Agrociencia

STANDARDIZED PRECIPITATION INDEX FORECASTING IN NORTH-CENTRAL MEXICO USING TRANSFORMER MODELS

Rafael **Magallanes-Quintanar**[1*], Carlos Eric **Galván-Tejada**[1], Jorge Issac **Galván-Tejada**[1], Santiago de Jesús **Méndez-Gallegos**[2], Antonio **García-Domínguez**[1]

[1]Universidad Autónoma de Zacatecas. Unidad Académica de Ingeniería Eléctrica. Avenida Ramón López Velarde 801, Colonia Centro, Zacatecas, Zacatecas, Mexico. C. P. 98000.
[2]Colegio de Posgraduados Campus San Luis Potosí. Iturbide 73, Salinas de Hidalgo, San Luis Potosí, Mexico. C. P. 78600.
* Author for correspondence: tiquis@uaz.edu.mx

Citation: Magallanes-Quintanar R, Galván-Tejada Ce, Galván-Tejada JI, Méndez-Gallegos SJ, García-Domínguez A. 2026. Standardized precipitation index forecasting in north-central mexico using Transformer models. **Agrociencia.** https://doi.org/ 10.47163/agrociencia.v60i4.3641

Editor in Chief:
Dr. Fernando C. Gómez-Merino

Received: November 04, 2025.
Approved: June 05, 2026.
Published in Agrociencia: June 26, 2026.

ABSTRACT

Drought prediction is crucial for water resource management, agriculture, and climate adaptation in arid and semi-arid regions such as Zacatecas, Mexico. This study evaluates the advanced neural network architectures Long Short-Term Memory (LSTM), Vanilla Transformer, and Informer, for forecasting the Standardized Precipitation Index (SPI) using monthly precipitation data from 1964–2020 collected at 31 weather stations. SPI series were clustered into four regional climate zones. Models were implemented using the Nixtla *NeuralForecast* framework, and performance was assessed with Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Diebold-Mariano significance tests. The Informer model achieved the highest predictive accuracy, reducing average MSE by approximately 15 % relative to LSTM and consistently outper-forming the Vanilla Transformer in most regions. Statistical testing confirmed regional differences in model performance, suggesting that an adaptive, region-specific modeling approach is optimal for drought forecasting. These results demonstrate the robustness, efficiency, and transferability of Transformer-based models, particularly Informer, for operational drought monitoring under variable climatic conditions.

Keywords: Nixtla, Informer, Long Short-Term Memory, time-series, Standardized Precipitation Index, drought.

INTRODUCTION

Climate change poses a serious threat to global water resources. As the Earth's surface continues to warm, the future dynamics of precipitation and their effects on regional rainfall patterns remain uncertain (Ferreira *et al*., 2018). Furthermore, fluctuations in water availability, particularly the risk of future shortages, can adversely affect hydroelectric generation and agricultural operations (Karmalkar *et al*., 2011). Because historical rainfall and drought records are organized as time series, forecasting these variables represents an important challenge.

Time-series forecasting has been applied across numerous domains in which anticipating future trends based on historical data is essential. Key sectors benefiting from this approach include finance and economics (Siami-Namini and Namin, 2018), energy (Liu *et al*., 2023), healthcare (Kaushik *et al*., 2020), retail and e-commerce, manufacturing, and climate-related studies (Magallanes-Quintanar *et al*., 2022, 2024). In hydrology, for example, Long Short-Term Memory (LSTM) models have been successfully used for spatiotemporal forecasting of hydrological extremes, including drought events in river basins. Likewise, Transformer-based architectures have shown promise for national-scale drought forecasting across diverse climatic zones (Pathania and Gupta, 2025) and for hydrological drought prediction, with performance benchmarked against LSTM models (Amanambu *et al*., 2022).

In Mexico, droughts represent a significant threat to both water and food security. This concern is supported by the numerous droughts documented throughout the country's history (Florescano, 2000; García-Acosta *et al*., 2003), particularly in recent decades. For example, the most severe drought period of the last 70 years, in terms of its socio-economic impacts, occurred between 2011 and 2012. This event caused damage to approximately 1.5 million ha of cropland, the loss of more than 60 000 head of cattle, agricultural losses exceeding MXN 16 billion (approximately USD 1.3 billion), and widespread water shortages affecting up to 70 % of the national territory (Arreguín-Cortés *et al*., 2016).

Various methodologies have been developed to evaluate drought conditions, among which the Standardized Precipitation Index (SPI) is widely recognized for its simplicity and effectiveness in classifying climate regimes based on deviations in precipitation from long-term averages (McKee *et al*., 1993). Its exclusive reliance on precipitation time-series data contributes to its versatility and reliability for drought assessment across diverse temporal and geographic settings, particularly in vulnerable regions such as Mexico (Mahfouz *et al*., 2016). Building on this established utility, the present study applies SPI to evaluate drought conditions in Zacatecas, north-central Mexico, and provides a foundation for advanced forecasting using neural network architectures. The National Water Commission (CONAGUA) serves as the official body responsible for declaring drought conditions. This is accomplished through the Mexico Drought Monitor, which evaluates various drought parameters, including the SPI, the percentage of normal rainfall anomaly, and the vegetation health index, among others (Esquivel-Saenz *et al*., 2024).

In the context of climate change, the emergence of artificial intelligence (AI) models for drought prediction represents a substantial advance, reflecting their effectiveness and accuracy in drought assessment. In recent years, innovations in machine learning have led to remarkable improvements in operational efficiency, predictive accuracy, and accessibility. Notable examples include the development of attention-based architectures such as the Transformer (Vaswani *et al*., 2017) and its time-series-optimized variants (Su *et al*., 2025), the emergence of automated machine learning (AutoML) frameworks that reduce the need for manual hyperparameter tuning by

automating model selection and optimization processes (He *et al*., 2021), and the availability of open-source libraries such as NeuralForecast (Olivares *et al*., 2022), which lower barriers to the deployment of state-of-the-art neural forecasting models. These advances have enhanced the utility of AI methods for hydrological data analysis. Neural networks are widely recognized as an effective approach for data-driven learning and have demonstrated considerable success in modelling and forecasting nonlinear time series across multiple disciplines, particularly in water resources and hydrology (Ali *et al*., 2017). Models based on artificial neural networks (ANNs), particularly LSTM networks, have emerged as robust data-driven tools for forecasting monthly SPI values (Soh *et al*., 2018). Despite these advances, no studies have explored the integration of SPI and Transformer models in Mexico, particularly in arid and semi-arid regions such as Zacatecas, highlighting the novelty of the present research.

The Transformer model was originally developed to leverage attention mechanisms for the efficient modelling of sequential data, addressing key challenges in sequence learning for natural language processing tasks such as machine translation (Vaswani *et al*., 2017). More recently, Transformer-based architectures have demonstrated exceptional performance across a wide range of domains, including computer vision, speech processing, multimodal learning, reinforcement learning, and time-series forecasting (Su *et al*., 2025). Notably, Transformer variants, including the Informer model, have been successfully applied in hydrometeorological studies for streamflow prediction (Demiray and Demir, 2024) and drought assessment using precipitation indices (Ghobadi *et al*., 2025), highlighting their considerable potential for forecasting SPI time series.

This study applied LSTM and Transformer-based models (Vanilla Transformer and Informer) to develop and implement artificial neural network architectures for forecasting the regional SPI. Based on the above, the following hypothesis was proposed: Transformer-based architectures (Vanilla Transformer and Informer) will outperform LSTM networks in the regional forecasting of the SPI in Zacatecas, Mexico, owing to their superior ability to capture long-range temporal dependencies in hydroclimatic time series.

Accordingly, the objectives of this study were to develop artificial neural network architectures based on LSTM and Transformer methodologies; apply these models to forecast the regional SPI; perform a systematic comparative evaluation of the predictive performance of LSTM and Transformer models across all study regions; and assess the statistical significance of performance differences between the best-performing model and the alternative approaches.

MATERIALS AND METHODS

Study region and datasets

In this study, 31 monthly rainfall time series from weather stations across the state of Zacatecas, Mexico (Figure 1), were analyzed. Monthly precipitation data for the period

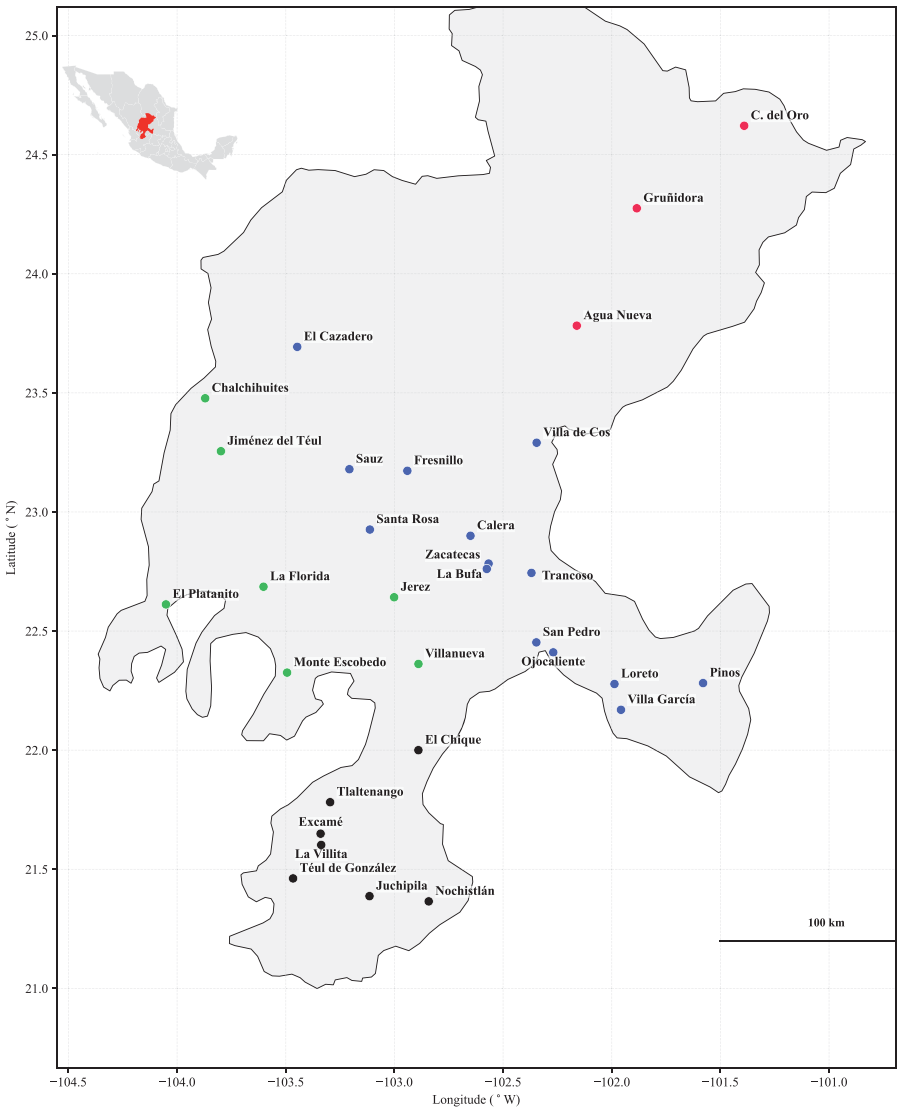

Figure 1. Geographic distribution of meteorological stations in Zacatecas, Mexico, and regional classification derived from cluster analysis. Colors represent Cluster I (Semi-arid, red), Cluster II (Highplain, blue), Cluster III (Mountains, green), and Cluster IV (Canyons, black).

1964–2020 were obtained from the network managed by the National Meteorological Service of Mexico. Prior to analysis, the datasets underwent a comprehensive quality-control process to remove outliers, correct missing or erroneous entries, and verify temporal consistency through homogeneity tests designed to detect shifts in seasonal variability. The Pettitt test was applied to identify abrupt change points in the mean

of each series, while the Standard Normal Homogeneity Test (SNHT) was used to detect step changes potentially associated with station relocations or instrument replacement. All 31 series were retained for the 1964–2020 period. In the few cases where isolated missing values were identified (representing less than 5 % of the records in any series), imputation was performed by linear interpolation using data from neighboring stations, following standard procedures for rainfall data reconstruction in hydrological studies (Navarro-Céspedes *et al*., 2023).

Standardized precipitation index (SPI)

In this study, the SPI was calculated using monthly rainfall data from 31 weather stations across Zacatecas, Mexico, spanning the period 1964–2020. The datasets were rigorously cleaned and validated prior to analysis. This methodology facilitates the assessment of precipitation anomalies by comparing observed values with long-term regional averages at specific locations and time scales. Following the methodology described by Koudahe *et al*. (2017), SPI computation proceeded through the following steps.

Monthly rainfall data were fitted to a gamma distribution, whose probability density function is defined as:

$$g(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{\frac{-x}{\beta}} \text{ for } x > 0$$

where $g(x)$ is the probability density function, α is the shape parameter ($\alpha > 0$), and β is the scale parameter ($\beta > 0$). The gamma function is defined as:

$$\Gamma(\alpha) = \int_0^{\alpha} y^{\alpha-1} e^{-y} dy$$

The parameters α and β are estimated as follows:

$$\alpha = \frac{1}{4A}\left(1 + \sqrt{1 + \frac{4A}{3}}\right)$$

$$\beta = \frac{\overline{x}}{\alpha}$$

$$A = \ln(\overline{x}) - \frac{\sum \ln(\overline{x})}{n}$$

where n is the number of precipitation observations and $\overline{x}$ is the arithmetic mean over the time scale of interest.

The cumulative probability $G(x)$ of an observed amount of rainfall for a given month and time scale is obtained by integrating the probability density function:

$$G(x) = \int_0^x g(x)dx = \frac{1}{\overline{\beta}^{\overline{\alpha}}\Gamma(\overline{\alpha})}\int_0^x x^{\overline{\alpha}} e^{\frac{\overline{x}}{\overline{\beta}}} dx$$

By substituting $t = \overline{x}/\beta$, the equation can be expressed as the incomplete gamma function:

$$G(x) = \frac{1}{\Gamma(\overline{\alpha})}\int_0^x t^{\overline{\alpha}-1} e^{-1}\, dt$$

Nevertheless, the gamma distribution function is undefined for $x = 0$ and $q = P(x = 0) > 0$; where $P(x = 0)$ is the probability of zero precipitation. Hence, the actual probability of non-exceedance $H(x)$ should be calculated as follows:

$$H(x) = q + (1 - q)G(x)$$

where $H(x)$ is the actual probability of non-exceedance and q the probability of $x = 0$. If m is zero in a sample of size n, then q is estimated as:

$$q = \frac{m}{n}$$

Finally, to calculate the SPI, the cumulative probability distribution $H(x)$ is transformed into a standard normal variable Z, with $\mu = 0$ and $\sigma = 1$. The interpretation of wet and drought periods based on SPI values was established by McKee *et al*. (1993). For the purposes of this study, a drought event was operationally defined as a period during which the SPI-12 value remained continuously at or below -1.0 for at least two consecutive months, consistent with the moderate-to-extreme drought categories proposed by McKee *et al*. (1993) (Table 1). Events with SPI-12 values ≤ -2.0 were classified as extreme drought.

SPI calculations incorporate multiple time scales because precipitation variability affects different components of the hydrological cycle (Caloiero, 2017). The 12-month SPI is particularly suitable for evaluating drought impacts on aquifer recharge and groundwater levels. In this study, SPI values were calculated using a 12-month accumulation window with R software version 4.4.1 (R Core Team, 2024) and the Standardized Precipitation-Evapotranspiration Index (SPEI) package version 1.8.1 (Beguería and Vicente-Serrano, 2017).

Table 1. Classification of wet and drought conditions based on the Standardized Precipitation Index (SPI) values.

SPI value	Class
≥2.0	Extremely wet
1.5 to 1.99	Severely wet
1.0 to 1.49	Moderately wet
-0.99 to 0.99	Near normal
-1.49 to -0.99	Moderately dry
-1.99 to -1.49	Severely dry
≤ -2.0	Extremely dry

Cluster analysis

Cluster analysis is a robust statistical technique widely used to identify homogeneous climate zones through the analysis of meteorological data. This method systematically categorizes observations according to their inherent characteristics and mutual relationships. Its primary objective is to group observations such that those within the same cluster exhibit high internal similarity while remaining distinct from those in other clusters. The effectiveness of the clustering process is enhanced by maximizing within-cluster homogeneity and between-cluster heterogeneity (Pampuch *et al*., 2023). Among hierarchical clustering algorithms, Ward's method was selected because of its proven efficiency and its tendency to produce compact, homogeneous clusters by minimizing total within-cluster variance. This characteristic makes it particularly suitable for identifying regional patterns in SPI data. In Ward's method, the distance between two clusters is defined as the increase in the total within-cluster sum of squares resulting from their merger.

In this study, a hierarchical tree-clustering algorithm was implemented to classify monthly SPI time series. The Canberra distance metric was selected as the linkage criterion because of its high sensitivity to relative variations near the origin, making it effective for identifying subtle differences in low-magnitude data (Lance and Williams, 1967). Unlike Euclidean distance, which is primarily influenced by absolute differences and may be biased by high-precipitation events, the Canberra metric ensures that variations during dry periods, which are critical for drought characterization, receive appropriate weight. This property makes it a robust choice for the regionalization of arid and semi-arid climates. Cluster analysis was performed using R version 4.4.1, with the ape package used for hierarchical cluster construction and visualization (Paradis and Schliep, 2018).

Neural time series forecasting

Evidence from previous studies indicates that artificial neural networks (ANNs) may provide a more efficient solution than conventional statistical and econometric methods for modelling nonlinear time-series data (Farajzadeh *et al*., 2014). Among the techniques used for time-series forecasting, LSTM networks and convolutional neural networks have received particular attention (Villegas-Vega *et al*., 2025).

Long Short-Term Memory Networks (LSTM)

LSTM networks were introduced by Hochreiter and Schmidhuber (1997) to address the difficulties encountered by Elman Recurrent Neural Networks in processing temporal patterns within datasets. One of the principal advantages of LSTM networks is their ability to capture long-term temporal dependencies while remaining effective at identifying short-term patterns. The architecture of an LSTM network and its information-processing pipeline (Figure 2) comprise three fundamental components: the input gate, the forget gate, and the output gate. These gates regulate the flow of information into, out of, and within the memory cell. Consequently, LSTM networks are capable of preserving temporal information across a wide range of time steps.

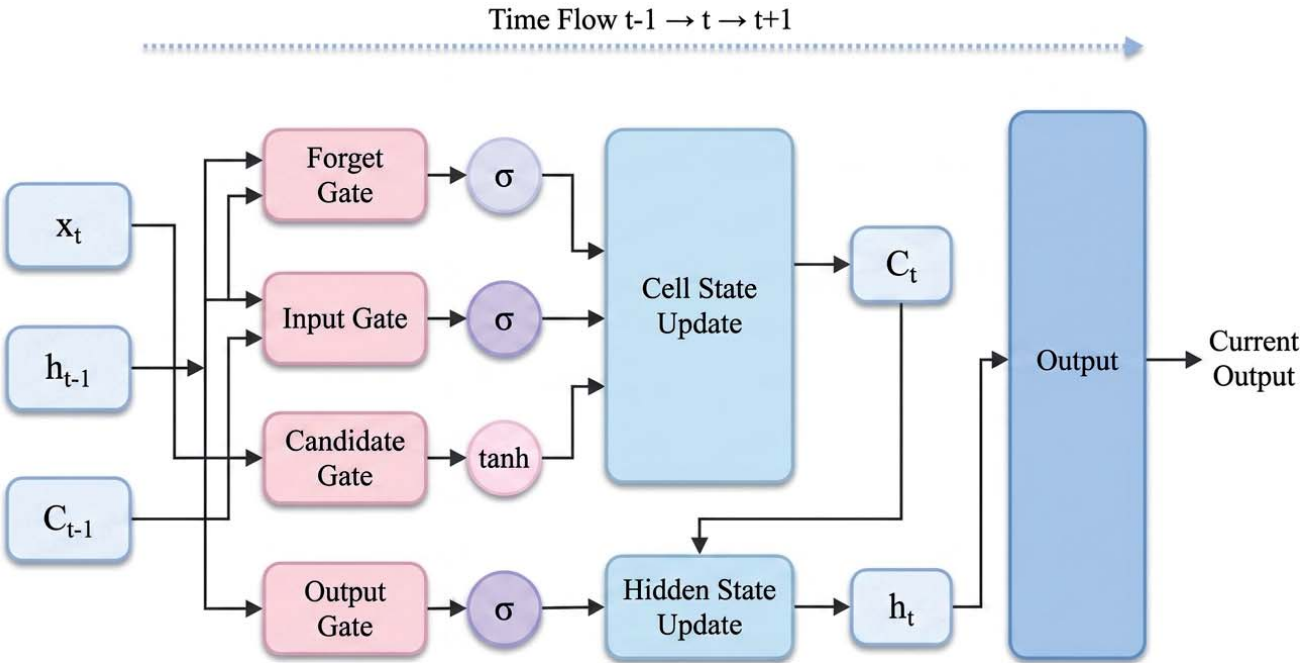

Figure 2. Schematic architecture and information flow of a Long Short-Term Memory (LSTM) unit.

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

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

$$\widetilde{C_t} = tanh(W_c[h_{t-1}, x_t] + b_c)$$

$$C_t = \sigma\left(f_t \times C_{t-1} + i_t \times \widetilde{C_t}\right)$$

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

where x_t is the input at time step t, h_{t-1} is the hidden state from preceding time step, and σ represents a logistic sigmoid function. The weight matrices *Wf*, *Wi*, *WC*, and *Wo*,

together with the bias terms b_f, b_i, b_c, and b_o are model parameters learned during the training phase.

Transformer model

In 2017, Google introduced the Transformer model (Vaswani *et al*., 2017), which uses attention mechanisms to process sequential data efficiently. The model was developed to address challenges associated with sequence-learning tasks in natural language processing, such as machine translation. It enables the transformation of an input sequence from one language domain into an output sequence in another.

The vanilla Transformer (Figure 3) follows the established architecture of neural sequence models and is based on an encoder-decoder framework (Vaswani *et al*., 2017). Both the encoder and decoder are composed of multiple identical blocks. Each encoder block contains a multi-head self-attention mechanism and a position-wise feed-forward network, whereas each decoder block incorporates cross-attention

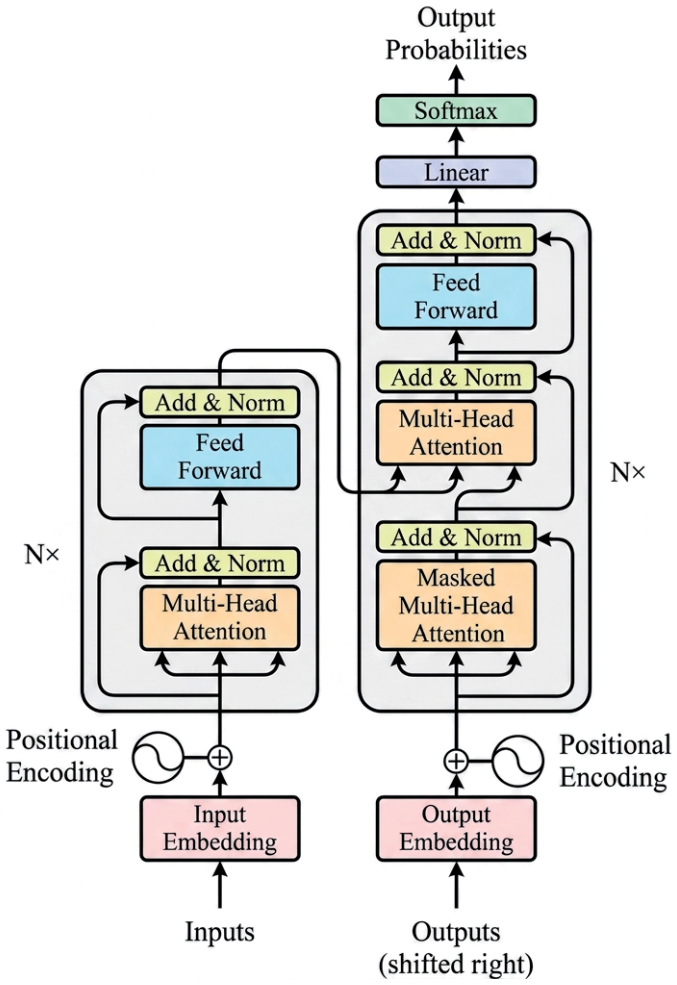

Figure 3. Architecture of the vanilla Transformer model (Vaswani *et al*., 2017).

mechanisms between the multi-head self-attention module and the position-wise feed-forward network.

If a segment of a time series is viewed as a sentence in one language and a subsequent segment as a sentence in another language, the multi-step forecasting problem can be reformulated as a sequence-learning task. Under this framework, the Transformer model is well suited for time-series forecasting. Following Vaswani *et al*. (2017), the core mathematical operations involved in processing time-series data with the Transformer model are described below.

Input representation. Time-series data are commonly represented as scalar- or vector-valued sequences. An input sequence X of length T is defined as:

$$X = [x_1 + x_2 + \cdots + x_T], x_T \in \mathbb{R}^d$$

where x_T is the input vector at time t, and d is the feature dimension.

To create an initial representation, input embeddings are combined with positional encoding to preserve temporal order:

$$Z^{(0)} = E(X) + P$$

where $E(X)$ maps x_T to a higher-dimensional space, and P is the positional encoding matrix.

Multi-head self-attention. For each layer l, the self-attention mechanism computes a weighted representation of the input:

$$Q = Z^{l-1}W_Q, K = Z^{l-1}W_K, V = Z^{l-1}W_V$$

where $W_Q, W_K, W_V \in \mathbb{R}^{d \times d_k}$ are learned projection matrices, and Q, K, and V represent the query, key, and value matrices, respectively.

The attention weights are computed as:

$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)$$

and the output of the self-attention mechanism is:

$$O = AV$$

In multi-head attention, several attention heads are computed in parallel and subsequently concatenated and linearly projected:

$$O^{\text{multi-head}} = \text{Concat}(O_1, O_2, \ldots, O_h)W_O$$

where h is the number of heads and $W_O \in \mathbb{R}^{hd_k \times d}$

Feedforward layer. Each layer contains a position-wise feed-forward network that is applied independently at each time step:

$$Z^{(l)} = FFN\left(O^{\text{multi-head}}\right)$$

where

$$FFN(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2$$

and W_1, W_2, b_1, and b_2 are learned weights and biases.

Layer normalization and residual connections. Both the self-attention and feed-forward modules are surrounded by residual connections and layer normalization:

$$Z^{(l)} = \text{LayerNorm}\left(Z^{(l-1)} + \text{SelfAttention}\left(Z^{(l-1)}\right)\right)$$

$$Z^{(l)} = \text{LayerNorm}\left(Z^{(l)} + FFN\left(Z^{(l)}\right)\right)$$

Output representation. After L Transformer layers, the final output is:

$$Z = Z^{(L)}$$

Time series specific adjustments. For forecasting and regression tasks, a prediction head is used to map the output sequence to the target variable:

$$\hat{y}_t = Z_t W_{out} + b_{out}$$

For causal or autoregressive forecasting, a mask is applied to the attention weights to prevent access to future observations:

$$A_{ij} = \begin{cases} \text{softmax}\left(\dfrac{QK^T}{\sqrt{d_k}}\right) & \text{if } j \leq i \\ 0 & \text{otherwise} \end{cases}$$

Numerous variants of the Transformer architecture have been developed to address specific challenges in time-series modelling (Wen *et al*., 2023). These adaptations have demonstrated effectiveness in a wide range of applications, including classification (Li *et al*., 2021), anomaly detection (Tuli *et al*., 2022), and forecasting (Li *et al*., 2019).

In this study, LSTM (Hochreiter and Schmidhuber, 1997), Vanilla Transformer (Vaswani *et al*., 2017), and Informer (Zhou *et al*., 2021) models were implemented to forecast regional SPI time series derived from precipitation records collected across the state of Zacatecas, Mexico, during the period 1964–2020. Model development and implementation were carried out using the NeuralForecast library (Olivares *et al*., 2022), part of the Nixtlaverse ecosystem, which comprises a suite of open-source libraries designed to facilitate the development of accurate and computationally efficient neural forecasting models (Nixtla, 2025).

Forecasting performance

LSTM, Vanilla Transformer, and Informer models exhibit distinct strengths depending on the forecasting horizon. LSTM performs best in short- to medium-term forecasting and is particularly effective for one-step and recursive multi-step predictions. However, its ability to capture long-range temporal dependencies and perform direct multi-horizon forecasting is more limited. In contrast, the Vanilla Transformer is better suited to medium- and long-term forecasting because it can model global temporal dependencies and supports direct multi-horizon and sequence-to-sequence predictions. Although it often achieves greater accuracy over extended forecasting horizons, it generally requires larger datasets and greater computational resources.

The Informer model, a Transformer-based architecture optimized for efficiency, was specifically designed for long-term forecasting over extended sequences. By utilizing sparse attention mechanisms, Informer achieves faster computation and improved performance in multi-horizon prediction, although it may be less efficient for relatively simple or short-term forecasting tasks. To evaluate the predictive performance of the LSTM, Vanilla Transformer, and Informer models, three standard error metrics were used: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\left(SPI_{p_i} - SPI_{o_i}\right)\right|$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(SPI_{p_i} - SPI_{o_i}\right)^2$$

$$RMSE = \sqrt{MSE}$$

where SPI_o represents the observed SPI values and SPI_p represents the corresponding predicted values. Lower values of MAE, MSE, and RMSE indicate better predictive performance.

The Mean Absolute Error (MAE) measures the average absolute difference between predicted and observed values, providing an easily interpretable measure of overall model accuracy. Lower MAE values indicate smaller average errors and, consequently, greater predictive precision. The Mean Squared Error (MSE) represents the average squared difference between observed and predicted values, placing greater emphasis on larger errors and serving as a sensitive measure of model fit. Lower MSE values indicate improved forecasting performance. The Root Mean Squared Error (RMSE), defined as the square root of MSE, expresses prediction errors in the original scale of the data. Like MSE, RMSE is sensitive to large deviations, while providing an interpretable measure of the typical magnitude of forecast errors when the model is unbiased (Su *et al*., 2025).

From a statistical perspective, the optimal model is the one that achieves the lowest MSE, RMSE, and MAE values among the candidate models. However, it is necessary to determine whether the superiority of this model is statistically significant rather than a result of random variation. To address this, the Diebold-Mariano (DM) test was applied to evaluate whether differences in forecast accuracy between the best-performing model and the alternative models were statistically significant at the 5 % significance level ($p < 0.05$).

When training multilayer neural networks, the available data are commonly divided into three subsets. The training set is used to compute gradients and update network weights and biases during the learning process. The validation set is used to monitor model performance and optimize hyperparameters throughout training. The test set serves as an independent benchmark for assessing the model's ability to generalize to previously unseen data, thereby providing an unbiased evaluation of predictive performance.

In this study, the LSTM, Vanilla Transformer, and Informer models were trained using an SPI dataset consisting of 518 monthly observations, representing 80 % of the total sample. The remaining 130 observations (20 % of the data) were used for validation. In addition, an independent set of 12 observed SPI values was reserved as the test sample for evaluating predictive accuracy. This procedure was applied to each SPI time series corresponding to the regions identified through the clustering process. The overall data workflow is illustrated (Figure 4).

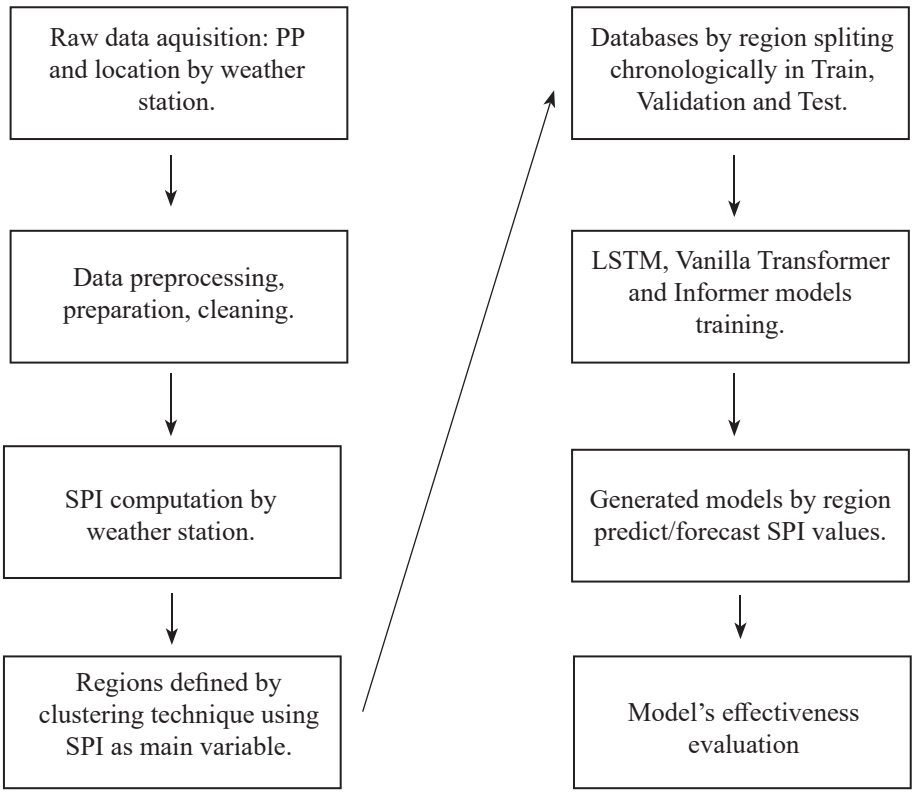

Figure 4. Overview of the analytical workflow from precipitation data processing for Standard Precipitation Index (SPI) forecasting and model evaluation.

RESULTS AND DISCUSSION

The dataset initially consisted of 31 individual rainfall time series, from which the corresponding SPI time series were derived. Cluster analysis was subsequently applied to the complete SPI dataset, resulting in the identification of four distinct regions (Semi-arid, Highplain, Mountains, and Canyons) characterized by similar SPI behavior and precipitation patterns (Figure 5). The regional SPI time series associated with these regions were then used as inputs for predictive modelling with the LSTM, Vanilla Transformer, and Informer architectures. SPI values were obtained from the training, validation, and testing phases, together with the SPI predictions generated by the three models for each region (Figure 6). Model evaluation based on MAE, MSE, and RMSE showed that both Transformer-based models consistently achieved lower error values than the LSTM model across all regions (Table 2).

Examination of the observed SPI-12 time series across the four regions revealed several historically documented drought events in Zacatecas. Using the operational criterion (SPI-12 ≤ -1.0 for at least two consecutive months), notable drought episodes were identified during 1994–1996, 2011–2012, and 2019–2020 in all regions, consistent

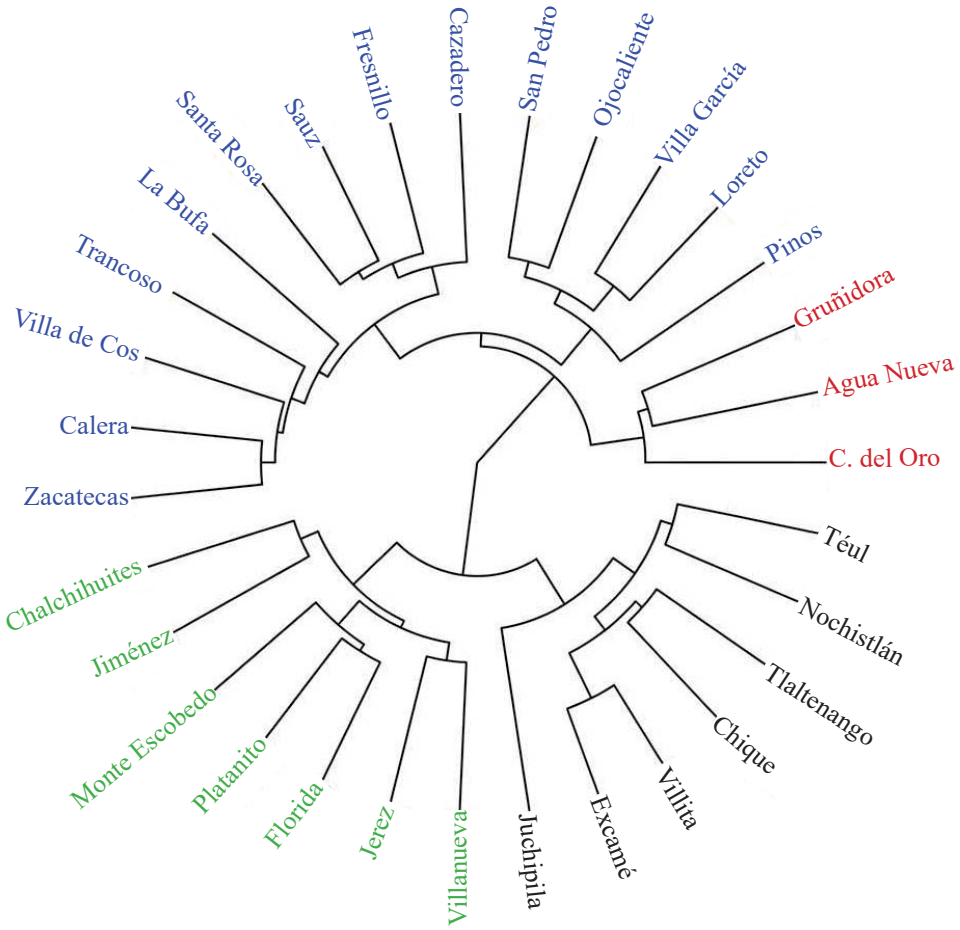

Figure 5. Dendrogram of 31 Standardized Precipitation Index (SPI) time series from Zacatecas, Mexico, generated using hierarchical clustering with Canberra distance. Four regions were identified: Cluster I, Semi-arid (red); Cluster II, Highplains (blue); Cluster III, Mountains (green); Cluster IV, Canyons (black).

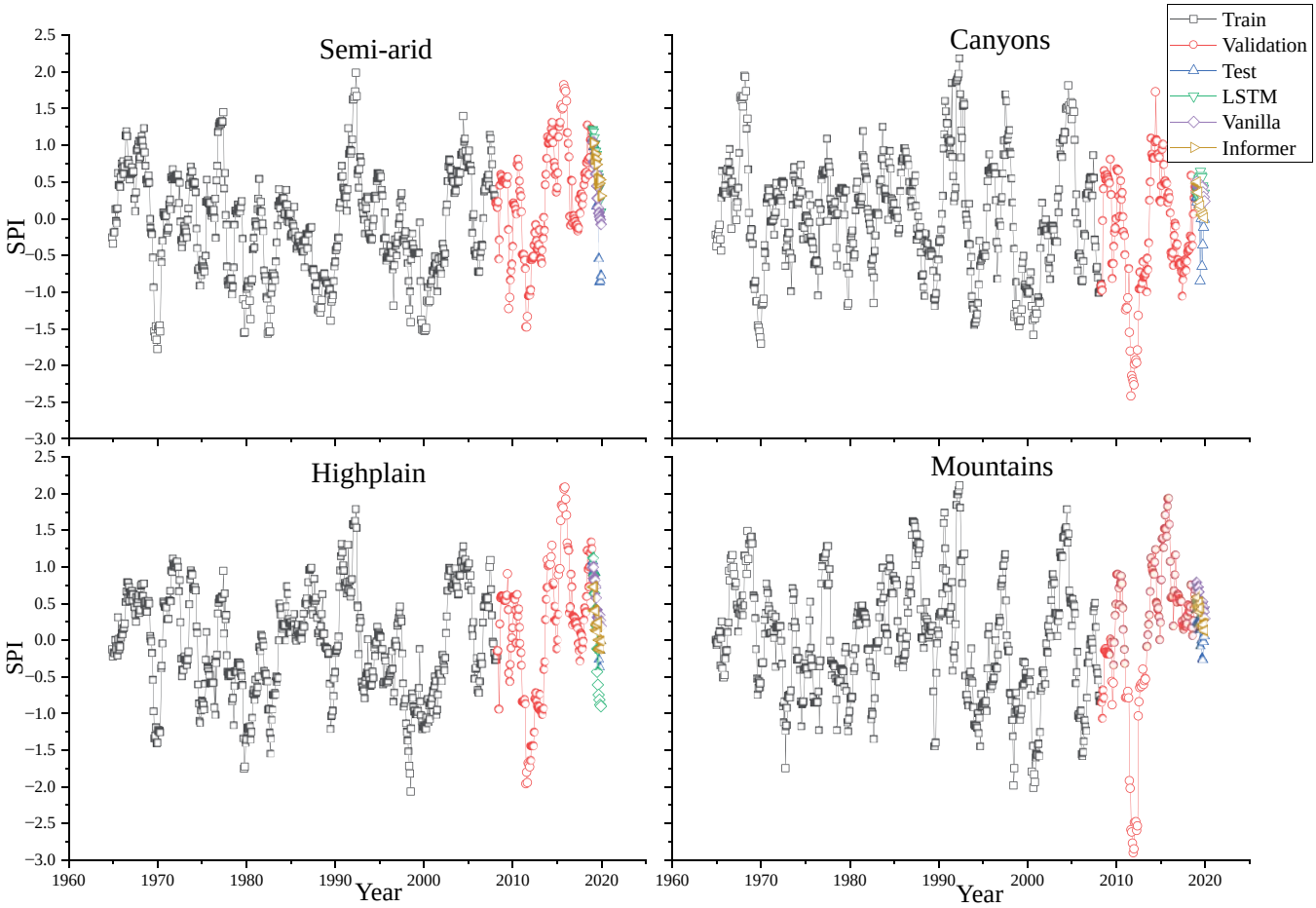

Figure 6. Time series of the regional Standardized Precipitation Index (SPI) for the territory of Zacatecas, encompassing both observed and projected data, in the period from 1964 to 2020.

Table 2. Performance metrics of the neural forecasting models during the validation phase for regional Standardized Precipitation Index (SPI) time series belonging to the state of Zacatecas, Mexico.

	Long Short-Term Memory			Vanilla Transformer			Informer		
Region	MSE	RMSE	MAE	MSE	RMSE	MAE	MSE	RMSE	MAE
Semi-arid	0.692	0.832	0.747	0.612	0.783	0.705	0.289	0.538	0.466
Highplain	0.314	0.560	0.474	0.228	0.478	0.405	0.214	0.462	0.403
Mountains	0.694	0.833	0.691	0.332	0.576	0.458	0.252	0.502	0.409
Canyons	0.589	0.767	0.713	0.791	0.890	0.813	0.488	0.698	0.631

with nationally documented drought events reported by Arreguín-Cortés *et al*. (2016). The Semi-arid and Canyons regions exhibited the most prolonged negative SPI-12 episodes, indicating greater sensitivity to precipitation deficits. The extreme drought of 2011–2012 (SPI-12 ≤ −2.0) was particularly severe in the Semi-arid and Highplain regions, consistent with the documented impacts on agricultural production and livestock in north-central Mexico.

The Informer model outperformed both the Vanilla Transformer and LSTM models across all evaluation metrics. For Mean Squared Error (MSE), the Informer achieved an average value of 0.311 across all regions, whereas the LSTM model recorded a higher average MSE of 0.572. A similar pattern was observed for Root Mean Squared Error (RMSE), with the Informer attaining an average value of 0.538 compared with 0.832 for the LSTM model. Mean Absolute Error (MAE) also supported the superior performance of the Informer, which achieved an average value of 0.477, compared with 0.656 for the LSTM model. Among the analyzed regions, the Highplain region exhibited the lowest MSE, RMSE, and MAE values, indicating the strongest predictive performance. In contrast, the Canyons region showed the highest error values across all metrics and, therefore, the lowest predictive accuracy.

The Diebold-Mariano (DM) test results provide a more comprehensive assessment of model performance beyond conventional accuracy metrics. Although the Informer model demonstrated strong overall performance, its superiority was not consistent across all climatic regions of Zacatecas (Table 3). Specifically, the Informer showed statistically significant improvements over the Vanilla Transformer in the Semi-arid region, highlighting its effectiveness under conditions of low rainfall variability. However, in the Canyons and Mountains regions, the LSTM and Vanilla Transformer

Table 3. Diebold-Mariano (DM) test statistics of the neural forecasting models during testing phase for regional Standardized Precipitation Index (SPI) time series for the state of Zacatecas, Mexico.

	Informer *vs* Long Short-Term Memory	
Region	DM statistic	*p*-value
Semi-arid	0.760	0.447
Highplains	-1.554	0.120
Mountains	-2.044	**0.041**
Canyons	-2.822	**0.005**
	Informer *vs* Vanilla Transformer	
Region	DM statistic	*p*-value
Semi-arid	-2.546	**0.011**
Highplains	0.553	0.580
Mountains	-4.911	**0.000**
Canyons	-2.672	**0.008**

models achieved better performance than the Informer. These results suggest that forecasting accuracy is context-dependent and influenced by regional climatic variability.

Forecasting the SPI directly from historical SPI time series, rather than forecasting precipitation and subsequently calculating SPI, provides a streamlined and statistically robust approach that minimizes the error propagation associated with multi-step forecasting. Indirect approaches can accumulate uncertainty because precipitation forecast errors propagate through the gamma-distribution fitting and parameter-estimation stages required for SPI calculation. Previous studies have shown that these compounding errors can alter drought classification by more than 0.2 SPI units in semi-arid environments (Zuo *et al*., 2022). In the present study, direct SPI forecasting using the Informer model produced low error values, with MAE ranging from 0.214 to 0.487, demonstrating both high predictive accuracy and computational efficiency while avoiding the accumulation of biases associated with precipitation-based forecasting approaches.

Furthermore, machine learning models, including neural networks and LSTM architectures, have demonstrated superior performance and generalization when trained on SPI values rather than raw precipitation data. This advantage arises from the normalized and statistically stable properties of SPI, which reduce the noise and variability inherent in precipitation records, thereby facilitating more robust learning and improving forecast consistency (Docheshmeh Gorgij *et al*., 2021).

Direct SPI forecasting is increasingly being incorporated into operational applications, including agricultural advisory services, drought preparedness programs, and governmental policy frameworks, particularly in regions with established early warning systems. Prominent examples include the National Drought Mitigation Center, the U.S. Drought Monitor, and the Global Drought Preparedness Network. Although many operational platforms continue to rely on SPI values derived from observed or forecasted precipitation, there is a growing trend toward the integration of direct SPI forecasts, particularly those generated using machine learning and seasonal climate prediction models. In addition, SPI-based forecasting facilitates integration with large-scale climatic drivers, such as El Niño-Southern Oscillation, and seasonal prediction systems (Hao *et al*., 2018). Collectively, these advantages support the increasing preference for SPI-based drought forecasting over approaches based solely on precipitation modelling.

LSTM networks are widely used for time-series forecasting because of their ability to capture long-term temporal dependencies while preserving short-term patterns (Villegas-Vega *et al*., 2025). Nevertheless, this architecture presents certain limitations. Traditional Multilayer Perceptron (MLP) neural networks, for example, are unable to adequately represent the sequential structure inherent in time-series data, often leading to reduced predictive accuracy (Farajzadeh *et al*., 2014). In contrast, Transformer models use a sequence-to-sequence architecture that provides the flexibility required to address complex sequence-learning tasks. By incorporating attention mechanisms,

Transformer-based approaches can efficiently identify and retain long-range dependencies within sequences (Lezmi and Xu, 2023), including those present in SPI time series.

The results obtained indicate that both Transformer-based models, particularly the Informer, achieved high predictive accuracy in forecasting the SPI across the regions of Zacatecas. The Informer model outperformed the Vanilla Transformer, likely because its sparse attention mechanism more effectively captures long-range dependencies in extended SPI time series (1964–2020), a characteristic associated with the highly variable precipitation patterns of the region. In addition, its computational efficiency and ability to focus on the most relevant temporal features make it particularly suitable for representing the complex climatic dynamics reflected in SPI variability.

Overall, the Informer model demonstrated strong predictive accuracy for SPI forecasting based on the MSE, RMSE, and MAE evaluation metrics. Specifically, the Informer reduced the average MSE by approximately 15 % relative to the LSTM model, highlighting the magnitude of its performance improvement. This result reflects the model's ability to capture long-term dependencies and temporal variability within SPI data while maintaining computational efficiency.

The Informer model exhibited statistically significant advantages over the Vanilla Transformer in the Semi-arid region, as confirmed by significance tests evaluating whether performance differences could be attributed to random variation. In contrast, the results indicated that the LSTM and Vanilla Transformer models outperformed the Informer in the Canyons and Mountains regions. These findings suggest that regional climatic heterogeneity influences model performance and that a regionalized model-selection strategy may improve drought forecasting accuracy. Specifically, the Informer may be more suitable for Semi-arid regions, whereas LSTM or Vanilla Transformer architectures may be preferable in more topographically complex areas. Such an approach could improve forecasting performance while strengthening drought preparedness and water-resource management efforts across Zacatecas.

These findings are consistent with previous studies that successfully applied artificial neural networks to forecast the monthly SPI (Magallanes-Quintanar *et al*., 2022, 2024; Villegas-Vega *et al*., 2025). In addition, the present study extends the work of Giddings *et al*. (2005) by introducing a methodology capable of identifying smaller and more precise climatic regions in Mexico using SPI. This research contributes to the international literature by adapting the Informer model to capture complex, region-specific precipitation dynamics and by providing a scalable and transferable framework for drought forecasting in other regions of the world.

Additionally, the Informer model implemented through the NeuralForecast framework (Olivares *et al*., 2022) represents a valuable and contemporary tool for time-series forecasting, including drought-related applications. Its combination of predictive accuracy, computational efficiency, and scalability makes it a promising approach for assessing climatic and agricultural risks associated with drought, particularly in arid and semi-arid regions that are increasingly affected by climate change-driven extremes.

CONCLUSIONS

Drought prediction has become increasingly important in meteorology, hydrology, water resources management, and sustainable agriculture because of the growing dependence of human activities on reliable water supplies. To address this challenge, artificial intelligence models based on the Long Short-Term Memory (LSTM), Vanilla Transformer, and Informer architectures were developed to forecast four monthly Standardized Precipitation Index (SPI) time series representing regional conditions across Zacatecas, Mexico. According to the evaluation metrics, the Informer model achieved the highest predictive performance across all study regions, reducing the average MSE by approximately 15 % relative to the LSTM model and showing comparable improvements over the Vanilla Transformer. These results were supported by the Diebold-Mariano test. However, additional significance tests suggest that a regionalized model-selection strategy, using the Informer in semi-arid regions and LSTM or Transformer models in more complex terrains, may further improve forecasting accuracy and reliability.

In the context of climate change, accurate SPI forecasting with the Informer model has considerable potential to enhance water-resource management in Zacatecas. As precipitation patterns become increasingly variable, reliable SPI forecasts can support proactive drought mitigation, improve agricultural planning, and contribute to sustainable water management. The model's ability to capture long-term dependencies within the 1964–2020 dataset highlights its applicability for decision-making in semi-arid and mountainous regions affected by persistent climatic variability.

Future research should evaluate the incorporation of additional climatological variables, such as temperature, evapotranspiration, and large-scale climate indices (e.g., El Niño-Southern Oscillation), to further improve SPI forecasting performance. These studies should also be extended to other climatic settings, including extreme drought periods beyond 2020 and contrasting environments such as tropical and coastal regions, to assess the robustness, adaptability, and generalizability of the proposed approach across diverse climate regimes.

REFERENCES

Ali Z, Hussain I, Faisal M, Nazir HM, Hussain T, Shad MY, Mohamd Shoukry A, Hussain Gani S, 2017. Forecasting drought using multilayer perceptron artificial neural network model. Advances in Meteorology 2017. https://doi.org/10.1155/2017/5681308

Amanambu AC, Mossa J, Chen YH, 2022. Hydrological drought forecasting using a deep transformer model. Water 14 (22): 3611. https://doi.org/10.3390/w14223611

Arreguín-Cortés FI, López-Pérez M, Ortega-Gaucin D, Ibañez-Hernández Ó. 2016. La política pública contra la sequía en México: avances, necesidades y perspectivas. Tecnología y Ciencias del Agua 7 (5): 63–76.

Beguería S, Vicente-Serrano SM, 2017. SPEI Calculator. Digital.CSIC. http://doi.org/10.20350/digitalcsic/8997

Caloiero T. 2017. Drought analysis in New Zealand using the standardized precipitation index. Environmental Earth Sciences 76 (16): 1–13. https://doi.org/10.1007/s12665-017-6909-x

Demiray BZ, Demir I, 2024. Towards generalized hydrological forecasting using transformer models for 120-hour streamflow prediction. arXiv. https://doi.org/10.48550/arxiv.2406.07484

Docheshmeh Gorgij A, Alizamir M, Kisi O, Elshafie A. 2021. Drought modelling by standard precipitation index (SPI) in a semi-arid climate using deep learning method: Long short-term memory. Neural Computing and Applications 34 (3): 2425–2442. https://doi.org/10.1007/s00521-021-06505-6

Esquivel-Saenz PJ, Ortiz-Gómez R, Zavala M, Flowers-Cano RS. 2024. Artificial neural networks for drought forecasting in the central region of the state of Zacatecas, Mexico. Climate 12 (9): 131. https://doi.org/10.3390/cli12090131

Farajzadeh J, Fakheri Fard A, Lotfi S. 2014. Modeling of monthly rainfall and runoff of Urmia Lake basin using "feed-forward neural network" and "time series analysis" model. Water Resources and Industry 7–8: 38–48. https://doi.org/10.1016/j.wri.2014.10.003

Ferreira RN, Nissenbaum MR, Rickenbach TM. 2018. Climate change effects on summertime precipitation organization in the Southeast United States. Atmospheric Research 214: 348–363. https://doi.org/10.1016/j.atmosres.2018.08.012

Florescano E. 2000. Breve historia de la sequía en México (Segunda edición). Consejo Nacional para la Cultura y las Artes: Ciudad de México, México. 252 p.

García-Acosta V, Pérez-Zevallos JM, Molina-del Villar A. 2003. Desastres agrícolas en México: catálogo histórico. Centro de Investigaciones y Estudios Superiores en Antropología Social: Ciudad de México, México. 506 p.

Ghobadi F, Tayerani Charmchi AS, Kang D. 2025. Enhancing long-term flood forecasting with SageFormer: A cascaded dimensionality reduction approach based on satellite-derived data. Remote Sensing 17 (3): 365. https://doi.org/10.3390/rs17030365

Giddings L, Soto M, Rutherford BM, Maarouf A. 2005. Standardized precipitation index zones for Mexico. Atmósfera 18 (1): 33–56.

Hao Z, Singh VP, Xia Y. 2018. Seasonal drought prediction: Advances, challenges, and future prospects. Reviews of Geophysics 56 (1): 108–141. https://doi.org/10.1002/2016rg000549

He X, Zhao K, Chu X. 2021. AutoML: A survey of the state-of-the-art. Knowledge-Based Systems 212: 106622. https://doi.org/10.1016/j.knosys.2020.106622

Hochreiter S, Schmidhuber J. 1997. Long short-term memory. Neural Computation 9 (8): 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Karmalkar AV, Bradley RS, Diaz HF. 2011. Climate change in Central America and Mexico: Regional climate model validation and climate change projections. Climate Dynamics 37 (3–4): 605–629. https://doi.org/10.1007/s00382-011-1099-9

Kaushik S, Choudhury A, Sheron PK, Dasgupta N, Natarajan S, Pickett LA, Dutt V. 2020. AI in Healthcare: Time-series forecasting using statistical, neural, and ensemble architectures. Frontiers in Big Data 3. https://doi.org/10.3389/fdata.2020.00004

Koudahe K, Kayode AJ, Samson AO, Adebola AA, Djaman K. 2017. Trend analysis in standardized precipitation index and standardized anomaly index in the context of climate change in southern Togo. Atmospheric and Climate Sciences 7 (4): 401–423. https://doi.org/10.4236/acs.2017.74030

Lance GN, Williams WT. 1967. A general theory of classificatory sorting strategies: 1. Hierarchical systems. The Computer Journal 9 (4): 373–380. https://doi.org/10.1093/comjnl/9.4.373

Lezmi E, Xu J. 2023. Time series forecasting with transformer models and application to asset management. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4375798

Li C, Yang J, Zhang P, Gao M, Xiao B, Dai X, Yuan L, Gao J. 2021. Efficient self-supervised vision transformers for representation learning. arXiv. https://doi.org/10.48550/arxiv.2106.09785

Li S, Jin X, Xuan Y, Zhou X, Chen W, Wang YX, Yan X. 2019. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. arXiv. https://doi.org/10.48550/arxiv.1907.00235

Liu H, Liu Y, Guo X, Wu H, Wang H, Liu Y. 2023. An energy consumption prediction method for HVAC systems using energy storage based on time series shifting and deep learning. Energy and Buildings 298: 113508. https://doi.org/10.1016/j.enbuild.2023.113508

Magallanes-Quintanar R, Galván-Tejada CE, Galván-Tejada JI, Gamboa-Rosales H, Méndez-Gallegos SDJ, García-Domínguez A. 2024. Auto-machine-learning models for standardized precipitation index prediction in north-central Mexico. Climate 12 (7): 102. https://doi.org/10.3390/cli12070102

Magallanes-Quintanar R, Galván-Tejada CE, Galván-Tejada JI, Méndez-Gallegos SJ, García-Domínguez A, Gamboa-Rosales H. 2022. Narx neural networks models for prediction of standardized precipitation index in Central Mexico. Atmosphere 13 (8): 1254. https://doi.org/10.3390/atmos13081254

Mahfouz P, Mitri G, Jazi M, Karam F. 2016. Investigating the temporal variability of the standardized precipitation index in Lebanon. Climate 4 (2): 27. https://doi.org/10.3390/cli4020027

McKee TB, Doesken NJ, Kleist J. 1993. The relationship of drought frequency and duration to time scales. *In* Proceedings of the 8th Conference on Applied Climatology. Anaheim, CA, USA, pp: 179–183.

Navarro-Céspedes JM, Hernández JH, Alcántara-Concepción PC, Morales-Martínez JL, Carreño-Aguilera G, Padilla-Benítez F. 2023. A comparison of missing values imputation methods applied to precipitation of two semi-arid and humid regions of México. Atmósfera 37. https://doi.org/10.20937/atm.53095

Nixtla. 2025. Nixtlaverse. https://nixtlaverse.nixtla.io/ (Retrieved: May 2026).

Olivares KG, Challú C, Garza F, Canseco MM, Dubrawski A. 2022. NeuralForecast: User friendly state-of-the-art neural forecasting models. PyCon: Salt Lake City, UT, USA.

Pampuch LA, Negri RG, Loikith PC, Bortolozo CA. 2023. A review on clustering methods for climatology analysis and its application over South America. International Journal of Geosciences 14 (9): 877–894. https://doi.org/10.4236/ijg.2023.149047

Paradis E, Schliep K. 2018. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35 (3): 526–528. https://doi.org/10.1093/bioinformatics/bty633

Pathania A, Gupta V. 2025. Interpretable transformer model for national scale drought forecasting: Attention-driven insights across India. Environmental Modelling and Software 187: 106394. https://doi.org/10.1016/j.envsoft.2025.106394

R Core Team, 2024. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Siami-Namini S, Namin AS. 2018. Forecasting economics and financial time series: ARIMA vs. LSTM. arXiv. https://doi.org/10.48550/arxiv.1803.06386

Soh YW, Koo CH, Huang YF, Fung KF. 2018. Application of artificial intelligence models for the prediction of standardized precipitation evapotranspiration index (SPEI) at Langat

River Basin, Malaysia. Computers and Electronics in Agriculture 144: 164–173. https://doi.org/10.1016/j.compag.2017.12.002

Su L, Zuo X, Li R, Wang X, Zhao H, Huang B. 2025. A systematic review for transformer-based long-term series forecasting. Artificial Intelligence Review 58 (3): 80. https://doi.org/10.1007/s10462-024-11044-2

Tuli S, Casale G, Jennings NR. 2022. TranAD: Deep transformer networks for anomaly detection in multivariate time series data. arXiv. https://doi.org/10.48550/ARXIV.2201.07284

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. 2017. Attention is all you need. Advances in Neural Information Processing Systems 30: 5998–6008.

Villegas-Vega R, Márquez-Grajales A, Mezura-Montes E, Salas-Martínez F, Ojeda-Misses MA, Romo-Gómez C. 2025. Optimization of LSTM networks through neuroevolution for drought forecasting in Mexico. Theoretical and Applied Climatology 156 (11): 562. https://doi.org/10.1007/s00704-025-05818-z

Wen Q, Zhou T, Zhang C, Chen W, Ma Z, Yan J, Sun L. 2023. Transformers in time series: A survey. *In* Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization: Macau, China, pp: 6778–6786. https://doi.org/10.24963/ijcai.2023/759

Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, Zhang W. 2020. Informer: Beyond efficient transformer for long sequence time-series forecasting. arXiv. https://doi.org/10.48550/ARXIV.2012.07436

Zuo DD, Hou W, Zhang Q, Yan PC. 2022. Sensitivity analysis of standardized precipitation index to climate state selection in China. Advances in Climate Change Research 13 (1): 42–50. https://doi.org/10.1016/j.accre.2021.11.004