

Article

Towards Data-Driven Weather Intelligence in Palestine: A Multi-Station Benchmark of Classical Machine Learning and Deep Learning Models

Mohammad Odeh and Ahmad Hasasneh * 

Department of Natural, Engineering and Technology Sciences, Faculty of Graduate Studies, Arab American University, Ramallah P600-699, Palestine; m.odeh28@student.aaup.edu

* Correspondence: ahmad.hasasneh@aaup.edu

Abstract

Precise weather forecasting plays a critical role in sectors such as agriculture, transport, energy management, and climate change adaptation, and machine learning and deep learning algorithms have been widely used for data-driven time series forecasting problems. In this work, we explore the application of machine learning and deep learning models for multi-weather variable forecasting in a dataset recorded over a period of ten years (2015–2025) for five weather stations in Palestine. The dataset comprises measurements for temperature, relative humidity, wind speed, precipitation, atmospheric pressure, and sunshine hours. To avoid the issue of temporal leakage, a chronological training, validation, and test set splitting approach was used in the evaluation experiments. The models used in this study include ARIMA, SARIMA, Random Forest, XGBoost, CNN, LSTM, GRU, ConvLSTM, CNN-GRU, and CNN-LSTM with station embeddings. Our experimental results indicate that the XGBoost model achieved the highest performance in predicting temperature and relative humidity ($R^2 = 0.953$ and $R^2 = 0.670$, respectively), while deep learning methods exhibited high accuracy across several weather features. The CNN-LSTM model was successfully able to learn temporal–spatial patterns via station embeddings, while recurrent neural networks performed impressively in forecasting sunshine hours and atmospheric pressure.

Keywords: weather forecasting; statistical; machine learning; deep learning; XGBoost; CNN-LSTM; Palestine; multi-variable prediction



Academic Editor: Olufemi A. Omitaomu

Received: 29 April 2026

Revised: 22 June 2026

Accepted: 22 June 2026

Published: 1 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

“Climate is what we expect; weather is what we get.” This quote, attributed to Mark Twain and associated with Heinlein [1,2], highlights the distinction between the terms “weather” and “climate.” Weather refers to the daily atmospheric conditions in a specific location, while climate describes the average weather conditions over a span of 30 years [3,4].

Forecasting weather conditions, which is a crucial task that significantly impacts people’s daily lives, has become increasingly important across various sectors, including agriculture, power generation, transportation, construction, and environmental pollution mitigation [5]. Decision-making in agriculture is informed by forecasts, influencing the timing for planting, harvesting, and irrigation to maximize crop yields and stabilize the food supply chain. In the energy sector, weather forecasting is used to optimize power generation and distribution and help manage fluctuations. In the transportation industry, aviation,

train, and marine schedules depend on weather forecasting, which plays a central role in mitigating delays and supporting safety protocols. The construction and infrastructure development sectors are also influenced by weather predictions, as bad weather can affect project quality and delay deadlines. Because of these factors, forecasting is a cornerstone in project management, as accurate weather forecasting keeps individuals safe during inclement weather and extreme temperatures, protects properties, and supports economic stability. The tourism and travel industries also depend on weather forecasting to ensure safety and allow timely precautions to be taken [6].

Although climate prediction has been utilized to support life on Earth, increasing levels of greenhouse gases like carbon dioxide have caused global warming and rising sea levels, creating challenges for the future of our planet.

Climate modeling and forecasting techniques can be utilized to develop mitigation strategies and gain significant understanding of the possible consequences of these phenomena. For instance, we can estimate sea level changes over the next decades to prevent disasters in coastal cities and inform urban planning in these cities.

Over extended periods of time, climate change can contribute to alterations in the geographical distribution of various species, threatening biodiversity. Advanced climate models incorporate various factors, including atmospheric conditions, oceanic currents, terrestrial ecosystems, and biospheric interactions, to provide a deep understanding of environmental transformations [7]. This comprehensive approach is essential for establishing global and regional policies to preserve ecological diversity. The economic sector is particularly vulnerable to the fluctuations caused by climate change: rising temperatures can lead to reductions in agricultural production and can negatively affect tourism. Longitudinal climate projections can be used by decision-makers in government and business to implement strategies for adapting to these changes.

Furthermore, long-term climate projections can contribute positively to managing and organizing sustainable resources such as water, forests, and land. Precise forecasting models can provide indicators for water scarcity in specific areas, thereby enabling the development of fair water management policies. Public health crises involving infectious diseases or heatwave incidents are an implication of climate change, and long-term climate models can support healthcare agencies in collecting the necessary data to allocate resources and develop response strategies. According to the definitions of weather and climate, forecasting applications can be divided into the following categories:

1. Weather prediction applications: These applications usually forecast weather conditions from hours up to seven days in advance, supporting humans with precise and sufficient information and allowing them to decide whether to wear a coat or delay outdoor exercise, for example. These forecasts become lower in accuracy when predicting further into the future. The essential elements of these forecasts include temperatures, sunshine, wind speed, humidity, precipitation, and atmospheric pressure. Because of the significance of weather forecasting, meteorologists have developed models and simulations using supercomputers and clusters in the form of High-Performance Computing (HPC) to create advanced and sophisticated weather forecasting models [8,9]. The observational input data for these models come from multiple sources, such as satellites, ground stations, and statistical techniques that can recognize patterns and calculate probabilities based on historical weather data. These weather predictions, which are more reliable than climate predictions, are influenced by the quality of input data, the accuracy of numerical models, and the initial atmospheric conditions. Weather predictions play a central role in many sectors, as mentioned previously, including decision-making circuits, transportation, and agriculture.

2. Climate prediction applications: These applications predict climate conditions at a time range extending from months to multiple years, in contrast with weather forecasts,

which concentrate on current atmospheric conditions. Climate predictions generally focus on extreme weather events like droughts, floods, and hurricanes, and encompass broader climate trends, such as global warming or localized climatic shifts. These projections use methods such as statistical models, which depend on historical data and seasonal patterns, dynamic models, which use complex mathematical equations rooted in physics, and integrated models, which combine many data sources and methodologies. The precision of climate projections is lower than that of weather forecasting due to the sophisticated, multi-scale, and multi-process interactions that climate systems include, in addition to the general lack of long-term weather forecast data. Economic variables, human activities, and changes in policy are also factors that affect the precision of climate projections. However, climate projections are essential in fields such as resource management, agricultural organization, catastrophe response, energy policy planning, and climate research and science.

Previous forecasting methods have primarily relied on statistical and physical models that can be complex and challenging to use, especially when dealing with nonlinear and time-inconsistent data. Although these tools are often costly and intricate, forecasts frequently prove inaccurate due to flawed initial observations or a lack of understanding of atmospheric processes, and deciphering such complicated models typically requires a significant amount of time [10]. Because of this, there is a need for more advanced methods and techniques [11], including machine learning and deep learning algorithms, which have shown effectiveness in weather prediction by adeptly capturing spatial features and temporal dependencies [8].

Recent breakthroughs in artificial intelligence, machine learning, and deep learning techniques have been used to greatly improve weather forecasting skills. Classical machine learning techniques such as Random Forest and XGBoost have shown excellent predictive ability by accurately capturing nonlinear interactions among meteorological variables while maintaining a relatively modest computational complexity. Simultaneously, deep learning architectures such as Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRUs), Convolutional Neural Networks (CNNs), and ConvLSTM models have been widely used to model complicated temporal and spatio-temporal correlations in meteorological data [12–17]. Hybrid designs such as CNN-LSTM integrate the feature extraction capacity of CNNs with the sequence learning strength of recurrent neural networks to better capture meteorological trends [12–18]. Even with these gains, it is still unclear whether deep learning models can consistently outperform state-of-the-art machine learning algorithms in predicting multiple weather variables and situations [18–20]. A detailed evaluation of statistical, machine learning, and deep learning models is needed to determine the most effective forecasting method for multi-variable weather prediction. In this work, we test ARIMA, SARIMA, Random Forest, XGBoost, CNN, LSTM, GRU, ConvLSTM, CNN-GRU, and CNN-LSTM models on a multi-station meteorological dataset obtained from Palestine.

In this paper, a detailed comparison framework for multi-variable weather forecasting using statistical, machine learning, and deep learning methodologies is presented. Our research is based on a ten-year meteorological dataset, obtained from five weather stations distributed across Palestine, including recordings of temperature, rainfall, sunshine duration, atmospheric pressure, relative humidity, and wind speed. The deep learning architecture uses station-level information by embedding representations to capture temporal and spatial variability, whereas machine learning models use artificial meteorological and temporal features. The forecasting methods that have been assessed are ARIMA, SARIMA, Random Forest, XGBoost, CNN, LSTM, GRU, ConvLSTM, and CNN-GRU. The main contribution of this study is the development of a reproducible benchmarking framework employing a previously unused Palestinian meteorological dataset. Moreover, we offer a

systematic comparison between classical machine learning and deep learning paradigms, providing a better understanding of the pros and cons of each for the prediction of diverse meteorological variables. Our findings help to improve prediction, support meteorological decision-making, and provide practical assistance for the selection of appropriate forecasting models for regional climate applications.

The rest of the paper is organized as follows: Section 2 provides a literature review of the existing approaches for weather forecasting. Section 3 describes the proposed model and methodology. Section 4 details the experimental setup and results, which are further discussed in Section 5. Finally, Section 6 concludes the paper and outlines future research directions.

2. Literature Review

Throughout the literature, numerous studies have investigated weather forecasting using different techniques and methods. While some studies have focused on classical machine learning (ML) methods [18,19], others have adopted deep learning approaches [17,20]. More recently, many studies have explored combined or hybrid models for improving forecasting accuracy and capturing complex temporal and nonlinear weather patterns [3,5,11,15,21–26]. A systematic summary of the existing weather forecasting studies is presented in Table 1.

Classical ML methods began when McCulloch and Pitts came up with the first neural network (NN) in 1943 [27]. The field grew until the early 1960s, when the initial algorithms were found to be slow and unreliable. A second wave of ML applications followed the advent of backpropagation in 1970 [28,29], which made it possible to build bigger NNs and teach them to find nonlinear relationships in data (e.g., [30]). Even though work on ML algorithms persisted, people quickly lost interest in them again because they did not always demonstrate a substantial speed boost and there were not enough computing resources to handle larger issues. Also, it was challenging to find the necessary amount of labeled data, which is needed for most data-driven ML methods (this was before the internet).

Three important things happened around 2010 that started the third wave of artificial intelligence, which is still going on today: graphical processing units (GPUs) made computing power much greater by allowing for massive parallel processing; Convolutional Neural Networks (CNNs) made it possible to analyze huge (image) datasets much more quickly; and large benchmark datasets became available on the internet. Highly complicated NNs with more than 106 parameters made a considerable difference in image recognition [31] and, soon after that, voice recognition [32], gaming [33,34], and video analysis and prediction [35,36].

Modern advances in meteorological prediction and time series forecasting, propelled by machine learning and especially deep learning, have significantly enhanced the accuracy and efficacy of forecasting models by automating the prediction process. Atmospheric scientists have identified supervised machine learning as the most promising methodological category, as it is predominantly addressed in recent articles in the field. If labeled data is available, it can be utilized as a training dataset to develop a function that associates specific inputs with corresponding outputs.

This function can then be applied to an alternative dataset, referred to as the testing dataset, to assess the model's performance. If the outcomes are satisfactory, it may be utilized for classification or regression in any required application. This method includes approaches such as decision trees, including Random Forest (RF) [37], XGBoost (XGB) [38], artificial neural networks (ANNs) [18], deep learning (DL) [19], and support vector machines (SVMs) [39]. The second category of machine learning is unsupervised learning, wherein algorithms lack labeled data for training and must determine alternative methods to partition a dataset or minimize its dimensions for subsequent analysis. Within this cate-

gory, the prevalent techniques utilized by atmospheric scientists are K-means Clustering (K-means) [20] and Principal Component Analysis (PCA) [40].

Some papers have used supervised machine learning regression algorithm models to predict and forecast weather conditions rather than traditional physical and statistical models, as in [21,22]. In [21], the authors outline a way of predicting the highest and lowest temperatures for the following seven days using data from the last two days, employing both a linear regression model and a modified version of a functional linear regression model. Their results revealed that the improved professional weather forecasting service algorithm introduced in this paper did a better job of predicting the weather for up to seven days than either model, and their model was better at predicting the weather for days or weeks into the future. Ref. [22] introduced an application that used machine learning support vector machines (SVMs) to predict weather conditions as a classification issue, which showed a better performance than traditional statistical models. The nonlinear regression method was more suitable for training the SVMs, and their performance was better compared with a Multi-Layer Perceptron (MLP) trained with a backpropagation algorithm.

An unsupervised machine learning predictive algorithm based on data mining was suggested by [23], with a main objective of determining changing trends in meteorological conditions. The algorithm utilized trends from past weather data to predict what the weather would be like in the future, employing a Hidden Markov Model to make predictions and K-means Clustering to obtain observations from weather conditions. The results indicated that the performance of the proposed model was enhanced when compared to a traditional Iterative Dichotomiser 3 (ID3) machine learning algorithm.

On the other hand, deep learning (DL) is a branch of machine learning that creates deep networks with more than 10 layers, along with methods for training and validating them. A Convolutional Neural Network (CNN) is a deep learning algorithm that emerged in 2017, which has been used to perceive weather conditions by successfully extracting spatial information from meteorological data and forecasting complex weather patterns.

For instance, Ref. [24] applied a multitask CNN model to extract features from weather time series data collected at multiple sites in order to predict amounts of rainfall based on multi-site features.

In 2018, the strength of Stacked Long Short-Term Memory (Stacked-LSTM) was demonstrated for modeling sequential data, including temporal dependencies crucial for weather prediction [25]. Studies have revealed that LSTM greatly increases prediction accuracy in temperature and precipitation forecasting by successfully addressing long-term relationships in time series data [17].

Hybrid or combined models have been widely used in various methods of weather condition forecasting. In [26], the authors proposed a hybrid numerical model that uses neural networks to describe the mechanics of weather forecasting. Furthermore, the authors of [41] predicted the weather by combining a deep neural network model and statistics from a group of weather-related variables with discriminatively trained predictive models. Another way of using hybrid or combined models in weather forecasting is shown in [42], which employed the idea of crowdsensing, which is when people share smartphone data to learn about environmental events. The authors developed an architecture called SenSquare, which collects data from IoT sources and crowdsensing platforms and shows it to subscribers in one place. While smart city environment monitoring uses this data, none of these works leverage the idea of assembling data from nearby sites.

The combination of CNN and LSTM in a hybrid model has demonstrated considerable potential in meteorological prediction. For instance, Ref. [5] effectively employed a CNN-LSTM model for historical temperature forecasting that exhibited superior performance

by addressing issues such as data incompleteness and high dimensionality. Ref. [43] similarly presented the Seoul Transportation Operation and Information Service (TOPIS), a hybrid model integrating CNN, LSTM, and Transposed CNN for forecasting network-wide congestion levels. This model proficiently extracts spatial and temporal data from images, exhibiting enhanced precision, recall, and accuracy over many prediction horizons. The PredNet model additionally demonstrated performance enhancement ranging from 2% to 12% in road-specific predictions.

Ref. [44] developed a CNN-LSTM model for forecasting the world annual average temperature over the next two decades, utilizing global temperature data from 1880 to 2022. Their findings underscored the model's advantages, as CNN diminishes data dimensionality and LSTM captures long-term connections, improving the prediction accuracy for extensive temperature datasets.

The authors in [16] integrated CNN and LSTM to anticipate PM2.5 concentration, illustrating that their CNN-LSTM model (APNet) surpasses existing machine learning techniques in accuracy by utilizing historical data, including rainfall, wind speed, and PM2.5 concentrations, for predictions. Additionally, Ref. [11] employed a CNN-LSTM hybrid model to forecast temperature in Delhi, demonstrating enhanced accuracy and stability relative to conventional approaches. Ref. [15] further substantiated the CNN-LSTM methodology for hourly temperature forecasting, demonstrating enhanced accuracy, particularly for extended time series datasets. Finally, Ref. [45] illustrated the efficacy of CNN-LSTM models for monthly climate prediction, markedly surpassing the performance of conventional ANN and RNN models, hence affirming the hybrid model's proficiency in managing intricate climatic data. In summary, current research has continually underscored the benefits of CNN-LSTM hybrid models for weather prediction tasks, demonstrating their significant enhancements compared to traditional and single-model methodologies [13,46,47].

Table 1. Summary of related studies on weather forecasting techniques.

Reference	Year	Study	Dataset	Model	Contribution	Research Gap
[5]	2024	Gong et al.	Historical temperature data	CNN-LSTM	Hybrid model for temperature forecasting	No spatial information; just temperature
[11]	2024	Li B et al.	Delhi, India	CNN-LSTM	Weather prediction (time series)	No spatial model
[13]	2025	Melnyk et al.	Odesa, Ukraine	CNN-LSTM	Time series analysis and climate forecasting	Limited spatial generalization
[15]	2022	Hou et al.	Hourly temperature data	CNN-LSTM	Temperature prediction for air	One variable prediction
[17]	2024	Shen et al.	Eastern China	CNN-LSTM multi-scale attention	Improved temperature forecasting accuracy	Focused on temperature indication only
[24]	2017	Qiu et al.	Rainfall data	CNN multitasking	Short-term rainfall prediction	Absence of temporal sequence modeling capacity
[25]	2018	Karevan & Suykens	Weather forecasting	LSTM (spatial-temporal)	Using spatial-temporal learning in temperature forecasting	CNN feature extraction does not exist
[41]	2015	Grover et al.	Weather data	Deep hybrid model, joint with statistics	Hybrid deep learning for weather forecasting	Regional variability limited discussion
[45]	2024	Guo et al.	Climate dataset	CNN-LSTM	Monthly climate prediction	Absence of spatial representation at the station level
[46]	2025	Wang et al.	Global weather data	Deep learning	Large-scale weather forecasting review	General framework without regional validation
[47]	2026	Kabir & Chakma	Bangladesh	HTC-LSTM-Attn	Multi-target weather forecasting with attention	Limited spatial representation

Our research extends these findings by introducing a strong CNN-LSTM hybrid model specifically designed for temperature and other weather component forecasting, overcoming the critical shortcomings noted in prior studies. Studies [21,22] both used supervised machine learning algorithms: The first one forecasted the highest and lowest temperatures for seven days depending on the past two days, which is a short time for prediction, and the authors did not show if their own improved professional weather forecasting service algorithm could do a better job of predicting the weather days or weeks into the future. The second paper used an SVM algorithm for classification without any predictions and compared its results with an MLP. The case in [23] was not better due to the use of a traditional statistical Hidden Markov Model for prediction and K-means for observing weather conditions, and their final results were compared to a third algorithm to check performance (ID13). Some studies used deep learning algorithms separately to predict weather conditions, as in [24,25], while other studies [26,41,43] used numerical and statistical models combined with neural networks to formulate a combined model to predict weather conditions, which suffered from slowness and inaccuracy. The idea of integrating a CNN and LSTM was introduced in studies such as [5,11,15,43,45], which used hybrid models to predict historical temperature data. Ref. [5] concentrated on sectors like agriculture, energy management, and urban planning, while [43] tried to improve prediction performance. In [11,15,44], the authors worked to enhance the accuracy of temperature prediction, while [16] predicted PM 2.5, rainfall, and wind speed to forecast climate change. Ref. [47] focused on the prediction of monthly climate change. These previous studies used old datasets or unknown sources, making their results less accurate and reliable. The main contributions of this research study are as follows:

- **Benchmarking framework:** A comprehensive benchmarking framework is employed to conduct a systematic comparison of the performance of statistical models, classical machine learning algorithms, and deep learning techniques on multi-variable weather forecasting tasks. The models considered in this research include ARIMA, SARIMA, Random Forest, XGBoost, CNN, LSTM, GRU, ConvLSTM, CNN-GRU, and CNN-LSTM models.
- **Multi-station forecasting with Palestinian meteorological data:** Our novel approach was tested on a real-life time series dataset that was generated by measuring meteorological features at five different stations located throughout Palestine: Jericho, Jenin, Ramallah, Nablus, and Hebron. Thus, this work is among the few attempts to provide an extensive benchmarking study for multi-station weather forecasting in Palestine.
- **CNN-LSTM model enhanced by station embedding:** Our research not only involves the benchmarking of various statistical and machine learning models but also evaluates a CNN-LSTM model that utilizes the concept of station embeddings for predicting multi-variable time series.
- **Practical evaluation for decision-making support:** In our research, we present a comprehensive comparative evaluation to determine the capabilities and limitations of various weather forecasting paradigms. The practical results provided by this study can be used in different sectors such as meteorology services, agriculture, renewable energy management, transportation, and urbanization.

Therefore, the main novelty of this research paper lies in its use of recently collected Palestinian meteorological data, its consideration of spatial station-level information by embedding the spatial data into the convolutional Long Short-Term Memory model, and its comparison of traditional statistical, machine learning, and deep learning algorithms in a reproducible way under identical conditions.

3. Methodology

Developing a prediction system for weather conditions primarily consists of several main steps, including data acquisition, data preprocessing and preparation, feature extraction, and deep learning and prediction model development, as shown in Figure 1.

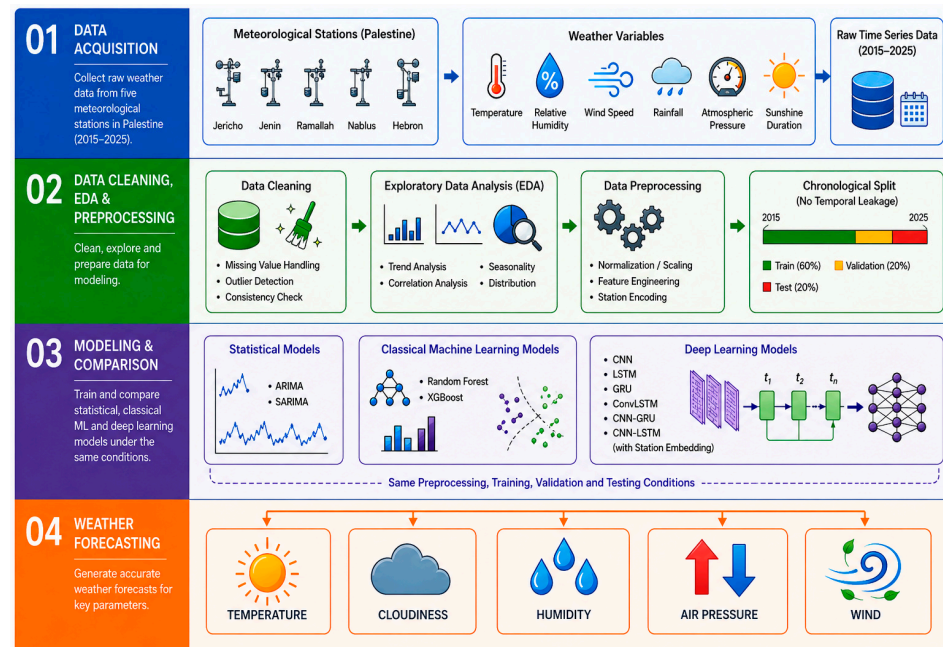


Figure 1. Overview of the proposed multi-station weather forecasting methodology, including data acquisition from Palestinian meteorological stations, data cleaning, exploratory data analysis, preprocessing, comparative modeling using statistical, classical machine learning and deep learning models, and final forecasting of key weather variables.

3.1. Data Acquisition

The dataset used in this study comprises daily main weather elements from Palestine, covering the years 2015 to 2025, including data on temperature, average rain, average sunshine, atmospheric pressure, and relative humidity. The dataset is a first-hand dataset authorized by the Palestinian meteorological authority, collected from the five stations belonging to the Palestinian meteorological authority that exist in five governorates in Palestine (Nablus, Jenin, Ramallah, Jericho, and Hebron). The dataset contains eight columns with 16,769 rows.

Station Coordinates

The station identifier used by the Palestinian meteorological authority, the exact geographical coordinates, the station elevation, and the WMO index number of each station are provided on an official page of the Palestinian meteorological authority [48]. Table 2 includes a detailed description of the five stations in Palestine.

Table 2. Station code, city, WMO, longitude, latitude, and elevation.

Station Code	City	Type/Index No. WMO.	Longitude	Latitude	Elevation (m)
HEB00008	Hebron	Synp 40240	35°06 E	31°32 N	1005 m
JEN00001	Jenin	Synp 40226	35°18 E	32°28 N	178 m
JER00005	Jericho	Synp 40247	35°27 E	31°51 N	−260 m
NAB00003	Nablus	Synp 40223	35°15 E	32°13 N	570 m
RAM00004	Ramallah	Synp	35°13 E	31°53 N	856 m

3.2. Data Cleaning, Exploration Data and Analysis (EDA), and Preprocessing

Each feature in the dataset that was used in this research has 16,769 rows in total. Rain and sunshine have fewer rows due to differences between seasons (summer, winter, spring, and autumn), as in the summer it rarely rains in Palestine.

3.2.1. Variables Units

The raw dataset contains six variables: air temperature, relative humidity, wind speed, rainfall, atmospheric pressure, and sunshine duration. Table 3 describes these variables and their units.

Table 3. Description of variables and their units.

Variable	Description	Unit
Temp	Air Temperature	°C
RH	Relative Humidity	%
Wind	Wind Speed	m/s
Rain	Rainfall	Mm
Press	Atmospheric Pressure	hPa
SUNSHINE	Sunshine Duration	Hours/day

3.2.2. Missing Data Ratio for the Meteorological Variables and Preprocessing Methods

The Palestinian meteorological dataset contains missing values. Table 4 summarizes the dataset's variables, the number of records for each variable, the number of missing values, the missing ratios, and the preprocessing methods (imputation strategies) applied for each variable.

Table 4. Statistics and corresponding imputation techniques for missing data.

Variable	Records	Missing Values	Missing Ratio (%)	Imputation Method
Temperature (Temp)	16,768	0	0.00	Linear Interpolation (Not Required)
Relative Humidity (RH)	16,794	1	0.0015	Linear Interpolation
Wind Speed	16,794	1	0.0012	Linear Interpolation
Atmospheric Pressure (PRESS)	15,976	4	0.0063	Linear Interpolation
Sunshine Duration (SUNSHINE)	12,067	23	0.0381	Monthly Climatology + Interpolation
Rainfall	2928	937	64,003	Linear Interpolation/Event Handling

Missing and incorrect observations were handled using variable-specific preprocessing procedures. While true zero rainfall data were left unchanged, as they indicate valid no-rain days, missing rainfall values were handled as a separate case from actual zero observations. Missing values of continuous variables including temperature, relative humidity, wind speed, and atmospheric pressure were adjusted using station-wise linear interpolation [49]. Zero pressure values that were physically invalid were replaced prior to interpolation. For sunshine duration, valid zero values were kept for cloudy or no-sunshine conditions. The long series of zero sunshine values at the Nablus station was treated as a missing measurement problem and replaced by monthly climatological sunshine averages calculated from the remaining stations. After preprocessing, the combined dataset had no missing values. The preprocessing algorithm (1) loads records from five weather stations; (2) parses date information; (3) handles missing values using linear interpolation; (4) encodes station IDs; (5) generates temporal

features (month, week, day of week); (6) applies Min–Max normalization; (7) creates 30-day sequences; and (8) splits data chronologically into training and test sets.

3.2.3. Distribution and Trending of Meteorological Features (Variables)

The distributions and yearly trends of the meteorological variables are presented in Figures 2 and 3. Figure 2 illustrates the distributions of six key meteorological variables, revealing diverse statistical behaviors across features. Temperature follows an approximately normal (Gaussian) distribution centered around moderate values (about 20–25 °C), indicating relatively stable conditions with limited variability. Similarly, atmospheric pressure exhibits a near-normal distribution tightly clustered around 950–1000 hPa, reflecting atmospheric stability. In contrast, relative humidity shows a broader and slightly multimodal distribution, with values spread across mid-to-high ranges (50–90%), suggesting variability influenced by seasonal or regional factors. Wind speed is highly right-skewed, with most observations concentrated at low speeds (0–4 m/s) and a long tail representing occasional stronger winds. Rainfall displays an extremely right-skewed distribution dominated by zero or near-zero values, indicating that dry conditions are far more frequent than precipitation events. Sunshine duration exhibits a bimodal and right-skewed pattern, with a significant peak at zero (cloudy conditions) and another at higher values (approximately 8–12 h), reflecting alternating clear and overcast days. For these heterogeneous and nonlinear data, CNN-LSTM is used to effectively capture both temporal variability and underlying patterns.

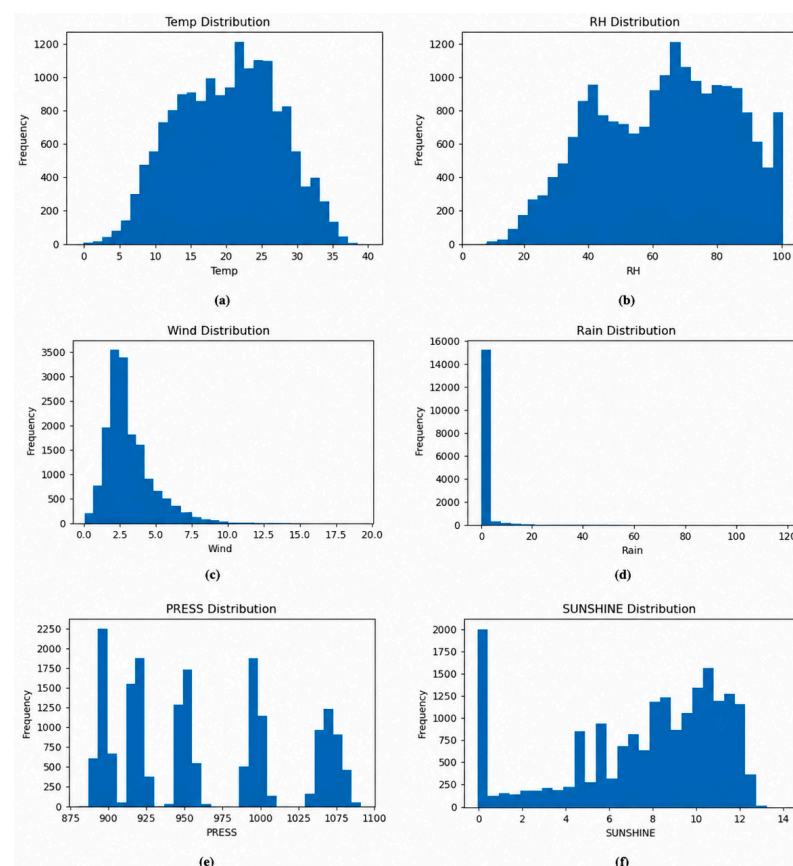


Figure 2. Distribution of the key meteorological variables used in this study: (a) temperature, (b) relative humidity, (c) wind speed, (d) rainfall, (e) atmospheric pressure, and (f) sunshine duration. The distributions show that temperature and atmospheric pressure exhibit relatively stable patterns, whereas wind speed and rainfall are highly skewed. Relative humidity and sunshine duration exhibit irregular and multimodal patterns, reflecting the variability and intermittency of atmospheric conditions.

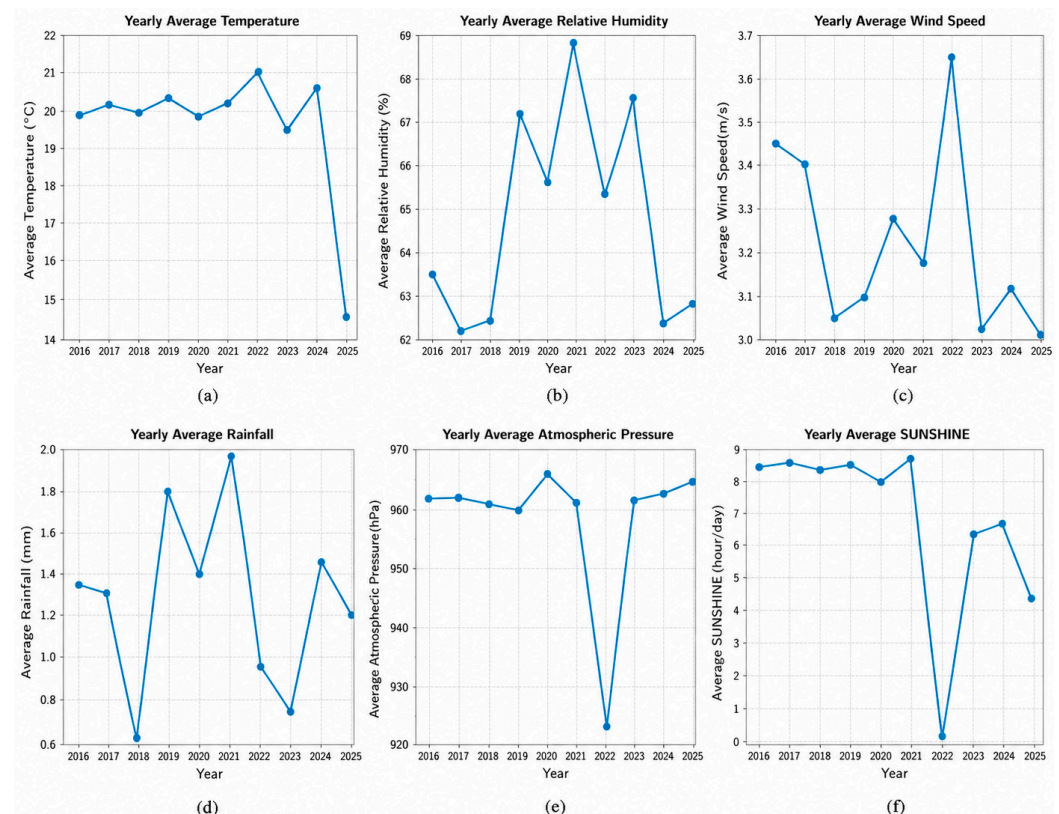


Figure 3. Yearly average trends of the meteorological variables used in this study from 2016 to 2025: (a) temperature, (b) relative humidity, (c) wind speed, (d) rainfall, (e) atmospheric pressure, and (f) sunshine duration. Temperature and atmospheric pressure remained relatively stable, while the other variables showed greater interannual variability, with a notable anomaly observed around 2022.

Figure 3 presents the yearly average trends of key meteorological variables over the study period, highlighting both stability and variability across features. Temperature remains relatively stable, with minor fluctuations around 19–20 °C for most years, followed by a noticeable drop in the final year. Relative humidity exhibits moderate variability, with peaks around 2020 and 2022 and lower values in earlier and later years, indicating changing atmospheric moisture conditions. Wind speed shows a fluctuating pattern, with a general increase toward 2021–2022, followed by a decline, reflecting variations in atmospheric dynamics. Rainfall demonstrates irregular variability, with notable peaks around 2018 and 2020 and lower values in drier years, emphasizing the intermittent nature of precipitation. Atmospheric pressure remains largely stable around standard values (950–100 hPa), except for a significant anomaly observed in 2022, which may indicate data irregularity or an extreme event. Sunshine duration is relatively consistent (around 6–7 h) for most years but shows a sharp drop in 2022, followed by a partial recovery, suggesting unusual climate conditions during that period. Overall, these trends indicate that, while some variables (temperature and pressure) maintain stability, others (rainfall, wind, humidity, and sunshine) exhibit notable interannual variability, reinforcing the need for robust temporal modeling approaches.

3.2.4. Chronological Splitting Approach

A chronological partitioning method was used to split the data and avoid time leakage. Since weather prediction is a time series forecast problem, random partitioning would introduce data from future periods into the training period, leading to overly optimistic results from the model. Hence, in the first step, the observations were arranged according to the station and date. As shown in Table 5, the period between 2015 and 2022 was used to

train the models, while the observations from 2023 served as the validation data for tuning the hyperparameters. The observations from 2024 were used solely for testing purposes.

Table 5. Chronological splitting approach summary.

Subset	Period	Purpose
Training	2015–2022	Model learning
Validation	2023	Hyperparameter tuning, early stopping
Testing	2024	Final future forecasting evaluation

3.2.5. Feature Engineering

Several feature engineering techniques were applied to the meteorological dataset to improve forecasting performance. First, the observations for each weather station were reordered chronologically and converted into fixed-length input sequences by applying a sliding window technique. We chose a sequence length of 30 consecutive days, which allowed the models to learn the temporal relationships from past data while forecasting the weather conditions for the next day.

Besides the six meteorological variables (temperature, relative humidity, wind speed, rainfall, atmospheric pressure, and sunshine duration), three temporal elements were extracted from the observation date: month, week of the year, and day of the week. These calendar-based variables provide seasonal and periodic information that enables the forecasting models to capture repeating weather parameters.

Each station identity was stored as a categorical characteristic to capture spatial diversity across the five meteorological stations. For the CNN-LSTM design, the encoded station IDs were also mapped to dense embedding vectors using a trainable embedding layer, which allows the network to learn latent station-specific properties. Each input sample was a feature matrix of size 30×9 , where 30 is the length and 9 represents the six meteorological variables and three temporal features.

3.2.6. Dataset Reproducibility

Table 6 summarizes the weather dataset characteristics and preprocessing procedures used in this study. It includes the data source, time period, monitored stations, meteorological variables, missing value handling strategy, normalization method, and sequence length used for developing the deep learning models.

Table 6. Weather data characteristics and preparatory method for CNN-LSTM.

Attribute	Description
Data Source	Palestinian Metrological Authority
Converge Period	January 2015–April 2025
Number of Stations	5
Regions	Jenin, Nablus, Ramallah, Jericho, Hebron
Temporal Resolution	Daily
Total records	16,769
Variable	Temperature, Relative Humidity, Wind, Rain, Pressure, Sunshine
Missing Values	Linear Interpolation
Scaling Method	Min–Max Normalization
Sequence Length	30 Days
Train/Validation/Test Split	Chronological

3.3. Forecasting Models

In this study, we compare statistical, classical machine learning, and deep learning models under identical settings of preprocessing, training, and evaluation to provide a comprehensive assessment of different forecasting methodologies.

3.3.1. Persistence Model

A persistence forecasting model was incorporated into the experimental evaluation to provide a basic and meaningful benchmark. Persistence implies that the forecast for the next time step is simply the observed value. This means that the weather conditions at time $(t + 1)$ are predicted from the observations at time (t) . Persistence forecasting is commonly used as a baseline in meteorology since many weather variables (e.g., temperature, atmospheric pressure) have substantial temporal autocorrelation, although it is a simple technique. This baseline is used as a reference to compare advanced forecasting models to assess whether they add any predictive value over and above merely repeating the most recent observation [50].

3.3.2. Statistical Models

To establish a baseline comparison, two frequently used statistical forecasting approaches were assessed, namely the Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) methods. ARIMA models are suitable for modeling linear dependencies across time in stationary time series, with autoregressive and moving average components, while SARIMA models enhance the ARIMA model by including seasonal effects, which are often observed in meteorological data. These models serve as standard forecasting benchmarks against which more advanced machine learning and deep learning approaches are measured [51,52].

3.3.3. Classical Machine Learning Models

Two types of ensemble-based machine learning methods are employed by meteorologists to forecast the weather: Random Forest and XGBoost. Random Forest creates multiple decision trees using bootstrap aggregation and averages the results to promote generalization and avoid overfitting. XGBoost is a gradient boosting approach that creates decision trees iteratively to decrease mistakes in prediction, with a regularization technique to improve performance. Both models can learn nonlinear correlations between meteorological variables and have achieved good performance in environmental prediction tasks [53,54].

3.3.4. Deep Learning Models

In order to explore the usefulness of neural network architectures for weather forecasting, multiple deep learning models were investigated, including CNN, LSTM, GRU, ConvLSTM, CNN-GRU, and CNN-LSTM. CNN models are good at capturing local feature patterns, while the recurrent architecture of such methods as LSTM and GRU is meant to capture long-term temporal dependencies in sequential data. ConvLSTM expands upon this ability by including convolution operations in the recurrent units, allowing the model to capture spatio-temporal correlations. CNN-GRU and CNN-LSTM are hybrid architectures that integrate convolution feature extraction with recurrent sequence learning, which can better depict complicated meteorological dynamics. In the proposed CNN-LSTM architecture, station embeddings are further integrated to capture station-specific spatial features and boost multi-station forecasting performance [12–17,55–57].

3.3.5. Hybrid Deep Learning Model

The proposed model combines the advantages of one-dimensional Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) networks to forecast future temperatures and additional elements. The architecture is depicted in Table 7, and a high-level architecture view of the hybrid deep learning model can be seen in Figures 4 and 5 [49]. The long-term interdependence [58] in the time series data hence improves the prediction accuracy.

Table 7. Architecture summary of the proposed CNN-LSTM model with station embedding for multi-variable weather forecasting.

Layer	Function	Output Shape
Seq_input	Sequential input (30 timesteps, 9 features)	(None, 30, 1)
Station_input	Single value (Station ID)	(None, 1)
Conv1D	Temporal feature extraction using 64 filters	(None, 28, 64)
LSTM	Learns temporal dependencies	(None, 64)
Embedding	Encodes station IDs into dense vectors	(None, 1, 4)
Global_Max_Pooling	Aggregates embeddings	(None, 4)
Concatenate	Combines CNN and station features	(None, 68)
Output	Dense output layer for predictions	(None, 5)

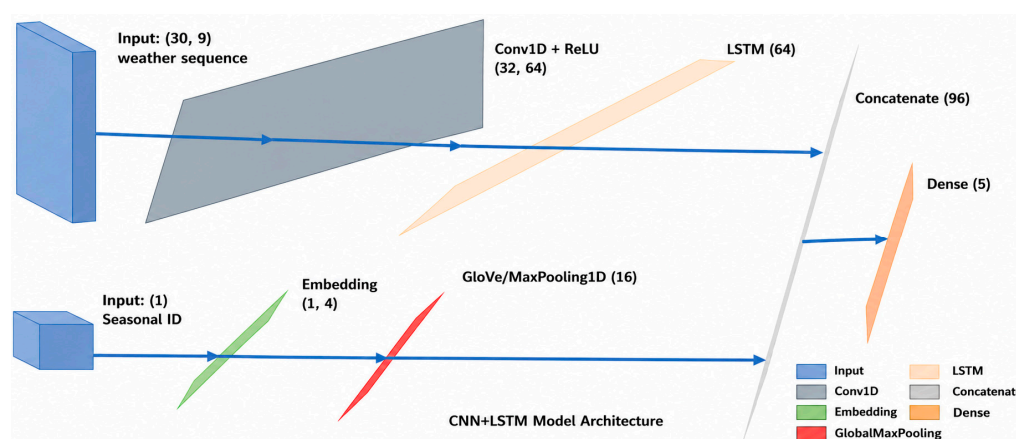


Figure 4. Architecture of the proposed spatial-temporal CNN-LSTM model for weather prediction. The model integrates temporal weather sequences through a Conv1D-LSTM pipeline and spatial station information via an embedding layer, followed by feature concentration and a dense output layer for multi-variable prediction.

Figure 4 summarizes the architecture of the model, describing each layer's output shape and associated parameters. The first input layer (Seq_input) has a shape of (None, 30, 9), accepting a length of 30 with nine features and containing no parameters. The second input layer (StationID) has a shape of (None, 1), accepting a single value.

The Conv1D layer, with 64 filters, produces an output shape of (None, 28, 64) and introduces 1792 parameters for convolutional operations. The LSTM layer, with 64 units, maintains an output shape of (None, 64) and consists of 33,024 parameters, enabling the

capture of long-term dependencies in the data. The embedding with output shape (None, 1, 4) converts station input into a four-dimensional vector with five station IDs, which gives 20 parameters. The Global_Max_Pooling layer with output shape (None, 4) aggregates embeddings and pools the embedding output over its dimension, resulting in a fixed-size output. The Concatenate layer output shape (None, 68) merges both main sequence features and station features for prediction. The output (Dense) layer makes the final prediction, producing five values (Temp, RH, Rain, PRESS, and SUNSHINE).

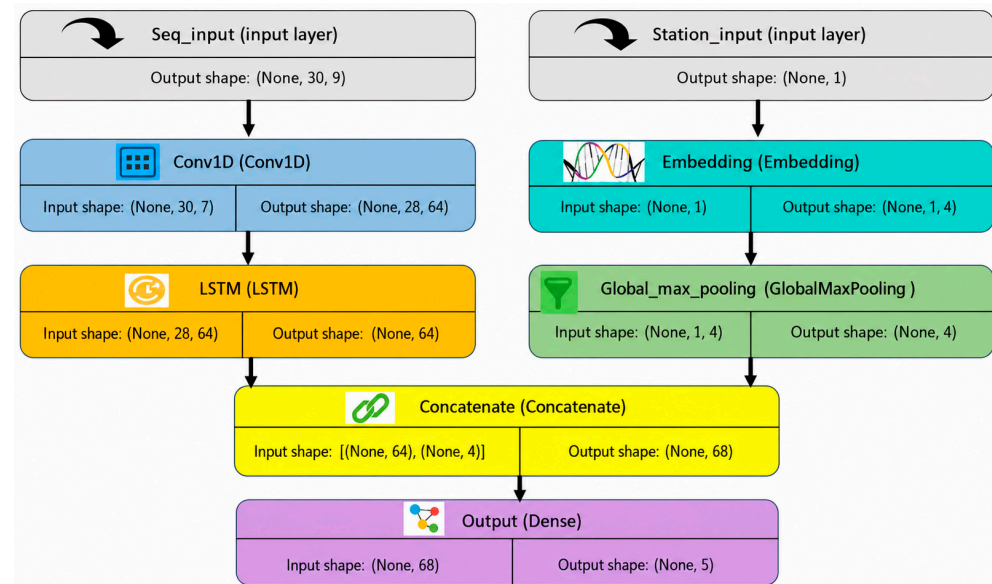


Figure 5. Detailed architecture of the proposed spatial-temporal CNN-LSTM model, showing dual-input processing of sequential weather data and station-based spatial features, followed by fusion and multi-output prediction.

Figures 4 and 5 illustrate the architecture of the proposed hybrid CNN-LSTM model designed to capture both temporal and spatial dependencies in weather data. The model consists of two parallel branches: the first processes temporal weather sequences with an input shape of (30, 7), representing 30 time steps and seven meteorological features. This sequence is passed through a ConvD layer with ReLU activation to extract local temporal patterns, followed by an LSTM layer with 64 units to model long-term dependencies in the time series. The second branch incorporates spatial information by encoding the station identifier through an embedding layer, which is then reduced using a GlobalMaxPooling1D operation to produce a compact feature representation. The outputs from both branches are concentrated into a unified feature vector (dimension 68), which is subsequently fed into a fully connected dense layer to generate predictions for multiple weather variables. This architecture effectively combines spatial and temporal feature learning, enabling the model to capture complex, nonlinear relationships in meteorological data and improve prediction performance.

4. Experimental Results

4.1. Experimental Setup

Table 8 presents the hardware and software configuration used for the development, training, and evaluation of all forecasting models in this study.

Table 8. Hardware and software setup for all models.

Component	Specification
Operating System	macOS Sequoia
Processor	Apple M1 Chip
CPU Cores	8-Core CPU
CPU	Apple M1 Integrated GPU
Memory (RAM)	8 GB/16 GB Unified Memory
Storage	SSD
Programming Language	Python 3.11
Development Environment	Jupyter Notebook (7.x)
Deep Learning Framework	TensorFlow / Keras
Machine Learning Libraries	Scikit-Learn, XGBoost
Statistical Modeling Library	Statsmodels
Data Processing Libraries	Pandas, NumPy
Visualization Libraries	Matplotlib (3.x), Seaborn (0.13.x)
Sequence Length	30 Days
Data Scaling	Min–Max Normalization
Train–Validation–Test Strategy	Chronological Split
Evaluation Metrics	MAE, MSE, RMSE, R ²

Table 9 summarizes the model-specific hyperparameter settings.

Table 9. Hyperparameters for all models.

Model	Key Parameters
ARIMA	ARIMA (5,1,2)
SARIMA	Seasonal ARIMA with optimized seasonal parameters
Random Forest	200 Trees, Random State = 42
XGBoost	200 Estimators, Learning Rate = 0.05, Max Depth = 6
CNN	Conv1D(64), Kernel Size = 3, Dense output
LSTM	SLSTM(64), Dense output
GRU	GRU(64), Dense output
CNN-LSTM	Conv1D(64) -> LSTM (64) -> Dense
ConvLSTM	ConvLSTM 2D + Dense

A chronological splitting approach was used for the training/validation/testing datasets of all models, as detailed in Section 3.2.4.

4.2. Persistence Model

A basic baseline for comparison was determined using a persistence forecasting model. The persistence model assumes that the weather conditions on the following day will be the same as the most recent observation. This method does not learn temporal patterns or nonlinear correlations but is extensively employed as a baseline in meteorological forecasting, as many weather variables are strongly temporally autocorrelated.

Table 10 shows the performance of the persistence model for the five target variables. The persistence baseline performed well for both temperature ($R^2 = 0.941$) and atmospheric pressure ($R^2 = 0.999$) due to the slow variance of these variables from day to day. However, the model showed a far worse performance for relative humidity ($R^2 = 0.556$), rainfall ($R^2 = -0.77$), and sunshine duration ($R^2 = 0.771$). The negative R^2 obtained for rainfall shows that the repetition of the previous day's rainfall is not adequate for reproducing the intermittent and varied nature of precipitation occurrences. These results show that, while persistence is a good reference for slowly shifting meteorological variables, sophisticated machine learning and deep learning models are needed for modeling complicated nonlinear dynamics and improving forecasting skills, specifically for rainfall and humidity.

Table 10. Performance results of the persistence model.

Target	MAE	MSE	RMSE	R^2
Temperature ($^{\circ}\text{C}$)	1.220	3.001	1.732	0.941
Relative humidity (%)	8.951	167.178	12.930	0.556
Rainfall (mm)	1.414	28.859	5.372	−0.077
Atmospheric Pressure (hPa)	1.551	4.231	2.057	0.999
Sunshine Duration	0.980	4.224	2.055	0.771

4.3. Statistical Models

The results of ARIMA and SARIMA show lower accuracy for prediction compared with the persistence baseline, machine learning, and deep learning, indicating poor performance for meteorological variables with nonlinear dynamics and rapid fluctuations like rainfall and sunshine duration. Table 11 demonstrates the limitations of statistical models for complicated multivariate weather forecasting problems and supports the use of more sophisticated learning-based methods.

Table 11. Performance of statistical models.

Model	Temperature R^2	Relative Humidity R^2	Rainfall R^2	Pressure R^2	Sunshine R^2
ARIMA	−1.339	−0.179	−0.070	−3.689	−3.542
SARIMA	−0.882	−0.145	−0.27	1.536	−2.359

4.4. Classical Machine Learning Models

The results for the Random Forest and XGBoost models demonstrate their strength in representing nonlinear feature interactions while being resilient to minimal training data and feature variability. Table 12 shows the performance of the two models for the five target variables. For temperature prediction, XGBoost showed the best overall accuracy, with (MAE = 1.144 $^{\circ}\text{C}$), (RMSE = 1.557 $^{\circ}\text{C}$), and ($R^2 = 0.953$ $^{\circ}\text{C}$). It additionally showed the best performance for relative humidity, with ($R^2 = 0.670$). Random Forest also performed very well in atmospheric pressure prediction, with ($R^2 = 0.999$), and provided the most accurate results for rainfall forecasting among the studied models, with ($R^2 = 0.205$). However, rain was still the most problematic variable to predict.

Table 12. Performance of classical machine learning models.

Target	Random Forest (R^2)	XGBoost (R^2)	Better Model
Temperature ($^{\circ}\text{C}$)	0.946	0.953	XGBoost
Relative humidity (%)	0.655	0.670	XGBoost
Rainfall (mm)	0.205	0.194	Random Forest
Atmospheric Pressure (hPa)	0.999	0.999	Random Forest
Sunshine Duration	0.418	0.673	XGBoost

4.5. Deep Learning Models

In this study, we examined different deep learning architectures, such as CNN, LSTM, GRU, ConvLSTM, CNN-GRU, and CNN-LSTM, for multi-variable weather forecasting using the same training, validation, and testing techniques. The MAE, MSE, RMSE, and coefficient of determination (R^2) were used to evaluate the models. Among the deep learning methods studied, the CNN performed the worst as shown in Table 13, which suggests that convolution feature extraction is not enough to capture long-term temporal correlations in meteorological time series data. On the other hand, recurrent neural network architectures such as LSTM and GRU captured sequential patterns and temporal correlations and performed significantly better in terms of prediction accuracy. The LSTM model had the best performance among the recurrent models for sunshine duration prediction ($R^2 = 0.831$) and had competitive results for temperature ($R^2 = 0.939$) and atmospheric pressure ($R^2 = 0.0995$). The GRU model achieved similar performance, but with a less sophisticated network structure and computational complexity. ConvLSTM and CNN-GRU also provided satisfactory results, especially for temperature and sunshine forecasting, which confirmed the efficiency of hybrid spatio-temporal learning algorithms. The suggested CNN-LSTM model with station embedding, which combined convolution feature extraction, recurrent temporal modeling, and geographical information encoded by station embeddings, demonstrated consistently good performance for numerous target variables. It did not provide the best results for all target variables against all competing models, but it provided a balanced forecast accuracy for temperature, relative humidity, rainfall, atmospheric pressure, and sunshine duration, and at the same time incorporated multi-station spatial information. The results show that hybrid deep learning models can capture complicated meteorological interactions and provide a stable framework for multi-variable weather prediction.

Table 13. Performance of deep learning models.

Model	Temp R^2	RH R^2	Rain R^2	PRESS R^2	SUNSHINE R^2
CNN	0.859	0.357	0.077	0.985	0.789
GRU	0.932	0.662	0.115	0.992	0.829
ConvLSTM	0.926	0.644	-	0.988	0.825
LSTM	0.939	0.658	0.115	0.995	0.831
CNN-LSTM	0.932	0.649	0.117	0.994	0.825
CNN-GRU	0.926	0.0.637	0.117	0.991	0.828

Figure 6 demonstrates that both training and validation losses exhibit a swift reduction during the initial epochs, indicating efficient early learning by the CNN-LSTM model. After roughly 5 epochs, both curves begin to stabilize, signifying that the model has predominantly assimilated the data. After around 15 epochs, both losses stabilize at low values (around 0.008), indicating reliable and consistent performance. A negligible disparity

between training and validation losses signifies low overfitting, indicating well-balanced model generalization. The swift initial decrease indicates effective data assimilation, while the ensuing stabilization verifies that the model attained an optimal or near-optimal solution without considerable overfitting. The convergence and minimal disparity between training and validation losses suggest that the model generalizes effectively to novel data. The loss curve demonstrates an exemplary training procedure, indicating successful convergence with little likelihood of overfitting. The CNN-LSTM model exhibits robustness and reliability, showcasing formidable predictive power for this application. Also, the training history of the proposed CNN-LSTM model is presented in Figure 7, which illustrates the evolution of the training and validation mean squared error (MSE) and mean absolute error (MAE) over the training epochs.

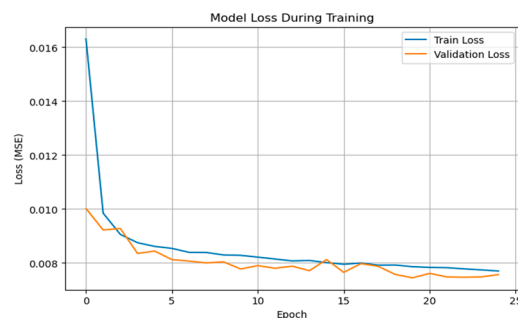


Figure 6. Training and validation loss curves of the proposed model over 25 epochs using the mean squared error (MSE). The curves show a rapid decrease in loss during the early epochs, followed by stable convergence, indicating effective model learning with no clear evidence of overfitting.

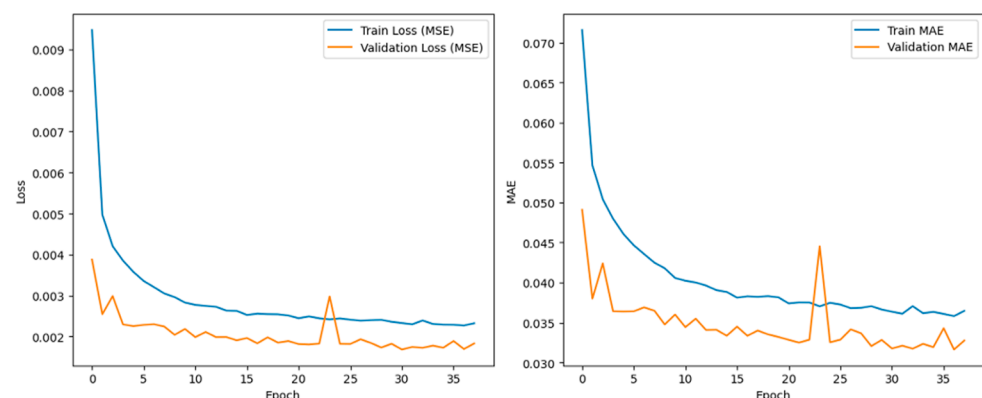


Figure 7. Training and validation performance of the proposed model in terms of mean squared error (MSE) and mean absolute error (MAE) over training epochs. Both metrics show a rapid decrease during the initial epochs, followed by gradual stabilization, indicating effective convergence and consistent generalization performance on the validation set.

The Five Meteorological Parameters

Figure 8 illustrates the actual versus expected values for five meteorological parameters for the CNN-LSTM model. For temperature (Temp), the forecasted values align closely with the actual temperature readings across all samples. The graph demonstrates significant convergence between observed and forecasted values, showing that the CNN-LSTM model delivers remarkably precise temperature forecasts and achieves exceptional performance. In the case of rainfall (Rain), forecasted precipitation levels adhere to general patterns yet exhibit inconsistencies, especially for extreme values (spikes). Although the model accurately reflects overall rainfall trends, it undervalues certain peak precipitation occurrences and is deficient in precisely forecasting peak rainfall events. Relative humidity (RH) predictions closely correspond to the actual readings, exhibiting minimal discrepancies. The model

precisely represents variations in relative humidity, with a strong prediction accuracy. For atmospheric pressure (PRESS), predictions often correspond well with actual values, although they exhibit significant discrepancies during frequent high–low oscillations. The model exhibits some difficulties in managing abrupt fluctuations in pressure. Although the model demonstrates a satisfactory performance, opportunities exist for enhancement in forecasting sudden pressure fluctuations. Finally, sunshine forecasts align closely with actual values, exhibiting minimum error and strong consistency. The forecasted duration of sunshine nearly aligns with the actual readings, demonstrating the model’s high accuracy for forecasting sunshine duration.

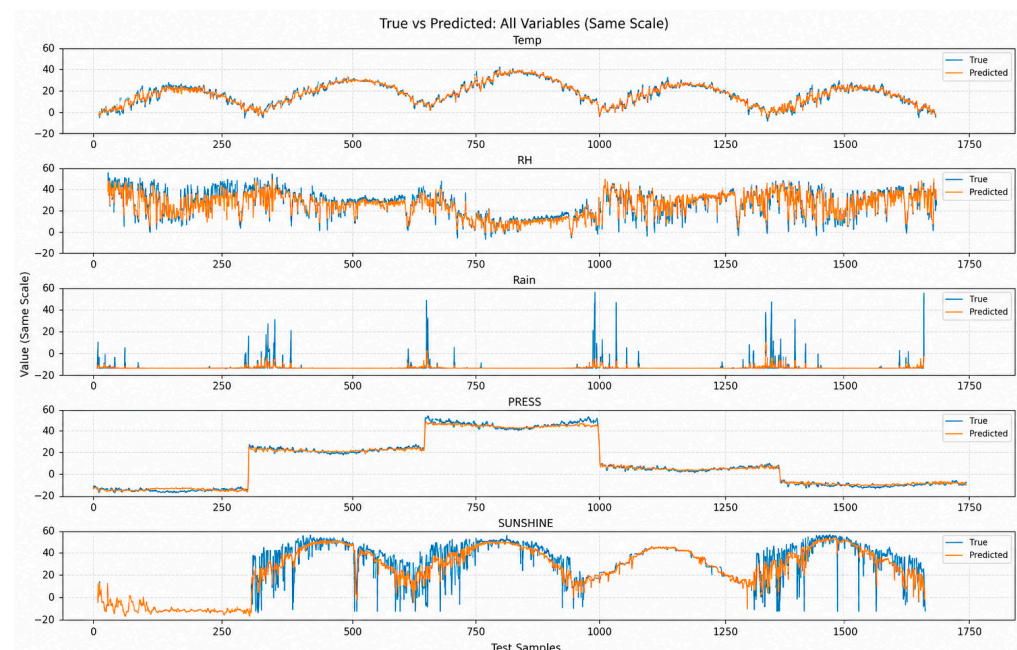


Figure 8. Actual versus predicted values for five meteorological variables, including temperature, relative humidity, rainfall, atmospheric pressure, and sunshine duration, over the test samples. The figure illustrates the model’s ability to capture temporal patterns across different weather parameters, with varying prediction accuracy depending on the variability and complexity of each variable.

The CNN-LSTM model excels at forecasting temperature, relative humidity, and sunshine duration and has considerable accuracy, reliability, and robustness across several meteorological factors, signifying its appropriateness for weather forecasting applications.

A comparison between the proposed CNN-LSTM framework and past weather forecasting approaches is provided in Table 14 below. The table focuses on some of the notable differences between the proposed approach and past research approaches in terms of the type of forecasting, data used, predicted meteorological quantities, evaluation criteria, and model results.

Table 14. Comparative analysis of the proposed CNN-LSTM model and previous weather forecasting studies.

Study	Dataset	Model	Variable	MAE	RMSE	R ²	Notes
[5]	Historical weather dataset	CNN_LSTM	Temperature	0.12	0.15	0.89	Hybrid model
[11]	Delhi, India	CNN_LSTM	Temperature	0.08	0.11	0.91	Urban climate
[15]	Hourly weather data	CNN_LSTM	Temperature	0.07	0.09	0.92	Hourly forecasting
[45]	Global temperature	CNN_LSTM	Temperature	0.05	0.08	0.93	Long-term forecasting
Proposed work	Five Palestinian stations	CNN-LSTM +Station Embedding	Multi-variable	1.38	1.81	0.93	Spatial-temporal forecasting

4.6. Comparison of Models Performance (Comparative Analysis)

As illustrated in Table 15, a complete comparison of statistical, classical machine learning, and deep learning models was performed to evaluate their performance for multi-variable weather forecasting. The methods considered were persistence, ARIMA, SARIMA, Random Forest, XGBoost, CNN, LSTM, GRU, ConvLSTM, CNN-GRU, and CNN-LSTM. The statistical models (Persistence, ARIMA, and SARIMA) provided useful baseline benchmarks but generally showed limited forecasting performance for variables with complex nonlinear dynamics. The persistence model achieved competitive results for slowly fluctuating variables, such as temperature and atmospheric pressure, due to their high temporal continuity, while ARIMA and SARIMA models showed a reduced prediction power for most target variables.

Table 15. Summary of the evaluated forecasting models, their categories, and main strengths and limitations.

Category	Model	Pros and Cons
Baseline	Persistence	Efficient for temperature and pressure, limited in rain
Statistical	ARIMA	Weak in modeling nonlinear weather dynamics
Statistical	SARIMA	Limited compared to ML and DL in capturing seasonality
Classical ML	Random Forest	Efficient for pressure and rainfall prediction
Classical ML	XGBoost	Efficient for temperature and relative humidity
Deep Learning	CNN	Limited compared to DL models
Deep Learning	LSTM	Very powerful temporal modeling (sunshine)
Deep Learning	GRU	Similar to LSTM with reduced complexity
Deep Learning	CovLSTM	Excellent spatio-temporal learning
Deep Learning	CNN-GRU	Good forecasting accuracy
Deep Learning	CNN-LSTM	Consistent overall performance and robust multi-variable forecasting with station embedding

Classical machine learning techniques, XGBoost and Random Forest, performed excellently and consistently outperformed the statistical baselines. XGBoost was the best model in terms of accuracy for temperature and relative humidity prediction, while the Random Forest model performed well, notably for rainfall and atmospheric pressure prediction. These results demonstrate the ability of ensemble tree-based approaches to capture nonlinear correlations among meteorological data.

The deep learning models outperformed the CNN-only architecture, indicating the relevance of recurrent mechanisms for learning temporal dependencies, and LSTM was the best performer in predicting sunshine duration. The suggested CNN-LSTM model integrated convolutional feature extraction, recurrent temporal learning, and station embedding to capture both temporal and geographical information. Although CNN-LSTM did not obtain the best score for all the individual target variables, it provided a stable performance for all forecasting tasks while allowing for the simultaneous prediction of multiple variables and learning from several stations.

In general, our comparative study shows that there is no single model which is the best for all meteorological variables. Classical machine learning approaches including XGBoost produced the best accuracy for numerous goals, whereas hybrid deep learning architectures were resilient and consistent in performance, effectively modeling complicated spatio-temporal correlations. These results indicate that the choice of forecasting model should be determined according to the specific forecasting goal, data characteristics, and operational constraints.

5. Discussion

Our experiments show that both classical machine learning and deep learning models perform significantly better than standard statistical forecasting methods for multi-variable weather prediction. The efficacy of statistical approaches such as ARIMA and SARIMA was limited, particularly for variables with nonlinear and highly dynamic behavior, demonstrating that linear time series models are not enough for capturing the complexity of meteorological processes. XGBoost had the best performance among the classical machine learning models in terms of forecast accuracy for temperature and relative humidity, whereas Random Forest demonstrated good performance, especially for atmospheric pressure and rainfall prediction. These results show that ensemble tree-based approaches are well suited for modeling nonlinear feature interactions and can be used as strong baselines for weather forecasting applications. The results also suggest that machine learning models can be competitive with more complicated neural network architectures by using properly engineered features and chronological training procedures. Deep learning models showed considerable improvements over the CNN-only architecture by using temporal dependencies in consecutive weather observations. LSTM and GRU had a consistently robust predictive accuracy, underscoring the significance of recurrent neural networks for learning long-term weather trends. The suggested CNN-LSTM model combined convolutional feature extraction with recurrent learning and station embedding, which allowed the simultaneous modeling of temporal dynamics and spatial variability across many weather stations. While not the best model in terms of accuracy for each target variable, CNN-LSTM attained a balanced performance for all prediction tasks and provided a single framework for multi-variable and multi-station forecasting. The most difficult task was rainfall prediction: although the absolute errors were small, the coefficient of determination was much lower for rainfall than for other meteorological variables. This tendency is expected because rainfall exhibits highly intermittent, sparse, and nonlinear features, with long periods of minimal precipitation interspersed by infrequent extreme episodes. Therefore, the low values of MAE are accompanied by low R^2 values. Future work should explore methods of enhancing rainfall-specific prediction, including weighted loss functions, probabilistic forecasting, event-based evaluation metrics, and hybrid classification–regression frameworks. Our comparison with the persistence baseline also yields key insights. Persistence provides competitive results for slowly fluctuating variables such as temperature and air pressure due to the great temporal continuity of these variables. However, the more advanced machine learning and deep learning models showed clear improvements for variables with more variability, such as relative humidity, rainfall, and sunshine duration, indicating their ability to capture complex nonlinear temporal relationships beyond simple carry-forward assumptions.

From an application standpoint, reliable weather forecasting is essential for many operational activities such as meteorological services, agricultural planning, renewable energy management, transportation scheduling, and urban planning. Our findings demonstrate that both classical machine learning and deep learning approaches can offer useful forecasting capacity, while the most appropriate model depends on the target variable and practical deployment needs. Additionally, the use of data from five geographically separated sites illustrates the feasibility of regional forecasting frameworks for Palestine and provides a replicable benchmark for future research.

Apart from typical meteorological forecasting, the prediction of weather variables has crucial implications for urban planning, energy management, and climate adaptation. Accurate prediction of temperature, humidity, solar radiation, and atmospheric variables can help to optimize energy-efficient cooling solutions and increase human thermal comfort in both indoor and outdoor contexts. Recent studies have shown that environmental pre-

diction can provide low-energy cooling solutions and human-centered climate adaptation through intelligent control strategies [59]. Furthermore, hybrid CNN-LSTM architectures have been effectively utilized in multimodal forecasting applications outside of weather forecasting, such as in energy-related decision-making, where heterogeneous data sources are merged to improve forecasting accuracy [60]. These results are in line with those of the present study and demonstrate the wider applicability of advanced machine learning and deep learning models to assist in sustainable infrastructure, renewable energy management, and smart city applications. The forecasting approach proposed in this study could therefore provide useful decision assistance for energy planning, agricultural management, transportation systems, and climate-resilient urban development in Palestine.

6. Conclusions and Future Work

In this paper, we conducted an extensive analysis of statistical models, classical machine learning techniques, and deep learning algorithms for multi-variable weather prediction using real-world meteorological data collected at five different weather stations in Palestine. The experimentation process was based on several different models, including persistence, ARIMA, SARIMA, Random Forest, XGBoost, CNN, LSTM, GRU, ConvLSTM, CNN-GRU, and CNN-LSTM. All methods were evaluated using the same preprocessing steps and chronological train-validation-test splits.

Our results showed that traditional statistical approaches offer reasonable forecasts, although their performance level is usually inferior to machine learning and deep learning models. According to our findings, the most accurate predictions for temperature and relative humidity are provided by XGBoost, whereas Random Forest can be recommended for atmospheric pressure and precipitation prediction. Additionally, our findings suggest that recurrent networks (LSTM and GRU) proved to be more efficient for time series modeling, compared to the CNN model. Finally, the developed CNN-LSTM architecture with station embeddings demonstrated promising forecast results.

We also found that there was no best forecasting model suitable for all kinds of weather conditions, since performance varies depending on the type of weather phenomenon and its associated data properties. Precipitation prediction continued to be the most difficult task due to the sparsity, nonlinearity, and unpredictability of rainfall patterns.

Some future directions of this work include investigating more sophisticated approaches to precipitation forecasting, such as using weighted loss functions, probabilistic forecasting approaches, quantile regression, and hybrid classification and regression approaches. Moreover, more sophisticated deep learning approaches such as attention mechanisms, Transformers [61,62], temporal convolutional networks, encoder-decoder networks, and graph neural networks with weather stations' geographical locations could also be explored further. In addition, multimodal weather data, including satellite images and output from numerical weather prediction, can be incorporated into the approach to enhance the model's forecasting capability. Another key direction would be to extend this work by incorporating more weather stations in neighboring regions and releasing benchmarking data and code.

Author Contributions: Conceptualization, M.O. and A.H.; methodology, M.O. and A.H.; software, M.O.; validation, A.H.; formal analysis, A.H.; investigation, M.O.; resources, M.O. and A.H.; data curation, M.O.; writing—original draft preparation, M.O.; writing—review and editing, A.H.; visualization, M.O.; supervision, A.H.; project administration, A.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset can be found on: <https://github.com/modeh4/Palestine-Weather-Forecasting> (accessed on 21 June 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Heinlien, R.A. *Heinlein—Time Enough for Love*, 3rd ed.; Berkley Publishing Corporation: New York, NY, USA, 1974.
2. Climate Is What We Expect, Weather Is What We Get: An Interview with Cobus Van Breda—Australian Institute for the Conservation of Cultural Material. Available online: <https://aiccm.org.au/network-news/climate-is-what-we-expect-weather-is-what-we-get-an-interview-with-cobus-van-breda/> (accessed on 8 August 2025).
3. Pike, R.G. *Compendium of Forest Hydrology and Geomorphology in British Columbia*; Ministry of Forests and Range: Victoria, BC, Canada, 2010.
4. Dole, R.M. Linking Weather and Climate. In *Synoptic—Dynamic Meteorology and Weather Analysis and Forecasting*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 297–348. [\[CrossRef\]](#)
5. Gong, Y.; Zhang, Y.; Wang, F.; Lee, C. Deep Learning for Weather Forecasting: A CNN-LSTM Hybrid Model for Predicting Historical Temperature Data. *Appl. Comput. Eng.* **2024**, *99*, 168–174. [\[CrossRef\]](#)
6. Chen, L.; Han, B.; Wang, X.; Zhao, J.; Yang, W.; Yang, Z. Machine Learning Methods in Weather and Climate Applications: A Survey. *Appl. Sci.* **2023**, *13*, 12019. [\[CrossRef\]](#)
7. Agapiou, A. Remote Sensing Heritage in a Petabyte-Scale: Satellite Data and Heritage Earth Engine© Applications. *Int. J. Digit. Earth* **2017**, *10*, 85–102. [\[CrossRef\]](#)
8. Kim, S.; Hong, S.; Joh, M.; Song, S.-K. DeepRain: ConvLSTM Network for Precipitation Prediction Using Multichannel Radar Data. *arXiv* **2017**. [\[CrossRef\]](#)
9. Fan, M.; Imran, O.; Singh, A.; Ajila, S.A. Using CNN-LSTM Model for Weather Forecasting. In *Proceedings of the 2022 IEEE International Conference on Big Data, Big Data 2022, Osaka, Japan, 17–20 December 2022*; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2022; pp. 4120–4125.
10. Jakaria, A.H.M.; Hossain, M.; Rahman, M.A. Smart Weather Forecasting Using Machine Learning: A Case Study in Tennessee. *arXiv* **2020**, arXiv:2008.10789. [\[CrossRef\]](#)
11. Li, B.; Qian, Y. Weather Prediction Using CNN-LSTM for Time Series Analysis: A Case Study on Delhi Temperature Data. *arXiv* **2024**, arXiv:2409.09414.
12. Wang, Y.; Goodall, J.L.; Kumar, C.; McSpadden, D.; Barbosa, S.A.; Roy, B.; Shahabi, A.; Tahvildari, N. A Hybrid CNN-LSTM Surrogate Model for Hyper-Resolution Spatiotemporal Flood Forecasting in Norfolk, Virginia. *J. Hydrol. Reg. Stud.* **2026**, *64*, 103234. [\[CrossRef\]](#)
13. Melnyk, S.; Vasiutynska, K.; Korduba, I.; Trach, Y.; Trach, R.; Butenko, D.; Chyliński, F.; Wrzesiński, G. Modeling and Forecasting of the Local Climate of Odesa Using CNN-LSTM and the Statistical Analysis of Time Series. *Sustainability* **2025**, *17*, 8424. [\[CrossRef\]](#)
14. Ladjal, B.; Nadour, M.; Bechouat, M.; Hadroug, N.; Sedraoui, M.; Rabehi, A.; Guermoui, M.; Agajie, T.F. Hybrid Deep Learning CNN-LSTM Model for Forecasting Direct Normal Irradiance: A Study on Solar Potential in Ghardaia, Algeria. *Sci. Rep.* **2025**, *15*, 15404. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Hou, J.; Wang, Y.; Zhou, J.; Tian, Q. Prediction of Hourly Air Temperature Based on CNN-LSTM. *Geomat. Nat. Hazards Risk* **2022**, *13*, 1962–1986. [\[CrossRef\]](#)
16. Dai, L.; Zhou, W. Prediction of PM_{2.5} Concentration in Lanzhou City Based on CNN-LSTM. *Adv. Appl. Math.* **2023**, *12*, 1003–1012. [\[CrossRef\]](#)
17. Shen, J.; Wu, W.; Xu, Q. Accurate Prediction of Temperature Indicators in Eastern China Using a Multi-Scale CNN-LSTM-Attention Model. *arXiv* **2024**, arXiv:2412.07997. [\[CrossRef\]](#)
18. Abbot, J.; Marohasy, J. The Application of Machine Learning for Evaluating Anthropogenic versus Natural Climate Change. *GeoResJ* **2017**, *14*, 36–46. [\[CrossRef\]](#)
19. Davenport, F.V.; Diffenbaugh, N.S. Using Machine Learning to Analyze Physical Causes of Climate Change: A Case Study of U.S. Midwest Extreme Precipitation. *Geophys. Res. Lett.* **2021**, *48*, e2021GL093787. [\[CrossRef\]](#)
20. Fang, Y.; Chen, H.; Lin, Y.; Zhao, C.; Lin, Y.; Zhou, F. Classification of Northeast China Cold Vortex Activity Paths in Early Summer Based on K-Means Clustering and Their Climate Impact. *Adv. Atmos. Sci.* **2021**, *38*, 400–412. [\[CrossRef\]](#)
21. Holmstrom, M.; Liu, D.; Vo, C. Machine Learning Applied to Weather Forecasting. *Meteorol. Appl.* **2016**, *10*, 1–5.
22. Radhika, Y.; Shashi, M. Atmospheric Temperature Prediction Using Support Vector Machines. *Int. J. Comput. Theory Eng.* **2009**, *1*, 55–58. [\[CrossRef\]](#)

23. Kumar, R.; Khatri, R. A Weather Forecasting Model Using the Data Mining Technique. *Int. J. Comput. Appl.* **2016**, *139*, 4–12. [CrossRef]
24. Qiu, M.; Zhao, P.; Zhang, K.; Huang, J.; Shi, X.; Wang, X.; Chu, W. A Short-Term Rainfall Prediction Model Using Multi-Task Convolutional Neural Networks. In Proceedings of the IEEE International Conference on Data Mining, ICDM 2017, New Orleans, LA, USA, 18–21 November 2017; pp. 395–404. [CrossRef]
25. Karevan, Z.; Suykens, J.A.K. Spatio-Temporal Stacked LSTM for Temperature Prediction in Weather Forecasting. *arXiv* **2018**, arXiv:1811.06341v1.
26. Krasnopolsky, V.M.; Fox-Rabinovitz, M.S. Complex Hybrid Models Combining Deterministic and Machine Learning Components for Numerical Climate Modeling and Weather Prediction. *Neural Netw.* **2006**, *19*, 122–134. [CrossRef] [PubMed]
27. McCulloch, W.S.; Lrerr, W.; Pitts, H. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133. [CrossRef]
28. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]
29. Linnainmaa, S. Taylor expansion of the accumulated rounding error. *BIT Numer. Math.* **1976**, *16*, 146–160. [CrossRef]
30. Rumelhart, D.E.; McClelland, J.L.; the PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*; MIT Press: Cambridge, MA, USA, 1986; Chapter 8. Available online: https://stanford.edu/~jlmcc/papers/PDP/Volume%201/Chap8_PDP86.pdf (accessed on 21 June 2026).
31. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
32. Amodei, D.; Ananthanarayanan, S.; Anubhai, R.; Bai, J.; Battenberg, E.; Case, C.; Casper, J.; Catanzaro, B.; Cheng, Q.; Chen, G.; et al. Deep Speech 2: End-to-End Speech Recognition in English and Mandarin. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 20–22 June 2016; pp. 173–182.
33. Oh, J.; Guo, X.; Lee, H.; Lewis, R.; Singh, S. Action-Conditional Video Prediction Using Deep Networks in Atari Games. *Adv. Neural Inf. Process. Syst. (NeurIPS)* **2015**, *28*, 286–294. [CrossRef]
34. (PDF) Mastering the Game of Go with Deep Neural Networks and Tree Search. Available online: https://www.researchgate.net/publication/292074166_Mastering_the_game_of_Go_with_deep_neural_networks_and_tree_search (accessed on 7 July 2025).
35. Mathieu, M.; Couprie, C.; LeCun, Y. Deep Multi-Scale Video Prediction beyond Mean Square Error. *arXiv* **2016**, arXiv:1511.05440.
36. Lee, A.X.; Zhang, R.; Ebert, F.; Abbeel, P.; Finn, C.; Levine, S. Stochastic Adversarial Video Prediction. *arXiv* **2018**, arXiv:1804.01523.
37. Lvsersonl, L.R.; Prasad, A.M.; Liaw, A. New Machine Learning Tools for Predictive Vegetation Mapping after Climate Change: Bagging and Random Forest Perform Better than Regression Tree Analysis. In *Landscape Ecology of Trees and Forests, Proceedings of the Twelfth Annual IALE (UK) Conference, Cirencester, UK, 21–24 June 2004*; U.S. Department of Agriculture: Washington, DC, USA, 2004; pp. 317–320.
38. Huang, L.; Kang, J.; Wan, M.; Fang, L.; Zhang, C.; Zeng, Z. Solar Radiation Prediction Using Different Machine Learning Algorithms and Implications for Extreme Climate Events. *Front. Earth Sci.* **2021**, *9*, 596860. [CrossRef]
39. Tripathi, S.; Srinivas, V.V.; Nanjundiah, R.S. Downscaling of Precipitation for Climate Change Scenarios: A Support Vector Machine Approach. *J. Hydrol.* **2006**, *330*, 621–640. [CrossRef]
40. Zong, L.; Yang, Y.; Gao, M.; Wang, H.; Wang, P.; Zhang, H.; Wang, L.; Ning, G.; Liu, C.; Li, Y.; et al. Synoptic Drivers of Co-Occurring Summertime Ozone and PM_{2.5} Pollution in Eastern China. *Atmos. Chem. Phys.* **2020**, *20*, 13001–13021.
41. Grover, A.; Kapoor, A.; Horvitz, E. A Deep Hybrid Model for Weather Forecasting. In Proceedings of the 21th ACM SIGKDD International Conference On Knowledge Discovery And Data Mining, Sydney, Australia, 10–13 August 2015. [CrossRef]
42. Ranjan, N.; Bhandari, S.; Zhao, H.P.; Kim, H.; Khan, P. City-Wide Traffic Congestion Prediction Based on CNN, LSTM and Transpose CNN. *IEEE Access* **2020**, *8*, 81606–81620. [CrossRef]
43. Montori, F.; Bedogni, L.; Bononi, L. A Collaborative Internet of Things Architecture for Smart Cities and Environmental Monitoring. *IEEE Internet Things J.* **2018**, *5*, 592–605. [CrossRef]
44. Zhang, Y. Research on Global Temperature Prediction Based on CNN-LSTM Model. *Adv. Appl. Math.* **2024**, *13*, 302–312. [CrossRef]
45. Guo, Q.; He, Z.; Wang, Z. Monthly Climate Prediction Using Deep Convolutional Neural Network and Long Short-Term Memory. *Sci. Rep.* **2024**, *14*, 17748. [CrossRef] [PubMed]
46. Wang, B.; Zhu, S.; Wu, L.; Liu, Y.; Lu, X.; He, K.; Liu, Y.; Liu, W. Global Weather Forecasting through Deep Learning: Applications and Challenges. *Technol. Rev. Carbon Neutrality* **2025**, *1*, 9550021. [CrossRef]
47. Kabir, M.A.; Chakma, C. Redefining Multi-Target Weather Forecasting with a Novel Deep Learning Model: Hierarchical Temporal Convolutional Long Short-Term Memory with Attention (HTC-LSTM-Attn) in Bangladesh. *PLoS ONE* **2026**, *21*, e0342431. [CrossRef] [PubMed]
48. Palestinian Meteorological. Available online: <https://www.pmd.ps/en> (accessed on 17 January 2026).
49. Handling Missing Data in Time Series: 5 Methods—Growth-Onomics. Available online: <https://growth-onomics.com/handling-missing-data-in-time-series-5-methods/> (accessed on 8 July 2025).

50. Wilks, D.S. *Statistical Methods in the Atmospheric Sciences*, 2nd ed.; Academic Press: San Diego, CA, USA, 2011.
51. Alharbi, F.R.; Csala, D. A Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) Forecasting Model-Based Time Series Approach. *Inventions* **2022**, *7*, 94. [[CrossRef](#)]
52. Wilson, G.T. *Time Series Analysis: Forecasting and Control*, 5th Edition, by George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung, 2015. Published by John Wiley and Sons Inc., Hoboken, New Jersey, pp. 712. ISBN: 978-1-118-67502-1. *J. Time Ser. Anal.* **2016**, *37*, 709–711. [[CrossRef](#)]
53. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2016, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [[CrossRef](#)]
54. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
55. Adefemi, K.O.; Mutanga, M.B.; Alimi, O.A. A Hybrid CNN–GRU Deep Learning Model for IoT Network Intrusion Detection. *J. Sens. Actuator Netw.* **2025**, *14*, 96. [[CrossRef](#)]
56. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *Adv. Neural Inf. Process. Syst. (NeurIPS)* **2015**, *28*, 802–810.
57. Dey, R.; Salem, F.M. Gate-Variants of Gated Recurrent Unit (GRU) Neural Networks. In Proceedings of the 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), Boston, MA, USA, 6–9 August 2017.
58. Kiranyaz, S.; Ince, T.; Gabbouj, M. Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 664–675. [[CrossRef](#)] [[PubMed](#)]
59. Wang, Z.; Matsushashi, R. Low-Energy Miniature Wearable Air-Conditioner with Direct Cold-Air Delivery: A Novel Water–Electricity Hybrid Energy System. *Build. Environ.* **2026**, *297*, 114590. [[CrossRef](#)]
60. Wang, Z.; Mae, M.; Yamane, T.; Ajisaka, M.; Nakata, T.; Matsushashi, R. Enhanced Day-Ahead Electricity Price Forecasting Using a Convolutional Neural Network–Long Short-Term Memory Ensemble Learning Approach with Multimodal Data Integration. *Energies* **2024**, *17*, 2687. [[CrossRef](#)]
61. Masri, S.; Hasasneh, A.; Tami, M.; Tadj, C. Exploring the Impact of Image-Based Audio Representations in Classification Tasks Using Vision Transformers and Explainable AI Techniques. *Information* **2024**, *15*, 751. [[CrossRef](#)]
62. Tami, M.; Masri, S.; Hasasneh, A.; Tadj, C. Transformer-Based Approach to Pathology Diagnosis Using Audio Spectrogram. *Information* **2024**, *15*, 253. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.