A deep learning approach for image time series forecasting: Study case, United States drought monitor

Manuel Medrano-Diaz [a], Hector Rodriguez-Rangel [a,*], Vicenç Puig-Cayuela [b], Juan J. Flores [c], Rodrigo Lopez-Farias [d,e], Carlos Lara-Alvarez [f]

[a] *División de Estudios de Posgrado e Investigación, Tecnológico Nacional de México, Instituto Tecnológico de Culiacán, Av. Juan de Dios Bátiz, No. 310, Culiacán, 80220, Sinaloa, Mexico*
[b] *Institut de Robotica i Informática Industrial (CSIC- UPC), Universitat Politécnica de Catalunya, Barcelona, 08028, Spain*
[c] *Computer Science Department, 1202 University of Oregon, 1477 E. 13th Ave., Eugene, OR 97403-1202, USA*
[d] *Centro de Investigación en Ciencias de Información Geoespacial A.C., Contoy 137, Lomas de Padierna, Tlalpan, 14240, Ciudad de Mexico, Mexico*
[e] *Secretaría de Ciencia, Humanidades, Tecnología e Innovación, Insurgentes Sur, 1528, Benito Juarez, 03940, Ciudad de Mexico, Mexico*
[f] *Centro de Investigación en Matemáticas, Calle Lasec y Andador Galileo Galilei, Quantum, 98160, Zacatecas, Mexico*

ABSTRACT

Image time series (ITS) are a chronologically ordered images set which contains spatial and temporal patterns. The United States drought monitor provides a map collection of drought severity spatial distribution across the regions which changes over the time. This work aims to use the drought map ITS to extract inner spatiotemporal features patterns and forecast the spatial drought severity distribution classes for a future horizon that ranges from one to twelve weekly time steps by using a convolutional long short-term memory network (ConvLSTM). This approach offers a new perspective by using a set of images (ITS) as input for a deep learning model to predict the spatial drought in an image that represents the next time step with the drought distribution. The design also allow us to implement a recursive multi-step forecasting strategy to generate an horizon up to twelve (h) weekly drought maps. The obtained results shows that the proposed approach achieves an overall of macro F1-score metric of 0.9953 in generating the next drought map $t+1$ and 0.6965 in generating the up to $12th$ drought map. ConvLSTM is more accurate in general when it is compared with convolutional neural networks, video visual transformers networks and the naïve baseline model. These findings demonstrate the approach efficacy in identifying spatiotemporal features for reliable drought map forecasts, providing a new valuable tool for drought prediction as a visual representation.

1. Introduction

Today, deep learning tools for information retrieval are essential for solving relevant problems in several applied disciplines. In geosciences, problems like climate change, the study of people's diseases in the geographical space, agriculture production, or natural disasters affect the communities and daily lives of people. How their inner characteristics are represented differs from each other due to their own behavior. The study of the different observations becomes important to minimize the effects of future catastrophic events.

Images are one of the most popular data representations; many machine learning applications and solutions are based on this information. Using a drought map image representation for map forecasting, offers the opportunity to anticipate the drought that affects the worldwide population. Forecasts allow planning and preparation for water management, improving water distribution for residential and industrial needs. Images offer a visual representation of the observed problem; this visualization is more understandable for the human eye than numerical values. However, when computational models are used to fit this kind of data, the required models are complex, increasing the parameters exponentially in size due to the increase in relations that are a function of the size of the multidimensional input.

Images are 2-dimensional (2D) matrix data structures whose dimensions correspond to the map's width and height. An image has w columns, r rows and c bands. This visual information presents and describes rich spatial features that are human interpretable. The presented study case consists in studying the forecast performance of several deep

* Corresponding author.
E-mail addresses: manuel.md@culiacan.tecnm.mx (M. Medrano-Diaz), hector.rr@culiacan.tecnm.mx (H. Rodriguez-Rangel), vicenc.puig@upc.edu (V. Puig-Cayuela), jflore10@uoregon.edu (J.J. Flores), rlopez@centrogeo.edu.mx (R. Lopez-Farias), carlos.lara@cimat.mx (C. Lara-Alvarez).

learning representative models over the United States drought image time series.

A drought map represents the extent and severity of drought conditions in particular regions or areas, showing areas affected by different severity levels, ranging from abnormally dry to exceptionally dry conditions.

Drought is a natural phenomenon that impacts globally, in the scarcity of the water resources, threatens ecosystems, agricultural activity, and human wealth of living in general.

A drought monitor creates and presents drought maps using data from various sources, including information on precipitation, soil moisture, water levels in rivers, lakes, and reservoirs, and other key drought indicators. The resulting data are evaluated by human experts who assign severity levels to each map region (Center, 2025). These maps emerge as human-created graphical representations illustrating drought intensity in specific geographic areas, which are organized chronologically over predetermined time intervals. Map sequences inherently encapsulate the passage of time as an implicit crucial feature. This temporal aspect enables the use of bidimensional historical data to fit a model and forecast future events.

An emerging area that has received relatively limited attention involves the inclusion of the use of spatiotemporal information simultaneously to produce forecasts. The use of adequate models on image time series enables the possibility of applying forecasting methods to generate subsequent maps. This uncharted territory allows the extraction of valuable information using inner drought patterns over time. A forecasting model assumes that the maps correspond to time instants separated by a predetermined time interval. This additional feature infuses the drought map images with a temporal dimension. An image sequence corresponding to the same location and captured at constant intervals is called an image time series (ITS).

An ITS is a sequence of images denoted by a 3-dimensional (3D) structure in which temporal information (t) is part of the space, which means that the data have a $2D + t$ structure (Chelali et al., 2021). The maps' two dimensions represent the image's inherent spatial characteristics, representing geographic information. Including the temporal dimension t, infuses this composition with a spatio-temporal character, opening a gateway to understanding how spatial patterns evolve. Using this information for ITS forecasting will provide the generation of future drought maps representation.

Inferring spatiotemporal information from drought maps is a complex challenge. It needs to capture both the temporal evolution and the complex spatial relationships in the data. The interaction between these two dimensions introduces additional complexity because drought patterns vary in time and geographic location.

Time series forecasting is a discipline that uses statistical and computational techniques used to estimate future values by extracting patterns from historical data (Little, 2013). Forecasting techniques use a temporal window (w) to analyze historical data and make projections that anticipate future events, trends, or conditions. These estimates are associated with a specific period for which the prediction is made, known as the forecast horizon (h). The forecast horizon of interest changes depending of the research question and temporal resolution; when the forecast of the immediate future is important, the horizon h is short, e.g., anticipate the climate conditions for everyday commutes. In contrary, the forecast for a longer horizon benefits the planification in a mid or long-term of activities, such as crop planning, urban planning, climate change, awareness, among others.

Combining spatiotemporal problem in the time series forecasting context increases the analysis complexity of the problem. Using spatiotemporal data, increases the computational time and power needed to process the drought maps. For example, let $X = \{x_0, x_1, \dots, x_n\}$ represent a time series of n images, where x_t represents an image at instant time t. In addition, to forecast the next image x_{n+1}, the learning model can use a previous temporal window of size w indicating the subset of images $[x_{n-w}, \dots, x_n]$ to produce the forecast x_{n+1}.

The work reported in this manuscript uses convolutional long-short term memory 2D (ConvLSTM2D). This kind of artificial neural network extracts spatiotemporal information by combining the convolutional neuronal network (CNN) for the extraction of spatial features and the memory capabilities of long-short term memory (LSTM) network for temporal features. The input of a ConvLSTM2D layer is represented by a 3D tensor I_t of size $p = \text{width} \times \text{height}$, representing an image x_t at time instant t, and $I_t \in R^p$.

The output from that set of ConvLSTM2D layers is a hidden state s_n, which has a data representation of the input. This hidden state is used by a Conv2D layer that generates the following drought map $\hat{x}_{n+1}$. To generate subsequent h drought maps, the Conv2D model is fed the generated drought map $\hat{x}_{n+1}$. There exist different strategies to generate the next time steps with a given forecast horizon h (i.e., recursive, direct, multiple input multiple output (MIMO), etc.).

A recursive strategy for multi-step forecasting generates the following $[y_{n+1} - y_{n+h}]$ drought maps. This strategy consists of training an f model to forecast a single step in time. The recursive strategy presented in Bontempi et al. (2013) is defined in Eq. (1),

$$\hat{y}_{t+1} = f(y_t, \dots, y_{t-w+1}) + r_{t+1} \tag{1}$$

where $t \in \{w, \dots, n-1\}$, n is the total number of observations, w is the window size, and r is the modeling error, disturbance, or noise. When forecasting h steps in the future, one step is the first forecast using the $f(y_t)$ model. The resulting value is fed to the model as a real measurement and predicts another step in time. The forecasting process repeats this step until the horizon is reached.

Artificial neural networks (ANN) architectures has been tested, such as feed forward (FF), LSTM, and CNN. However, these architectures cannot integrally correlate the spatiotemporal characteristics; they do so independently. This study focuses on using ConvLSTM2D layers to correlate spatiotemporal characteristics from previous w ($[x_{n-w}, \dots, x_n]$) drought maps to forecast the upcoming h ($[\hat{y}_{n+1}, \dots, \hat{y}_{n+h}]$) maps using a recursive strategy.

However, before using ConvLSTM2D networks, an image processing technique was performed to simplify the task for the learning model, resulting in image categorization. The forecasting results will be subject to a comprehensive performance analysis using a confusion matrix and macro precision, recall, and F1-score metrics. Those results were compared to the CNN, video visual transformer (ViViT) and naïve forecasting models. Overall, this works aims to extract inner spatiotemporal features from a set of drought maps images to fully forecast and generate the next h maps images through the use of a model-base ConvLSTM2D layers and a recursive multi-step forecasting strategy.

Solving the problem of drought map forecasting up to twelve future drought maps entailed several challenges:

- **Matrix-based forecasting:** The most commonly used time series forecasting is a vector-based information structure with inner temporal features. In this work, the information of the drought includes both time and space, making a necessary analysis of the correlations between space and time and estimating the following drought map. This matrix-based forecast was performed by ConvLSTM2D-based models that allow the extraction of the inner spatiotemporal information from the United States (US) drought maps, allowing the identification of changes in the geographical distribution of the drought impact for future estimations.
- **Spatiotemporal correlation:** Drought maps present a distribution of the severity level of drought across the geographical area of the given observation. Each map is created once weekly, meaning that these drought observations change the severity level and distribution at each step. ConvLSTM2D-based models find the inner correlation between the geographical distribution of the drought (space) across the changes in the historical data (time) for estimating the next drought level and its distribution.

- **Categorical drought map evaluation:** Establishing categorical-based forecasting using the drought severity level representation for enhancing drought map forecast and generation. The drought map was processed to have 7 categories on the network input and was post-processed to correlate the obtained categories. The outputs were evaluated through a multi-label confusion matrix for comprehensive analysis.
- **Matrix multi-step forecasting:** Design and apply a multi-step forecasting methodology for ITS spatiotemporal information extraction, prediction, and generation for the upcoming h next drought maps. A recursive multi-step forecasting strategy was applied to obtain multiple future drought maps, feeding the network with the newly generated map to get a longer horizon map.

- **Available hardware:** Although the hardware used in the experimentation is relatively limited, it was enough to validate the experimental results.
- **Data dimensionality:** The used drought map ITS is a high dimensional structure driven by a 3-dimension distribution (width, height, and time). The above, combined with the available hardware, limits the possibilities of using a complex ConvLSTM2D-based model to solve the problem. Different pre-processing methods were applied to the drought maps to reduce data complexity. The obtained maps after this process are a matrix with the spatial distributions of drought levels in the range value from 0 to 1.

2. Related work

This section presents a review of relevant work related to the present contribution. The study of the drought phenomenon has recently increased, as there is an actual increase in drought activity around the world. The offering of different solutions to this problem has become a really important topic in forecasting. The work (Oyarzabal et al., 2025) shows an overview of drought forecasting applications, where most works focus on statistical or regression methods as a solution, and fewer works use remote sensing data as a source for analysis and forecasting. This offers a research opportunity to develop a drought forecasting solution. It is worth noting that the works presented below served as a comparative analysis for the developed solution (i.e., Yin et al., 2025), or solutions to problems in other domains were adopted to address the presented problem (i.e., Yu et al., 2024).

This section starts by describing the contributions to drought forecasting using statistical and artificial intelligence models. Then, continue describing works focused on spatiotemporal extraction and forecasting techniques to get estimations of drought and non-drought problems. This is followed by a review of ITS forecasting studies using different methods to generate some form of image. Finally, the last subsection describes other works that use a data source to generate a drought map.

2.1. Statistical and artificial intelligence models for drought forecasting

Works that focuses on drought forecasting commonly uses statistical or artificial intelligence models to perform this task. Using this type of models allows the extraction of behavioral patterns from a data set to determine its future behavior. In the area of forecasting, the main objective is to analyze the characteristics of a time series. These characteristics are related to autocorrelation, trend and seasonality, which will be useful to determine future events in the data.

A set of works uses mainly autoregressive integrated moving average (ARIMA) and hybrid ARIMA-based models to successfully forecast, like Wu et al. (2021) made standardized precipitation index (SPI) predictions, Alquraish et al. (2021) compare models for SPI forecasting, and Xu et al. (2022) made evapotranspiration SPI (SPEI) forecasting.

Artificial neural networks (ANN) have also shown significant results in drought forecasting, like the work from Khan et al. (2018) that made annual SPI and standardized water storage index (SWSI) forecasting in Malaysia, and Valipour et al. (2024) focused on short-term SPI forecasting. Ozan Evkaya and Sevinç Kurnaz (2021) made an study for SPI forecasting using statistical methods and neural networks, as well as Mokhtar et al. (2021) that uses different SPI data specifications and machine learning models. On the other hand, Brust et al. (2021) introduces Droughcast, a machine learning framework which uses primary information from the US drought monitor, satellite images and drought indexes, to make reliable mid-term forecasts up to 12 weeks ahead for the drought state and flash drought prevention.

The previous studies demonstrate that using both statistical and machine learning approaches for drought forecasting are effective. The ARIMA-based models are effective in prediction of indices, while neural networks showed strong performance in capturing nonlinear patterns. Most notable, Brust et al. (2021) show the effectiveness of combining satellite drought maps with drought indices for forecasting.

2.2. Spatiotemporal extraction and forecasting

The analysis of spatiotemporal elements allows acquiring a deeper knowledge of spatial patterns. It requires a more detailed and extensive analysis due to the large number of elements that are interacting with each other. Some of the works focus on determining the direction of the elements in space, others focus on measuring the changes of the whole space through time. However, what is mainly sought is to extract the correlations between temporal and spatial information. Chehibi et al. (2022) calculate the flood risks, He et al. (2024) made multivariate time series prediction, and Zhang et al. (2024) made traffic flow prediction.

Spatiotemporal extraction has also been used for landscape evolution modeling. For example, Geng et al. (2024) proposed a model for land use/cover change, and Xiao et al. (2019) made sea surface temperature (SST) prediction.

In drought related domains, Zhang et al. (2023a) developed a short-term SPEI to assess drought in China, Ferchichi et al. (2024) addresses drought prediction in Africa, and Marusov et al. (2024) attempted to generate one-year-ahead drought predictions using monthly climate data. Chavez et al. (2023) made a NDVI spatiotemporal analysis to detect anomalies by the climate change in the Andes.

Understanding spatiotemporal analysis helps to comprehend how the environmental patterns evolve over space and time. Capturing the dynamic interactions between spatial elements and temporal trends allows to identify meaningful correlations often missed in single analysis. However, the whole spatiotemporal analysis requires processing large volumes of high-dimensional data and detecting complex dependencies. This perspective helps to define the type of model, and which spatiotemporal capabilities are necessary for ITS drought map prediction.

2.3. Image time series forecasting

This area involves two main topics, forecasting and image processing. The main objective is to analyze the patterns present in images to determine the behavior of the elements and make an accurate forecast. On the other hand, an analysis of different data sequences could also be performed to obtain the behavioral patterns of the data and visualize them in images as a result. An interesting application that falls within this area is the analysis of the patterns present in the images to determine the future behavior of that image, generating as output an anticipated visualization of the events that will happen.

ITS forecasting is used in various disciplines to obtain successful predictions. Many of these studies adopt various approaches to predicting images, contributing to exploring this field by identifying methods and techniques applicable to this research (Lara-Alvarez et al., 2024). This review of the literature conducted a large study to find ITS forecasting-related works from 2010 to 2024, where 14 of the studies are related to drought image forecasting. A prevalent strategy

involves leveraging hybrid models, often integrating other data types for improved forecasting results.

Satellite images are commonly used due to their rich information content. In the study of Zambrano et al. (2018) predicted agricultural productivity, on the study of Yuan et al. (2020) forecasted burned areas from satellite ITS data, and Zhang et al. (2023b) predict the el niño-southern oscillation (ENSO) phenomenon. There are also several studies that focus on predicting natural disasters, like Zhang et al. (2020) that predicted land surface albedo using MODIS images.

Some studies use ITS to contribute significant research for various purposes. For instance, Firozjaei et al. (2021) predict land surface temperature (LST) and normalized difference vegetation index (NDVI) maps, Mafi-Gholami et al. (2020) predict mangrove-covered areas, Ma et al. (2018) forecast the number of metro passengers in the Beijing metro network, Yu et al. (2024) creates a local and long-range ConvLSTM model for wind speed prediction, and Xu et al. (2024) focuses on improving landslide displacement forecasting. Das et al. (2023) predict landslide using multiple sources, while Rana and Mahanta (2023) made an spatial flash flood prediction with GIS data sources.

ITS forecasting is a powerful tool that captures complex temporal patterns from visual data, making it valuable for multiple applications. A common and effective strategy, as shown in the previous works, involves hybrid solutions combining ITS with complementary data sources. However, it is also possible to only use images for forecasting, by capturing spatiotemporal patterns already presented in the data. This approach allows to predict spatial distributed phenomena over time, using the rich information content in satellite images. With the integration of temporal patterns in the spatial structures, the forecasting models have a better source of information, enabling the capture of environmental variability and improving the image forecasts.

2.4. Other works

Some studies generate future drought maps using diverse data sources. For instance, Mokarram et al. (2021) analyzed 12 key parameters to forecast drought maps for the years 2030 and 2040, Yoo et al. (2020) used a clustering algorithm to construct drought maps through spatiotemporal analysis of hydro-meteorological time series data, and Guillory et al. (2023) generated drought monitor maps using multi-index approach for long and short-term drought forecasting. In contrast, Yin et al. (2025) successfully generated drought maps through the analysis of the US drought monitor data and other data sources (getting up to 47 indexes). They use a specialized learning model called USDMNet that can analyze multiple-sourcing data simultaneously for now drought map generation.

2.5. Related work analysis

The works presented are condensed in Table 1, showing an overview of the correlation between different types of data sets and their applications. To understand the dataset and models names, Table 2 shows the meaning of all the acronyms used in this section. Through the analysis of related work, the table presents statistical models, neural networks, hybrid models, and other specialized techniques to address specific datasets, providing valuable insight into the state of art.

The works presented show a great diversity of applications both in the drought area and in similar areas. Based on the analysis of these works, there are a lot of opportunity areas to improve the existing methods in drought forecasting. By adding some layers of complexity of the actual problem presented in drought maps and the analyzed works, such as incorporating multi-temporal satellite imagery and spatiotemporal dependencies, more accurate and context-aware forecasting models can be applied. The main differences between this work and the others relies in the application difficulty.

Previous research helps to find an opportunity area for this work-study case; none of the exposed works in the literature review uses only images to successfully forecast a full future image applied to satellite images for drought maps. They also develop complex specific mechanisms and methodologies to achieve their results. Some of them also separate the extraction of spatial and temporal patterns, which can imply some limitations to their solutions. In this work, the use of ConvLSTM networks allows the extraction of the inner spatiotemporal features of a set of images to generate the next drought map images. The two closest works to this are the works from Brust et al. (2021) and Yin et al. (2025), which use the data from the US drought monitor, and other sources, for forecasting or drought map generation. However, making a direct comparison between their results and ours is not fair, since the methods and scopes are different from each other, because they establish their own conditions. This offers an opportunity to investigate learning models that are capable of extract spatiotemporal features and future image generation. Also, multi-step forecasting methods allows to amplify the forecasting range and generate mid-term future images.

The presented work shows an easy to follow method based on an existing deep learning model called ConvLSTM, which consider spatiotemporal relationships natively. The advantages of this application are the low amount of time needed for training, as well as it is easy to apply and adapt to different contexts. In fact, most of the presented related works propose specific solutions to specific problems, which makes it difficult to adapt to new problems.

3. Materials and methods

3.1. Problem definition

Let $\mathrm{I} = x_{t=1}^{w}$ be a drought map time series of length N, i.e., $|\mathrm{I}| = w$, where each x_t denotes the drought map observed at time t. Each x_t is a spatial grid (image) defined over a single, connected landmass (i.e., no disconnected regions or islands), where every pixel takes values from a finite, totally ordered set D. The elements of D represent quantized levels of drought severity, obtained by partitioning a continuous range of drought indicators into a finite, totally ordered set, with lower values indicating milder or no drought conditions and higher values indicating more severe drought levels. The objective is to forecast a future drought map $\hat{x}_{w+\tau}$ for some horizon $\tau \in \mathbb{N}$ with $\tau \geq 1$. In the particular case where $\tau = 1$, the goal is to predict the immediate next time step.

3.2. Data pre-processing

Satisfying problem preconditions. This stage aims to ensure that each drought map image contains a fully connected landmass and that drought conditions are represented by an ordered set D.

To focus on relevant areas, a mask is applied to identify the landmass, considering only pixels with meaningful information. On the other hand, the quantization process partitions the original pixel range for maps represented using conventional pixel values and maps these values to a finite set D. This process produces the ordered drought level, reducing the range of values and providing a simplified representation of the intensity of the drought across the dataset.

Obtaining image time series. Drought maps are typically provided as a single sequence of images. To generate multiple forecasting steps, a common strategy is to apply a sliding temporal window of size w, defining the size of the sequence of consecutive images. At each step of the sliding window, a new forecast is created by stacking w consecutive images as the input.

Fig. 1 illustrates an example where a sequence of $w + 4$ images produces a dataset with three instances of the problem. Specifically, from the sequence $[x_1, \ldots, x_{w+4}]$, we obtain a dataset consisting of three instances,

$$\{\mathrm{I}_1 \mapsto y_1 = x_{w+1}, \mathrm{I}_2 \mapsto y_{w+2} = x_{w+2}, \mathrm{I}_3 \mapsto y_{w+3} = x_{w+3}\},$$

where $\mathrm{I}_j = [x_j, \ldots, x_{j+w-1}]$.

Table 1
Related works, data sources, and their corresponding applied models.

Work	Aspects	
	Data source	Applied model
Wu et al. (2021)	Monthly SPI from 3 meteorological sites	WT, ARIMA, LSTM, ARIMA-LSTM, GM, DGM
Alquraish et al. (2021)	SPI from Bisha Valley, Saudi Arabia	HMM-GA, ARIMA-GA, ARIMA-GA-ANN
Xu et al. (2022)	SPEI (6, 12, and 24-month)	ARIMA, SVR, LSTM, ARIMA-SVR, Least Squares SVR, ARIMA-LSTM
Khan et al. (2018)	Annual SPI, SWSI in Malaysia	WANN
Valipour et al. (2024)	Short-term daily precipitation for 1, 2, and 3 day from 28 US sites	WPSOANFIS, WGMDH, W-LSTM
Ozan Evkaya and Sevinç Kurnaz (2021)	SPI (3, 6, 9, 12-months) from Bursa	NAR, NARX
Mokhtar et al. (2021)	SPI (3, 6-months) from Qinghai-Tibet Plateau	RF, XGB, LSTM, CNN
Brust et al. (2021)	CONUS 2003–2020 as SMAP 9-km and 15 drought indexes	Seq2Seq model based on GRU encoder/decoder
Chehibi et al. (2022)	Flood extent mapping (Sentinel-2, MODIS, and Lansat-8) from Chad lake (2003–2019)	Interval methods and arithmetic operations
He et al. (2024)	Multivariate time series (PeMS04, PeMS08, CORAv2.0)	DMF-STNet, STGNNs
Zhang et al. (2024)	Traffic flow	FEHGCARN
Geng et al. (2024)	China Land Use	LUCC, Cellular Automata
Xiao et al. (2019)	Sea surface temperature	ConvLSTM
Zhang et al. (2023a)	SPEI-5d from China 1962–2018	ARIMA, Random Forest, RNN, LSTM, ConvLSTM
Ferchichi et al. (2024)	Multivariate sensing data, SMI	GAN, CNN-LSTM
Marusov et al. (2024)	PDSI, TerraClimate Monthly	ConvLSTM, Transformers
Chavez et al. (2023)	NDVI, SPI and PDSI	Pearson and Spearman correlation coefficient
Zambrano et al. (2018)	zcNDVI, SPI (1, 3, 6, 12 and 24-month), PDO, Multivariate ENSO, growing length, latitude, longitude	Optimal linear regression, Deep learning model
Yuan et al. (2020)	Satellite ITS of burned areas from Heilongjiang province, MODIS	LSTM encoder–decoder
Zhang et al. (2023b)	Simulated LASG/IAP LICOM3, remote sensing data	ED-PredRNN, ConvLSTM
Zhang et al. (2020)	Land surface Albedo of MODIS images transforming into IMFs	EEMD, NARnet
Firozjaei et al. (2021)	LULC, LST, NDVI	CA-Markov
Mafi-Gholami et al. (2020)	Mangrove-covered areas, rainfall, sea-levels	CA, CA-Markov, statistical models
Ma et al. (2018)	Metro passengers from Beijing	CNN, Bi-LSTM
Yu et al. (2024)	Wind speed from 253 virtual wind turbines	LLConvLSTM, Deformable Conv V2, Coordinate Attention, RDCM, DCAM
Wu et al. (2024)	LaSOT, GOT-10K, TrackingNet, VOT2018, OTB100, VAU123	DTS, 2D CNN, 3D CNN, SFM, DAM
Xu et al. (2024)	InSAR landslide displacement	SAU-DTAE, LTAE, SampEN
Das et al. (2023)	LSMs, FR, IOE, SI, MIV, EBF from Sikkim, India	FR, IOE, SI, MIV, EBF models
Rana and Mahanta (2023)	Plan curvature, elevation, NDVI, slope, stream power index, TWI distance from river and road, LULC, STI from Uttarakhand, India	WOE, ANN, SVM, KNN
Mokarram et al. (2021)	Altitude, aridity index, erosion, groundwater depth, land use, potential evapotranspiration, precipitation days, precipitation, slope, soil texture, soil salinity, distance to river	CA-Markov, Fuzzy-AHP
Yoo et al. (2020)	Hydro-meteorological time series	Clustering algorithm
Guillory et al. (2023)	CDI, SPI-3, NDVI, evaportranspiration and soil moisture	Normalized Linear combination with PCA, AussieGRASS
Yin et al. (2025)	Multi-source images and indexes from different perspectives (Precipitation, Temperature, Evapotranspiration, Soil moisture, Vegetation, Others)	USDMNet that blends U-net, Siamese network, and Convolutional Block Attention Module

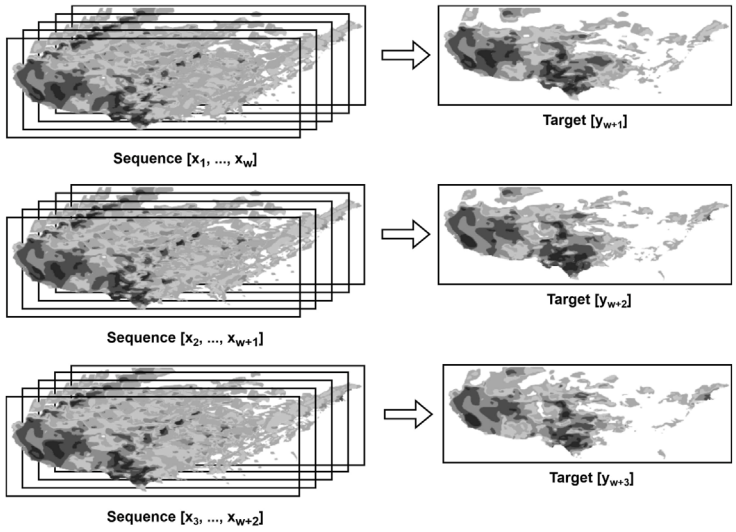


Fig. 1. Dataset of drought map time series consisting of input image sequences and corresponding target drought maps.

Table 2
List of acronyms and their meanings in related studies.

Acronym	Meaning	Acronym	Meaning
ANN	Artificial neural network	ARIMA	Auto-regressive integrated moving average
CA	Cellular automata	CDI	Combined drought indicator
CNN	Convolutional neural network	CONUS	Contiguous United States
DCAM	Dense coordinate attention module	DGM	Discrete gray model
DMF	Dynamic multi-fusion	DTS	Deformable transformer and spatiotemporal information
EEMD	Ensemble empirical mode decomposition	ENSO	Multivariate El Niño-Southern Oscillation
FEHGCARN	Fourier-enhanced heterogeneous graph convolution attention recurrent network	Fuzzy-AHP	Fuzzy analytic hierarchy process
GA	Genetic algorithm	GAN	Generative adversarial network
GM	Gray model	HMM	Hidden Markov model
LICOM3	LASG/IAP Climate System Ocean Model version 3	LSTM	Long short-term memory
NAR	Nonlinear auto-regressive	NARX	NAR with external inputs
RF	Random forest	XGB	Extreme gradient boost
LTAE	Lightweight temporal attention encoder	LUCC	Land use/cover change
LULC	Land use and land cover	LST	Land surface temperature
MODIS	Moderate resolution imaging spectroradiometer	NARnet	Nonlinear auto-regressive neural network
NDVI	Normalized difference vegetation index	PDSI	Palmer drought severity index
PDO	Pacific decadal oscillation	PCA	Principal component analysis
RDCM	Residual deformable convolution module	RNN	Recurrent neural network
SAU-DTAE	Self-adaptive UNet with decomposed temporal attention encoder	SampEN	Sample entropy
SFM	Spatiotemporal fusion model	SMAP	Soil moisture active and passive
SPI	Standardized precipitation index	SPEI	Evapotranspiration standardized precipitation index
STGNN	Spatiotemporal graph neural network	SVR	Support vector regression
SWSI	Standardized water storage index	WANN	Wavelet transform artificial neural network
WGMDH	Wavelet group method of data handling	WPSO ANFIS	Wavelet particle swarm optimization adaptive neuro-fuzzy inference system
WT	Wavelet transform	zcNDVI	Seasonal cumulative normalized difference vegetation index
LSMs	Landslide susceptibility maps	FR	Frequency ratio
IOE	Index of entropy	SI	Statistical Index
MIV	Modified Information Value	EBF	Evidential Belief Function
WOE	Weight of evidence	KNN	K nearest neighbor
TWI	Topographic wetness index	STI	Sediment transport index

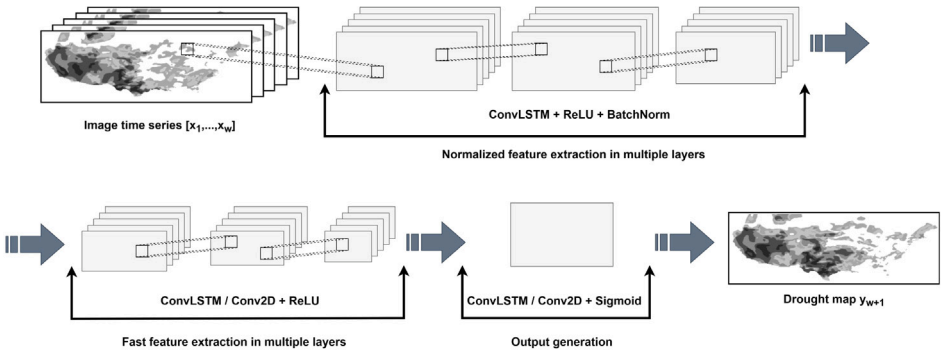


Fig. 2. Proposed architecture.

3.3. Proposed approach

The proposed architecture to obtain a prediction is shown in Fig. 2. In this work, a ConvLSTM2D-based architecture is used; this type of layer can handle spatiotemporal data structures to get inner features that will be used to forecast a full image.

3.3.1. Overview of the ConvLSTM2D

ConvLSTM2D networks (Shi et al., 2015) are useful to obtain spatiotemporal features on a set of temporal images. Due to the capabilities of ConvLSTM2D layer, they can extract spatial information of the actual image while considering information of the same spatial zone from the past. This structure combines the extraction mechanism from the convolution while storing the extracted features. Fig. 3 shows the inner structure of a ConvLSTM.

3.3.2. Architecture

Stacking ConvLSTM layers helps build a hierarchy of spatial features. To improve the performance of LSTM our proposal has three phases:

Normalized feature extraction. This phase is composed of one or more blocks, each consisting of a ConvLSTM layer, followed by a batch normalization layer and a ReLU activation function. Placing this phase at the beginning of the network architecture contributes to more effective training. Convolutional layers, such as ConvLSTM, can produce outputs with widely varying distributions, which may slow down training and destabilize the learning process. Batch normalization mitigates this issue by stabilizing the distribution of activations across layers and training iterations. In addition to improving stability, normalization provides a mild regularization effect that can help reduce overfitting. Furthermore, during backpropagation, normalization helps maintain well-conditioned gradients by keeping activations within a suitable range. This reduces the likelihood of

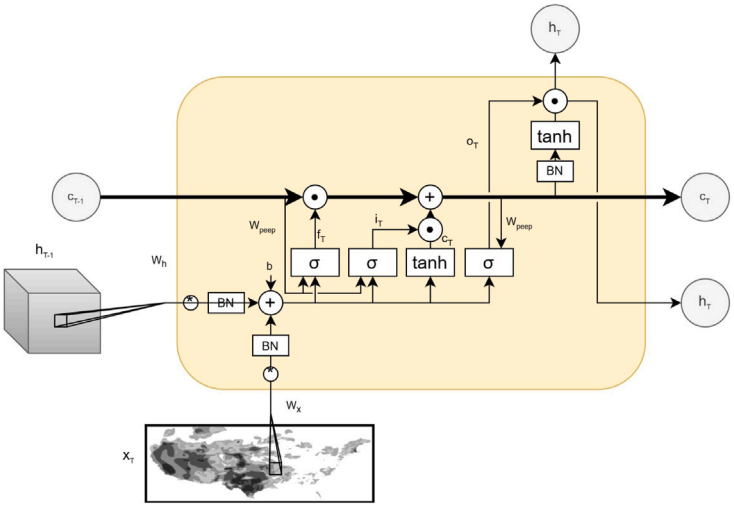


Fig. 3. ConvLSTM2D inner structure.
Source: Based on Shi et al. (2015).

vanishing or exploding gradients, particularly when combined with ReLU activations, and facilitates more reliable and efficient optimization in deep networks.

Fast feature extraction. Omitting normalization layers near the output can help preserve the natural scale and distribution of the learned features, which is particularly important for ensuring accurate predictions in tasks where the output values carry meaningful magnitudes. Additionally, avoiding normalization at this stage prevents also is compatible with the output layer. This design choice also reduces unnecessary computational overhead, as the stabilizing effects of normalization are less critical in the final layers. Furthermore, removing normalization near the output can help maintain stronger gradient signals during backpropagation, supporting more effective learning in earlier layers of the network.

Output layer. At the output layer, its important to not retrieve the data sequence. The weights inside the ConvLSTM layers will perform operations in the data sequence at the end to extract the relevant information inside. The model will summary the information of the hidden state to generate an output that tries to look that the target drought map. This design ensures that the output from the model is a 2-dimensional matrix with the drought severity distributed in the space. The sigmoid activation ensures that the output matrix values has a range from 0 to 1.

3.3.3. Multi-step forecast

Different strategies exist that perform a Multi-step forecast, for instance, recursive, direct, or MIMO. This work uses the recursive strategy to forecast to a given horizon (h). The idea is forecast $\hat{y}_{n+1}$ using the time series $\mathrm{I} = x_{t=1}^{w}$. Then, $\hat{y}_{n+1}$ is used to predict a second forecast. This way, the model will be fed with $[x_{n-w+1}, \ldots, x_{n+1}]$ to generate the next step x_{n+2}, until the full horizon is forecasted, generating the images $[x_{n+1}, \ldots, x_{n+h}]$. Therefore, the forecast is performed recursively until the specified horizon is achieved.

3.3.4. Loss function

We propose using the *mean absolute error* (MAE) loss function,

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|, \tag{2}$$

which applies a uniform, linear penalty to all errors, treating small and large errors equally. Such a characteristic makes MAE particularly suitable for sequence prediction tasks, as it promotes consistency across predictions and avoids excessive sensitivity to outliers. In contrast, traditional loss functions like mean squared error (MSE) and root mean squared error (RMSE) disproportionately emphasize larger errors, potentially leading to suboptimal performance. This focus on large errors can cause instability during training and hinder the model's ability to generalize effectively.

4. Experimental studies

4.1. Dataset

Study area. The work presented in this manuscript uses observations of the US drought monitor map to make future drought map forecasts. The data are based on a week-to-week drought map image. The maps are obtained from a time interval from January 4, 2000, to October 8, 2024. In that time interval, 1,293 samples were obtained through the US drought monitor services.

The US drought monitor provides the drought map in the format of geographic information system (GIS), which is obtained through a web map service (WMS). This service provides images in portable network graphics (PNG) format of geo-referenced maps containing clean delimitation information. Each WMS file contains a drought monitor layer based on clean and re-projected shapefiles to the world geodetic system of 1984 (WGS84) projection (Center, 2025). This process provides with PNG format images of 640 × 480 pixels in RGB, where each pixel represents a square region of $\approx 12.5 \times 23$ km.

Using free geographic delimitation drought maps gives some advantages. These clean data provide a more faithful and continuous evolution of drought conditions in a geographic area. Delimitation can introduce artifacts and artificial constraints to the data, affecting the accurate prediction of drought-affected regions. Working with maps without delimitation provides a more comprehensive and cohesive view of drought spread, facilitating more precise analysis and informed decision-making in water resource management and drought impact mitigation. Fig. 4 shows the provided drought map from the US drought monitor and Fig. 5 the obtained free delimitation post-processed WMS map.

The drought monitor database provides a set of drought maps, a visual representation that displays the distribution of drought intensity in a specific geographical region. It is a tool for understanding and communicating the drought magnitude. Researchers, planners, and decision-makers can assess its impact and severity to take preventive actions in the community. The drought maps are made by collecting

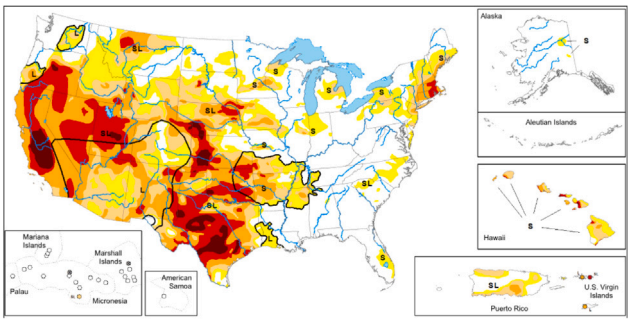


Fig. 4. Original drought map with delimitation.

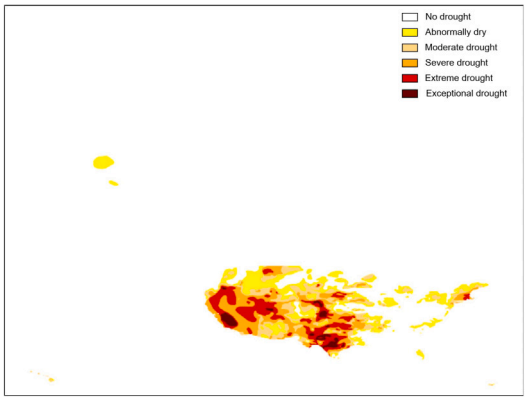


Fig. 5. WMS representation of US drought.
Source: From Center (2025).

and analyzing data, which is a convergence of some characteristics, such as precipitation, soil moisture, water levels in rivers, lakes, reservoirs, and other key indicators (Svoboda et al., 2002). An expert group of scientists uses these indicators to determine drought severity through different regions on the map (Center, 2025).

The given drought severity levels are represented on the map using values distributed in the different regions. Since there are only six different drought level indicators presented, this allows the given problem to be treated as a regression or categorical problem. The proposed solution is to treat the given problem like a regression for feature behavior pattern extraction for the forecast, but evaluating the generated drought maps like a pixel classification for precise evaluation of the category exact assignation. The above means that the drought map can be evaluated like a multi-class problem for pixel category evaluation.

4.2. Dataset pre-processing

Originally, the US drought map images contained all of the US territory, including parts such as Alaska, Puerto Rico, and Hawaii. This will present a problem due to the dispersion of the geographical areas; the generated drought map contains at least 75% of their information as a "No drought" category. Also, this area does not change over time. This can lead to an unbalanced data category, affecting the map generation and the given results of the map generations. To avoid the above, the drought map was cropped to include only the contiguous United States, as shown in Fig. 6, getting a more focused map on the interest region. The obtained set of new images goes from the original size of 640 × 480 to a new size of 360 × 120. This way, the balance of the different drought levels will have a better distribution.

The cropped maps are stored in a matrix structure size of 360 × 120 with numbers representing the drought categories evenly spaced between 0 to 1. Each pixel in this configuration represents ≈275 km^2 of land surface. However, not all the map surface is important for the model, since there are zones in the map that does not present any type of drought value. This zone is deactivated adding an extra category called "not of interest" with the assigned value of 0, as can be shown in Fig. 7.

Table 3
Grid search layers and hyperparameters range.

Hyperparameter	Value
Number of layers	1, 3, 5, 7
Type of layers	ConvLSTM, BatchNorm, Conv2D
Filter quantity	16, 32, 64
Filter size	(3,3), (5,5)
Window size	2–12

4.3. Methodology

Fig. 8 shows the main steps used to process, learn, and forecast future drought maps.

The main process consists of using a grid search algorithm for architecture and model hyperparameter selection. The deep learning model architecture and the window size parameter will be modified through a grid search. The training parameters will be the same for all learning models, and each will forecast the following h drought maps, which will be generated by a recursive forecasting strategy. Finally, the obtained maps will be evaluated and validated to get the experiment's final results. The results are used to select the conditions and hyperparameters of the deep learning model to enhance the model precision.

4.4. Architecture search and training

The grid search establishes different combinations of model conditions to identify the most effective of the assigned conditions, improving forecast accuracy. The main modifications included variations in the number and type of layers, as well as the number and size of the filters. Additionally, the window size parameter was also adjusted on the ITS dataset generation. The range of used values can be seen in Table 3.

This work proposes the exploration of several modifications to the model, e.g., the number of layers, kernels, kernel shape, and data window size. Table 4 shows several proposed model combinations that use these parameters as a guide for modifications. Each row represents a different model, and the columns represent the type of layer used in each part of these models. For instance the first model contains six layers: layer 1 is a ConvLSTM2D with 64 filters, using 5 × 5 convolution kernels and ReLU as activation function. All proposed models will be evaluated later on in the experimental section.

4.5. Evaluation metrics

The precision metric is valuable because it retrieves the number of instances that were correctly classified. Is defined by Eq. (3),

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{3}$$

On the other hand, the recall metric is the number of retrieved relevant instances that were correctly classified. Is defined by Eq. (4),

$$\text{recall} = \frac{\text{TP}}{\text{TP+FN}}. \tag{4}$$

Finally, the F1-score metric is based on the two previous metrics to calculate the model performance, it is defined by Eq. (5),

$$\text{F1-score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{2\text{TP}}{2\text{TP+FP+FN}}. \tag{5}$$

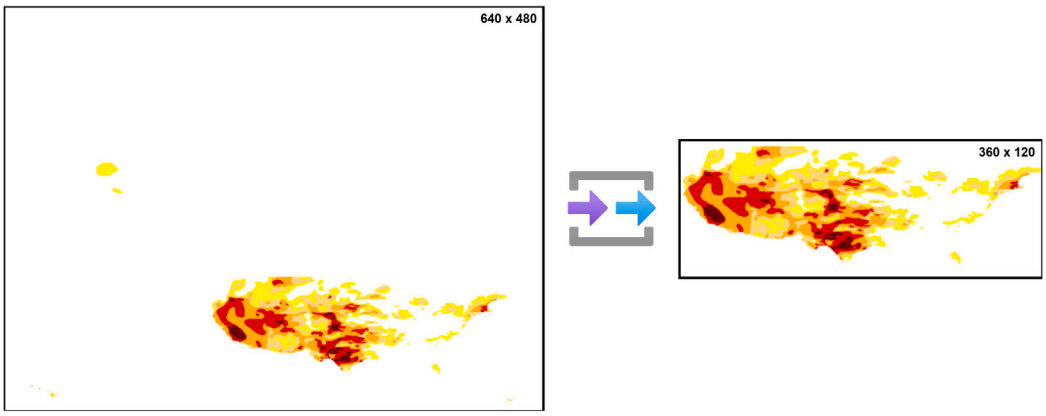


Fig. 6. Drought map crop from 640 × 480 pixels to 360 × 120 pixels.

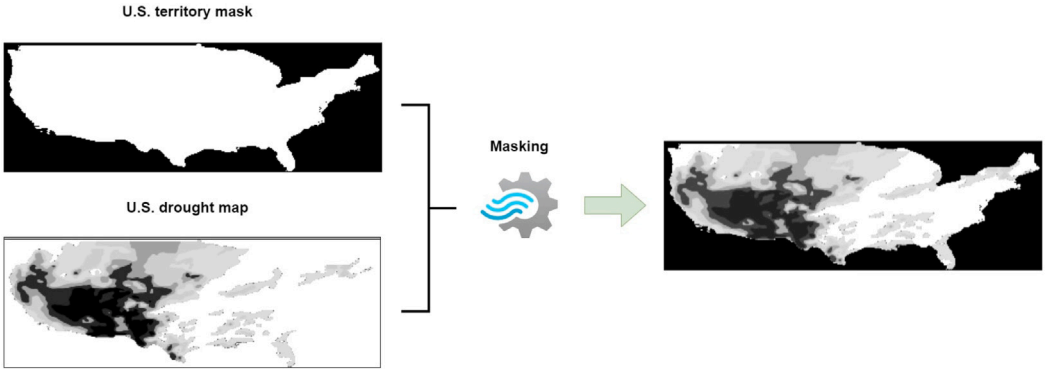


Fig. 7. Map mask applied to drought maps interest zone.

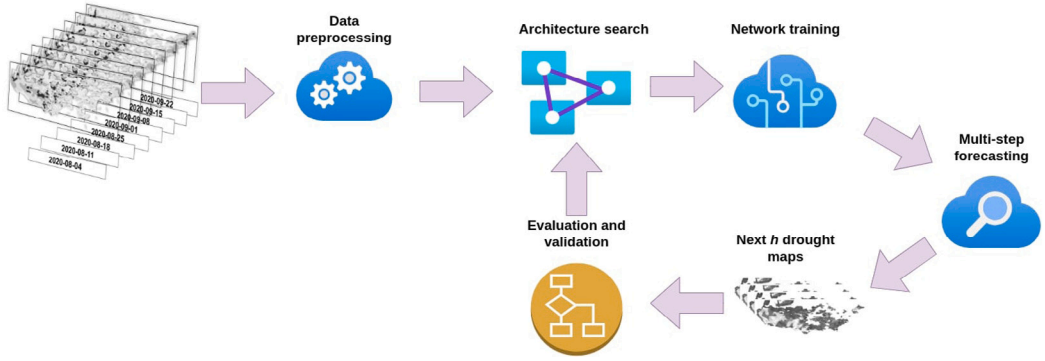


Fig. 8. Drought map forecasting methodology.

This evaluation is calculated for each category presented in the confusion matrix. To calculate the overall model accuracy, a macro score was used. To calculate the macro it is needed to obtain the mean value from all the classes for each metric. The macro metric is defined in Eq. (6),

$$\text{macro} = \frac{\sum_{i=1}^{n} \text{metric}_i}{n}. \tag{6}$$

This evaluation ensures to obtain the overall model accuracy for evaluation purposes. This will help to understand the performance of the model with the specified conditions.

4.6. Model evaluation and validation

There are different ways to evaluate model performance, and selecting the correct one will depend on the features of interest for this work. In this work, a pixel class evaluation is performed to verify the effectiveness of the forecasting model.

The model performs the forecast, generating images with values from [0–1]. The objective is to compare the original data using a set of images that the model did not use for the training phase, called test set. To perform this operation the forecasted images from the test set will be transformed to a quantized set containing drought levels.

Since drought map images present seven different pixel categories, one per drought index level and one for class “not of interest”, a multi-label confusion matrix is used to represent pixel categorization. Using this technique for evaluation allow us to acquire an accurate evaluation on each pixel for the generated drought maps, allowing to understand the performance of the model in generating new images using an unseeing test set.

With the confusion matrix is possible to get the number of pixel correct classifications, true positives (TP) is the number of times that

Table 4
Tested architectures.

ID	Layer							
	1	2	3	4	5	6	7	8
1	ConvLSTM (64,(5,5), ReLU)	Batch norm	ConvLSTM (64,(5,5), ReLU)	Batch norm	ConvLSTM (64,(3,3), ReLU)	Conv2D (32, (3,3), ReLU)	Conv2D (16,(3,3), ReLU)	Conv2D (1,(3,3), Sigmoid)
2	ConvLSTM (16,(5,5), ReLU)	Batch norm	ConvLSTM (16,(5,5), ReLU)	Batch norm	ConvLSTM (16,(3,3), ReLU)	Conv2D (1,(3,3), Sigmoid)	None	None
3	ConvLSTM (64,(5,5), ReLU)	Batch norm	ConvLSTM (64, (5,5), ReLU)	Batch norm	ConvLSTM (64,(3,3), ReLU)	ConvLSTM (1,(3,3), Sigmoid)	None	None
4	ConvLSTM (64,(5,5), ReLU)	Batch norm	ConvLSTM (64, (5,5), ReLU)	Batch norm	ConvLSTM (64,(5,5), ReLU)	Conv2D (1,(3,3), Sigmoid)	None	None
5	ConvLSTM (64,(5,5), ReLU)	Batch norm	ConvLSTM (64, (5,5), ReLU)	Batch norm	ConvLSTM (64,(3,3), ReLU)	Conv2D (64,(3,3), ReLU)	Conv2D (1,(3,3), Sigmoid)	None
6	ConvLSTM (64,(5,5), ReLU)	Batch norm	ConvLSTM (64, (5,5), ReLU)	Batch norm	ConvLSTM (64,(3,3), ReLU)	Conv2D (64,(3,3), ReLU)	Conv2D (32,(3,3), ReLU)	Conv2D (1,(3,3), Sigmoid)
7	ConvLSTM (16, (5,5), ReLU)	ConvLSTM (16, (5,5), ReLU)	ConvLSTM (16, (3,3), ReLU)	Conv2D (1,(3,3), Sigmoid)	None	None	None	None
8	ConvLSTM (1, (5,5), Sigmoid)	None	None	None	None	None	None	None

the given class is correctly classified in its pixel, and true negatives (TN) is the number of times that the given class is correctly not classified in a pixel. On the other hand, the confusion matrix also gets the number of pixel incorrect classification, false positives (FP) is the number of times that the given class is incorrectly classified in a pixel, and the false negative (FN) is the number of times that the given class is incorrectly not classified in the pixel. With these values it is possible to obtain different metrics, such as precision, recall and F1-score.

5. Results

This section explains the conditions and parameters during the experimentation process. It also shows the results for each experiment. The experimentation phase performs the grid search for each hyperparameter configuration of every ConvLSTM variation. The experiment results show the model sensitivity by using multiple layers and selecting different window size values. It will also prove that the proposed methodology produces high-quality short-term forecasts, in most cases the ConvLSTM architectures outperform the Naïve model, with the possibility of generating mid-term forecasts.

5.1. Model training and validation

The configurations presented in Table 4 were trained and evaluated using Tensorflow and Keras. All models were trained under the same hardware and training conditions. Table 5 shows the experimental hardware setup where the experiments were performed. On the other hand, the model was trained under the same hyperparameter conditions, using ADAM as the optimizer, and the mean average error (MAE) as the loss function for weight adjustment. In addition, the training uses an *early stopping* callback that observes if the model is not learning anymore. The values used for the training parameters are specified in Table 6.

Once each model was evaluated, the parameter window size was changed over the best-given model from Table 4. In this way, the exploration will give different results depending on the number of samples from the past needed to forecast the next h maps. This will allow determining a good quality model based on the limited search space for limited historical data.

The confusion matrix shows the results from the presented methodology for the h forecasted drought maps. Since the forecast images contain only seven categories, the confusion matrix obtained compares the seven classes on the forecasted h drought maps versus the real drought maps. The values obtained can be used to compare how a pixel is classified in the image. This multi-label confusion matrix will be used for model performance analysis.

Table 5
Hardware experimentation conditions.

Hardware	Description
CPU	Intel core i7 - 8700 3.2 GHz
RAM	2X 16 Gb DDR4 3200 MHz
GPU	2X NVIDIA GeForce RTX 2080 Ti 12 Gb
Operating system	Ubuntu 24.04.1

Table 6
Base hyperparameter values.

Parameter	Name	Value
Optimizer	Adam	Default
Loss function	MAE	Default
Batch size	Batch size	2
Epochs	Epochs	150
Callback	Early stopping	15
Train data	Train size	56%
Validation data	Validation size	14%
Test data	Test size	30%

5.2. Optimized architecture via grid search

To evaluate the ConvLSTM-based model for drought map forecast, it was trained with each of the different architecture configurations shown in Table 4. Each of the given learning model were trained to generate the next drought map $t + 1$ and each of them was trained using a window size range from 2 to 12. Once the model is trained, an unseen test dataset generates the drought maps for a fair evaluation. Each pixel in the images was classified in a given drought category through a confusion matrix. To show an example, Table 7 shows the evaluation results of the model with ID #3 using a window size of 6.

Through an analysis process, the confusion matrix presents the fundamental metrics TP, TN, FN and FN, used in this manuscript to calculate the precision, recall and F1-score metrics. To capture the classification error over the different categories, we report the macro

Table 7
Confusion matrix of the best-performing model (Model 3) with a window size of $w = 6$.

Res. category	Expected category						
	NOI[a]	ND[b]	D0	D1	D2	D3	D4
NOI	**6 212 772**	0	0	0	0	0	0
ND	121 273	**4 807 521**	1935	168	8	2	55
D0	0	90	**3 208 872**	389	14	0	0
D1	0	1	496	**926 443**	236	4	0
D2	0	0	44	718	**87 567**	135	4
D3	0	0	0	0	165	**508 196**	29
D4	4	0	0	0	4	308	**192 947**

[a] NOI (Not of interest).
[b] ND (No drought).

Table 8
Metrics of the best-performing model (Model 3) with window size $w = 6$.

Category	TP	TN	FP	FN	Precision	Recall	F1 score
NOI	6 212 772	9 736 351	121 277	0	0.9809	1	0.9903
ND	4 807 521	11 139 347	91	123 441	0.9992	0.9750	0.9873
D0	3 208 872	12 858 560	2475	493	0.9992	0.9998	0.9995
D1	926 443	15 141 945	1275	737	0.9986	0.9992	0.9989
D2	87 567	15 981 505	427	901	0.9952	0.9898	0.9925
D3	508 196	15 561 561	449	194	0.9991	0.9996	0.9994
D4	192 947	15 877 049	88	316	0.9995	0.9984	0.9990
Macro score					**0.9961**	**0.9945**	**0.9953**

approach. This approach allow us to measure the classification error for each class. Then we calculate the arithmetic mean. Table 8 shows the complete values for the model ID #3 using a window size of 6.

5.3. Drought map forecasting using the optimized architecture

Each learning model configuration from Table 4 was trained using a window size range from 2 to 12. This will allow us to show the difference between different ConvLSTM-based model configurations and how much data history is needed to perform a forecast correctly. The model's performance, varying the parameters with the grid search algorithm, presents the results in Fig. 9 as a 3D plot representation of the obtained F1-score per experimental setup.

As can be seen in Fig. 9(a) Heat map, shows a comparative overview of the different F1 score values obtained by the given grid search conditions. The architectural design, with the conditions of the window size value, affects the model's process learning. Using a window size value from 2 to 10 with the specified model architecture obtains a high F1-score value. However, using certain window size values, such as 11, negatively affects the learning process, affecting the generation of drought maps.

Also, using a simple ConvLSTM model configuration, such as Model with ID 1 (204 parameters), without a deep construction, also affects the accuracy of the generated map, since there are not enough network connections to store spatiotemporal relations. However, for the most part, architectural experiments show a high F1-score value. In this context, it is not necessary to construct a high-complexity model (≈1,500,000 parameters) of millions of parameters to obtain a highly accurate image prediction. This means that using a medium complexity model (≈97,000 parameters) with a not-high window size value is enough to get a high-accuracy for one step-ahead drought map generation.

The results show that keeping the model with the specified architectural design of Section 3.3 can mostly generate drought maps with a high accuracy value. Mainly by using deep learning networks based on multiple ConvLSTM2D layers with batch normalization and Conv2D. However, it can also be seen that when they are not using batch normalization layers, the model is unstable in a wider window size range. This also happens when the model does not have a deep architecture, since the accuracy of the model decreases significantly. Overall, the best model configuration is model #3 in generating the next drought map t+1 with a macro F1-score of 0.9953. Combining all pre-processing algorithms and multiple stacking ConvLSTM2D with BatchNorm and Conv2D provides a high-quality drought map forecast for step $t + 1$. Enhancing the inner features of the drought map significantly helps the interpretability of the model, to the point that the ConvLSTM-based model efficiently extracts the spatio-temporal features to understand changing patterns and generate a drought map forecast.

The window size parameter, on the other hand, gives a more interesting result. The selection of this parameter with the different learning models can help the model improve or worsen. Meaning, this parameter is important in the learning process of the model since it is related to how much past information is needed to understand as best as possible the changing patterns in the data.

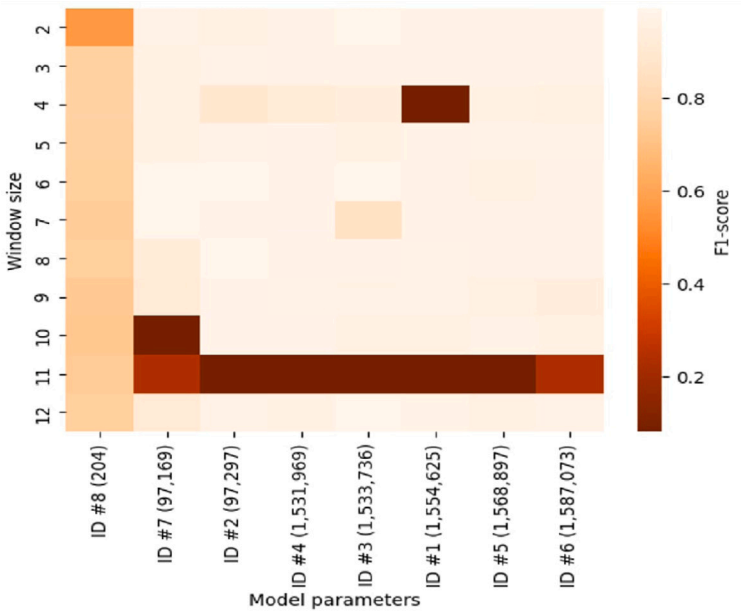
We compare the results of ConvLSTM with CNN, Video Visual Transformer (ViViT), and the Naïve base line model. The Naïve model uses the last image $[x_n]$ drought map as the next generated maps $[x_{n+1}, \dots, x_{n+h}]$. These results are shown in Fig. 10, the highest value indicates the better performance of the model.

As can be seen, ConvLSTM has the best results in generating the next drought map. To provide a reference for the visual forecast representation of the drought map, Fig. 11 shows the error in each generated map for each of the given models in generating the next drought map $t + 1$.

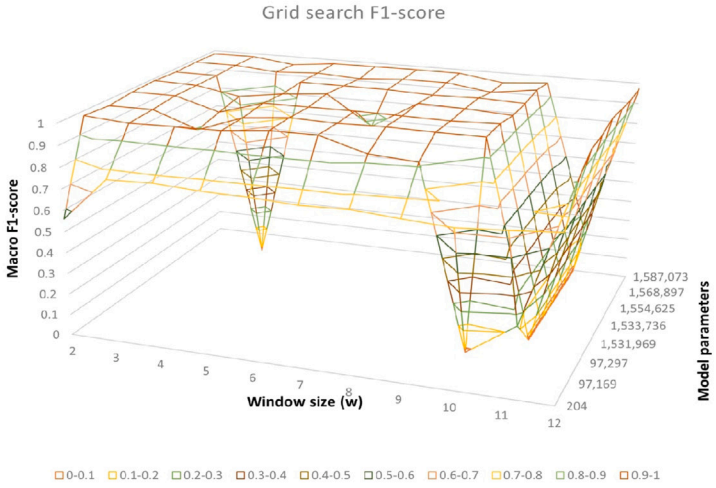
5.4. Horizon (h) drought maps forecasting results

To properly make a multi-step forecast, it is recommended to apply a strategy. The idea is to establish conditions that can alter the model or the data in specific ways to perform a longer horizon forecast. The objective of this work is to apply a recursive multi-step forecasting strategy to generate the next $[y_{t+1}, y_{t+2}, \dots, y_{t+h}]$ maps. Since the objective is to generate a short and mid-term forecast, a horizon of $h = 12$ was established. The best configurations of the proposed ConvLSTM models are used to create Fig. 12, which presents the performance in terms of the F1 score of each model for each forecast horizon.

Made a mid-term forecast is a difficult task, it needs to be consider that the model, using a recursive strategy, will have an accumulative error. This error will impact the model's performance, the longer the forecast. However, the obtained results have a constant cumulative error, maintaining a reasonable accuracy in the first 2–3 steps ahead. To compare the obtained horizon forecast with other models, CNN, ViViT,



(a) Heat map



(b) 3D visualization

Fig. 9. F1-score results for various parameter settings evaluated via grid search.

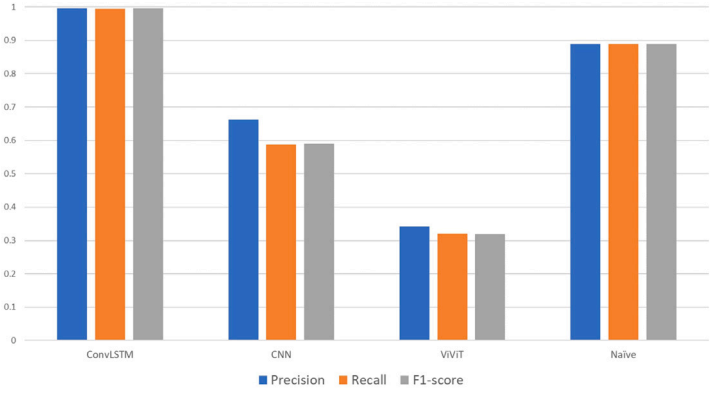


Fig. 10. Drought map forecast $t+1$ models type precision, recall and F1-score visual results.

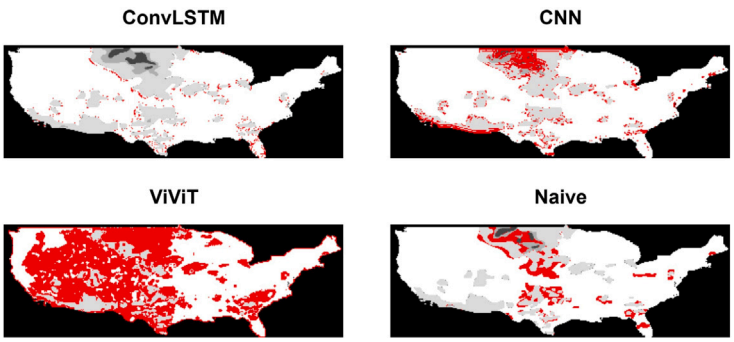


Fig. 11. Image $t+1$ difference comparison original and forecast visualized in red.

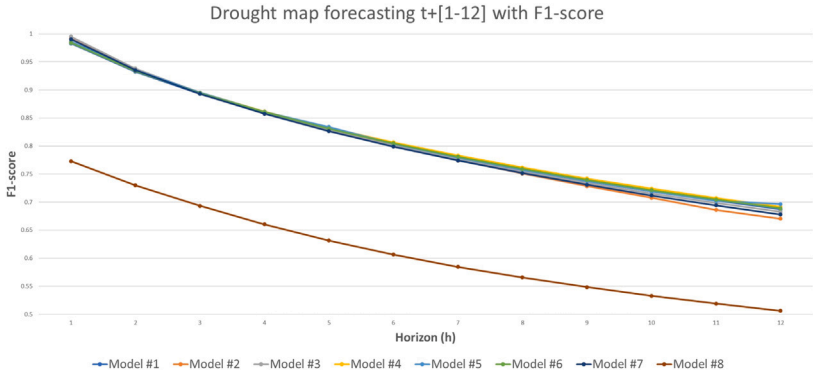


Fig. 12. Performance evaluation of the ConvLSTM at different forecasting horizon values.

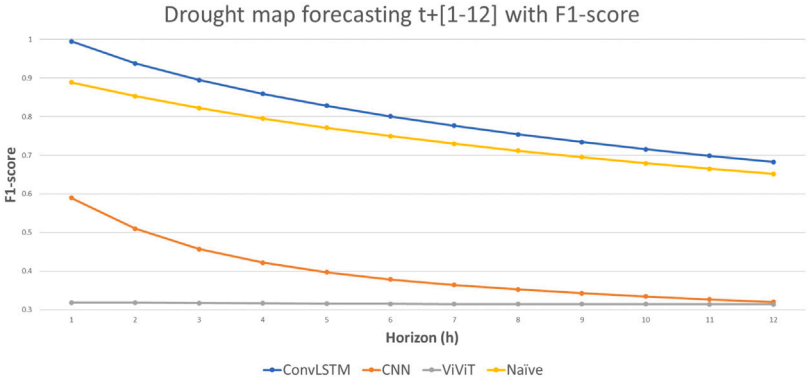


Fig. 13. Model comparative forecast from $t+[1-12]$ 3D visual results with macro F1-score.

and naïve models were used with the same horizon. The results are visualized in Fig. 13.

As can be seen, the proposed ConvLSTM-based model obtains the best results in generating a full horizon of 12 drought maps. This can be visually interpretable by generating the maps and comparing the errors. Fig. 14 shows the generation error of each model in the future drought map horizon from one to twelve steps ahead.

6. Discussion

The results show the differences between the proposed ConvLSTM-based architectures in Table 4. This shows that the proposed approach in Section 3.3 generates the following $t+1$ drought map with high accuracy. It can be demonstrated using the results in Fig. 9 that the architectural complexity of ConvLSTM is very important to generate maps with a high macro F1 score value of up to 0.9953. The architecture design also affects the precision of the model. Using a low-complexity model does not have enough inner-model connections to store the spatio-temporal relations from the data, affecting the quality of the generated map. A medium complexity model is the best option since it will require less memory and computational power to correctly solve the problem and generate a drought map with high precision. Even if using high-complexity models provides a high F1 score value, the extra computational memory, power, and time needed do not differ significantly from medium-complexity models.

However, these results show how the window size parameter affects the ConvLSTM models. As can be seen in Fig. 9, the most unstable value to use is $w = 11$, showing that the combination of any architecture with this value prevents the model from successfully extracting the spatiotemporal relations. However, the heat map (Fig. 9(a)) shows that in most cases the use of almost any combination with the proposed solution gives a high F1 score in the generation of the drought map. This means that it is not necessary to explore more complex model architectures with high window size values, since the problem can be solved in the specified region.

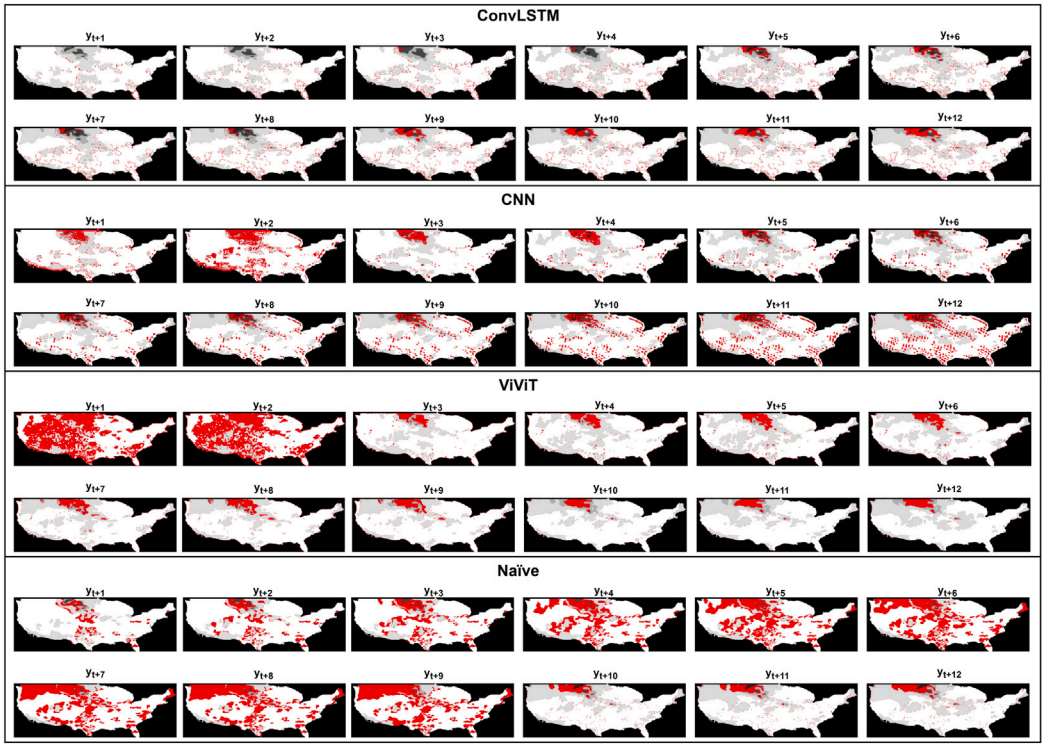


Fig. 14. Image differences comparison for ConvLSTM, CNN, ViViT and Naïve models results, difference between original and forecast visualized in red.

By using a recursive multi-step forecasting strategy, the horizon of $h = 12$ future drought maps could be generated. Most of the ConvLSTM-based models generated the forecast horizon with high accuracy. The results shown in Fig. 12 show that there is a constant cumulative error in generating the drought maps. This is to be expected due to the way the recursive strategy works. At the same time, it is shown how some models that generated drought maps with lower accuracy in the first steps become more stable as the forecast horizon progresses. The best macro F1-score value obtained reached 0.6965 when generating the complete horizon.

The ConvLSTM-based model was compared with the CNN, ViViT and Naïve models. Fig. 10 shows that only the ConvLSTM could outperform the Naïve model in generating the next $t + 1$ drought map. Also, comparing the generation of the map itself, Fig. 11 shows the visual differences obtained by each of the models. Thus, the accuracy of the ConvLSTM model in generating a drought map, and how the other models generate it, is demonstrated.

On the other hand, when comparing the models in the horizon generation, Fig. 13 shows how each of the models obtains less accuracy as the horizon increases, except for the ViViT model, which remains stable from the generation of the first map $t + 1$. By showing the visual errors in Fig. 14, it is shown how each of the models generates the forecast horizon. It can be seen that the proposed ConvLSTM-based model is the only one that outperforms the naïve model in generating the $t + [1 - 12]$ drought maps. An interesting observation, is that the ViViT model generates with low accuracy the $t + [1 - 2]$ maps, but the t+[3−12] time steps were generated with good accuracy, which explains the stability of the model in Fig. 13.

7. Conclusion

This work shows different topics to process drought image maps to successfully generate (i.e., forecast) accurate drought map images for a forecasting horizon h. The preprocessing applications were a key element for data cleaning and objective segmentation, resulting in cleaner and more accurate forecasted maps. The most important part of that process was applying the map masking, allowing the network to focus on the changing drought region.

The ConvLSTM2D-based model effectively captured spatiotemporal dependencies in drought maps. This allows the deep learning model to understand the inner changing patterns in data history to successfully generate a full future map with high accuracy. Using only ConvLSTM2D layers allows spatial extraction features through the convolution process, and temporal feature extraction through the LSTM behavior. Compared to other, more complex models, algorithms, and methods described in the related work section for spatiotemporal extraction and forecasting.

The presented methodology solves the problem by combining the classification processing techniques and the regression solving problem techniques. Due to the specifications of the given task, the drought map can be seen as a classification problem since it presents 7 different categories for the drought intensity. However, solving this problem with regression learning models gives a more accurate forecast and drought map generation, since the learning model generates a real value in the forecasting after it evaluates each drought map in history. By solving the problem using these techniques, the obtained drought map has a high accuracy and low error percentage.

The recursive multi-step forecasting method implemented in this study facilitated the generation of sequential drought forecasts for short-to-medium-term horizons. However, this strategy exhibited limitations like the progressive accumulation of errors, resulting in the amplification of uncertainties across forecast steps. To mitigate this effect, future investigations will explore alternative forecasting paradigms, including direct forecasting and MIMO strategies. While these alternative methodologies are anticipated to demand increased computational resources, they hold the potential to substantially reduce cumulative error propagation.

The presented methodology requires less complexity, since it relies on fundamental deep learning components while avoiding excessively deep or complex structures that may hinder interpretability and scalability. The design makes it highly adaptable to other study cases by adjusting the input data format, and fine-tuning hyperparameters.

Given that many environmental and geospatial datasets share similar patterns, the model can be easily transferred to other applications, such as flood, land cover or air quality forecasting with minimal modifications.

The actual work shows a comprehensive overview of the use of a ConvLSTM-based model for drought map forecasting. The mechanism can be enhanced in future work by implementing different statistical or artificial intelligence algorithms or methods that can properly extract inner spatiotemporal features of the ITS data. Including, the possibility of using newer technologies such attention mechanisms to enhance the drought map generation accuracy.

CRediT authorship contribution statement

Manuel Medrano-Diaz: Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Hector Rodriguez-Rangel:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Conceptualization. **Vicenç Puig-Cayuela:** Writing – review & editing, Supervision, Resources, Project administration. **Juan J. Flores:** Writing – review & editing, Methodology. **Rodrigo Lopez-Farias:** Writing – review & editing, Methodology. **Carlos Lara-Alvarez:** Writing – review & editing, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was conducted as part of the development of a PhD thesis. The authors would like to thank the *Secretaría de ciencia, humanidades, tecnología e innovación* (SECIHTI) that promotes and strengthens the development of science and innovation in the country and providing scholarships to PhD students. To the *Tecnológico Nacional de México, Instituto Tecnológico de Culiacán* for providing adequate facilities for academic development, especially to the division of Academic and Graduate Studies. To the *Universitat Politécnica de Catalunya* for the adequate facilities and hardware access that helps to have better development of the works.

Data availability

The dataset analyzed during the current study is available https://droughtmonitor.unl.edu/DmData/GISData.aspx.

References

Alquraish, M., Ali. Abuhasel, K., S. Alqahtani, A., Khadr, M., 2021. SPI-based hybrid hidden Markov-GA, ARIMA-GA, and ARIMA-GA-ANN models for meteorological drought forecasting. Sustain. 13 (22), 12576. http://dx.doi.org/10.3390/su132212576.

Bontempi, G., Ben Taieb, S., Le Borgne, Y.-A., 2013. Machine learning strategies for time series forecasting. Bus. Intell.: Second. Eur. Summer Sch. EBISS 2012, Bruss. Belg. July 15- 21, 2012, Tutor. Lect. 2 62–77.. http://dx.doi.org/10.1007/978-3-642-36318-4_3.

Brust, C., Kimball, J.S., Maneta, M.P., Jencso, K., Reichle, R.H., 2021. DroughtCast: A machine learning forecast of the United States drought monitor. Front. Big Data 4, 773478. http://dx.doi.org/10.3389/fdata.2021.773478.

Center, N.D.M., 2025. United States drought monitor. (Accessed 6 March 2025.) https://droughtmonitor.unl.edu/About/AbouttheData.aspx.

Chavez, R.O., Meseguer-Ruiz, O., Olea, M., Calderón-Seguel, M., Yager, K., Meneses, R.I., Lastra, J.A., Núñez Hidalgo, I., Sarricolea, P., Serrano-Notivoli, R., et al., 2023. Andean peatlands at risk? Spatiotemporal patterns of extreme ndvi anomalies, water extraction and drought severity in a large-scale mining area of atacama, northern Chile. Int. J. Appl. Earth Obs. Geoinf. 116, 103138.

Chehibi, M., Ferchichi, A., Farah, I.R., 2022. Representing and modeling spatio-temporal uncertainty using belief function theory in flood extent mapping. Expert Syst. Appl. 209, 118212. http://dx.doi.org/10.1016/j.eswa.2022.118212.

Chelali, M., Kurtz, C., Puissant, A., Vincent, N., 2021. Deep-STaR: Classification of image time series based on spatio-temporal representations. Comput. Vis. Image Underst. 208, 103221. http://dx.doi.org/10.1016/j.cviu.2021.103221.

Das, J., Saha, P., Mitra, R., Alam, A., Kamruzzaman, M., 2023. GIS-based data-driven bivariate statistical models for landslide susceptibility prediction in upper tista basin, India. Heliyon 9 (5).

Ferchichi, A., Chihaoui, M., Ferchichi, A., 2024. Spatio-temporal modeling of climate change impacts on drought forecast using generative adversarial network: A case study in Africa. Expert Syst. Appl. 238, 122211. http://dx.doi.org/10.1016/j.eswa.2023.122211.

Firozjaei, M.K., Sedighi, A., Firozjaei, H.K., Kiavarz, M., Homaee, M., Arsanjani, J.J., Makki, M., Naimi, B., Alavipanah, S.K., 2021. A historical and future impact assessment of mining activities on surface biophysical characteristics change: A remote sensing-based approach. Ecol. Indic. 122, 107264. http://dx.doi.org/10.1016/j.ecolind.2020.107264.

Geng, J., Cheng, C., Shen, S., Dai, K., Zhang, T., 2024. STAPLE: A land use/-cover change model concerning spatiotemporal dependency and properties related to landscape evolution. Environ. Model. Softw. 178, 106059.

Guillory, L., Pudmenzky, C., Nguyen-Huy, T., Cobon, D., Stone, R., 2023. A drought monitor for Australia. Environ. Model. Softw. 170, 105852.

He, Z., Xu, L., Yu, J., Wu, X., 2024. Dynamic multi-fusion spatio-temporal graph neural network for multivariate time series forecasting. Expert Syst. Appl. 241, 122729. http://dx.doi.org/10.1016/j.eswa.2023.122729.

Khan, M., Muhammad, N., El-Shafie, A., 2018. Wavelet-ANN versus ANN-Based model for hydrometeorological drought forecasting. Water 10 (8), http://dx.doi.org/10.3390/w10080998.

Lara-Alvarez, C., Flores, J.J., Rodriguez-Rangel, H., Lopez-Farias, R., 2024. A literature review on satellite image time series forecasting: Methods and applications for remote sensing. Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. e1528.. http://dx.doi.org/10.1002/widm.1528.

Little, T.D., 2013. The Oxford handbook of quantitative methods in psychology: Vol. 2: statistical analysis. Vol. 2, Oxford University Press, http://dx.doi.org/10.1093/oxfordhb/9780199934898.001.0001.

Ma, X., Zhang, J., Du, B., Ding, C., Sun, L., 2018. Parallel architecture of convolutional bi-directional LSTM neural networks for network-wide metro ridership prediction. IEEE Trans. Intell. Transp. Syst. 20 (6), 2278–2288. http://dx.doi.org/10.1109/TITS.2018.2867042.

Mafi-Gholami, D., Zenner, E.K., Jaafari, A., Bui, D.T., 2020. Spatially explicit predictions of changes in the extent of mangroves of Iran at the end of the 21st century. Estuar. Coast. Shelf Sci. 237, 106644. http://dx.doi.org/10.1016/j.ecss.2020.106644.

Marusov, A., Grabar, V., Maximov, Y., Sotiriadi, N., Bulkin, A., Zaytsev, A., 2024. Long-term drought prediction using deep neural networks based on geospatial weather data. Environ. Model. Softw. 179, 106127.

Mokarram, M., Pourghasemi, H.R., Hu, M., Zhang, H., 2021. Determining and forecasting drought susceptibility in southwestern Iran using multi-criteria decision-making (MCDM) coupled with CA-Markov model. Sci. Total Environ. 781, 146703. http://dx.doi.org/10.1016/j.scitotenv.2021.146703.

Mokhtar, A., Jalali, M., He, H., Al-Ansari, N., Elbeltagi, A., Alsafadi, K., Abdo, H.G., Sammen, S.S., Gyasi-Agyei, Y., Rodrigo-Comino, J., 2021. Estimation of SPEI meteorological drought using machine learning algorithms. IEEe Access 9, 65503–65523.

Oyarzabal, R.S., Santos, L.B., Cunningham, C., Broedel, E., de Lima, G.R., Cunha-Zeri, G., Peixoto, J.S., Anochi, J.A., Garcia, K., Costa, L.C., et al., 2025. Forecasting drought using machine learning: a systematic literature review. Nat. Hazards 1–29.

Ozan Evkaya, O., Sevinç Kurnaz, F., 2021. Forecasting drought using neural network approaches with transformed time series data. J. Appl. Stat. 48 (13–15), 2591–2606.

Rana, M.S., Mahanta, C., 2023. Spatial prediction of flash flood susceptible areas using novel ensemble of bivariate statistics and machine learning techniques for ungauged region. Nat. Hazards 115 (1), 947–969.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., Woo, W.-c., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Adv. Neural Inf. Process. Syst. 28, http://dx.doi.org/10.48550/arXiv.1506.04214.

Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., Rippey, B., Tinker, R., Palecki, M., Stooksbury, D., et al., 2002. The drought monitor. Bull. Am. Meteorol. Soc. 83 (8), 1181–1190. http://dx.doi.org/10.1175/1520-0477-83.8.1181.

Valipour, M., Khoshkam, H., Bateni, S.M., Jun, C., 2024. Machine-learning-based short-term forecasting of daily precipitation in different climate regions across the contiguous United States. Expert Syst. Appl. 238, 121907. http://dx.doi.org/10.1016/j.eswa.2023.121907.

Wu, R., Wen, X., Yuan, L., Xu, H., Liu, Y., 2024. Visual tracking based on deformable transformer and spatiotemporal information. Eng. Appl. Artif. Intell. 127, 107269.

Wu, X., Zhou, J., Yu, H., Liu, D., Xie, K., Chen, Y., Hu, J., Sun, H., Xing, F., 2021. The development of a hybrid wavelet-ARIMA-LSTM model for precipitation amounts and drought analysis. Atmos. 12 (1), 74. http://dx.doi.org/10.3390/atmos12010074.

Xiao, C., Chen, N., Hu, C., Wang, K., Xu, Z., Cai, Y., Xu, L., Chen, Z., Gong, J., 2019. A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data. Environ. Model. Softw. 120, 104502.

Xu, D., Zhang, Q., Ding, Y., Zhang, D., 2022. Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting. Environ. Sci. Pollut. Res. 29 (3), 4128–4144. http://dx.doi.org/10.1007/s11356-021-15325-z.

Xu, M., Zhang, D., Li, J., Wu, Y., 2024. An adaptive spatial–temporal prediction model for landslide displacement based on decomposition architecture. Eng. Appl. Artif. Intell. 137, 109215.

Yin, G., Ren, N., Mao, X., Wang, J., He, W., Zhang, H., 2025. Automatic generation of weekly large-scale drought monitor maps using deep learning and multiple open-access datasets. http://dx.doi.org/10.2139/ssrn.5134185, Available At SSRN 5134185.

Yoo, J., So, B.-J., Kwon, H.-H., Kim, T.-W., 2020. Development of drought map based on three-dimensional spatio-temporal analysis of drought. KSCE J. Civ. Environ. Eng. Res. 40 (1), 25–33. http://dx.doi.org/10.12652/Ksce.2020.40.1.0025.

Yu, M., Tao, B., Li, X., Liu, Z., Xiong, W., 2024. Local and long-range convolutional LSTM network: A novel multi-step wind speed prediction approach for modeling local and long-range spatial correlations based on ConvLSTM. Eng. Appl. Artif. Intell. 130, 107613.

Yuan, Y., Lin, L., Huo, L.-Z., Kong, Y.-L., Zhou, Z.-G., Wu, B., Jia, Y., 2020. Using an attention-based LSTM encoder-decoder network for near real-time disturbance detection. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 13, 1819–1832. http://dx.doi.org/10.1109/JSTARS.2020.2988324.

Zambrano, F., Vrieling, A., Nelson, A., Meroni, M., Tadesse, T., 2018. Prediction of drought-induced reduction of agricultural productivity in Chile from MODIS, rainfall estimates, and climate oscillation indices. Remote Sens. Environ. 219, 15–30. http://dx.doi.org/10.1016/j.rse.2018.10.006.

Zhang, Q., Miao, C., Gou, J., Zheng, H., 2023a. Spatiotemporal characteristics and forecasting of short-term meteorological drought in China. J. Hydrol. 624, 129924. http://dx.doi.org/10.1016/j.jhydrol.2023.129924.

Zhang, X., Wang, Y., Wei, L., Jiang, J., Lin, P., Liu, H., 2023b. Spatiotemporal networks for ENSO forecasting with LICOM3 and remote sensing data. Eng. Appl. Artif. Intell. 125, 106641.

Zhang, W., Wang, H., Zhang, F., 2024. Spatio-temporal Fourier enhanced heterogeneous graph learning for traffic forecasting. Expert Syst. Appl. 241, 122766. http://dx.doi.org/10.1016/j.eswa.2023.122766.

Zhang, G., Zhou, H., Wang, C., Xue, H., Wang, J., Wan, H., 2020. Forecasting time series albedo using NARnet based on EEMD decomposition. IEEE Trans. Geosci. Remote Sens. 58 (5), 3544–3557. http://dx.doi.org/10.1109/TGRS.2019.2958048.