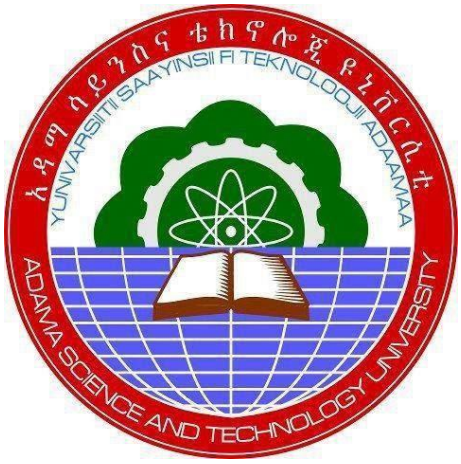


Adama Science and Technology University

School of Electrical Engineering and Computing

Department of Computer Science and Engineering

M.Sc Regular Program, Computer Vision SIG

Drought Prediction Using CNN and LSTM Techniques

1. Helina Tefera	PGR/35782/16
2. Mulugeta Berhe	PGR/35786/16
3. Wendirad Demelash	PGR/35791/16

Submitted to: Werku J. (Assoc. Prof)

June 2024

Adama, Ethiopia

Abstract

Drought is a complex natural disaster that will directly impact human living status, not only the living status of individuals but also the economic status. The best prediction of whether there is a drought in the upcoming seasons or not is important for effective water resource management and agricultural planning. Our study presents a comparison between multiple approaches to drought prediction using Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM) based approaches. We developed a model to forecast rainfall, maximum temperature, and minimum temperature, which are critical indicators of drought conditions. Our study leverages the strengths of Artificial Neural Network to estimate and predict drought for specific regions. The model was trained and validated using a comprehensive dataset comprising over 30 years of monthly precipitation, maximum temperature, and minimum temperature records from different geographical sites. Based on these forecasts, we utilized the Standardized Precipitation Evapotranspiration Index (SPEI) to predict drought conditions.

Our evaluation of various LSTM-based models, including Vanilla LSTM, Stacked LSTM, Bidirectional LSTM, Stacked Bidirectional LSTM, CNN-LSTM, and ConvLSTM, revealed that the Vanilla LSTM and ConvLSTM models performed best for short-term predictions, particularly for SPEI-3, with high scores across accuracy, precision, recall, and F1-Score metrics. However, all models showed a decline in performance for longer timescales, highlighting the challenge of maintaining accuracy over extended periods.

Contents

List of Figures

List of Tables

Acronyms

AI Aridity Index.

ANN Artificial Neural Network.

ARIMA AutoRegressive Integrated Moving Average.

BiLSTM Bidirectional LSTM.

CMI Crop Moisture Index.

CNN Convolutional Neural Networks.

ConvLSTM Convolutional LSTM.

DRI Drought Reconnaissance Index.

HTC Hydro-thermal Coefficient of Selyaninov.

KBDI Keetch-Byram Drought Index.

LSTM Long Short Term Memory.

MAE Mean Absolute Error.

MPE Mean Percentage Error.

MSE Mean Squared Error.

NDVI Normalized Difference Vegetation Index.

PDSI Palmer Drought Severity Index.

PET potential evapotranspiration.

PET Weighted Anomaly Standardized Precipitation.

ReLU Rectified Linear Unit.

RNN Recurrent Neural Network.

SDI Standardized Drought Index.

SPEI Standardized Precipitation Evapotranspiration Index.

SPI Standardized Precipitation Index.

SVM Support Vector Machines.

SVR Support Vector Regression.

1 Introduction

1.1 Background

Drought is a natural hazard characterized by a prolonged period of deficient rainfall, leading to a significant water shortage in the affected area. It can have several impacts on agriculture, water supply, and the environment, posing challenges to both human and ecological systems. Unlike other natural disasters, droughts develop slowly and can last for months or even years, making them particularly difficult to predict and manage. The complexity of drought arises from its varying definitions, which can be meteorological, agricultural, hydrological, or socioeconomic, each highlighting different aspects of the phenomenon. Accurate prediction of drought conditions is crucial for implementing effective water resource management and agricultural planning strategies, thereby mitigating the adverse effects of droughts on society.

In recent years, drought prediction has advanced significantly, introducing improvements in data collection, computational power, and modeling techniques which is promising and brings a lot of advancement for predicting drought earlier. There are traditional methods including statistical models and hydrological simulations that use historical weather data to predict future drought conditions. Now there are many recent papers on machine learning and deep learning methods have been increasingly adopted due to their ability and performance to handle large datasets and capture complex patterns for predicting the future using patterns from previous data provided [2].

Models like Support Vector Machines (SVM), Random Forests, and Artificial Neural Network (ANN) have shown promise in various studies. Additionally, remote sensing technologies provide valuable real-time predictions. Specifically, the integration of Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM) networks, known as like ConvLSTM models, has emerged as a powerful tool for drought prediction. These models combine the spatial feature extraction capabilities of CNNs with the temporal sequence learning ability of LSTMs, making them well-suited for climate data analysis. In addition to that Standardized Precipitation Evapotranspiration Index (SPEI) and other indexing techniques are commonly used to quantify drought severity and duration, providing a standardized way to assess and compare drought conditions across different regions and periods.

1.2 Problem Statement

Drought is a natural disaster that has negatively impacted human lives such as agricultural productivity, water resources, and ecosystem health. Accurate predic-

Figure 1: Drought in the Borena zone is said to have killed over three million cattle (Photo: SM)

tion of drought conditions is very essential for effective management strategies. Recently, the Oromia region, particularly the Borena zone, has been severely affected by drought, indicating the need for reliable prediction models[9].

1.3 Objectives

The main objective of our research is to determine an advanced drought prediction model using Long Short Term Memory (LSTM) networks. Specifically, our study aims to forecast critical climatic variables like predict monthly rainfall, maximum temperature, and minimum temperature using historical climate data spanning over 33 years from various sites. We predicted drought conditions and evaluated them with different matrices which will be leveraging the strengths of Artificial Neural Network networks to capture temporal dependencies in the climate data. We provide regional insights by focusing on specific regions, to address the specific challenges and needs of drought-prone areas which will help that specifying region by contributing to better water resource management and agricultural planning.

2 Related Works

2.1 Statistical Models

Statistical models have long been used for time series forecasting and drought prediction. Among these, AutoRegressive Integrated Moving Average (ARIMA) is a traditional statistical model for predicting drought conditions. Support Vector Regression (SVR), another statistical model, has also been utilized for similar purposes. These models provide a baseline for comparing the performance of more advanced methods in drought prediction [10].

2.2 Neural Network Models

Recently, CNN variants have gained prominence in image processing and pattern recognition, particularly in satellite-based scenarios. A study employed the EfficientNet architecture to classify drought conditions using a satellite image database from the Kolar region of Karnataka. The performance of EfficientNet was compared with state-of-the-art CNN variants, including the original CNN, AlexNet, and VGGNet. The findings underscore the effectiveness of CNN variants, particularly EfficientNet, in enhancing the accuracy of drought classification using satellite images [4].

2.3 Hybrid Models

Hybrid models, which combine different modeling approaches, have shown superior performance in drought prediction. A study compared the prediction accuracy of three hybrid models (ARIMA-SVR, LS-SVR, and ARIMA-LSTM) with that of individual models (ARIMA, SVR, and LSTM). The hybrid ARIMA-LSTM model achieved the highest prediction accuracy across all timescales, particularly at 6, 12, and 24 months. The model demonstrated the lowest prediction accuracy at SPEI1 with a 1-month lead time but the highest accuracy at SPI24 and SPEI24. The prediction accuracy of the hybrid models decreased as the prediction advance period increased, making the 1-month lead time models more accurate than the 2-month lead time models. Spatial distribution diagrams consistently revealed that the hybrid ARIMA-LSTM model outperformed other models, indicating its effectiveness in both short-term and long-term drought predictions [10].

Additionally, the LSTM model has been used to predict the normalized difference vegetation index (NDVI), a key indicator for quantifying vegetation. The LSTM model, known for its precision over traditional RNNs, has been employed to predict future drought conditions using historical temperature and precipitation data recorded over months[6].

2.4 Drought prediction in Ethiopia

Classifying Drought in Ethiopia Using Machine Learning [7] This study applied machine learning techniques to classify drought conditions in Ethiopia, focusing on temperature and precipitation. A genetic algorithm with 10-fold cross-validation was used for variable selection, reducing 27 potential predictors to seven. The classification process employed logistic regression and Primal Estimated sub-Gradient Solver (Pegasos) for Support Vector Machine (SVM), using both hinge and log cost functions. The models trained for drought classification included logistic regression and Pegasos SVM. These models were trained using a 10-fold cross-validation approach to optimize the predictor subset and model parameters. The logistic regression model achieved 81.14% accuracy, while the Pegasos SVM model

with log cost function achieved 83.44% accuracy. The study highlighted the significant role of temperature anomalies and El Niño in drought classification, with the optimized models providing effective classification with high accuracy. A primary limitation of the study was its reliance on historical data, which may not fully capture future climate variability.

Long-term SPI Drought Forecasting in the Awash River Basin [3]. This study compared five data-driven models for long-term drought forecasting using the Standardized Precipitation Index (SPI). The models included ARIMA, artificial neural networks (ANNs), and support vector regression (SVR). Wavelet transforms were used to preprocess the inputs for ANN and SVR models, forming wavelet-ANN (WA-ANN) and wavelet-SVR (WA-SVR) models. The models were trained using historical precipitation data to forecast SPI values at 6 and 12-month lead times. The wavelet-transformed models aimed to enhance the prediction accuracy by capturing both temporal and frequency domain characteristics of the input data. The WA-ANN models outperformed the other models in forecasting SPI values, demonstrating superior performance with lower RMSE, MAE, and higher R2 values. The coupling of wavelet transforms with ANNs provided a significant improvement in forecasting accuracy over traditional models. The primary limitation was the complexity of the wavelet transform and ANN models, which require extensive computational resources and expertise to implement. Additionally, the models' performance heavily depends on the quality and length of the input time series data.

SPI Drought Forecasting in Ethiopia Using SWAM [2]. This study utilized the SPI for drought monitoring and forecasting, applying both traditional and machine learning techniques. The focus was on using the SPI index to predict drought conditions, leveraging machine learning models for enhanced forecasting accuracy. The models were trained on historical precipitation data to predict SPI values over various lead times. The study explored different machine learning approaches, including SVR and ANN, to improve the accuracy of SPI forecasts. The machine learning models, particularly those incorporating wavelet transforms, demonstrated improved performance in forecasting SPI values. The study highlighted the effectiveness of machine learning techniques in handling the non-linear and non-stationary nature of climate data. Similar to the other studies, the reliance on historical data and the complexity of the models were significant limitations. The models' accuracy could be affected by data quality issues and the inherent challenges of forecasting in regions with highly variable climate conditions.

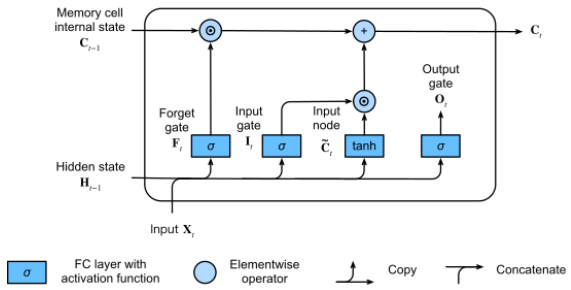


Figure 2: Architecture of a LSTM Unit.

3 Methodology

For this study, we have built multiple LSTM models with the combination of Convolutional Neural Networks (CNN) architecture to evaluate the performance of the drought prediction. The predicted result from each model is used to determine Standardized Precipitation Evapotranspiration Index (SPEI). In this section, we will discuss the details of each model architecture and meteorology-based standardized drought indices.

3.1 Model Architectures

3.1.1 Long Short Term Memory (LSTM)

Long Short Term Memory (LSTM) networks, introduced by Sepp Hochreiter and Jürgen Schmidhuber in 1997[5], are a type of Recurrent Neural Network (RNN) designed to address the limitations of traditional RNN, such as vanishing and exploding gradients. LSTMs are particularly effective in learning and remembering long-term dependencies in sequential data due to their unique architecture, which includes memory cells and gating mechanisms (input, forget, and output gates). These features allow LSTMs to selectively retain and update information, making them highly suitable for tasks involving time series prediction, natural language processing, and other applications requiring the modeling of temporal dynamics.

The Long Short Term Memory (LSTM) network [Figure2] consists of several key components that allow it to effectively capture long-term dependencies in sequential data. These components are the forget gate, input gate, cell state, and output gate.

The **forget gate** is responsible for deciding what information to discard from the cell state. It uses a `sigmoid` activation function, which outputs a value between 0 and 1, where 0 means *completely forget* and 1 means *completely keep*. The formula for the forget gate is:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

where σ is the `sigmoid` function, W_f is the weight matrix, h_{t-1} is the previous hidden state, x_t is the current input, and b_f is the bias. The `sigmoid` function is

defined as:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

The **input gate** determines what new information to add to the cell state. It also uses a `sigmoid` function to decide which values to update, combined with a `tanh` function to create a vector of new candidate values, $\tilde{C}_t$, that could be added to the state. The formulas are:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{4}$$

Here, W_i and W_C are weight matrices, b_i and b_C are biases. The `tanh` function is defined as:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{5}$$

The **cell state** C_t is updated based on the forget gate and the input gate. The forget gate controls which information from the previous cell state C_{t-1} should be retained, and the input gate updates the cell state with the new candidate values $\tilde{C}_t$. The cell state is updated using the following formula:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{6}$$

The **output gate** determines the output of the LSTM unit. It uses a `sigmoid` function to decide which parts of the cell state should be output, combined with the `tanh` function applied to the cell state to limit the values between -1 and 1. The formulas are:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{7}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{8}$$

Here, W_o is the weight matrix, b_o is the bias, and h_t is the hidden state, which serves as the output of the LSTM cell. There is a multiple variant of LSTM model, for this study, we will cover the following models.

Vanilla LSTM is a basic LSTM model with a single hidden layer of LSTM units and an output layer used to make predictions. Its simplicity lies in using one layer to capture long-term dependencies in sequential data. This straightforward architecture processes input sequences sequentially, updating hidden and cell states at each step.

Stacked LSTM is an advanced vanilla LSTM 3.1.1 architecture that consists of multiple LSTM layers arranged in a sequence, one on top of the other. This design enables the model to learn and represent more complex temporal patterns and dependencies within sequential data. Each LSTM layer processes the output

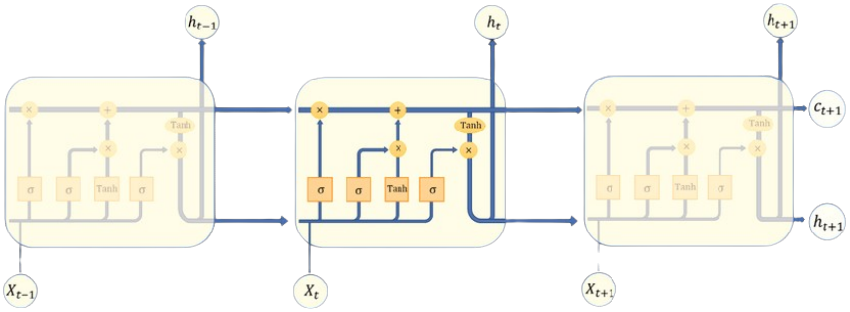


Figure 3: Architecture of Vanilla LSTM

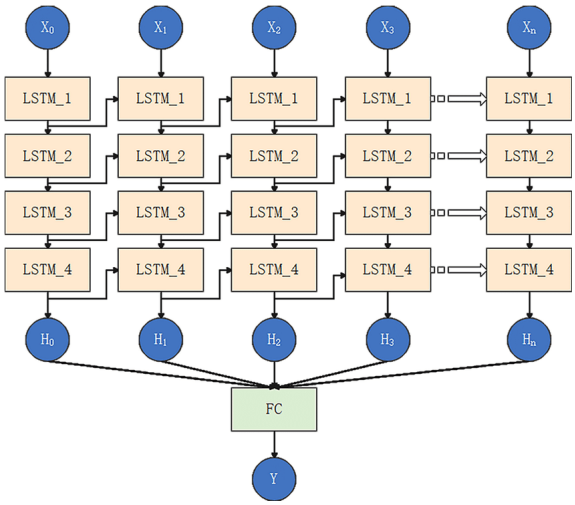


Figure 4: Architecture of Stacked LSTM

sequences from the previous layer, allowing the network to build up increasingly abstract and high-level features.

Bidirectional LSTM (BiLSTM) is a type of LSTM network that processes data in both forward and backward directions, enabling it to capture information from both past and future contexts simultaneously.

Stacked Bidirectional LSTM includes multiple layers of bidirectional LSTM units. Each bidirectional layer processes the combined output of the previous bidirectional layer.

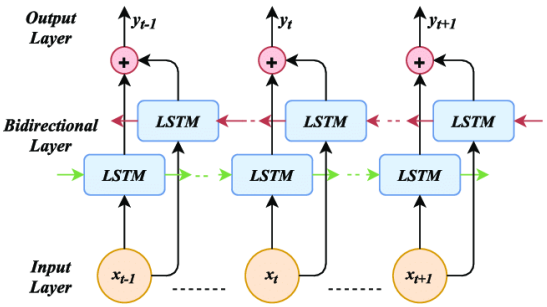


Figure 5: Architecture of BiLSTM

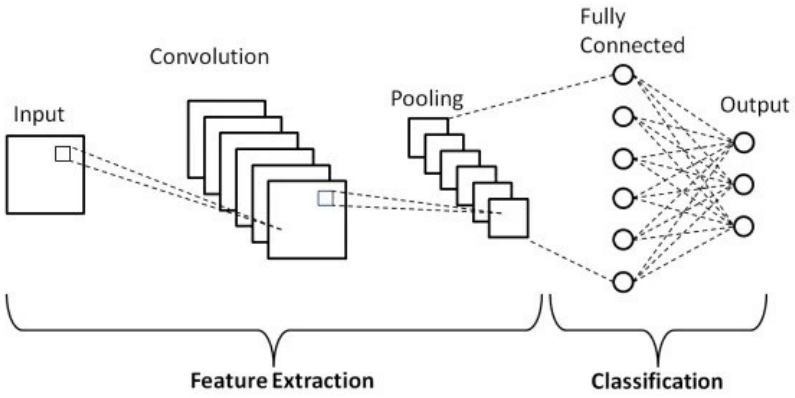


Figure 6: Architecture of CNN model

Types of LSTM Models for Time Series Forecasting

Univariate LSTM models are designed to handle time series data where there is only one variable being predicted based on its past values. These models are useful for forecasting a single time-dependent variable by learning patterns from its historical data. For instance, predicting future stock prices based on past stock prices is a typical application of univariate LSTM models.

Multivariate LSTM models extend the univariate approach by handling multiple variables simultaneously. These models predict one or more dependent variables using historical data from multiple correlated time series. For example, in weather forecasting, a multivariate LSTM might predict temperature based on past values of temperature, humidity, pressure, and wind speed.

Multi-step LSTM models are designed to forecast multiple future time steps at once, rather than just the next time step.

Multivariate multi-step LSTM models combine the capabilities of multivariate and multi-step LSTM models. They predict multiple future time steps of one or more dependent variables using historical data from multiple time series. This type of model is particularly useful in complex scenarios such as climate modeling, where the aim might be to predict various weather conditions (temperature, humidity, etc.) over the next several days based on multiple past observations.

3.1.2 Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNN) is a type of deep learning model designed to analyze visual data. Its architecture is inspired by the visual cortex of animals, and it is particularly effective for image and video recognition tasks. CNNs are composed of several key layers, each with a specific role in processing and understanding

the input data.

The **Convolutional Layer** is the first essential component of a CNN. It applies a set of filters (kernels) to the input image, performing convolution operations to produce feature maps. These feature maps capture various patterns such as edges, textures, and shapes at different locations in the input image. The convolution operation involves sliding the filter over the input and computing dot products between the filter and patches of the input.

To introduce non-linearity into the model, CNNs utilize activation functions, with the **Rectified Linear Unit (ReLU)** being the most common. The `ReLU` function applies an element-wise operation, converting all negative values in the feature maps to zero while keeping positive values unchanged.

After convolution and activation, the network typically includes a **Pooling Layer**, which reduces the spatial dimensions of the feature maps. Pooling operations, such as max pooling, take the maximum value from patches of the feature map, effectively downsampling the data. This reduction in size helps decrease the number of parameters, mitigates overfitting, and makes the network invariant to small translations and distortions in the input image.

CNNs also contain **Fully Connected Layers** (or **Dense Layers**) towards the end of the architecture. These layers take the high-level feature maps generated by the convolutional and pooling layers and interpret them to make a final decision or prediction. Each neuron in a fully connected layer is connected to every neuron in the previous layer.

An essential aspect of CNNs is the **Flattening** operation, which converts the 2D feature maps into a 1D vector before passing them to the fully connected layers. This transformation is necessary because the dense layers expect a 1D input, and it allows the network to combine spatial features learned in previous layers.

3.1.3 Hybrid of CNN and LSTM models

CNN-LSTM Combining Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM) networks can be particularly effective for predicting droughts, as these models can leverage both spatial and temporal data features. Here are the main ways to integrate CNNs and LSTMs for drought prediction:

Using the Output of the CNN as the Input to the LSTM In this method, the CNN processes spatial data, such as satellite images or spatial climate data, to extract features. These extracted features are then flattened and fed

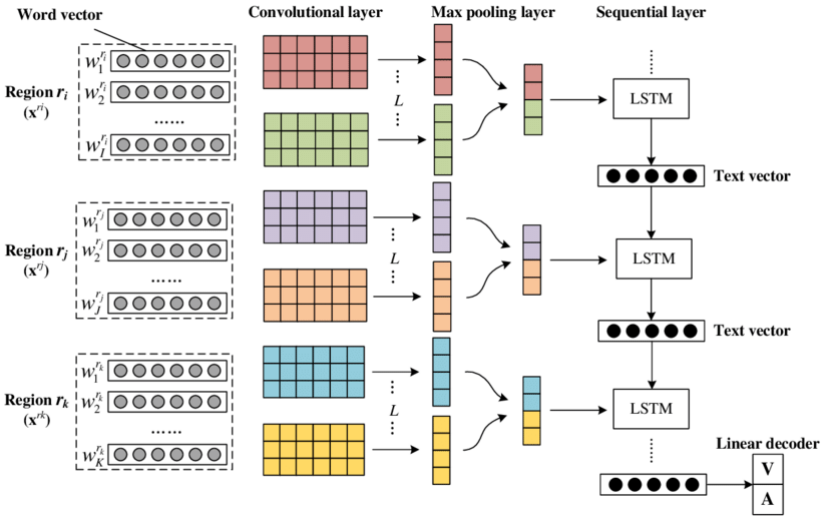


Figure 7: Architecture of CNN-LSTM hybrid model

into the LSTM, which models the temporal dependencies to predict future drought conditions.

- Example: For drought prediction, satellite images of vegetation indices (such as NDVI) are processed by the CNN to extract spatial features. These features are then passed to the LSTM, which uses past time steps to predict future drought events.

Using the Output of the LSTM as the Input to the CNN In this approach, the LSTM first processes sequential data, such as time series of climate variables, and generates output sequences that are then fed into the CNN. This method can be useful when it is essential to model temporal dynamics before extracting spatial patterns.

- Example: Climate data time series, including variables like temperature and precipitation, are first processed by the LSTM to capture temporal patterns. The output sequences from the LSTM are then input into the CNN to extract spatial patterns that are significant for predicting droughts.

Parallel Architecture with CNN and LSTM Operating Independently In this architecture, both CNN and LSTM operate on the input data independently. The outputs from both networks are then concatenated and passed to a fully connected layer. This method allows the model to capture both spatial and temporal features simultaneously and can be particularly powerful for complex tasks.

- Example: For drought prediction, satellite images are processed by the CNN to extract spatial features, while time series data of climatic variables are processed by the LSTM to capture temporal dependencies. The outputs of both networks are concatenated and fed into a fully connected layer to predict drought conditions.

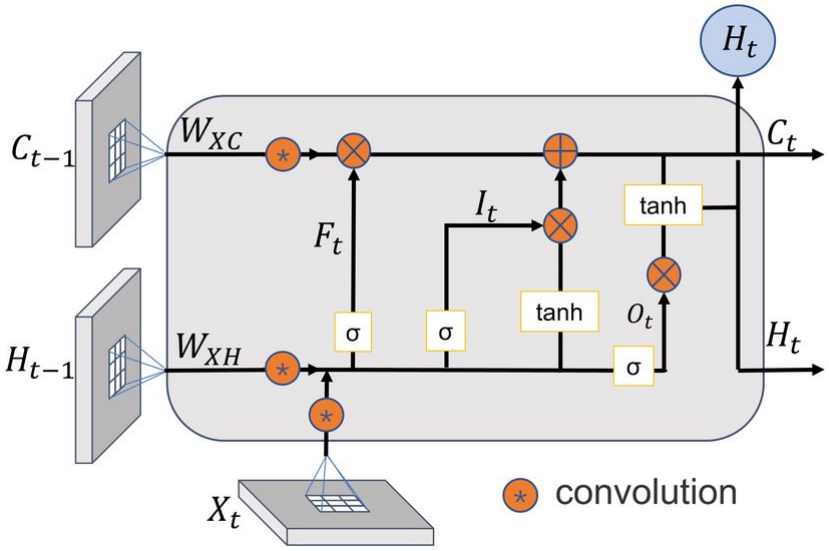


Figure 8: Architecture of ConvLSTM

3.1.4 Convolutional LSTM

Convolutional LSTM (ConvLSTM) is an advanced neural network architecture that merges the strengths of Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM) networks into a single model. This hybrid architecture is particularly well-suited for tasks involving spatiotemporal data, where both spatial and temporal patterns need to be captured and analyzed. ConvLSTM models are employed in a variety of applications, including weather forecasting, video analysis, and environmental monitoring, to leverage both spatial and temporal dependencies in the data.

ConvLSTM models address the challenge of handling spatially and temporally correlated data by incorporating convolutional operations within the LSTM units. In traditional LSTM networks, the input, output, and state transitions are managed through fully connected layers, which are not ideal for spatial data. ConvLSTM replaces these fully connected layers with convolutional layers, enabling the model to maintain the spatial structure of the data while learning temporal dependencies. This is crucial for applications such as video analysis and weather prediction, where the spatial relationships within the data are just as important as the temporal sequences.

In ConvLSTM, the input to each cell is a sequence of frames, and the convolutional operations are applied to these frames to extract spatial features at each time step. The recurrent nature of LSTM cells then captures the temporal dependencies across these frames. This combination allows ConvLSTM to effectively model the spatiotemporal dynamics present in the data.

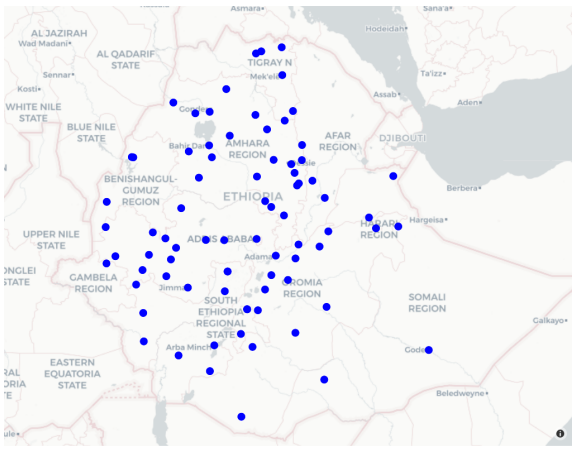


Figure 9: Data distribution across Ethiopian locations

One of the key applications of ConvLSTM is in weather forecasting, particularly for tasks such as precipitation nowcasting. In this context, ConvLSTM models process sequences of radar images to predict short-term weather conditions, such as rainfall intensity and movement. The convolutional layers extract spatial features from the radar images, while the LSTM layers capture the temporal evolution of these features to make accurate forecasts.

The structure of a ConvLSTM cell includes gates that control the flow of information into and out of the cell, similar to traditional LSTM cells. However, these gates use convolutional operations to process the data, preserving the spatial structure. The input gate controls the extent to which new information flows into the cell state, the forget gate decides which information from the cell state should be discarded, and the output gate generates the output of the ConvLSTM cell. The cell state and hidden state transitions are computed using convolutional operations, allowing the model to handle spatial data effectively.

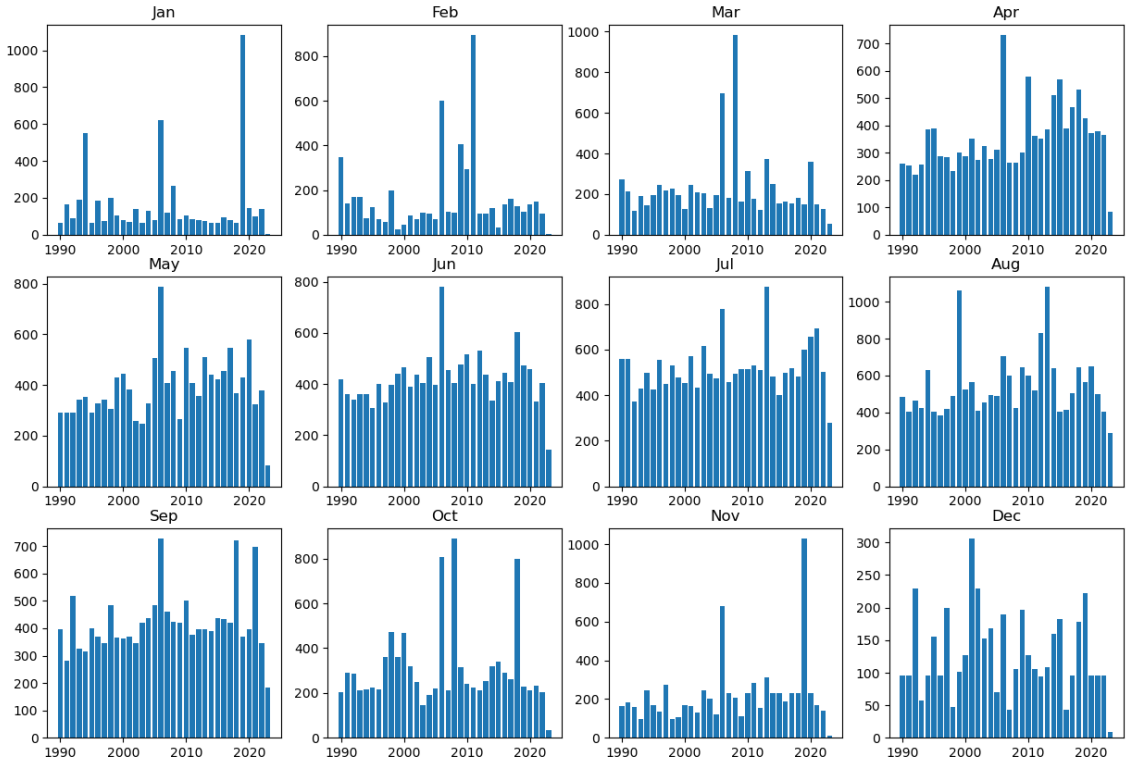
3.2 Data Processing

3.2.1 Data Collection

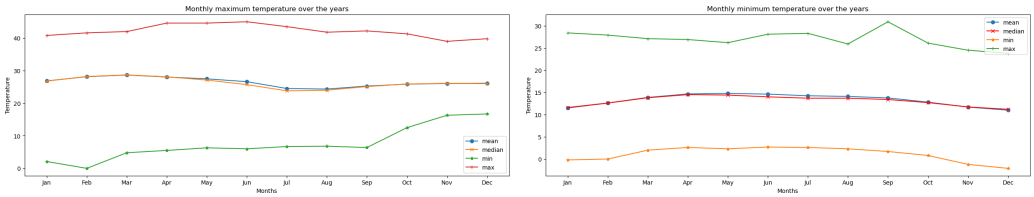
In this study, we have used monthly meteorological data from Ethiopian National Meteorology Agency. The data includes the precipitation and temperature data of 76 cities and 81 geographical sites in Ethiopia. The dataset contains 12 months of data on monthly maximum temperature, minimum temperature, and precipitation.

The following are all columns provided with the dataset:

- *NAME* - The legal location name.
- *YEAR* - The year in which the climate element is recorded.
- *GH_ID* - Unique identifier of the meteorological site.



(a) Precipitation data over the years



(b) Maximum and minimum temperature data over the years

Figure 10: Monthly data distribution over the years

- *GEOGR1* - The geographical latitude of the meteorological site.
- *GEOGR2* - The geographical longitude of the meteorological site.
- *GEOGR2* - The geographical longitude of the meteorological site.
- *ELEVATION* - Elevation of the specified location.
- *Element* - Type of climate data
  - *PRECIP* - Precipitation information of specified year for each month
  - *TMPMIN* - The minimum information recorded per month for the year.
  - *TMPMAX* - The maximum information recorded per month for the year.
- Months from *January* to *December* with the value of the specified element.

3.2.2 Data Pre-processing

The first step in building a machine learning model is pre-processing the given data. In our case, the data is organized in a tabular format, with each row representing a data point for a specific year, geographical location, and climate element. The data is collected from multiple geographical locations in the Ethiopia region. We have performed the following data-cleaning steps:

- The analysis of the average distance between working sites of one state is very close across the country, which gives us an insight into filling the missing geographical points (latitude and longitude) with the equidistant relative to the surrounding sites.
- Having the geographical coordinates, and filling in the missing elevation data is straightforward. We find the actual elevation data using dCode.fr API.
- The statistical measurements of the data distribution, displayed in figure 10, indicate the majority of the data points are accumulated into the center of the data. This helps us to decide how we handle the missing data on an average group of climate data.
- The data was organized in a way that each row represents a climate element of a specific region in a specific year. The data will be incomplete if one of the elements is missing for a specific year. We drop rows with incomplete elements. In addition, we dropped a row with a meteorological site of data less than `180` months which is not enough to accomplish a time series forecasting.

3.3 Feature Engineering

After cleaning the data the next step was making the data good for feature extraction which will help the LSTM learn from the time series data, so the model can forecast the next maximum temperature, minimum temperature, and monthly precipitation for prediction of drought. The first step of our feature engineering was converting the given dataset into a time series format by transposing the columns of months into rows and putting the climate elements as columns with site ID and other locations. After this, we will have a time series data sorted by year and the information for each year. We select all three elements as a future due to the reason that different standardized drought indices use different outputs [8].

DATE	GH_ID	GEOGR2	GEOGR1	ELEVATION	PERCIPT	TMPMAX	TMPMIN

Table 1: Modified dataset columns.

The time series data alone is not effective in describing the climate pattern. So we grouped the data based on their specific site id (*GH_ID*, so the model can learn the pattern from the long-range specific location.

This data will be used to predict the feature climate elements which will be evaluated against the actual data based on standardized drought indices. The necessary information on the possible index is listed below.

3.4 Model Training

3.4.1 Evaluation Metrics

We are training a forecasting model to predict key climatic variables such as precipitation, maximum temperature, and minimum temperature. The primary objective is to use these forecasts for accurate drought prediction. To validate the model's performance, we employ the Mean Squared Error (MSE) as the validation metric for both the loss and the validation loss. The MSE is calculated as follows:

$$\texttt{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{9}$$

where y_i represents the actual values, and $\hat{y}_i$ denotes the predicted values.

In addition to MSE, we also compare the results based on Mean Absolute Error (MAE) and Mean Percentage Error (MPE). The MAE is given by:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{10}$$

and the MPE is defined as:

$$\text{MPE} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{y_i}\right) \times 100 \tag{11}$$

Our models are used to predict data for 3, 6, 9, and 12 months into the future. Based on these predictions, we calculate the Standardized Precipitation Evapotranspiration Index (SPEI) and classify drought severity.

After obtaining the SPEI values, we compare the labeled drought conditions with the actual results. Based on this comparison, we compute various performance metrics, including total accuracy, precision, recall, and the F1-score. Mathematically the metrics are defined as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{12}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{13}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{14}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{15}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

By evaluating each approach using these metrics, we aim to identify the most accurate and reliable forecasting model for drought prediction.

3.4.2 Standardized Drought Indices

The Standardized Drought Index (SDI) is a general term that can refer to any drought index standardized to enable comparisons across different regions and timescales. It often involves normalizing variables like precipitation and temperature. SDI provides a common framework for drought assessment, facilitating regional and temporal comparisons, and can be applied to various drought indices. SDI is useful for comparative studies and regional drought monitoring where standardization is necessary.

In our study, we evaluated several drought indices that utilize both precipitation (`P`) and temperature (`T`) to determine their effectiveness for drought prediction. The indices considered include Standardized Precipitation Evapotranspiration Index (SPEI), Keetch-Byram Drought Index (KBDI) Weighted Anomaly Standardized Precipitation (PET), Aridity Index (AI), Crop Moisture Index (CMI), and Drought Reconnaissance Index (DRI).

Standardized Precipitation Evapotranspiration Index (SPEI) incorporates both precipitation and temperature data to calculate potential evapotranspiration (PET) and assess the balance between water input and output. SPEI captures the impact of temperature on drought conditions and can be computed over various timescales, allowing for flexibility in assessing both short-term and long-term droughts. However, it requires a serially complete dataset for both temperature and precipitation, which may limit its use in regions with insufficient data. Standardized Precipitation Evapotranspiration Index (SPEI) is particularly useful for identifying the impact of climate change on drought by considering the effects of increasing temperatures.

The Drought Reconnaissance Index (DRI) includes a simplified water balance equation considering precipitation and PET. It has three outputs: the initial value, the normalized value, and the standardized value, which can be directly compared to SPI. DRI is more representative than SPI because it considers the full water balance instead of precipitation alone. However, PET calculations can be

subject to errors when using temperature alone, and monthly timescales may not react quickly enough for rapidly developing droughts.

The Hydro-thermal Coefficient of Selyaninov (HTC) uses temperature and precipitation values and is sensitive to dry conditions specific to the climate regime being monitored. It is flexible enough to be used in both monthly and decadal applications. HTC is simple to calculate and the values can be applied to agricultural conditions during the growing season. However, the calculations do not take into account soil moisture, which limits its comprehensiveness in drought assessment.

The Keetch-Byram Drought Index (KBDI) uses daily maximum temperature and daily precipitation to monitor fire danger due to drought. It calculates moisture deficiency in the upper layers of the soil and indicates the amount of precipitation needed to saturate the soil. KBDI is useful for both fire danger and agricultural contexts, as it expresses moisture deficiency and is simple to use. However, it assumes a limit of available moisture and requires specific climatic conditions for drought to develop.

The Weighted Anomaly Standardized Precipitation (PET) index utilizes gridded precipitation and temperature data to monitor drought in tropical regions. PET is valuable for its use of gridded data, making it effective for large-scale drought monitoring. However, it may not be as effective in regions with sparse data availability.

The Aridity Index (AI) is often used in climate classifications and considers both precipitation and temperature data. Aridity Index (AI) helps in understanding the balance between precipitation and evapotranspiration, providing insights into arid and semi-arid conditions. However, it may not be suitable for detailed drought monitoring due to its broader application in climate classification.

The Crop Moisture Index (CMI) is calculated weekly and uses weekly precipitation and mean temperature. It responds quickly to changing conditions, making it suitable for monitoring short-term agricultural droughts. However, it was developed specifically for grain-producing regions and may not accurately reflect long-term drought conditions.

Among the indices reviewed, the Standardized Precipitation Evapotranspiration Index (SPEI) stands out due to its comprehensive incorporation of temperature effects alongside precipitation, providing a balanced approach to understanding water

availability and drought severity. Its flexibility across different timescales makes it particularly advantageous for comprehensive drought assessment.

These indicators are often used to provide a comprehensive understanding of drought conditions and their impacts on different sectors. Among those tools When we have maximum and minimum temperature and precipitation data, the Standardized Precipitation Evapotranspiration Index(SPEI) [1] is a particularly good drought indicator. The SPEI combines the sensitivity of PDSI to changes in evaporation demand (caused by temperature fluctuations and trends) with the simplicity of calculation and the multi-temporal nature of the SPI. The SPI can not identify the role of temperature increase in future drought conditions, and independently of global warming scenarios can not account for the influence of temperature variability and the role of heat waves. The SPEI can account for the possible effects of temperature variability and temperature extremes beyond the context of global warming. Two reasons why SPEI is good for our case:

- **Temperature Inclusion**: Unlike the Standardized Precipitation Index (SPI), the SPEI incorporates temperature data, which allows it to account for the effects of temperature on water balance. Higher temperatures increase evapotranspiration, which can exacerbate drought conditions.
- **Comprehensive Assessment**: SPEI combines both precipitation and potential evapotranspiration(PET), providing a more holistic measure of drought by considering the water demand in addition to the water supply.

All the needed element for calculating SPEI is provided by our trained model. After calculating SPEI if the result is an indicator of wetter than average but if is negative the value indicates drier than average. How are we going to calculate SPEI with the given data, we calculate SPEI with those three elements. The SPEI uses the monthly (or weekly) difference between precipitation and PET. This represents a simple climatic water balance which is calculated at different time scales to obtain the SPEI. With a value for PET, the difference between the precipitation (P) and PET for the month i is calculated.

$$PET_i = 16 \left(\frac{10 \cdot T_i}{I} \right)^a \tag{16}$$

- PET_i is the potential evapotranspiration for month i.
- T_i is the mean monthly temperature in degrees Celsius for month i.
- I is the annual heat index, calculated as the sum of monthly heat indices $\left(\frac{T_i}{5}\right)^{1.514}$.

- a is an empirical exponent calculated as $6.75\times10^{-7}T^3-7.71\times10^{-5}T^2+1.792\times10^{-2}T+0.49239$.

$$D_i = P_i - PET_i \tag{17}$$

Selection of the most suitable statistical distribution to model the D series was difficult, given the similarity among the four distributions (Pearson III, Lognormal, Log-logistic and General Extreme Value). The selection was based on the behavior at the most extreme values. Log-logistic distribution showed a gradual decrease in the curve for low values, and coherent probabilities were obtained for very low values of D, corresponding to 1 occurrence in 200 to 500 years. Additionally, no values were found below the origin parameter of the distribution.

$$f(x) = \frac{\beta}{\alpha}\left(\frac{x-\gamma}{\alpha}\right)^{\beta-1}\left(1+\left(\frac{x-\gamma}{\alpha}\right)^{\beta}\right)^{-2} \tag{18}$$

where α, β and γ are scale, shape and origin parameters, respectively, for D values in the range ($\gamma > D < \infty$). Parameters of the Log-logistic distribution can be obtained following different procedures. Among them, the L-moment procedure is the most robust and easy approach (Ahmad et al., 1988). When L-moments are calculated, the parameters of the Log-logistic distribution can be obtained following Singh et al. (1993):

$$\beta = \frac{2W_1 - W_0}{6W_1 - W_0 - 6W_2} \tag{19}$$

$$\alpha = \frac{(W_0 - 2W_1)\beta}{\Gamma\left(1+\frac{1}{\beta}\right)\Gamma\left(1-\frac{1}{\beta}\right)} \tag{20}$$

$$\gamma = W_0 - \alpha\Gamma\left(1+\frac{1}{\beta}\right)\Gamma\left(1-\frac{1}{\beta}\right) \tag{21}$$

where $\Gamma(\beta)$ is the gamma function of β.

In Vicente-Serrano et al. (2010), when the log-logistic α, β and γ distribution parameters were calculated, the probability-weighted moments (PWMs) method was used, based on the plotting-position approach (Hosking, 1990), where the PWMs of order s are calculated as:

$$w_s = \frac{1}{N}\sum_{i=1}^{N}(1-F_i)^s D_i \tag{22}$$

where N is the number of data, F_i is a frequency estimator following the approach of Hosking (1990) and D_i is the difference between Precipitation and Potential Evapotranspiration for the month i. Nevertheless, we have found that using the plotting position formulae the standard deviation of the SPEI series changes noticeably as a function of the SPEI time scale, which affects the spatial comparability of the SPEI values. On the contrary, if the PWMs are obtained using the unbiased estimator

given by Hosking (1986), the SD of the series does not change among the different SPEI time scales. The unbiased PWMs are obtained according to:

$$w_s = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{N-i}{N-1}\right)^s D_i \tag{23}$$

The method also solves the problem of the no solution of the SPEI model in some regions of the world. For these reasons, we recommend SPEI calculation using unbiased PWMs.

The probability distribution function of D according to the Log-logistic distribution is then given by:

$$F(x) = \left[1 + \left(\frac{\alpha}{x-\gamma}\right)^{\beta}\right]^{-1} \tag{24}$$

With $F(x)$ the SPEI can easily be obtained as the standardized values of $F(x)$. For example, following the classical approximation of Abramowitz and Stegun (1965):

$$SPEI = W - \frac{C_0 + C_1 W + C_2 W^2}{1 + d_1 W + d_2 W^2 + d_3 W^3} \tag{25}$$

where

$$W = -2\ln(P) \tag{26}$$

for $P \leq 0.5$, P being the probability of exceeding a determined D value, $P = 1 - F(x)$. If $P > 0.5$, P is replaced by $1 - P$ and the sign of the resultant SPEI is reversed. The constants are: $C_0 = 2.515517$, $C_1 = 0.802853$, $C_2 = 0.010328$, $d_1 = 1.432788$, $d_2 = 0.189269$, $d_3 = 0.001308$. The average value of the SPEI is 0, and the standard deviation is 1. The SPEI is a standardized variable, and it can therefore be compared with other SPEI values over time and space. An SPEI of 0 indicates a value corresponding to 50% of the cumulative probability of D, according to a Log-logistic distribution.

After completing the calculation the result can be used to classify drought sensitivity. Table 2 shows available ranges for classification.

3.5 Experiment Setup

3.5.1 Hyper-parameters

We have adjusted different hyperparameters to effectively configure and train LSTM models for drought prediction. One key parameter is the number of previous time steps used to predict the next one, set to `3`, `6`, `9`, `12`, and `24` based on the number of months need to be predicted. The target variables, including pre-

Table 2: SPEI Ranges and Drought Categories

SPEI Range	**Drought Category**
SPEI $\geq$ 2.0	Extremely Wet
1.5 $\leq$ SPEI $<$ 2.0	Very Wet
1.0 $\leq$ SPEI $<$ 1.5	Moderately Wet
-1.0 $<$ SPEI $<$ 1.0	Near Normal
-1.5 $\leq$ SPEI $\leq$ -1.0	Moderately Dry
-2.0 $\leq$ SPEI $<$ -1.5	Severely Dry
SPEI $<$ -2.0	Extremely Dry

cipitation, maximum temperature, and minimum temperature, define the specific outputs the model is designed to predict, shaping the configuration of the output layer accordingly.

In addition to the target variables, the input features used for prediction are also specified, such as precipitation, maximum temperature, and minimum temperature.

The proportion of data allocated for training, typically set at 90% or 80%, ensures that the model has sufficient data to learn from while reserving a portion for testing to allow robust evaluation. A normalization step, using a scaler, is crucial for standardizing the data so that all features contribute equally to the learning process and prevent any single feature from dominating due to scale differences.

The model iterates over the entire training dataset one hundred times, allowing for thorough learning. The batch size, often set to 32 or 1, dictates the number of samples processed before updating the model's parameters, balancing computational efficiency and convergence speed.

A validation split, set at 10%, ensures that a portion of the training data is used for validation, helping to monitor the model's performance and avoid overfitting. To further enhance training efficiency, we use callbacks such as early stopping and learning rate reduction.

3.5.2 Model Components

In this study, we used several LSTM-based models are used for drought prediction, each with distinct components and configurations tailored to capture temporal dependencies and improve prediction accuracy.

Vanilla LSTM model is designed with simplicity and effectiveness in mind. It includes a single LSTM layer with `100` units, which processes input sequences to capture temporal dependencies. Following this, the model flattens the output to prepare it for the dense layers. The model includes two Dense layers, each with `128` units and `ReLU` activation functions, interspersed with Dropout layers to prevent overfitting. The final output is reshaped to match the required prediction format.

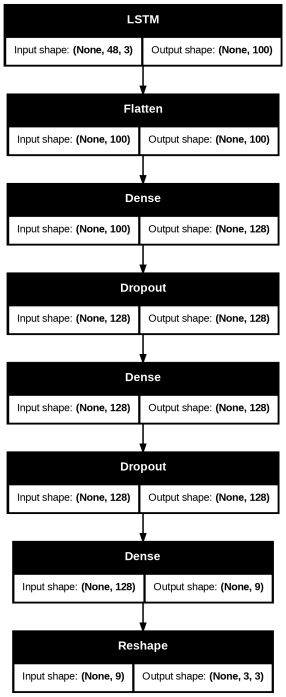


Figure 11: Vanilla LSTM Model component

The Stacked LSTM model enhances the Vanilla LSTM by stacking multiple LSTM layers to capture more complex patterns. It starts with an LSTM layer with `150` units, followed by another with `100` units, and a final LSTM layer with `50` units. Each LSTM layer processes the sequential data, with the output from one layer serving as the input to the next. After the LSTM layers, the model includes Flatten and Dense layers, similar to the Vanilla LSTM, to produce the final output.

The Bidirectional LSTM model processes the input sequence in both forward and backward directions, capturing dependencies from both directions. This model includes a single Bidirectional LSTM layer with `100` units. The bidirectional approach enhances the model's ability to understand context from both past and future data points within the sequence. The output from the Bidirectional LSTM layer is then flattened and passed through Dense layers with Dropout to generate the final prediction.

The Stacked Bidirectional LSTM combines the strengths of bidirectional processing with multiple LSTM layers. It starts with a Bidirectional LSTM layer with 100 units, followed by another Bidirectional LSTM layer with 50 units. This configuration allows the model to capture complex temporal patterns from both directions. Similar to the other models, the output is flattened and processed through Dense layers with Dropout to produce the final output.

The CNN-LSTM model integrates convolutional layers with LSTM layers to capture spatial and temporal dependencies. It begins with an input layer, followed

by time-distributed Conv1D layers with 64 filters and a kernel size of 1, which apply convolution operations across the input sequences. A TimeDistributed MaxPooling1D layer reduces the spatial dimensions, followed by a TimeDistributed Flatten layer to prepare the data for the LSTM layer. A single LSTM layer with 50 units processes the temporal dependencies, followed by Flatten and Dense layers with Dropout to produce the final prediction.

The ConvLSTM model employs ConvLSTM2D layers to capture spatiotemporal dependencies directly. It starts with an input layer, followed by a ConvLSTM2D layer with 64 filters and a kernel size of (1, 2), which applies convolution and LSTM operations simultaneously. The output is flattened and passed through Dense layers with Dropout to generate the final prediction.

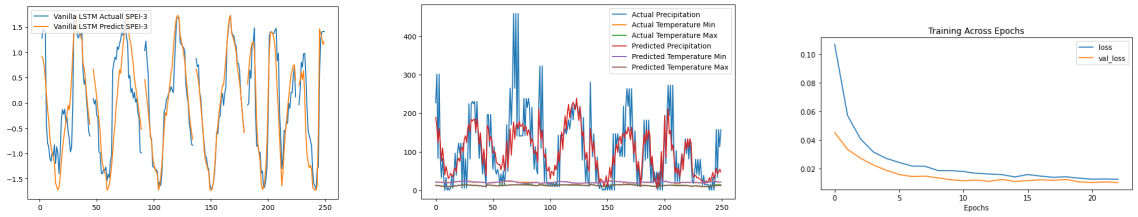
These models are compiled using the Adam optimizer and Mean Squared Error (MSE) as the loss function. The inclusion of Dropout layers in the dense sections of the models helps mitigate overfitting.

4 Result and Discussion

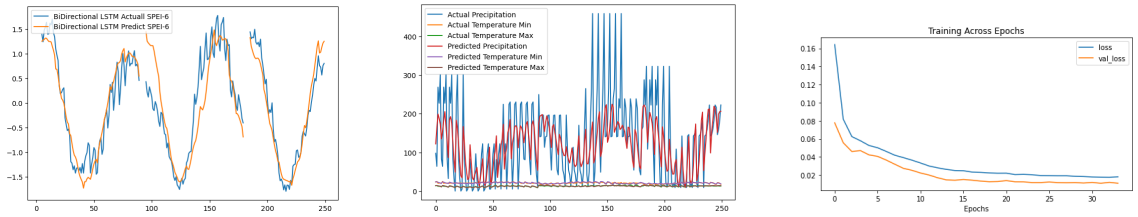
4.1 Model Evaluation

The results are presented in the table 3. compare the performance of various LSTM-based models in predicting drought conditions using the Standardized Precipitation Evapotranspiration Index (SPEI)) at different timescales (`3`, `6`, `9`, `12`, and `24` months). The performance metrics considered are Accuracy, Precision, Recall, and F1-Score.

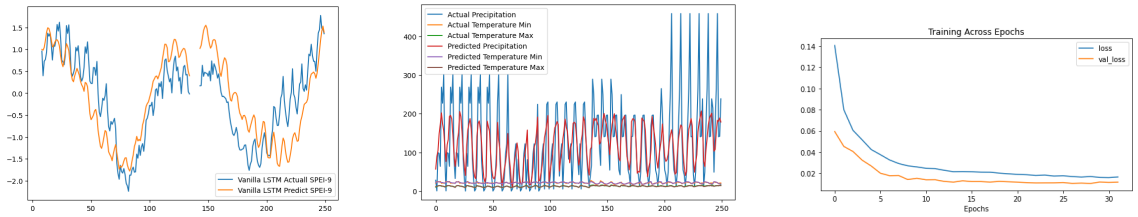
- **Vanilla LSTM** model demonstrates robust performance across all SPEI timescales, maintaining relatively high scores in Accuracy, Precision, Recall, and F1-Score. For SPEI-3, it achieves an Accuracy and F1-Score of 0.80, indicating strong short-term predictive capabilities. However, as the timescale extends to SPEI-24, the performance slightly declines, with an Accuracy of 0.65 and an F1-Score of 0.66. This decline suggests that while Vanilla LSTM is effective for short-term predictions, its performance diminishes for longer-term forecasts.
- **Stacked LSTM** model exhibits a different trend, with significantly lower performance metrics across all timescales. For instance, the Accuracy for SPEI-3 is only 0.40, and it gradually improves to 0.47 for SPEI-24. Similarly, the F1-Score starts at 0.42 for SPEI-3 and reaches 0.47 for SPEI-24. This indicates that the Stacked LSTM model struggles with both short-term and long-term predictions, we belive the reason is overfitting or insufficient learning of temporal dependencies.



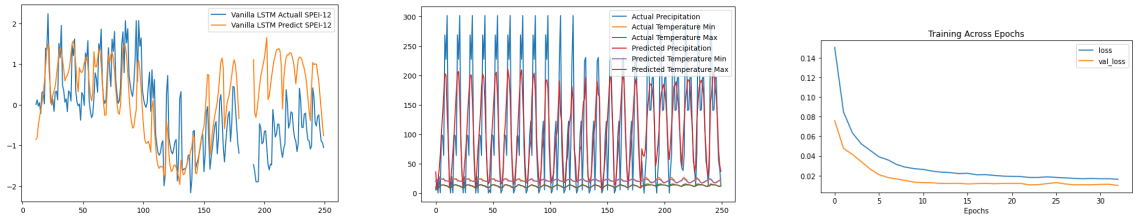
(a) SPEI-3 - Vanilla LSTM



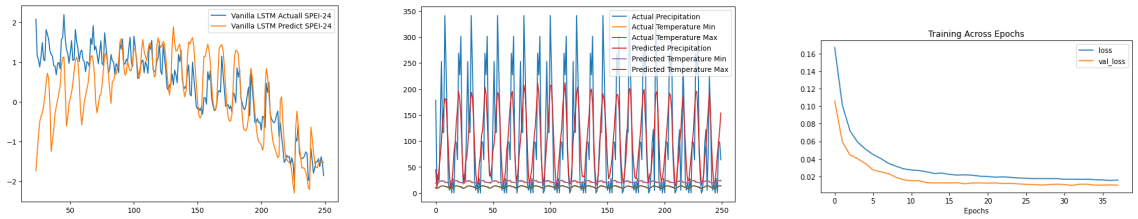
(b) SPEI-6 - BiDirectional LSTM



(c) SPEI-9 - Vanilla LSTM



(d) SPEI-12 - Vanilla LSTM



(e) SPEI-24 - Vanilla LSTM

Figure 12: Model training result best from each season

- **Bidirectional LSTM** model shows strong performance for short-term predictions, with an Accuracy of 0.76 and an F1-Score of 0.76 for SPEI-3. However, the performance declines significantly for longer timescales, with an Accuracy of 0.55 and an F1-Score of 0.56 for SPEI-24. This pattern suggests that BiLSTM is effective at capturing short-term temporal dependencies but less so for longer-term dependencies.

- **Stacked Bidirectional LSTM** model has moderate performance metrics, with an Accuracy of 0.73 and an F1-Score of 0.73 for SPEI-3. Similar to the BiLSTM model, its performance decreases for longer timescales, achieving an Accuracy and F1-Score of 0.54 for SPEI-24. This indicates that while stacking bidirectional layers provides some benefits, it still faces challenges in maintaining performance over longer prediction horizons.

- **CNN-LSTM** model performs well for short-term predictions, with an Accuracy of 0.74 and an F1-Score of 0.74 for SPEI-3. However, its performance declines for longer timescales, with an Accuracy of 0.51 and an F1-Score of 0.52 for SPEI-24. This suggests that while the CNN component enhances short-term predictions, the model struggles to maintain accuracy for extended periods.

- **ConvLSTM** model shows strong performance for short-term predictions, with an Accuracy and F1-Score of 0.77 for SPEI-3. Its performance declines for longer timescales, similar to other models, with an Accuracy of 0.53 and an F1-Score of 0.53 for SPEI-24. This indicates that ConvLSTM is effective at capturing complex patterns in short-term data but faces challenges with long-term predictions.

4.2 Feature Directions

In the context of advancing drought prediction models, several avenues for future research and development can be explored. First, enhancing the model's ability to handle and integrate multi-source data, including satellite imagery, soil moisture content, and other environmental variables, could significantly improve prediction accuracy. This would involve developing hybrid models that combine the strengths of Convolutional Neural Networks (CNNs) for spatial data and Long Short-Term Memory (LSTM) networks for temporal data.

There is potential in applying transfer learning techniques to leverage pre-trained models on larger, more diverse datasets, which could be fine-tuned for specific re-

Table 3: Model training result

Model	Metrics	SPEI-3	SPEI-6	SPEI-9	SPEI-12	SPEI-24
Vanilla LSTM	Accuracy	0.80	0.71	0.69	0.67	0.56
	Precision	0.80	0.71	0.69	0.68	0.58
	Recall	0.80	0.71	0.69	0.67	0.56
	F1-Score	0.80	0.71	0.69	0.68	0.67
Stacked LSTM	Accuracy	0.40	0.42	0.44	0.46	0.47
	Precision	0.44	0.39	0.41	0.46	0.48
	Recall	0.40	0.42	0.44	0.46	0.47
	F1-Score	0.42	0.41	0.42	0.46	0.47
BiLSTM	Accuracy	0.76	0.72	0.68	0.58	0.55
	Precision	0.77	0.71	0.69	0.59	0.56
	Recall	0.76	0.72	0.68	0.58	0.55
	F1-Score	0.76	0.71	0.68	0.58	0.56
Stacked BiLSTM	Accuracy	0.73	0.67	0.67	0.54	0.54
	Precision	0.74	0.67	0.67	0.55	0.55
	Recall	0.73	0.67	0.67	0.54	0.54
	F1-Score	0.73	0.67	0.67	0.54	0.54
CNN - LSTM	Accuracy	0.74	0.67	0.56	0.56	0.51
	Precision	0.75	0.68	0.55	0.56	0.52
	Recall	0.74	0.67	0.56	0.56	0.51
	F1-Score	0.74	0.67	0.55	0.56	0.52
ConvLSTM	Accuracy	0.77	0.71	0.66	0.54	0.53
	Precision	0.77	0.72	0.66	0.55	0.54
	Recall	0.77	0.71	0.66	0.54	0.53
	F1-Score	0.77	0.71	0.66	0.54	0.53

gional drought prediction. This approach would help in addressing the data scarcity issue in certain regions by utilizing knowledge from broader datasets.

Moreover, the incorporation of real-time data assimilation methods can be beneficial. Implementing frameworks that can continuously update and refine the models with real-time data inputs would enhance their responsiveness and accuracy in dynamic environments.

Additionally, exploring the use of Generative Adversarial Networks (GANs) for synthetic data generation could provide a means to augment training datasets, thereby improving model robustness and performance. This technique could be particularly useful for simulating extreme weather events.

Further, integrating socio-economic factors and water resource management strategies into the predictive models can provide a more holistic view of drought impacts and facilitate more comprehensive mitigation planning. This interdisciplinary approach would ensure that predictions are not only accurate but also actionable.

We believe these future directions, the field of drought prediction can continue to evolve, providing more reliable and timely forecasts that are essential for mitigating the adverse impacts of droughts on communities and ecosystems.

5 Conclusion

In this study, we explored the effectiveness of various LSTM-based models for drought prediction using the Standardized Precipitation Evapotranspiration Index (SPEI) at different timescales. The models evaluated include Vanilla LSTM, Stacked LSTM, Bidirectional LSTM, Stacked Bidirectional LSTM, CNN-LSTM, and ConvLSTM. Each model was assessed based on key performance metrics: Accuracy, Precision, Recall, and F1-Score.

The results indicate that the Vanilla LSTM and ConvLSTM models perform best for short-term drought predictions, particularly for SPEI-3, with high scores in all metrics. These models effectively capture temporal dependencies and provide robust short-term forecasts. However, their performance declines as the prediction horizon extends to SPEI-24, highlighting a common challenge in long-term prediction accuracy.

The Stacked LSTM model showed significantly lower performance across all timescales, suggesting potential issues with overfitting or insufficient temporal learning. The Bidirectional LSTM models performed well in capturing short-term dependencies but also faced a decline in long-term prediction accuracy. The CNN-LSTM model demonstrated the benefit of integrating spatial and temporal data for short-term predictions but struggled with long-term accuracy.

Overall, while each model has its strengths, there is a clear need for further refinement to improve long-term prediction capabilities. Future work should focus on integrating multi-source data, employing transfer learning, incorporating real-time data assimilation, and exploring synthetic data generation with GANs. Additionally, an interdisciplinary approach that includes socio-economic factors and real-world validation can enhance the practical applicability of these models.

By advancing these research directions, we can develop more accurate and reliable drought prediction models, providing essential tools for effective water resource management and mitigation strategies to combat the adverse impacts of droughts on communities and ecosystems.

References

[1] James Adams. *climate_indices, an open source Python library providing reference implementations of commonly used climate indices*. May 2017. URL: https://github.com/monocongo/climate_indices.

[2] A. Belayneh, J. Adamowski, and B. Khalil. "Short-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet transforms and machine learning methods". In: *Sustainable Water Resources Management* 2.1 (Dec. 2015), pp. 87–101. ISSN: 2363-5045. DOI: 10.1007/s40899-015-0040-5. URL: http://dx.doi.org/10.1007/s40899-015-0040-5.

[3] A. Belayneh et al. "Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models". In: *Journal of Hydrology* 508 (Jan. 2014), pp. 418–429. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2013.10.052. URL: http://dx.doi.org/10.1016/j.jhydrol.2013.10.052.

[4] Shilpa Shashikant Chaudhari, Vandana Sardar, and Prosenjit Ghosh. "Drought classification and prediction with satellite image-based indices using variants of deep learning models". In: *International Journal of Information Technology* 15.7 (Aug. 2023), pp. 3463–3472. DOI: 10.1007/s41870-023-01379-4. URL: https://doi.org/10.1007/s41870-023-01379-4.

[5] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: *Neural Computation* 9.8 (Nov. 1997), pp. 1735–1780. ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735. URL: https://doi.org/10.1162/neco.1997.9.8.1735.

[6] Yusef Kheyruri, Ahmad Sharafati, and Aminreza Neshat. "Predicting agricultural drought using meteorological and ENSO parameters in different regions of Iran based on the LSTM model". In: *Stochastic Environmental Research and Risk Assessment* 37.9 (June 2023), pp. 3599–3613. ISSN: 1436-3259. DOI: 10.1007/s00477-023-02465-6. URL: http://dx.doi.org/10.1007/s00477-023-02465-6.

[7] Michael B. Richman, Lance M. Leslie, and Zewdu T. Segele. "Classifying Drought in Ethiopia Using Machine Learning". In: *Procedia Computer Science* 95 (2016). Complex Adaptive Systems Los Angeles, CA November 2-4, 2016, pp. 229–236. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2016.09.319. URL: https://www.sciencedirect.com/science/article/pii/S1877050916324929.

[8] Mark D Svoboda and Brian A Fuchs. *Handbook of drought indicators and indices*. en. 2016.

[9] *The drought is driving up cases of child hunger and malnutrition in Borena. — unicef.org*. https://www.unicef.org/ethiopia/stories/drought-driving-cases-child-hunger-and-malnutrition-borena. [Accessed 21-06-2024].

[10] Dehe Xu et al. "Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting". In: *Environmental Science and Pollution Research* 29 (Jan. 2022). DOI: 10.1007/s11356-021-15325-z.