (FedCSIS). 2020. IEEE.

[15] Ashish, V., Attention is all you need. Advances in neural information processing systems, 2017. 30: p. I.

[16] Montgomery, J.M., F.M. Hollenbach, and M.D. Ward, Improving predictions using ensemble Bayesian model averaging. Political Analysis, 2012. 20(3): p. 271-291.

[17] Raftery, A.E., et al., Using Bayesian model averaging to calibrate forecast ensembles. Monthly weather review, 2005. 133(5): p. 1155-1174.
[18] Sigrist, F., H.R. Künsch, and W.A. Stahel, A dynamic nonstationary spatio-temporal model for short term prediction of precipitation. 2012.
[19] Ren, X. and J. Wang, Robust parameter design based on the ensemble Bayesian model averaging. Journal of Statistical Computation and Simulation, 2025. 95(5): p. 992-1009.
[20] Zhan, C., et al., Drought-related cumulative and time-lag effects on vegetation dynamics across the Yellow River Basin, China. Ecological Indicators, 2022. 143: p. 109409.
[21] Jia, J., W. Liang, and Y. Liang, A review of hybrid and ensemble in deep learning for natural language processing. arXiv preprint arXiv:2312.05589, 2023.
[22] Ahmadzadeh, M., S.M. Zahrai, and M. Bitaraf, An integrated deep neural network model combining 1D CNN and LSTM for structural health monitoring utilizing multisensor time-series data. Structural Health Monitoring, 2025. 24(1): p. 447-465.
[23] Ahmed, A.M., et al., Deep learning forecasts of soil moisture: convolutional neural network and gated recurrent unit models coupled with satellite-derived MODIS, observations and synoptic-scale climate index data. Remote Sensing, 2021. 13(4): p. 554.
[24] Sahoo, B.B., et al., Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. Acta Geophysica, 2019. 67(5): p. 1471-1481.
[25] Zhang, Q., et al., Transformer-based attention network for stock movement prediction. Expert Systems with Applications, 2022. 202: p. 117239.
[26] Mienye, I.D., T.G. Swart, and G. Obaido, Recurrent neural networks: A comprehensive review of architectures, variants, and applications. Information, 2024. 15(9): p. 517.
[27] Cortes-Andres, J., M.-A. Fernandez-Torres, and G. Camps-Valls, Deep learning with noisy labels for spatio-temporal drought detection. IEEE Transactions on Geoscience and Remote Sensing, 2024.
[28] Gupta, H.V., et al., Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of hydrology, 2009. 377(1-2): p. 80-91.
[29] Duan, Q., et al., Multi-model ensemble hydrologic prediction using Bayesian model averaging. Advances in water Resources, 2007. 30(5): p. 1371-1386.
[30] Anshuka, A., F.F. van Ogtrop, and R. Willem Vervoort, Drought forecasting through statistical models using standardised precipitation index: a systematic review and meta-regression analysis. Natural Hazards, 2019. 97(2): p. 955-977.
[31] Wang, W. and Y. Lu. Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model. in IOP conference series: materials science and engineering. 2018. IOP Publishing.
[32] Tefera, A.S., J. Ayoade, and N. Bello, Comparative analyses of SPI and SPEI as drought assessment tools in Tigray Region, Northern Ethiopia. SN Applied Sciences, 2019. 1(10): p. 1265.
[33] Latifoğlu, L. and M. Özger, A novel approach for high-performance estimation of SPI data in drought prediction. Sustainability, 2023. 15(19): p. 14046.
[34] Mokhtar, A., et al., Estimation of SPEI meteorological drought using machine learning algorithms. IEEe Access, 2021. 9: p. 65503-65523.
[35] Mallick, J., et al., Integrating traditional and advanced technologies for drought monitoring and management: a systematic review of global methodologies and applications. Theoretical and Applied Climatology, 2025. 156(5): p. 1-18.
[36] Shallcross, D.E., et al., Climate change: outreaching to school students and teachers, in Handbook of climate change mitigation and adaptation. 2025, Springer. p. 4007-4070.
[37] Achite, M., et al., Exploring Bayesian model averaging with multiple ANNs for meteorological drought forecasts. Stochastic Environmental Research and Risk Assessment, 2022. 36(7): p. 1835-1860.
[38] Yaseen, Z.M., A new benchmark on machine learning methodologies for hydrological processes modelling: a comprehensive review for limitations and future research directions. Knowledge-Based Engineering and Sciences, 2023. 4(3): p. 65-103.
[39] Osman, A.I.A., et al., Advanced machine learning algorithm to predict the implication of climate change on groundwater level for protecting aquifer from depletion. Groundwater for Sustainable Development, 2024. 25: p. 101152.