Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Evaluating Machine Learning Models for GLDAS Air Temperature Prediction on North Western Coast of Egypt in a Changing Climate

Moataz H. Khalil

moatezhassankhalil@gmail.com

Informatics Research Institute (IRI), City of Scientific Research and Technological Applications (SRTA-City)

Sami Z. Mohamed

Arid Lands Cultivation Research Institute (ALCRI), City of Scientific Research and Technological Applications (SRTA-City)

Article

Keywords:

Posted Date: March 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8855112/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Evaluating Machine Learning Models for GLDAS Air Temperature Prediction on North Western Coast of Egypt in a Changing Climate

Moataz H. Khalil[1*], Sami Z. Mohamed[2]

[1]Computer Based Engineering Applications Department, Informatics Research Institute (IRI), City of Scientific Research and Technological Applications (SRTA-City), Alexandria, 21934, Egypt

[2]Land and Water Technologies Department, Arid Lands Cultivation Research Institute (ALCRI), City of Scientific Research and Technological Applications (SRTA-City), Alexandria, 21934, Egypt

*Corresponding Author: moatezhassankhalil@gmail.com

Abstract

Accurate air temperature forecasting is essential for climate risk management, resilient agricultural planning, and sustainable development, yet it remains challenging in regions with sparse meteorological station networks. In Egypt, the limited spatial coverage of ground-based observations constrains the reliable characterization of local climate variability and predictive capacity. This research investigates the potential of integrating NASA's Global Land Data Assimilation System (GLDAS) remote sensing product with machine-learning (ML) techniques to overcome data scarcity and improve daily air temperature forecasting along the Egyptian North-Western Coast. GLDAS v2.1, 0.25° resolution, Near-surface air temperature was evaluated against in situ observations for the period 2000 – 2020 at four locations: Alexandria, Marsa Matrouh, Sallum, and Siwa Oasis, confirming its high fidelity ($R^2 \geq 0.886$) and reliability of reanalysis data for local-scale forecasting. A correlation-based feature selection strategy was applied to identify the most informative lagged temperature predictors, followed by the implementation of Linear Regression (LR), Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) models. Results demonstrate that the SVR model consistently achieved superior performance across all sites, delivering the lowest Mean Squared Error (0.659 – 1.610°C) and highest Nash-Sutcliffe Efficiency (0.9428 – 0.9685). The research establishes a robust, scalable framework that transforms freely available global satellite data into precise local forecasts. This workflow provides a critical tool for enhancing climate adaptation strategies in agriculture, water management, and early warning systems in arid, data-deficient regions, overcoming the delay in the availability of climate models online.

Background & Summary

Inadequate spatial coverage of weather stations limits the ability to capture localized climate variations, leading to inaccuracies in predicting extreme events such as droughts, floods, and heat waves. This deficiency hampers the development of effective strategies for mitigating the impacts of these events on agricultural productivity and urban infrastructure. Moreover, reliable meteorological data is needed to assess climate trends and variability, which is necessary for long-term planning and adaptation to changing climatic conditions. Consequently, in a changing climate, the ability to accurately predict weather patterns, especially air temperature changes, is critical for understanding and mitigating the impacts of climate change for early

warning and risk management systems that are essential for development [1]. As of 2024, there are approximately only 30 meteorological stations operated by the Egyptian Meteorological Authority (EMA) [2]. These stations are distributed across various regions of Egypt depending on the airport's allocations, making data availability a challenge and, therefore, forecasting and prediction. With limitations in meteorological stations and monitoring systems, remote sensing offers a cost-effective solution to acquire climate data. The Global Land Data Assimilation System (GLDAS) is one of the largest global datasets available, providing important information on land surface variables [3]. This system was developed by the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC) and the National Oceanic and Atmospheric Administration (NOAA)/National Centers for Environmental Prediction (NCEP) [3]. The NASA Global Land Data Assimilation System (GLDAS) aims to combine satellite data and ground-based observations (OBS) using advanced land surface modeling and assimilation techniques to produce accurate land surface states and fluxes. GLDAS operates multiple land surface models independently from the atmosphere and integrates much observation-based data. It works at high global resolutions, ranging from 0.25° to 1°, facilitated by the Land Information System (LIS) [3,4]. However, accurate forecasting of air temperature is the primary foundation for applications in environmental planning, including agriculture, energy, urban development, and other fields. However, this remains challenging due to the complex, multidimensional, and chaotic nature of weather systems. The growing abundance of weather data offers new opportunities for improved temperature forecasting. However, it also introduces challenges in unraveling the complex relationships among multiple weather variables and building reliable models to capture hidden patterns.

In parallel with machine learning (ML), it is revolutionizing weather forecasting by enabling more accurate temperature predictions [5,6]. Continued advances in ML algorithms, data collection, and computational power are steadily improving the accuracy and reliability of temperature prediction models. By improving the usefulness of large weather datasets with sophisticated ML techniques, researchers can develop efficient models that enhance understanding of air temperature dynamics and improve forecast accuracy [5,6]. In general, ML approaches are broadly categorized into supervised and unsupervised methods. For continuous targets like temperature, supervised regression models are most appropriate [7]. Nonetheless, applying ML to weather data entails two main challenges. The first challenge is capturing the complex functional relationship between numerous input features and the output values [8]. With many weather variables across locations and times, the feature space can become huge; a targeted feature selection strategy simplifies the model without sacrificing important information, thereby improving predictive performance. The second challenge is rigorous error analysis while evaluating a model’s performance on both training and unseen data to diagnose shortcomings and refine the model. Aside from the challenges, statistical parameters are mandatory for evaluating the model; however, the Mean Squared Error (MSE) is a fundamental regression metric that quantifies the average squared difference between predicted and actual values. MSE can be decomposed into

bias and variance, which are the two primary sources of prediction error [9]. Bias error results from overly simplistic models that underfit the data, while variance error arises from overly complex models that overfit and generalize poorly [9]. Managing the bias–variance trade-off is therefore necessary; as model complexity increases, bias typically decreases and variance increases. An optimal model complexity minimizes total error by balancing these competing factors [9]. The relationship between complexity and error can manifest in different patterns: a monotonic decrease if bias dominates, a U-shaped curve if variance dominates, or a double-descent curve in which error declines again beyond a certain complexity threshold [10]. These dynamics underscore the need to control model complexity to achieve robust performance carefully.

Recent advances in data-driven environmental prediction have demonstrated the growing utility of deep learning and hybrid machine-learning frameworks across diverse meteorological and climate applications, including a visibility-forecasting system for Indonesian airports developed using single- and multi-layer Long Short-Term Memory (LSTM) architectures [11]. By augmenting time-series visibility data with auxiliary meteorological variables, the multi-layer LSTM achieved a notably low RMSE (0.0775), highlighting the value of incorporating pressure as an important intermediate predictor. Expanding on neural sequence-modeling approaches, a research [5] benchmarked MLP, LSTM, and CNN+LSTM models for short- and medium-range temperature prediction at John F. Kennedy International Airport. Using a multivariate seven-day historical input window, the CNN+LSTM consistently exhibited superior predictive skill for both 1-day and 10-day horizons, underscoring the advantage of integrating convolutional feature extraction with recurrent temporal modeling. In the climate domain, an exploration of the relationship between greenhouse gas emissions and climate variability using a suite of linear and advanced ML techniques [12]. Classical repressor (SGD, Huber, RANSAC, Ridge, Lasso) were contrasted with LSTM, CNN (multihead and multichannel), and hybrid autoencoder-ConvLSTM architectures. Trained on observational records from the U.K. National Centre for Earth Observation and Copernicus, the AE-ConvLSTM achieved the lowest RMSE for temperature prediction, demonstrating the increasing relevance of deep spatiotemporal learning for quantifying GHG-driven climate responses. Long-term summer air-temperature prediction and heatwave detection were further examined in [13], which introduced 3 AI pipelines combining CNNs with data-fusion strategies, conventional ML models, and recurrence-plot–enhanced CNNs. The RP-CNN framework, which embeds temporal dynamics into a compact visual representation, yielded the strongest performance, particularly in identifying extreme temperature peaks linked to heatwave events, and presents an emerging priority in climate-risk analytics. For short-term urban forecasting, a development of Bi-LSTM models for two sites in London [14], trained on five years of meteorological observations. The models produced competitive accuracies for 24- and 72-hour predictions, achieving RMSE values of 1.45°C for temperature at Kew Gardens and 2.26°C at Heathrow, while maintaining robust humidity forecasts. These findings illustrate the scalability of sequence-based deep learning across heterogeneous temporal horizons. Complementing these efforts, a proposal for a machine-

learning-augmented bias-correction framework for near-surface temperature forecasts in Poland, integrating RF, multiple linear regression, and ANNs with NWP outputs (ALARO, AROME, and COSMO). The RF model reduced RMSE by up to 27% relative to NWP baselines, while ANNs achieved marginal gains over RF [15]. Each model displayed context-specific strengths; the RF effectively corrected warm nighttime biases in summer, whereas ANNs better captured rapid cooling events in early spring, highlighting the importance of model diversity in post-processing operational forecasts.

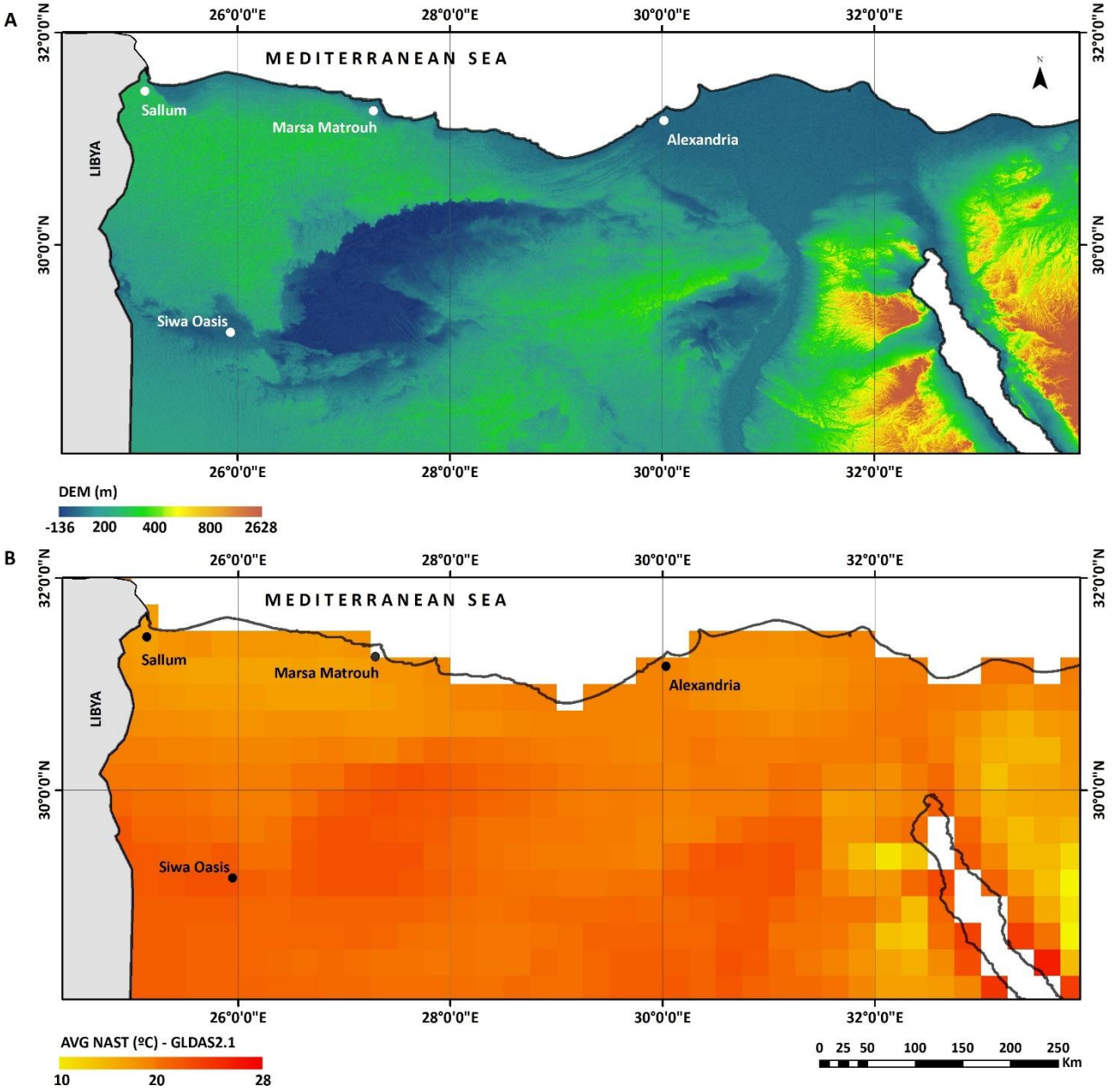


Fig. 1 (A) Cities geographic location with the region digital elevation model (DEM), (B) GLDAS average daily near air surface temperature (NAST)

In our research, we evaluate several ML models for predicting GLDAS 2.1 (0.25° spatial resolution) near-surface air temperature (2m) and local observed air temperature values along the Egyptian North-Western Coast from 2000 to 2020. We apply three modeling techniques: Linear Regression (LR), Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) neural networks, in four cities: Alexandria, Marsa Matrouh, Sallum, and Siwa Oasis, as shown in Figure 1 the geographic locations and the average daily near air surface temperature of the region for the region from 2000 to 2020. The choice of those locations was based on the availability of real measured data for the same time frame. For each location, a feature selection process identifies the most relevant predictors, and we train a tailored model on each dataset. The analytical approach we followed yielded several key outcomes: (1) Identification of the optimal feature set for each location, (2) Development of tailored ML models for each location dataset combination, (3) Comparative performance analysis to determine the best model for each city, and (4) Detailed error analysis of each model's predictions, examining MSE as well as bias and variance.

Methods

The strategic objective of this research is to evaluate and compare the predictive performance of three machine-learning models: Linear Regression (LR), Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) networks, in forecasting the average daily air temperature of the GLDAS air temperature. Following a multi-stage, structured, and reproducible methodology (Figure 2), cross-validation of GLDAS values against real weather stations, followed by a correlation-based feature selection process, was used to identify the most informative historical temperature inputs for prediction. The three models were then trained to predict the target day's mean temperature using lagged observations of prior daily temperatures, to identify the most accurate and robust modeling framework for air temperature time-series prediction at each target location along the Egyptian North-Western Coast.

Data source:

Analysis of near-surface air temperature over a 21-year period (January 2000 to December 2020) using two complementary datasets: ground-based observations (OBS) and model-based reanalysis data from the Global Land Data Assimilation System (GLDAS). The benchmark data consist of daily mean temperature measurements from meteorological weather stations. The model data were collected from the GLDAS Version 2.1 Noah product (GLDAS_NOAH025_3H_2.1), a joint initiative of NOAA, NASA, and associated centers that generates land surface fields by assimilating observations into numerical models [3, 4]. This dataset, available online at https://disc.gsfc.nasa.gov/datasets/GLDAS_NOAH025_3H_2.1/summary, provides near-surface air temperature at a 0.25° spatial resolution and 3-hourly temporal resolution from 2000 to present. Its reliability for climatological analysis is supported by prior extensive validation against station data [16,17]. To ensure direct comparability with the daily OBS data, the native 3-hourly GLDAS records were processed using R. For

each station location, the GLDAS temperature time series were extracted from the global NetCDF-4 files and aggregated to daily mean values using an arithmetic average to fit the OBS records.

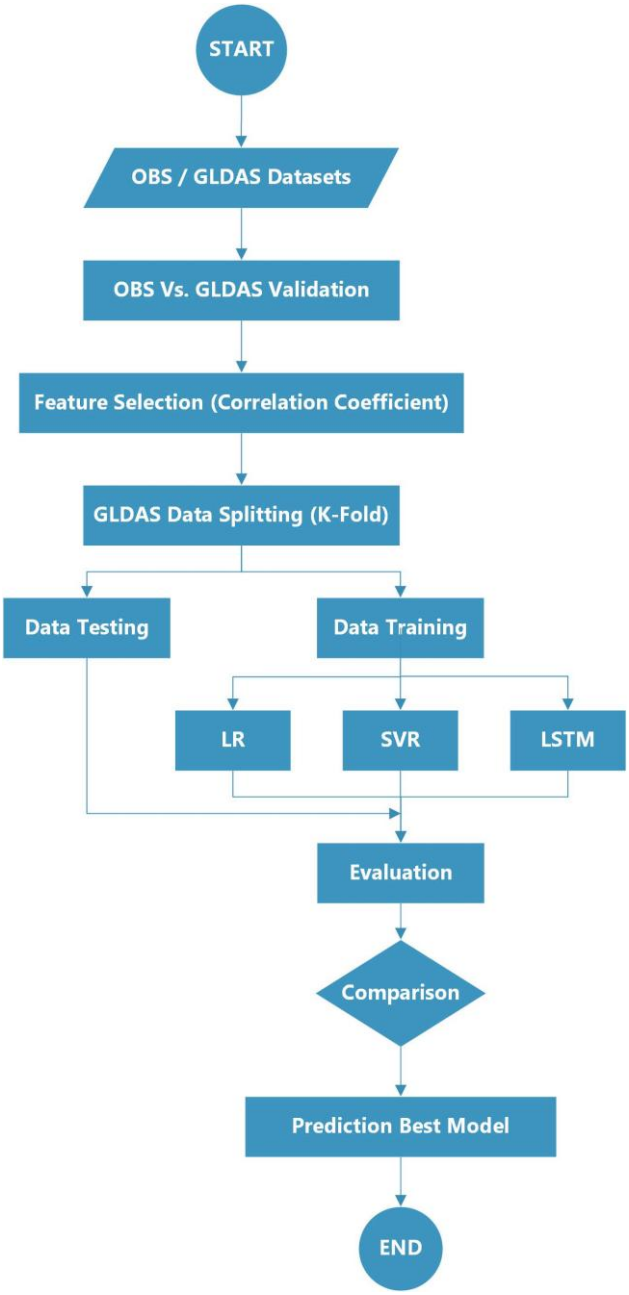


Fig. 2 The structures methodology flowchart

GLDAS Data Validation:

The GLDAS-2.1 air temperature data were validated against OBS as an initial step. While this dataset has demonstrated high validity in other regions, evaluation within the specific study area is essential due to potential performance variations caused by local topography and environmental conditions. The validation

involved a comprehensive spatial and statistical comparison, following a standard spatio-temporal alignment, preprocessing, and cross-correlation-based evaluation methodology.

Feature Selection:

Feature selection in the workflow is a statistical screening process that identifies lagged daily air temperature variables with a strong linear association with the target daily temperature. Each feature corresponds to the average air temperature of a preceding day, and the feature set consists of multiple lagged observations used as predictors in the machine learning models. In other words, "features" represent the number of preceding daily temperature observations used to forecast the target day's temperature. For instance, predicting the temperature on day 10 using five features would incorporate the average daily temperatures from days 5 through 9. Utilizing the filter-based feature selection methodology, by Pearson Correlation Coefficient [18], to quantify the linear relationship between each candidate feature (historical temperature) and the target variable (future temperature). The strength of correlation is calculated using the following equation 1:

$$\boldsymbol{r} = \frac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2 \sum_{i=1}^{N}(y_i-\bar{y})^2}} \qquad \textbf{Eq 1}$$

Where N is the dataset size, $x\ and\ \bar{x}$ are the vector of the feature and the mean value of the feature vector, $y, and\ \bar{y}$ are the vector of the target prediction and the mean of the target prediction vector. The calculated value of r compares with the threshold equal to 0.9 ($r \geq 0.9$) to choose as a feature for machine learning models.

Machine Learning Models:

Linear Regression (LR):

Using the R Studio platform, Linear Regression (LR) was operated as a baseline regression model to predict continuous daily temperature values using a linear relationship between the target variable and one or more independent features [19]. The model assumes that the predicted temperature can be expressed as a linear combination of input variables, typically estimated via ordinary least squares to minimize the sum of squared residuals. Forecasting the average daily temperature for GLDAS dataset, the independent variables comprised a temporal window of temperatures from preceding days. The equation [2] below gives the general linear regression formula [20].

$$\boldsymbol{y = k_0 + k_1 * x_1 + k_2 * x_2 + \ldots\ldots\ldots + k_m * x_m} \qquad \textbf{Eq 2}$$

Where y is the target predicted value, k_0 is the intercept section (the location where it intersects an axis Y), $k_1 + k_2 \ldots, k_m$ are coefficient parameters for equation, $x_1, x_2 \ldots, x_m$ are the values of the features, m is the size of features.

The complexity of the LR model is directly governed by the number of independent features X [21]. Therefore, the computational complexity during the inference (prediction) phase is linear with respect to the feature count. This can be expressed as equation 3: $\boldsymbol{Time\ Complexity} \approx \mathbf{O}(\boldsymbol{X})$ Where X is the size of the features.

Support Vector Regression (SVR):

Support Vector Regression (SVR) is an extension of the Support Vector Machine (SVM) algorithm, adapted for regression analysis of time series data [22, 23]. The core objective of SVR is to identify the non-linear function of the GLDAS air temperature that maps input features to a continuous target variable (predicted). This is achieved by determining a hyperplane that fits the data while maximizing the margin around the predicted values, defined by a subset of the training data known as support vectors. The model formulation results in a convex optimization problem. The general SVR function is expressed as equation [4]:

$$F = w^T \varphi(x) + b \qquad \textbf{Eq 4}$$

Where F is the predicted target value, w is the weight vector, $\varphi(x)$ is a function mapping the input feature vector x to a higher-dimensional space, and b is the bias term. The parameters w and b are derived by minimizing the regularized risk function $R(F)$ given by equation [5]:

$$\boldsymbol{R(F)} = \frac{1}{2}\|\boldsymbol{w}\|^2 + \boldsymbol{C}\ \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{L_\varepsilon(y_i, F_i)} \qquad \textbf{Eq 5}$$

Where C is a regularization parameter that controls the trade-off between model complexity and the penalty for deviations larger than ε, n is the dataset size, and $L_\varepsilon(y_i, F_i)$, is the ε-insensitive loss function which ignores errors within a margin ε of the true value y_i.This optimization can be reformulated by introducing positive slack variables ξ, ξ^*, leading to the standard minimization problem in equation [6]:

$$\boldsymbol{minimize}\left(\frac{1}{2}\ \|\boldsymbol{w}\|^2 + \boldsymbol{C}\ \sum_{i=1}^{n}(\boldsymbol{\xi_i} + \boldsymbol{\xi_i^*})\right) \qquad \textbf{Eq 6}$$

Where ξ, ξ^* are two positive slack variables.

To model nonlinear relationships, SVR utilizes kernel functions. This study employs the Radial Basis Function (RBF) kernel, which is defined as equation [7]:

$$\boldsymbol{K(x, y)} = \boldsymbol{e^{-\gamma\|x-y\|^2}} \qquad \textbf{Eq 7}$$

Where γ is the kernel parameter determining the influence of individual training samples.

Several factors influence the complexity of the resulting SVR model: (1) the number of support vectors ($\boldsymbol{n_{SV}}$), (2) the dimensionality of the input features ($\boldsymbol{X}$), and (3) the choice of kernel and its parameters. Since this research exclusively uses the RBF kernel, the computational complexity during the prediction phase is primarily governed by $\boldsymbol{n_{SV}}$ and $\boldsymbol{X}$. Following [24], this time complexity can be approximated as equation [8]:

$$\boldsymbol{Time\ Complexity} \approx \mathbf{O}(\boldsymbol{n_{SV} * X}) \qquad \textbf{Eq 8}$$

Long Short-Term Memory (LSTM) Network:

For time series forecasting of GLDAS air temperature, we implemented a Long Short-Term Memory (LSTM) network. LSTMs are specialized recurrent neural networks (RNNs) designed to capture long-term temporal dependencies, making them suitable for this application [24]. Their architecture uses gating mechanisms, input, forget, and output gates, to regulate information flow and retain critical context from earlier in a sequence, effectively mitigating the vanishing gradient problem common in standard RNNs [11]. This allows the model to account for preceding days' temperatures when making predictions. The main operations of the LSTM cell follow established formulations [25].

For a given timestep t with input x_t and previous hidden state h_{t-1}, the gates, cell state C_t, and output h_t are computed as follows (equation [9-14]):

$$\boldsymbol{i_t = ReLU(W_i x_t + U_i h_{t-1} + b_i)} \quad \textbf{Eq 9}$$

$$\boldsymbol{z_t = ReLU(W_Z x_t + U_Z h_{t-1} + b_Z)} \quad \textbf{Eq 10}$$

$$\boldsymbol{f_t = ReLU(W_f x_t + U_f h_{t-1} + b_f)} \quad \textbf{Eq 11}$$

$$\boldsymbol{C_t = i_t \otimes z_t + f_t \otimes C_{t-1}} \quad \textbf{Eq 12}$$

$$\boldsymbol{o_t = ReLU(W_o x_t + U_o h_{t-1} + V_o C_t + b_o)} \quad \textbf{Eq 13}$$

$$\boldsymbol{h_t = o_t \otimes C_t} \quad \textbf{Eq 14}$$

Where, W, U, and V are weight matrices, b are bias vectors, $\otimes$ denotes the Hadamard product, and $ReLU$ is the chosen activation function.

We used the ReLU activation function and the Adam optimizer. While the network's depth (number of layers) and width (neurons per layer) were held constant, model tuning focused on two key hyperparameters: the input sequence length (the number of prior days used as features) and the number of training epochs. These were optimized to enhance learning while actively mitigating overfitting, guided by literature on LSTM training dynamics [24, 26]. The training time complexity for our experiments is therefore primarily a function of these two variables as per equation 15.

$$\boldsymbol{Time\ Complexity \approx O(X * E)} \quad \textbf{Eq 15}$$

Where, X is the size of features is the number of days implemented by the model to predict the target day temperature, E is the number of Epochs.

ML Models Performance Evaluation:

The predictive performance of the SVR, LSTM, and LR models for the GLDAS air temperature data was evaluated using independent test datasets. Model accuracy and generalization capability were assessed using the coefficient of determination (R^2), Mean Square Error (MSE), and Nash–Sutcliffe Efficiency (NSE). The

coefficient of determination (R^2) [27] quantifies the proportion of variance in observed air temperature explained by the model predictions. It serves as a goodness-of-fit metric between simulated and observed values, (equation 16). The Nash–Sutcliffe Efficiency (NSE) evaluates model performance by comparing the variance of the prediction errors to the variance of the observed time series, as defined in equation 17. NSE values range from $-\infty$ to 1, where values approaching 1 indicate optimal model performance. An NSE value of zero corresponds to predictive skill equivalent to the mean of observations, while negative values indicate inferior performance [27]. Mean Square Error (MSE) measures the average squared difference between predicted and observed values and provides a direct estimate of the model's error magnitude, as shown in equation 18 [9]. However, the MSE can be decomposed into bias and variance components, as shown in equation 19. Bias reflects systematic model error resulting from oversimplified assumptions and is commonly associated with underfitting (equation 20), whereas variance represents sensitivity to noise and data variability, often leading to overfitting (equation 21) [9].

$$\boldsymbol{R^2 = \left(\frac{\sum_{i=1}^{n}(O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n}(\bar{O} - O_i)^2}\sqrt{\sum_{i=1}^{n}(\bar{P} - P_i)^2}}\right)^2} \qquad \textit{Eq 16}$$

$$\boldsymbol{NSE = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(\bar{O} - O_i)^2}} \qquad \textit{Eq 17}$$

$$\boldsymbol{MSE = \frac{\sum_{i=1}^{n}(O_i - p_i)^2}{n}} \qquad \textit{Eq 18}$$

$$\boldsymbol{MSE = bias^2 + variance + e} \qquad \textit{Eq 19}$$

$$\boldsymbol{bias^2 = (E[\bar{P} - O])^2} \qquad \textit{Eq 20}$$

$$\boldsymbol{variance = E((\bar{P} - P)^2)} \qquad \textit{Eq 21}$$

Where $O, P, \bar{O}, \bar{P}$ are the observed GLDAS dataset, the predicted data, the mean of the observed GLDAS dataset, and the mean of the predicted data respectively, n is the dataset size, e refers to a very small error, E refers to mean operation.

Analysis and Results

GLDAS Validation and Statistics:

To assess the suitability of the Global Land Data Assimilation System (GLDAS2.1) air temperature data for forecasting, which is much needed in Egypt for resilient agricultural expansion in an arid and changing climate, a direct validation against ground-based observations (OBS) was conducted for the study time frame (2000–2020). The validation metrics, the coefficient of determination (R^2) and Root Mean Square Error (RMSE) demonstrate a strong overall correlation between the GLDAS forced model and OBS across the study region, with site-specific variations.

Table. 1 Basic statistical analysis for OBS and GLDAS

City	OBS Elevation (m)*	Features	OBS	GLDAS
Alexandria	18	Max	32.8	31.54
		Min	8.3	9.536
		Range	24.5	22.01
		Mode	27.8	18.3
		Mean	21.2	21.23
		Std	5.213	4.569
Marsa Matrouh	25	Max	33.5	31.79
		Min	7.1	8.835
		Range	26.4	22.96
		Mode	26.4	16.67
		Mean	20.42	20.65
		Std	5.048	4.555
Sallum	40	Max	34.2	33
		Min	4.6	7.12
		Range	29.6	25.88
		Mode	25.6	16.59
		Mean	19.68	20.2
		Std	5.63	4.91
Siwa Oasis	-15	Max	38.6	38.77
		Min	7	6.967
		Range	31.6	31.8
		Mode	30.3	20.35
		Mean	23.02	22.85
		Std	7.061	6.838

** OBS: Weather station elevation*

The highest correlation was found for Marsa Matrouh (R^2 = 0.960, RMSE = 1.11°C) and Alexandria (R^2 = 0.954, RMSE = 1.23°C), indicating that GLDAS captures temporal temperature dynamics with high fidelity and minimal error at these coastal sites. Similarly, Siwa Oasis exhibits a very high correlation (R^2 = 0.950), though with a slightly higher RMSE of 1.59°C, suggesting a consistent bias, potentially related to the model's representation of the oasis microclimate. In contrast, Sallum shows a lower, though still strong, correlation (R^2 = 0.886) coupled with the highest RMSE (2.00°C). A suggestion of this performance is consistent with the previously identified time-series offset and discrepancy in minimum temperatures at this location, likely indicating a localized representativeness challenge in the model due to the topography or a scale mismatch between the point observation (weather station - OBS) and the GLDAS grid cell.

In synthesis, the high R^2 values (≥ 0.886) confirm that GLDAS reliably replicates the observed temporal variance and seasonal trends, a fundamental prerequisite for its use in forecasting. While the RMSE highlights site-specific biases, their magnitude (≤ 2.00°C) falls within an acceptable range for large-scale climatic applications and planning studies. The robust statistical validation, particularly for Alexandria, Marsa Matrouh, Sallum and

Siwa Oasis, provides a solid foundation for employing GLDAS data as a reliable input for temperature forecasting models in northwestern Egypt.

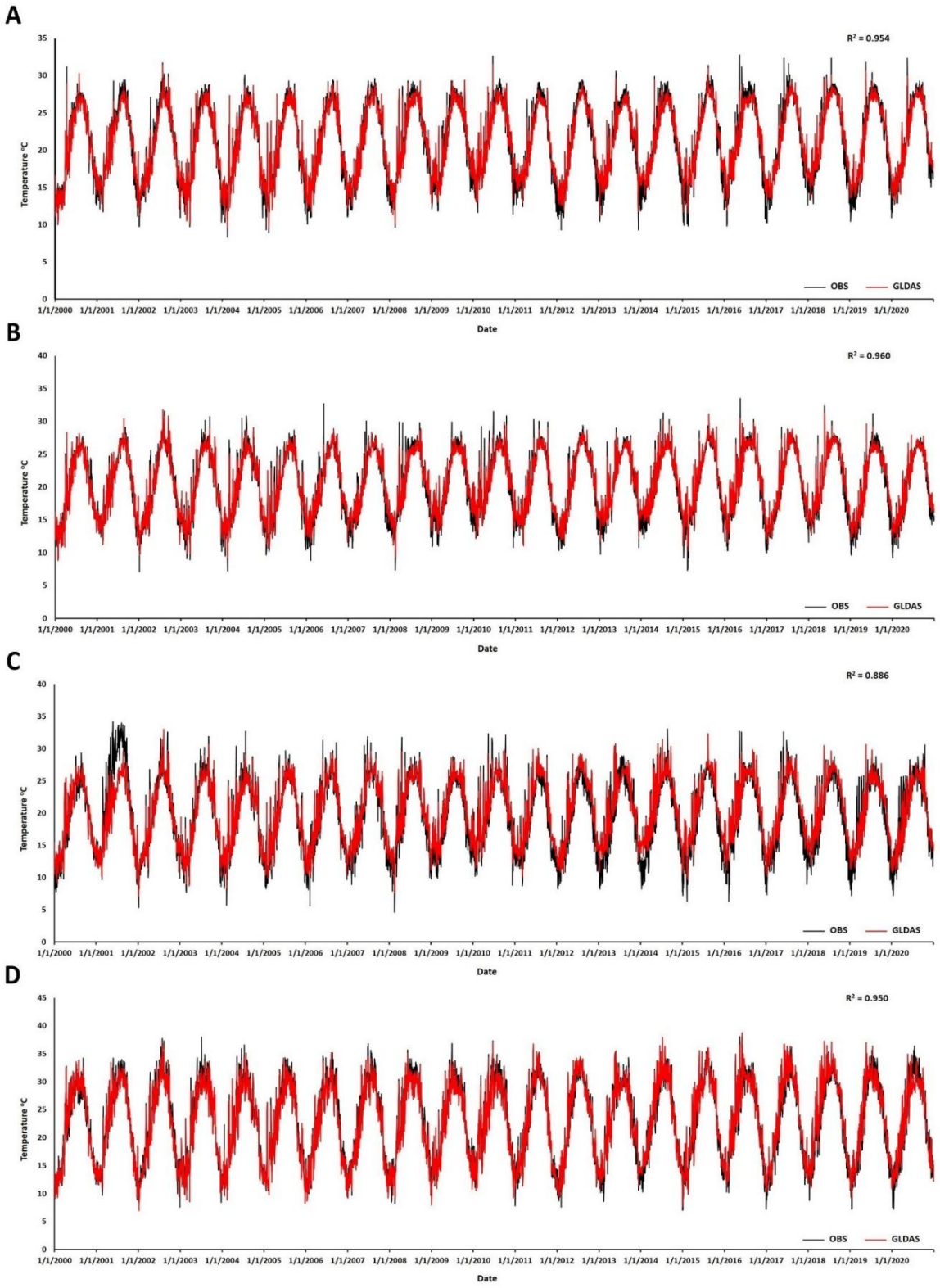


Fig. 3 OBS vs. GLDAS Near Air Surface Temperature for (A) Alexandria, (B) Marsa Matrouh, (C) Sallum and (D) Siwa Oasis

However, a comparative statistical analysis was conducted between the OBS and GLDAS models for the four study locations. The summary statistics, presented in Table 1, along with the corresponding time-series plots (Figures 3), demonstrate a generally high degree of concordance between the two datasets. An initial, systematic difference is noted in data precision: the OBS dataset is limited to one decimal place, while GLDAS values are provided to five decimal places. Despite this formatting discrepancy, the absolute differences in key statistical metrics are relatively small. The maximum temperature difference between GLDAS and OBS ranges from 0.1°C (Siwa Oasis) to 1.8°C (Marsa Matrouh), with all four locations exhibiting differences ≤1.8°C. Also, differences in minimum temperatures vary from 0.04°C (Siwa Osais) to 2.5°C (Sallum). The larger discrepancy in Sallum's minimum value is visually attributable to a pronounced divergence at the beginning of the record (2001-2002), after which the time-series curves converge and become more similar. The mean bias (GLDAS minus OBS) further quantifies this agreement, ranging from a negligible 0.023°C in Alexandria to 0.516°C in Sallum. The aforementioned offset directly influences the elevated mean bias in Sallum in minimum temperatures. Significantly, across all locations and metrics (max, min, mean), the absolute differences remain within a narrow range of ≤2.5°C, even in the most divergent case. The consistent, small magnitude of these statistical differences, corroborated by the strong visual covariation in the time-series plots, confirms that the GLDAS model captures the region's thermal climatology and temporal dynamics.

Feature Selection for Lagged Temperature Predictors:

The feature selection procedure was performed to quantify the number of preceding daily average air temperature observations that significantly contribute to accurately predicting the subsequent day's temperature. Pearson's correlation coefficient was employed as the selection criterion, using a fixed (static) threshold of 0.90 to retain only strongly correlated lagged features. For each location, the ten days preceding the target day were initially considered the candidate feature space, from which only lagged temperatures that met or exceeded the threshold were selected for model input. The results reveal apparent spatio-temporal variability in the temporal dependency structure of air temperature across the studied locations. For the GLDAS dataset, six previous days were retained as significant predictors for Alexandria, indicating a relatively strong and persistent temporal autocorrelation. In contrast, Marsa Matrouh, and Siwa Oasis exhibited more moderate memory effects, with three preceding days meeting the correlation threshold at each site. Sallum showed the shortest effective temporal dependency, where only two prior days demonstrated sufficiently strong correlation with the target temperature. However, the optimal number of lagged temperature inputs is location-dependent, reflecting differences in local climatic variability. The selected features were subsequently used as inputs for all machine learning models to ensure consistency across experiments and to reduce redundancy and overfitting in the predictive framework.

Forecasting Performance of Machine Learning Models:

The forecasting performance of three machine learning models, Linear Regression (LR), Support Vector Regression (SVR), and Long Short-Term Memory (LSTM), was evaluated using daily GLDAS air temperature data for four locations in Egypt: Alexandria, Marsa Matrouh, Sallum, and Siwa Oasis. For each location, feature selection was used to determine the optimal number of lagged daily temperature predictors, which varied with local temperature persistence. Model performance was assessed using Mean Squared Error (MSE), the coefficient of determination (R^2), the Nash–Sutcliffe Efficiency (NSE), and bias–variance decomposition diagnostics. The forecasting results are presented and analyzed separately for each of the four study locations to provide a detailed examination of the unique characteristics of each site and the optimal complexity and behavior of the machine learning models.

For the first location, Alexandria, the pre-feature selection for Alexandria identified six lagged daily temperatures as significant predictors. Consequently, six model configurations of increasing complexity were evaluated (Figure 4 (1)). The LR model performance exhibited a monotonic reduction in MSE from 0.964 °C to 0.679 °C as complexity increased. The bias–variance decomposition revealed that this error reduction was primarily driven by a decrease in variance, despite a slight increase in bias for the most complex model. The optimal six-feature LR model achieved an R^2 of 0.9676 and an NSE of 0.9674. In contrast, the SVR model's performance was non-monotonic with respect to complexity, with MSE reaching a minimum of 0.659°C at four features before deteriorating. This optimal configuration corresponded to the fewest support vectors, indicating an efficient nonlinear representation. The corresponding bias–variance analysis showed variance reduction up to the fourth model, after which bias began to increase. This model performed the highest R^2 and NSE values (0.9685). Likewise, the LSTM model: Performance varied with complexity, with the lowest MSE (0.723 °C) achieved using six features. However, the highest R^2 and NSE (0.9690) were attained by the five-feature model. The bias–variance diagnostics indicated that variance dominated the total error, suggesting LSTM performance is more sensitive to feature variance representation than to systematic error minimization. March 2017 was chosen for all comparisons for the direct comparison as March had the maximum temperature variation of the optimal models from each algorithm (Figure 4 (1-D)) demonstrated that SVR outperformed both LR and LSTM, achieving the lowest MSE with fewer predictors. The bias–variance comparison confirmed that SVR had the lowest variance, underscoring its superior generalization capability.

Moving west to Marsa Matrouh, feature selection retained three lagged predictors (Figure 4 (2)). The LR model performance improved modestly with increasing model complexity; however, the three-feature model achieved an MSE of 1.034 °C, an R^2 of 0.9541, and an NSE of 0.9539. Bias–variance diagnostics indicated concurrent reductions in both error components. Contrarily, the SVR model achieved optimal performance with all three features, reducing the MSE to 0.920 °C. This configuration corresponded to an intermediate number of support vectors, suggesting a balance between model flexibility and stability. Variance was the

dominant component of the error structure, and the model yielded an R^2 of 0.9592 and an NSE of 0.9590. Likewise, the LSTM model performed best with three features, achieving an MSE of 0.968 °C. Bias remained negligible, with performance improvements attributable primarily to variance reduction. The comparative analysis (Figure 4 (2-D)) consistently ranked SVR as the best-performing model, followed by LSTM and LR. SVR had the lowest variance error, which explains its superior predictive stability with the GLDAS.

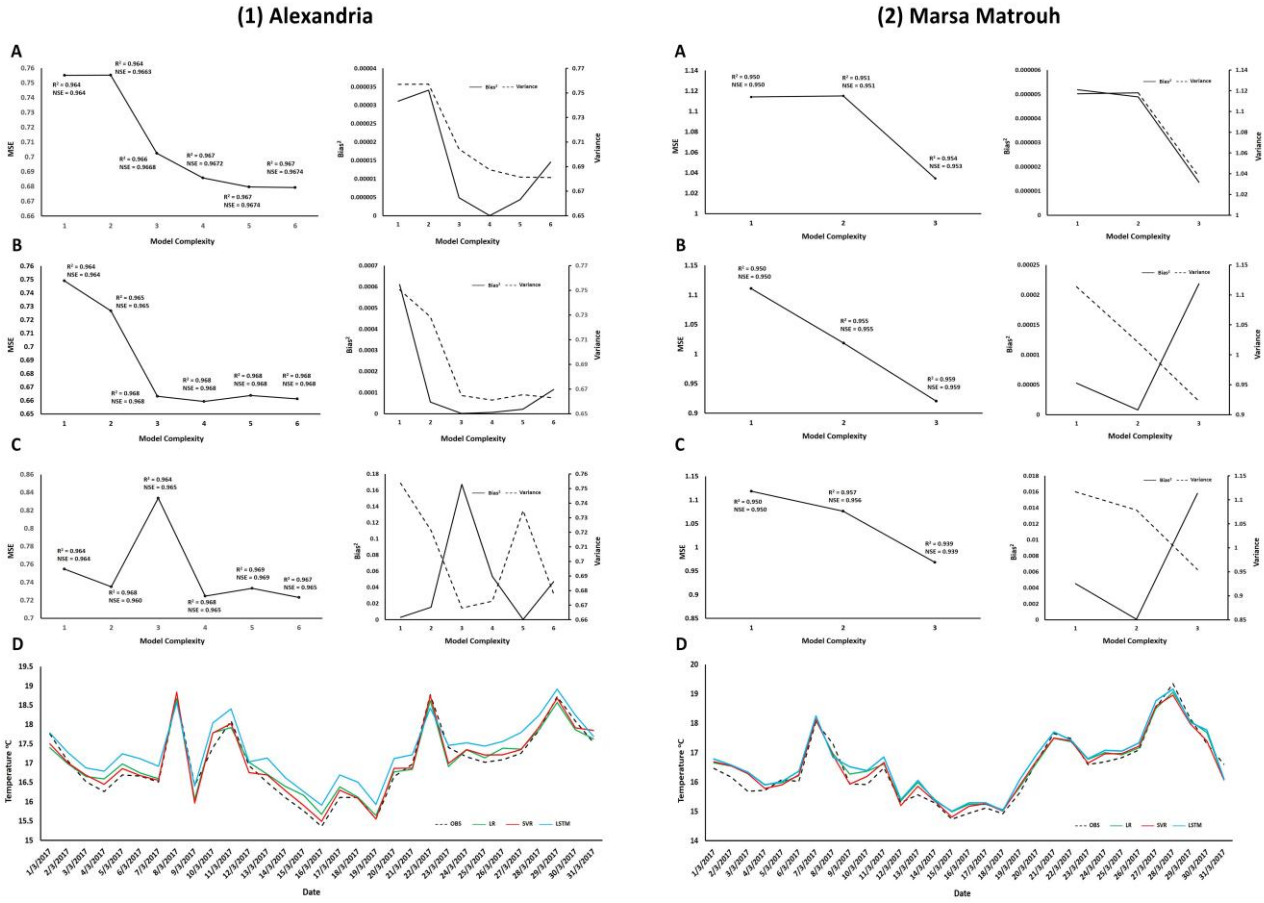


Fig. 4 Models evaluations and comparison for Alexandria and Marsa Matrouh – (A) LR Model, (B) SVR Model, (C) LSTM Model and (D) Models Comparison

For Sallum, the feature selection identified two lagged predictors (Figure 5 (1)). The LR model showed minimal performance gains with increased complexity, with MSE decreasing only slightly to 1.440 °C in the two-feature configuration. Variance decreased marginally while bias increased insignificantly. In contrast, the SVR model demonstrated more apparent improvement, with MSE reduced to 1.329 °C using two features. This improvement overlapped with a reduction in the number of support vectors and concurrent decreases in both bias and variance, resulting in an R^2 of 0.9429 and an NSE of 0.9428. While the LSTM model's performance was similar to LR, it achieved a slightly lower MSE (1.346 °C), likely due to its inherent memory effects. As with other locations, variance reduction dominated the error behavior. The comparative results (Figure 5 (1-D))

confirm that SVR is the most effective model for Sallum, providing a superior nonlinear representation and generalization, as evidenced by its bias–variance profile.

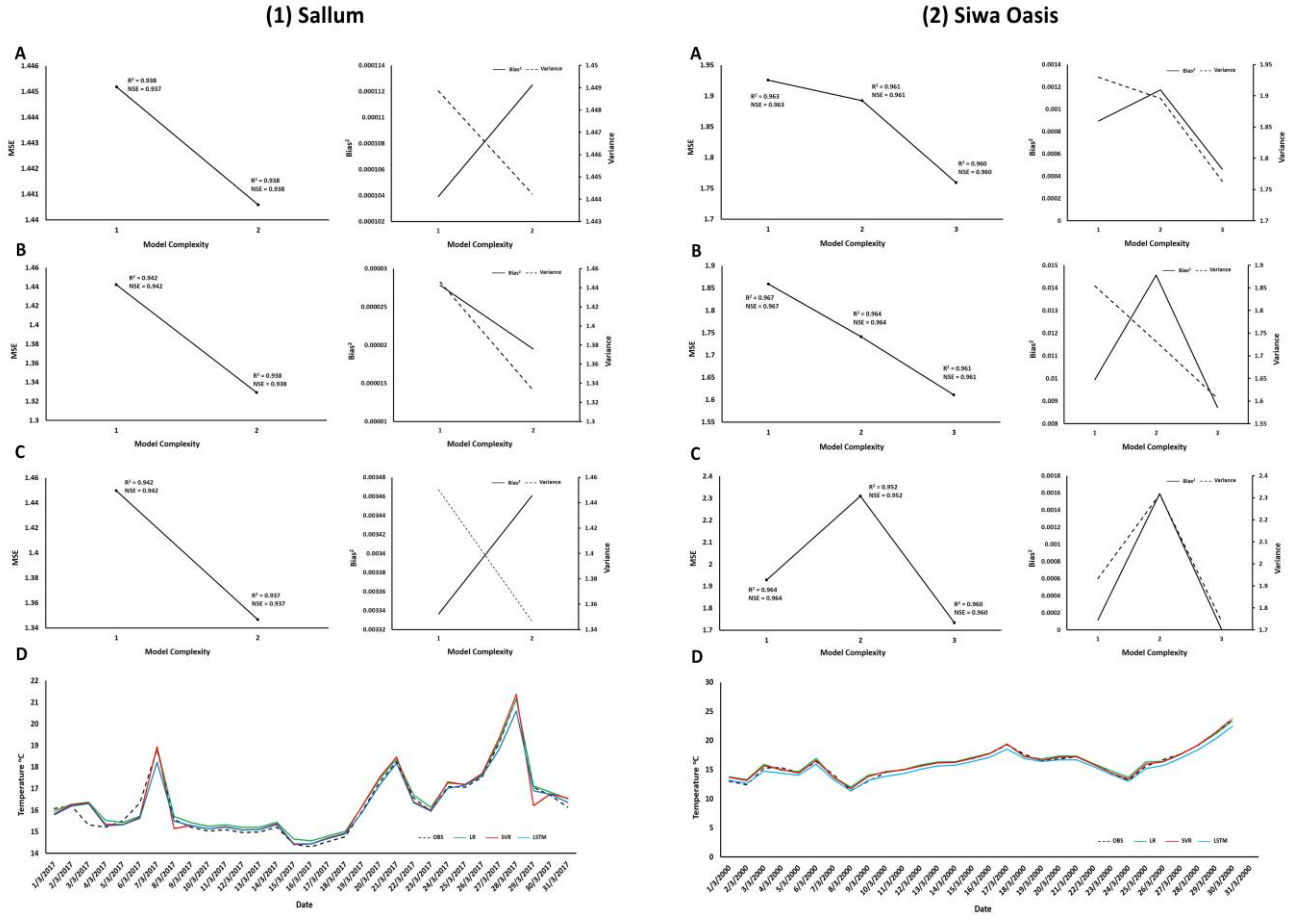


Fig. 5 Models evaluations and comparison for Sallum and Siwa Oasis – (A) LR Model, (B) SVR Model, (C) LSTM Model and (D) Models Comparison

Going to the heart of the western desert, for Siwa Oasis: Three lagged predictors were retained (Figure 5 (2)). In the LR model, the MSE decreased with complexity, reaching 1.759 °C with three features. Bias remained negligible, while variance decreased steadily, yielding an R^2 of 0.9638 and an NSE of 0.9637. In contrast, the SVR model achieved optimal performance with three features, reducing MSE to 1.610 °C. This configuration used the fewest support vectors and exhibited the lowest combined bias–variance error, resulting in the highest scores for both R^2 (0.9670) and NSE (0.9668. Moreover, the LSTM model performance degraded at intermediate complexity (two features) but was restored when all three predictors were included, achieving an MSE of 1.734 °C. This final configuration effectively minimized both bias and variance. The model comparison (Figure 5 (2-D)) confirms that SVR is the most resilient method for Siwa's Oasis GLDAS air temperature regimes, with its superior MSE performance driven by effective variance reduction.

Under the ML performance analysis across the four studied locations, SVR consistently delivered the best forecasting performance. It achieved the lowest MSE (ranging from 0.659 °C to 1.610 °C), the highest R^2 and NSE values, and the most optimistic bias–variance trade-off, often requiring fewer lagged predictors than LR and LSTM.

Conclusion

The convergence of GLDAS 2.1 Near Air Surface Temperature open-access and carefully designed machine-learning frameworks constitutes a powerful, scalable solution for temperature prediction in regions constrained by observational gaps. The demonstrated accuracy, consistency, and interpretability of the results provide a strong foundation for extending this framework to early warning systems, climate risk management, and sectoral applications such as agriculture, water resources, and energy planning. As climate variability intensifies and data demands grow, approaches that influence globally available reanalysis products and data-driven intelligence will be indispensable for building resilient, forward-looking climate services, particularly in arid and developing regions.

The results highlight a correlation-based feature selection framework that reveals apparent spatial heterogeneity in temperature memory effects, underscoring the need to account for local climatic persistence when designing predictive models explicitly. By adjusting the input feature space to each location, the modeling framework achieved both parsimony and performance, reducing redundancy while preserving essential temporal information. This approach directly addresses one of the central challenges in climate-oriented ML applications: balancing model complexity with generalization capability. From a predictive standpoint, the comparative analysis across Linear Regression, Support Vector Regression, and Long Short-Term Memory networks provides compelling evidence that model choice matters as much as data quality. While all three models benefited from high-quality GLDAS inputs, Support Vector Regression consistently delivered superior performance across coastal and desert environments alike. Its ability to achieve lower errors with fewer predictors, coupled with a favorable bias–variance trade-off, demonstrates its robustness for operational temperature forecasting in heterogeneous climatic settings. Further research is needed to support long-term extrapolation and spatial interpolation to address data uncertainty gaps.

References

1. Wang, H., Yang, J., Chen, G., Ren, C., & Zhang, J. Machine learning applications on air temperature prediction in the urban canopy layer: A critical review of 2011–2022. *Urban Climate*, *49*, 101499 (2023).
2. World Meteorological Organization. World weather—Egypt. *World Meteorological Organization website:* https://worldweather.wmo.int/en/country.html?countryCode=EGY (2024).
3. Rodell, M., Houser, P. R., Jambor, U. E. A., Gottschalck, J., Mitchell, K., Meng, C. J., ... & Toll, D. The global land data assimilation system. *Bulletin of the American Meteorological society*, *85*(3), 381-394 (2004).

4. Rui, H., Beaudoing, H., and Loeser, C.: README document for NASA GLDAS version 2 data products, NASA Goddard Earth Science Data Information and Services Center (GES DISC), Greenbelt, Maryland, USA, available at: https://hydro1.gesdisc.eosdis.nasa.gov/data/GLDAS/README_GLDAS2.pdf, (2022).

5. Roy, D. S. Forecasting the air temperature at a weather station using deep neural networks. *Procedia computer science*, *178*, 38-46 (2020).

6. Anjali, T., Chandini, K., Anoop, K., & Lajish, V. L. Temperature prediction using machine learning approaches. In *2019 2nd International conference on intelligent computing, instrumentation and control technologies (ICICICT)* (Vol. 1, pp. 1264-1268). IEEE (2019).

7. Safia, M., Abbas, R., & Aslani, M. Classification of weather conditions based on supervised learning for Swedish cities. *Atmosphere*, *14*(7), 1174 (2023).

8. Büyükkeçeci, M., & Okur, M. C. A comprehensive review of feature selection and feature selection stability in machine learning. *Gazi University Journal of Science*, *36*(4), 1506-1520 (2023).

9. Doroudi, S. The bias-variance tradeoff: How data science can inform educational debates. *AERA open*, *6*(4), 2332858420977208 (2020).

10. Yang, Z., Yu, Y., You, C., Steinhardt, J., & Ma, Y. Rethinking bias-variance trade-off for generalization of neural networks. In *International Conference on Machine Learning* (pp. 10767-10777). PMLR (2020).

11. Salman, A. G., Heryadi, Y., Abdurahman, E., & Suparta, W. Single layer & multi-layer long short-term memory (LSTM) model with intermediate variables for weather forecasting. *Procedia Computer Science*, *135*, 89-98 (2018).

12. Ibrahim, S. K., Ziedan, I. E., & Ahmed, A. Study of climate change detection in North-East Africa using machine learning and satellite data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 11080-11094 (2021).

13. Fister, D., Pérez-Aracil, J., Peláez-Rodríguez, C., Del Ser, J., & Salcedo-Sanz, S. Accurate long-term air temperature prediction with machine learning models and data reduction techniques. *Applied Soft Computing*, *136*, 110118 (2023).

14. Zenkner, G., & Navarro-Martinez, S. A flexible and lightweight deep learning weather forecasting model. *Applied Intelligence*, *53*(21), 24991-25002 (2023).

15. Stachura, G., Ustrnul, Z., Sekuła, P., Bochenek, B., Kolonko, M., & Szczęch-Gajewska, M. Machine learning based post-processing of model-derived near-surface air temperature–A multimodel approach. *Quarterly Journal of the Royal Meteorological Society*, *150*(759), 618-631 (2024).

16. Han, S., Liu, B., Shi, C., Liu, Y., Qiu, M., & Sun, S. *Evaluation of CLDAS and GLDAS datasets for near-surface air temperature over major land areas of China. Sustainability, 12, 4311* (2020).

17. Yu, L., Qiu, G. Y., Yan, C., Zhao, W., Zou, Z., Ding, J., ... & Xiong, Y. A global terrestrial evapotranspiration product based on the three-temperature model with fewer input parameters and no calibration requirement. *Earth System Science Data Discussions*, *2022*, 1-33 (2022).

18. Malathi, T., & Manimekalai, M. Feature selection techniques for weather forecasting models using machine learning techniques. *Journal of Electrical Engineering and Technology*, *11*(4), 443-455 (2020).

19. Patil, T., & Shah, D. K. Weather forecasting analysis using linear and logistic regression algorithm. *Int. Res. J. Eng. Technol.(IRJET)*, *8*(6), 2557-2564 (2021).

20. Gupta, I., Mittal, H., Rikhari, D., & Singh, A. K. Mlrm: A multiple linear regression based model for average temperature prediction of a day. *arXiv preprint arXiv:2203.05835* (2022).

21. Date, P., & Potok, T. Adiabatic quantum linear regression. *Scientific reports*, *11*(1), 21905 (2021).

22. Yin, G., Yoshikane, T., Yamamoto, K., Kubota, T., & Yoshimura, K. A support vector machine-based method for improving real-time hourly precipitation forecast in Japan. *Journal of Hydrology*, *612*, 128125 (2022).

23. Amroune, M. Support vector regression-bald eagle search optimizer-based hybrid approach for short-term wind power forecasting. *Journal of Engineering and Applied Science*, *69*(1), 107 (2022).

24. Hochreiter, S., & Schmidhuber, J. Long short-term memory. *Neural computation*, *9*(8), 1735-1780 (1997).

25. Zhang, C. J., Zeng, J., Wang, H. Y., Ma, L. M., & Chu, H. Correction model for rainfall forecasts using the LSTM with multiple meteorological factors. *Meteorological Applications*, *27*(1), e1852 (2020).

26. Rotman, M., & Wolf, L. Shuffling recurrent neural networks. In Proceedings of the *AAAI Conference on Artificial Intelligence* (2020).

27. Amaranto, A., Munoz-Arriola, F., Solomatine, D. P., & Corzo, G. A spatially enhanced data-driven multimodel to improve semiseasonal groundwater forecasts in the High Plains aquifer, USA. *Water Resources Research*, *55*(7), 5941-5961 (2019).

Funding Declaration

This research was conducted under the Precision Agronomic System for the New Delta Through 2030 (PAS2030) project, funded by the Science, Technology, and Innovation Funding Authority (STDF), under the Applied Sciences Research Grants #45957.

Authors Contributions

Both authors contributed to the conception, writing, and editing of the manuscript. Moataz H. Khalil conducted the data analysis and machine learning modeling, while Sani Z. Mohamed performed the data acquisition and validation.

Competing Interests

The authors declare no competing interests.

Data Availability

513 The data sets used and analyzed during the current study are available from the corresponding author upon
514 reasonable request.
515

516