

Research Article

Subseasonal Prediction of Summer Temperature in West Africa Using Artificial Intelligence: A Case Study of Senegal

Annine Duclaire Kenne ^{1,2} **Mory Toure** ^{3,4} **Lema Logamou Seknewna** ^{1,5}
and Herve Landry Ketsemen ⁶

¹*African Institute for Mathematical Sciences, M'Bour, Senegal*

²*African Masters of Machine Intelligence, M'Bour, Senegal*

³*Agence Nationale de l'Aviation Civile et de la Meteorologie (ANACIM), Dakar, Senegal*

⁴*Universite Cheikh Anta Diop (UCAD), Dakar, Senegal*

⁵*Centre Universitaire de Formation et de Recherche (CUFR), Mayotte, France*

⁶*University of Yaounde 1, Yaounde, Cameroon*

Correspondence should be addressed to Annine Duclaire Kenne; akenne@aimsammi.org

Received 18 May 2023; Revised 1 January 2024; Accepted 7 February 2024; Published 27 February 2024

Academic Editor: Yaxin Bi

Copyright © 2024 Annine Duclaire Kenne et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Despite the rapid growth of machine learning (ML) and its far-reaching applications in various fields such as healthcare, finance, and urban heat management, there are still some unresolved challenges in the field of climate change. Reliable subseasonal forecasts of summer temperatures would be a great benefit to society. Although numerical weather prediction (NWP) models are better at capturing relevant sources of predictability, such as temperatures, land, and sea surface conditions, the subseasonal potential is not fully exploited. One such challenge is accurate subseasonal temperature forecasting using cutting-edge ML technology. This study aims to assess and predict the changes in subseasonal temperature during the summer season (from March to June) in Senegal on 2-weeks time scales. Six ML techniques, including linear regression (LR), decision tree (DT), support vector machine (SVM), artificial neural network (ANN), long short-term memory (LSTM), and gated recurrent units (GRU), are used. The experiments utilize a multivariate approach by incorporating variables of the ERA-5 dataset from 1981 to 2022. The results compared all the performances of the methods to assess their overall effectiveness in forecasting air temperature (t2m) values over 2 weeks. Our analysis demonstrates that the GRU model outperforms the other ML models, achieving a Nash–Sutcliffe efficiency (NSE) score of 74.68% and a mean absolute percentage error (MAPE) of 2.51%. The GRU model effectively captures long-term dependencies and exhibits superior performance in temperature forecasting. Furthermore, a comparison between the observed and predicted values confirms the accuracy of the GRU model in aligning with actual temperature trends. Overall, this study contributes an impactful deep learning model to the field of subseasonal temperature forecasting in West Africa (Senegal), which offers local authorities the capability to anticipate climatic events and enact preventive measures accordingly.

1. Introduction

Monitoring and forecasting subseasonal temperature play a crucial role in investigating future climate patterns. The impacts of climate change and global warming have been identified as the greatest global health threat of the twenty-first century, exposing billions of people live to risks [1, 2], increasing the risks of water-related disasters (e.g., floods and severe droughts) [3, 4], affecting the water quality in

many regions in the world [5], and might also have an impact on the urban climate environment, which can affect the urban climate, environment, and sustainable development [6–8].

Consequently, it requires a reassessment of how we address the protection of vulnerable populations [9, 10]. The World Health Organization (WHO) acknowledges the negative health impacts associated with climate change [11]. Moreover, regions with weaker healthcare systems face

greater challenges in adapting, preparing, and responding to the escalating health risks arising from a changing climate [12, 13].

One of the manifestations of climate change is the increase in average temperatures and the occurrence of extreme temperature events, commonly referred to as heat waves [14]. This rise in temperature has been observed around the world and is expected to become more intense and frequent due to human activities [15, 16]. However, in less-developed countries like Senegal, there is a lack of comprehensive studies on extreme temperatures and their specific health impacts [17]. Senegal, a country in West Africa, is facing the effects of climate change, which have a significant impact on the health of its population [18, 19]. Sweltering temperatures, which are becoming more frequent and intense, represent one of the most evident impacts of climate change [20]. They can cause health problems such as heat strokes, dehydration, cardiovascular and respiratory diseases, and even death, especially among the most vulnerable individuals, such as the elderly and children [21]. According to the literature, an intensification and exacerbation of prolonged periods of extreme temperatures could be expected in the coming years, making the study and understanding of these phenomena crucial for preventing harmful effects on human health [22].

Accurate forecasting of summer temperatures with subseasonal lead times has become increasingly important for various applications, including agriculture, energy demand, and, most notably, public health [23]. Subseasonal prediction can provide valuable insights for planning and decision-making processes, especially in protecting vulnerable populations from heat-related risks [24]. Therefore, forecasting such temperatures with 100% might not be possible, but the prediction error or computational speed can be minimized [25].

In the past, machine learning (ML) methods have proven to be the most promising tools for their capabilities in capturing complex patterns, nonlinearity, and accuracy to predict natural phenomena such as temperature, precipitation, drought, flood, and soil temperature [2–4, 26–28]. ML-based neural networks such as WNN have also proven their ability to model photovoltaic power plants [29]. Subseasonal temperature forecasts allow for estimating temperatures for the upcoming weeks, offering the opportunity to proactively plan prevention and adaptation measures.

Linear regression (LR) is one of the oldest and most common ML methods used in different areas for regression tasks. LR is applicable to single or multiple variable problems [30–32]. Decision tree (DT) regression is another common ML method used in the literature to solve regression problems [33, 34]. Support vector machine (SVM) regression is also a well-known and typical ML method for determining the link between features and targets [32, 35–37]. Artificial neural network (ANN), one of the most popular ML methods, has been used to solve many ML problems in different areas [38–41]. It is capable of identifying nonlinear patterns in the functional connection between features and targets. Deep neural network methods, named long short-term memory (LSTM) and gated

recurrent unit (GRU), also used in this study, have proven effective in time series prediction by capturing patterns from sequential data and retaining them in internal state variables [42–44]. The ability of these methods to retain crucial information over time enables them to effectively classify, process, and predict complex dynamic sequences.

The aim of this study is to demonstrate the capability and effectiveness of ML methods for modeling subseasonal summer temperatures in the Sahel (Senegal). Although most studies have already been conducted using traditional complex approaches and single machine learning methods for temperature prediction, few studies have discussed the impact of changes in summer air temperatures on a subseasonal scale. Furthermore, these studies discussed the change in summer temperature using a univariate approach, the numerical weather prediction (NWP) approach, simple ML methods (e.g., RF and LR), or simply used artificial neural networks (e.g., ANN) to predict air temperature. Among these studies, only a few researchers have combined multiple ML methods, compared their performance, or used deep learning methods with multivariate approaches.

Forecasting changes in summer temperature on a subseasonal scale using different (advanced) ML models is therefore a new and crucial task. In this investigation, this topic is further investigated on the basis of previous studies. This study primarily improves the detection and forecasting of subseasonal temperatures during the summer season from March to June (MAMJ) in the Senegal region by comparing the performance of six machine learning and deep learning models, including LR, DT, SVM, ANN, LSTM, and GRU based on multivariate approaches.

The remainder of this paper is summarized as follows: Section 2 reviews related work. Section 3 describes the study area, the data, the methods, and the evaluation metrics used. Section 4 describes the experiments and results. Section 5 describes the discussion of the results, and Section 6 draws some conclusions and suggestions for future work.

2. Related Work

Machine learning, especially deep neural networks, has emerged as a powerful tool for subseasonal forecasting, as it can effectively capture complex and nonlinear relationships between climatic variables [45]. Recent studies such as Guigma et al. [46] and Domeisen et al. [47] have examined the ability to forecast heat waves or extreme temperatures in the Sahel with subseasonal lead time and the importance of atmospheric tropical variability modes using reanalysis data from the European Center for Medium-Range Weather Forecasts (ECMWF). They found that temperatures in the Sahel can be predicted up to 3–4 weeks in advance with better accuracy and that the predictive capacity can be improved by incorporating atmospheric tropical variability modes such as the Madden-Julian Oscillation (MJO) and the African Easterly Jet (AEJ) [46, 48, 49].

Van Straaten et al. [50] and Benet et al. [51] used machine learning-based explainable forecasting (random forest) to discover subseasonal drivers of high summer temperatures in western and central Europe. They found that

atmospheric circulation patterns, soil moisture, and sea surface temperature anomalies are important predictors of high summer temperatures [52]. Zhang et al. [44] demonstrated the effectiveness of artificial neural networks (ANN) in predicting LULC and land surface temperature (LST).

Benet et al. [51] investigated the subseasonal forecasting of summer heat waves in Central Europe using linear and random forest (RF) machine learning models. They found that the RF model outperformed the linear model and that anomalies in sea surface temperature were important predictors of central European summer heat waves. Sharaff and Roy [53] and Anjali et al. [54] analyzed the performance of multilinear regression (MLR), support vector machine (SVM), ANN, and regression tree methods to predict daily temperature values using data collected from Mumbai Chhatrapati Shivaji Airport (2001–2016) and Central Kerala (2007–2015). They compared the results based on mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and correlation coefficient metrics. ANN and MLR revealed the best performances on the Mumbai (MSE = 3.46) and Central Kerala datasets, respectively.

The linear regression (LR), random forest (RF), gradient boosting machine (GBM), decision tree (DT), ANN, and SVM regression methods have also shown their capability in estimating the daily air temperature [55]. Among these ML models, GBM showed the best performance (coefficient of determination $R^2 = 0.68$, MAE = 1.60, and RMSE = 2.03). Zhang et al. [56] developed a seasonal forecasting of the frequency of heat waves or extreme temperatures in the Eurasian summer using convolutional neural networks (CNN). They showed that the deep learning model can predict the frequency of heat waves with high accuracy up to 3 months in advance and that the North Atlantic Oscillation and the Western Pacific Subtropical High are important predictors of the frequency of heat waves in the Eurasian summer [57]. Lu et al. [58] developed a data-driven global subseasonal forecast model (GSFM) for intraseasonal oscillation components using deep learning. They found that GSFM outperformed the traditional statistical model and that the MJO is a key driver of intraseasonal oscillation [48].

Weyn et al. [59] investigated subseasonal forecasting with a large ensemble of deep learning weather prediction models. The results of their study showed that the ensemble model outperformed the individual models and that sea surface temperature anomalies were important predictors of subseasonal weather. In summary, the literature suggests that deep learning models can effectively predict subseasonal summer temperature and that the inclusion of additional predictors such as atmospheric circulation patterns, soil moisture, sea surface temperature anomalies, and intraseasonal oscillations can improve prediction accuracy.

More research is needed to investigate the optimal combination of predictors and deep learning (DL) models for subseasonal summer temperature forecasting in specific regions, such as West Africa (Senegal). Such a study has not yet been investigated in this particular area using advanced machine learning techniques. Therefore, the main contribution of this work is to compare six machine learning

techniques named LR, DT, SVR, ANN, LSTM [60], and GRU [61, 62] based on the multivariate approach to forecast summer temperature in the region of West Africa (Senegal).

In the following sections, further explanation about the data and methods, experiments, and results is provided, followed by a discussion and conclusion with future work.

3. Study Area, Data, and Methods

3.1. Study Area. Senegal, located on the West coast of Africa [63], serves as a captivating study area for our research focused on the prediction of subseasonal summer temperatures with machine learning (ML) techniques. With a diverse landscape ranging from the arid regions of the Sahel to the verdant Casamance in the south, Senegal has unique climatic conditions that make accurate temperature prediction difficult. It is important to be able to anticipate these climate events and take the appropriate measures. The study seeks to leverage advanced ML techniques to enhance our understanding of subseasonal variations in summer temperatures, exploring patterns, trends, and influencing factors. The implications of successful prediction models extend beyond meteorological understanding, encompassing applications in agriculture, water resource management, and climate change adaptation. As Senegal grapples with the impacts of a changing climate, this research aims to contribute valuable information on the dynamics of subseasonal temperature, ultimately aiding in the development of more resilient and adaptive strategies for the region. Additionally, health authorities constantly require information on heat conditions, preferably with a lead time of at least 7 days to one month, in order to implement necessary measures to protect vulnerable populations [64]. Figure 1 highlights the location of the study area.

3.2. Data. The data used for this study are from the ERA-5 dataset provided by the European Center for Medium-Range Weather Forecasts (ECMWF, [65]). The parameters used, which are given in Table 1, were collected in the ERA-5 database from 1981 to 2022. The data with the highest correlation are selected using the Pearson correlation coefficient [66] and entered into the prediction model. Figure 2(a) shows the development of temperatures from 1981 to 2022 in the form of stripes, which show an exponential rise in temperature over West Africa (Senegal). Figure 2(b), on the other hand, shows the distribution of temperatures according to monthly averages. This distribution clearly shows that April, May, and June are the hottest months of the year. This is the reason why we have chosen the months of March to June for the predictions. Figure 3 shows the anomaly of the temperature, the trends, and the climatological reference in Senegal from 1981 to 2022, which is interpreted as follows:

- (i) The blue curve represents the temperature anomaly compared to the climatology (1991–2020). It shows how temperatures vary relative to the climatic average during this period. Positive values indicate temperatures above the average, while negative values indicate temperatures below the average.

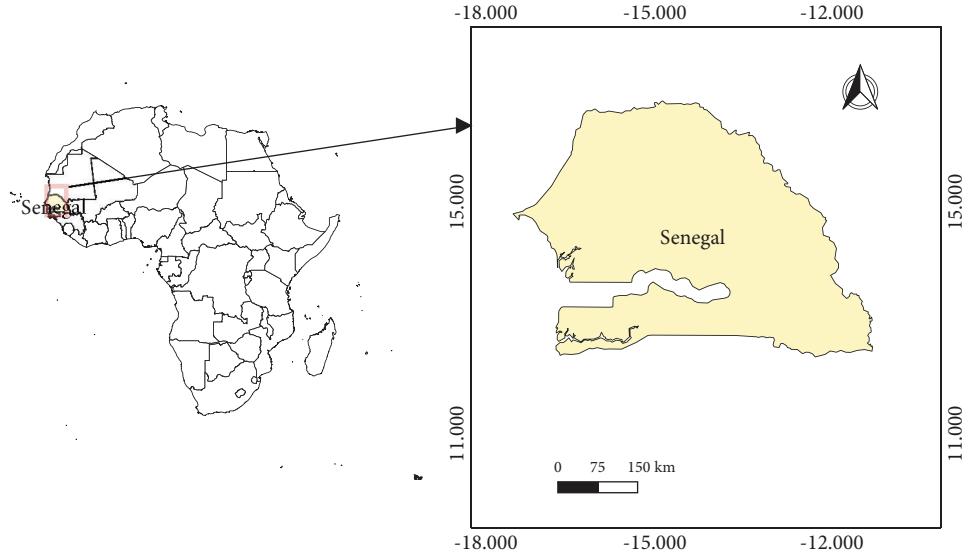


FIGURE 1: Map of the study area in West Africa (Senegal).

- (ii) The horizontal dashed line at $y = 0$ represents the climatological reference. Points above this line indicate temperatures higher than the climatic average, while points below indicate temperatures lower than the average.

In summary, the figure illustrates how temperatures in Senegal have evolved over the years, highlighting variations compared to the climatic average and indicating the general trend using the linear regression.

3.3. Machine Learning Methods. Machine learning (ML) can improve subseasonal forecasts by incorporating large amounts of historical data and learning patterns that can be used to make more accurate predictions. ML techniques [67] can capture predictability on subseasonal time scales and can outperform climatological baselines [51]. Machine learning models, called linear regression (LR), decision tree (DT), support vector machine (SVR), artificial neural network (ANN), long short-term memory (LSTM), and gated recurrent unit (GRU), are used for prediction with a dataset of size 2749026, using data from 1981–2010, 2010–2015, and 2015–2022 for training, validation, and test sets, respectively. For these datasets, we use the shift function to transform the time series data into a supervised learning problem. Figure 4 shows an overview and a simple representation of the ML models (LR, DT, SVR, and ANN).

3.3.1. Linear Regression (LR). Linear regression (LR) is the most common predictive machine learning (ML) model to find a linear relationship between one or more predictors. The LR, described by equation (1), can be single regression or multilinear regression [30, 31]. The single regression minimizes a single objective function based on the single variable.

$$y = \beta x + \epsilon, \quad (1)$$

TABLE 1: Variables description of ERA-5 data.

Variable	Description	Unit
Div200	Divergence at 200 hpa	s^{-1}
Div700	Divergence at 700 hpa	s^{-1}
Div850	Divergence at 850 hpa	s^{-1}
Rh1000	Relative humidity at 1000 hpa	%
t2m	Air temperature at 2 m	$^{\circ}C$
td	Dew-point temperature at 2 m	$^{\circ}C$
t850	Temperature at 500 hpa	$^{\circ}C$
z500	Geopotential at 500 hpa	$m^2 s^{-2}$
vent10	Wind at 10 m	ms^{-1}

where y is the dependent variable, which can be either a continuous or categorical value, and x is the independent variable.

In this case, multilinear regression (MLR), also known as multiple regression, is used. MLR is the most common form of linear regression analysis. In contrast to simple linear regression, MLR establishes the relationship between the response variable (target) and the predictors. With multiple variable regression, the method not only uses a single objective function but also combines the individual objective functions of each variable using a weighting scheme [68]. The parameters L1 and L2 are used to measure the elasticity in the weighting scheme. When configured, the approach minimizes the difference between the line and the best-fit criterion, yielding the best-fit line for prediction. Multiple regression is represented by the following equation:

$$y_i = b_0 + b_1 x_{1,i} + b_2 x_{2,i} + \dots + b_k x_{k,i} + \epsilon, \quad (2)$$

where y is the number of observations x_i is the number of inputs b_0 is the constant or intercept coefficient for the first and n^{th} explanatory variables b_1, b_2, \dots, b_k are the slope coefficients for the first and n^{th} explanatory variables. $y_i, x_{1,i}, \dots, x_{n,i}$ represent the i^{th} observation of each variable $y, x_1, \dots, x_{k,i}$, respectively. ϵ is the remaining unexplained

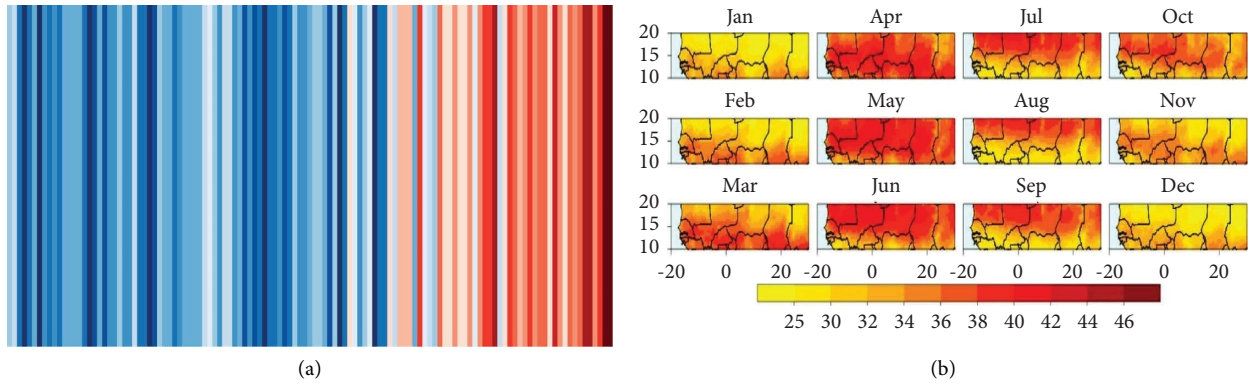


FIGURE 2: Evolution of Senegal's temperatures, 1981-2022. (a) Evolution of Senegal's temperatures. (b) Distribution of Senegal's temperatures.

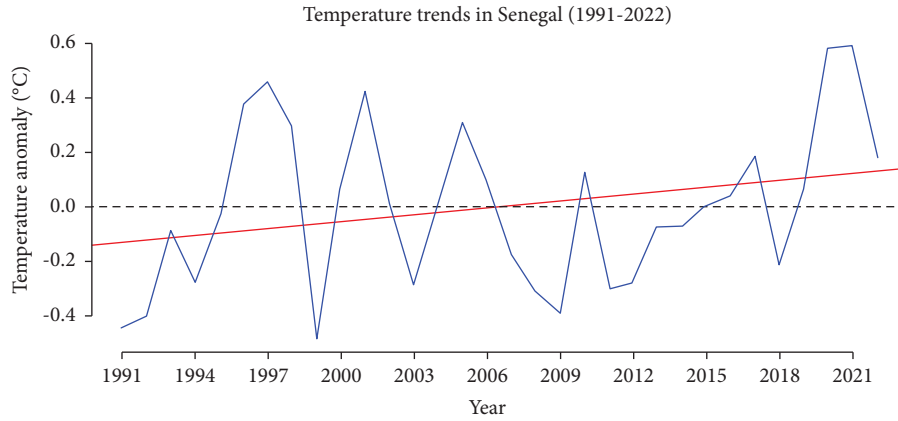


FIGURE 3: Temperature trends in Senegal from 1981 to 2022.

noise in the data, i.e., a random error term representing the remaining effects on y of variables not explicitly included in the model b_0 and b_i are unknown and, therefore, are estimated from the data. So, these values b_0 and b_1, b_2, \dots, b_k are

estimated by employing the least square criterion with minimum sum of square of error terms (MSS); this allows to find the values b_0, b_1, b_2, \dots and b_k that minimize MSS.

$$MSS = \frac{1}{n} \sum_{i=1}^n (y_i - b_0 - b_1 * x_{1,i} - b_2 * x_{2,i} - \dots - b_k * x_{k,i})^2 = \frac{1}{n} \sum_{i=1}^n e_i^2,$$

$$MSS = \frac{1}{n} \sum_{i=1}^n (y_i - y_{ipred})^2, \tag{3}$$

$$MSS = \frac{1}{n} \sum_{i=1}^n e_i^2,$$

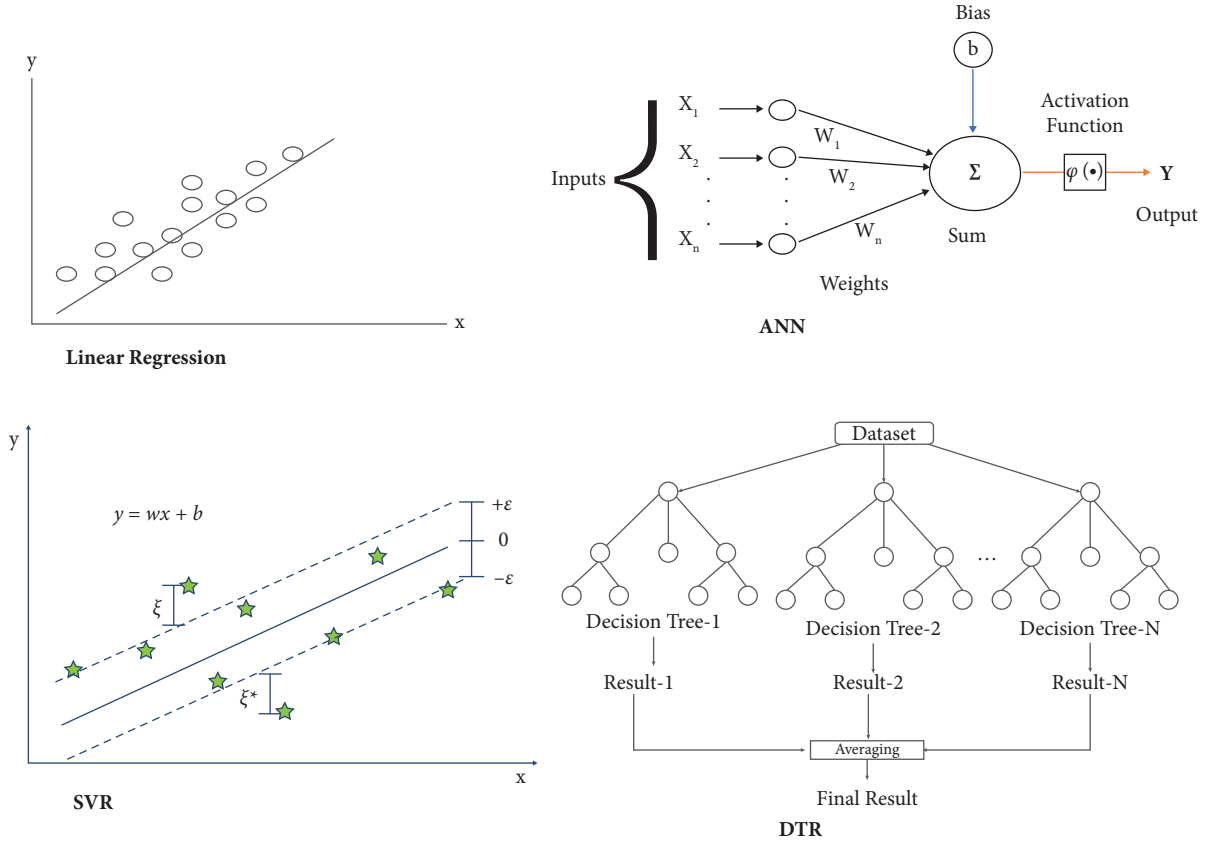


FIGURE 4: Overview of the applied LR, SVR, DTR, and ANN models with Y denoting the estimated T2m and X the matrix of the observed features.

y represent the observation y_{ipred} : simulated values $b_0, b_1, b_2, \dots, b_k$ must satisfy the following equation:

$$\frac{\partial \text{MSS}}{\partial b_j} = 2 \sum_{i=1}^n e_i \frac{\partial e_i}{\partial b_j}, \quad j = 0, 1, \dots, k. \quad (4)$$

Different values of the hyperparameter alpha are tested during the training of the LR model and the values alpha = 7 are selected for the final model.

3.3.2. Decision Tree Regression (DTR). The decision tree regression (DTR) model [69] is a machine learning algorithm used for predicting continuous numerical outcomes. DTR works by recursively partitioning the dataset into subsets based on the values of input features, creating a tree-like structure [70]. At each internal node of the tree, a decision is made based on a specific feature, and the dataset is split into two or more branches [71]. This process is repeated until a stopping criterion, such as maximum depth or a minimum number of samples per leaf, is reached. The leaf nodes of the tree represent the final predicted values. The prediction for a new instance is then the average of the target values in the leaf node to which it belongs.

3.3.3. Support Vector Machine. Support vector machine (SVM) regression, a machine learning technique, finds the hyperplane that best separates the optimal trade-off between fitting the data closely and maintaining a margin. The vector lengths and variance between the feature and the plane are minimized. Additionally, SVM regression can incorporate kernel functions to map the input features into a higher-dimensional space, allowing it to capture nonlinear relationships. The kernels include Euclidean, Gaussian, Exponential, and Dirichlet kernels [72]. The following equation describes the general form of SVR. The SVR method is trained with four different kernels: linear ($k(x, z) = x^T z$), rbf ($k(x, z) = \exp(-(x - z)^2 / \sigma^2)$), poly, and sigmoid. The rbf kernel is then selected for the final model.

$$f(x) = \mu + w^T X, \quad (5)$$

where μ is the base or bias w is the weight X is the number of inputs $y = f(x)$ is the target

Equation (5) is also called the equation of a hyperplane. The SVR model is trained by solving the optimization problem states in equation (6) according to the constraints described in equations (7)–(9).

$$\min \left(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^I (\xi_i + \xi_i^*) \right), \quad (6)$$

subject to

$$y_i - w^T \phi(x_i) - b \leq \epsilon + \xi_i, \quad i = 1, \dots, I, \quad (7)$$

$$-y_i - w^T \phi(x_i) + b \leq \epsilon + \xi_i^*, \quad i = 1, \dots, I, \quad (8)$$

$$\xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, I, \quad (9)$$

where w controls the smoothness of the model ξ_i, ξ_i^* are slack variables ϕ is the function of projection of the input space to the feature space b is the parameter of bias y_i is the target value to be estimated

3.3.4. Artificial Neural Network (ANN). The artificial neural network (ANN) is a machine learning technique commonly used in temperature prediction [38, 44, 54]. The error backpropagation learning algorithm, commonly known as the least-mean-square algorithm (L.M.S. algorithm) [30, 73] is used by the ANN to capture the nonlinear pattern between the input and target series. The widely used multilayer feed-forward ANN, called a multilayer perceptron (MLP) which consists of one input layer, one or more hidden layers of computational nodes, and one output layer, is used in this study due to its popularity [74–76]. The ANN trained with a hidden layer (equation (10)) size of $\{(50, 50)\}$, the ReLu ($\text{relu}(x) = \max(0, x)$) activation function, and the widely used ADAM optimizer [77] is selected as the final model configuration.

$$\bar{y}_k(t) = \phi_0 \left[\sum_{j=1}^m w_{kj} * \phi \left(\sum_{i=1}^N w_{ij} x_i + b_{j0} + b_{k0} \right) \right], \quad (10)$$

where $\bar{y}_k(t)$ is the predicted k th output (t2m) at time step t . w_{ji} is the weight that connects the i th neuron in the input layer and the j th neuron in the hidden layer. ϕ_0 is the activation function for the output neuron. ϕ is the activation function of the hidden neuron. w_{kj} is the weight that connects j th neuron in the hidden layer and the k th neuron in the output layer. b_{j0} is the bias for the hidden j th hidden neuron. b_{k0} is the bias for the k th neuron. m is the number of hidden neurons. N is the number of samples. $x_i(t)$ is the i th output variable at the time step used.

3.3.5. Recurrent Neural Networks (RNNs). The RNNs [78, 79], one of the deep learning (DL) methods, are deep forward neural networks (DFNN) with one-way information flow during the training phase. Some experience ([80]) has shown that RNNs can encounter significant challenges when training sequences with long-term dependencies, where temporal information between input and output sequences spans a long period of time. Accordingly, this can lead to difficulties related to backpropagation errors as the time span of long-term dependencies increases. As a result of the vanishing gradient problem in RNN training, updates to the weights become insignificant when the error is backpropagated [81].

Unlike canonical RNN, the LSTM (Section 3.3.6) and GRU (Section 3.3.7) models do not have these difficulties because they have extended memory components, such as memory cells and gates, that store long sequences of information over long periods of time. For time series, as in its generality, a simple encoder-decoder for multilevel forecasting with RNNs such as LSTM and GRU is used. Let us assume a set of N time series. The setting consists of a multivariate time series $(x_t, y_t), t = 1, \dots, T$, with a feature vector $x_t \in \mathfrak{R}^N$ and a target variable $y_t \in \mathfrak{R}$. Assuming that at time t , the values at the time points $(t+1, \dots, t + \text{HORIZON})$ that depend on the previous T values in the time series are predicted. Let us denote $D = (X_t, Y_t)_{t=1}^N$ as the training dataset. The input feature matrix X_{t+1} contains all past and current features $x_{i=i-1}^{t+1}$, as well as all past targets $y_{i=i-1}^{t+1}$. Thus, the function to approximate is given by the following equation:

$$Y_t = f(y_{t+k-1}, \dots, y_{t+k-1}, X_{t-1}, \dots, X_{t-T}), \quad (11)$$

where y_t is the t2m at point t and X_t is the additional regressor at point t . For this case study, $T = 25$ and $k = \{1, 2\}$. The architecture of our deep neural network models (LSTM 3.3.6 and GRU 3.3.7) derived from the general structure of RNN is given in Figure 5.

3.3.6. Long Short-Term Memory (LSTM). The LSTM unit [60, 82, 83] has three different gates, namely, the input gate, the forget gate, and the output gate. The forget gate determines what relevant information is required from the previous steps. The input gate determines what relevant information can be added from the current step, and the output gate completes the next hidden state. The internal structure of an LSTM unit cell according to [84] is shown in Figure 6. The LSTM cell model is defined by the following mathematical functions:

$$\begin{aligned} f_t &= \text{sig}(x_t W_f + h_{t-1} U_f + b_f), \\ i_t &= \text{sig}(x_t W_i + h_{t-1} U_i + b_i), \\ o_t &= \text{sig}(x_t W_o + h_{t-1} U_o + b_o), \\ \tilde{C}_t &= \tanh(x_t W_c + h_{t-1} U_c + b_c), \\ C_t &= \tanh(f_t * C_{t-1} + i_t * \tilde{C}_t), \\ h_t &= o_t * \tanh(C_t), \end{aligned} \quad (12)$$

where t is the time step; $x_t \in \mathfrak{R}^d$ is the input; $h_t \in (-1, 1)^h$ is the hidden state vector (output vector of the LSTM unit); h_{t-1} is the previous hidden state; b_i is the bias vector at t . W_f is the weight matrix between forget gate and input gate. $i_t \in (0, 1)^h$ is the input gate at t ; $o_t \in (0, 1)^h$ is the output gate activation vector at t ; $C_t \in (-1, 1)^h$ is the cell activation vector or value generated by tanh; $c_t \in \mathfrak{R}^d$ is cell state information; W_p and U_p contain the weights of the input and recurrent connections, where the subscript p refers to the input gate i , the output gate o , the forgetting gate f , or the memory cell unit c . In addition, $W \in \mathfrak{R}^{h \times d}$, $U \in \mathfrak{R}^{h \times h}$, and $b \in \mathfrak{R}^h$ are weighting matrices and bias vector parameters to be learned during the training phase,

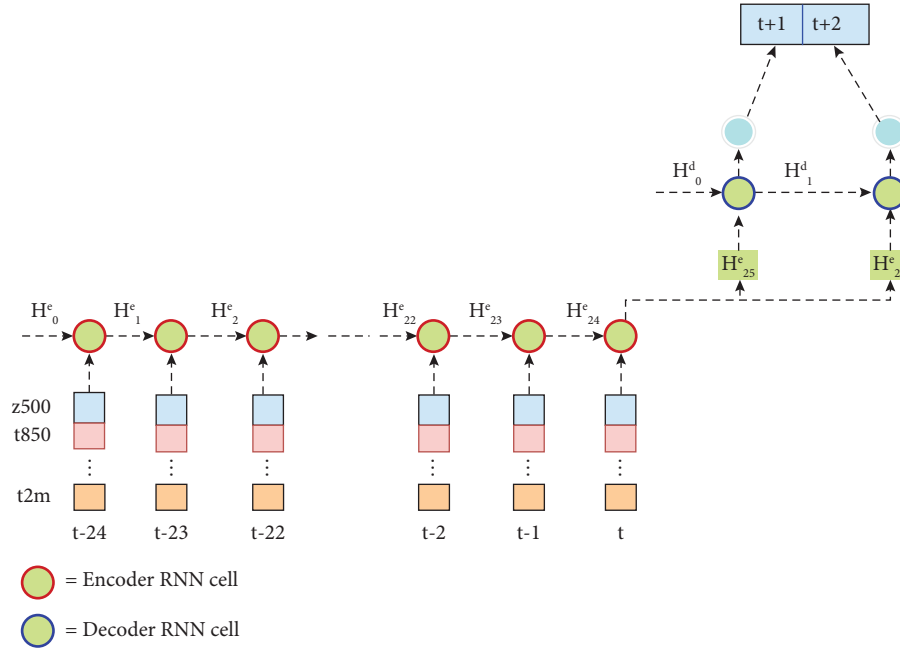


FIGURE 5: Simple encoder-decoder for multistep forecast.

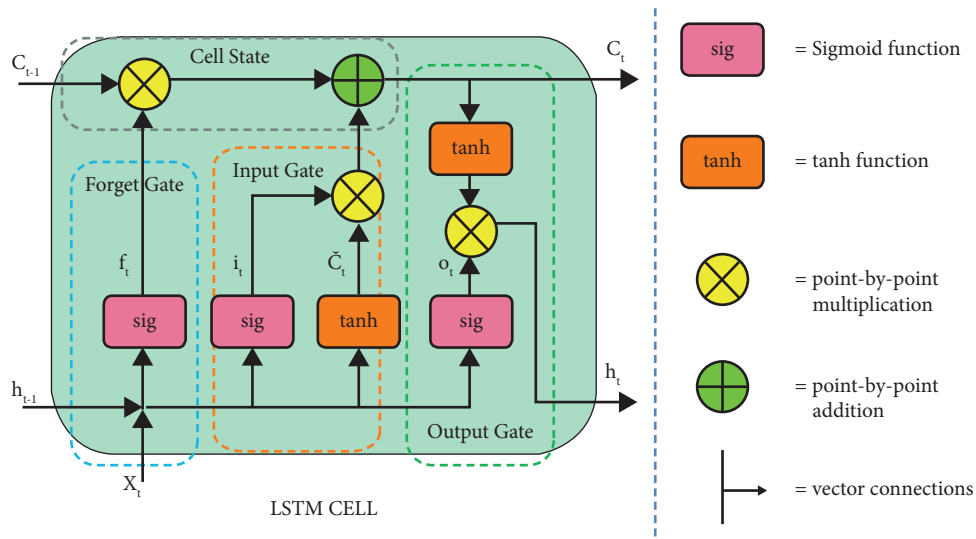


FIGURE 6: The cell structure of a long short-term memory (LSTM) unit.

where the indices d and h refer to the number of input features and the number of hidden units. $*$ denotes element-wise multiplication. $\text{sig}()$ and $\text{tanh}()$ are the logistic sigmoid and hyperbolic tangent functions with the following mathematical formula:

$$\begin{aligned} \text{sig}(x) &= \frac{1}{1 + \exp(-x)}, \\ \text{tanh}(x) &= \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}. \end{aligned} \tag{13}$$

3.3.7. *Gated Recurrent Unit (GRU)*. The gated recurrent unit (GRU) introduced by [61, 62] is like a special type of LSTM-based recurrent neural network that has fewer parameters than the LSTM [85]. Its internal unit is similar to the internal unit of the LSTM [62, 86] except that it has a separate cell state C_t ; it has only one hidden state (h_t) and is faster to train due to its simpler architecture. Therefore, other new systems based on GRU models have been developed, such as the multi-GRU prediction system after the study of the authors in [87]. The GRU [82, 83] helps avoid the vanishing gradient problem that usually occurs in

LSTM and requires less computation to update the external states. The update gate determines how much of the previous state's information is retained in the current state, while the reset gate determines whether to combine the current state with the previous information. The structure of the GRU unit cell [88] is shown in Figure 7. The mathematical functions used to characterize the GRU unit are as follows:

$$\begin{aligned} z_t &= \sigma(x_t W_z + h_{t-1} U_z + b_z), \\ r_t &= \text{sig}(x_t W_r + h_{t-1} U_r + b_r), \\ \tilde{h}_t &= \tanh(x_t W + r_t * h_{t-1} U + b), \\ h_t &= (1 - z_t * \tilde{h}_t + z_t * h_{t-1}), \end{aligned} \quad (14)$$

where σ is the sigmoid activation function; z_t is the update gate; r_t is the reset gate; h_t denotes the candidate hidden layer; W_z , W_r , and W are the weighting matrices for the corresponding connected input vector; U_z , U_r , and U are the weighting matrices of the previous time step; b_z , b_r , and b are biases. The following paragraph describes the approach to model development.

3.3.8. Multivariate Approach. The machine learning framework with linear regression (LR), decision tree (DT) regression, support vector machine (SVM) regression, artificial neural networks (ANN), recurrent neural networks based on long short-term memory (LSTM), and gated recurrent units (GRU) offers multilevel predictions that are generated for two time steps (horizons = 2 weeks). According to [89], forecasting is the prediction of time series from past data into future data on a longer time scale. Thus, in our multivariate approach, data from the entire West African (Senegal) study region are used to examine the relationship between the different futures. Thus, the selection of Pearson correlation features ([66]) is examined to characterize the degree of correlation of each feature with the target feature, air temperature (t2m). Based on the result of the Pearson correlation matrix, six features are selected. Then, using the past information, the next two weeks are predicted.

3.3.9. Model Training. The LR, DT regression, SVM regression (SVR), and ANN are implemented using the Sklearn framework. These models are trained and tested several times using the parameters and their value ranges provided in Table 2. The LSTM and GRU frameworks are written in Python and implemented using the Keras application of TensorFlow [90, 91]. The dataset is divided into a training set to train the model (1981–2010), a validation set (2010–2015) to improve model performance by fine-tuning the model after each epoch, and a test set (2015–2022) to determine the final result after training (Figure 8). The architecture of the deep learning models consists of one input layer, two bidirectional layers for LSTM and GRU, with fifty and hundred neurons in the hidden layers, respectively, and an output layer. All deep learning model configurations are trained several times with different activation functions (e.g., linear, tanh), numbers of

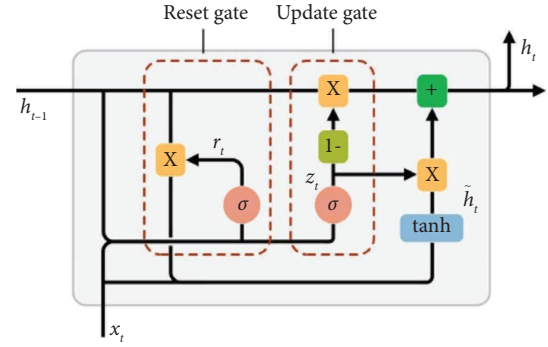


FIGURE 7: The cell structure of GRU unit.

epochs (e.g., 20, 50, 100), and neurons (e.g., 50, 100, 150). Additionally, the adaptive moment optimizer (Adam) is used for the following reasons:

- (i) Faster convergence
- (ii) Computational efficiency
- (iii) Low memory requirement
- (iv) Invariant to diagonal rescaling of gradients
- (v) It is well suited for problems with large data or parameters, according to the authors in [77].

The linear activation function is used for dense layers in deep learning methods. To evaluate the model's performance, the mean squared error (MSE) is used as a loss function that quantifies the squared difference between the observed and predicted values. However, the overfitting phenomenon may occur during the testing phase because of the large starting weight of the network. Hence, to overcome this issue, optimizing the model by using the early stopping approach might be necessary, which automatically stops training at the step when performance on the validation set starts to degrade. Ensuring each experiment with LSTM and GRU models starts with a unique set of weights and biases representing the LSTM and GRU model parameters, enabling the model to deal with nonstationarity in the data. Tables 2 and 3 show the hyperparameters used for each method and their value ranges.

3.4. Model Evaluation. The training set and the independent test set are evaluated to assess the performance of the applied machine learning models. However, the following evaluation metrics are most commonly used: (1) coefficient of determination (R^2), (2) mean absolute error (MAE), (MSE), (3) root mean square error (RMSE), (4) mean absolute percent error (MAPE), (5) Pearson correlation, and (6) Nash–Sutcliffe Efficiency (NSE). The R^2 measures the degree of correlation between the observed and predicted values; the MAE provides the absolute error information; the RMSE (square root of the MSE) measures the average magnitude of the errors between the predicted and actual values and how the residuals are distributed; the MAPE quantifies the absolute difference between the observed and predicted values divided by the observed values; and the NSE measures the magnitude of the residual variance with respect to the

TABLE 2: Summary of parameters used and range of values for LR, DT, SVM, and ANN models in this study.

Models	Parameters	Values
LR	Alpha	1, 2, 3, . . . , 50
	Solver	“Auto,” “svd,” “cholesky,” “lsqr,” “sparse_cg,” “sag,” “saga,” “lbfgs”
DT	Splitter	“Best,” “random”
	Max_depth	1, 3, 5, 7, 9, 11, 12
	Min_samples_leaf	1, 2, . . . , 10
	Min_weight_fraction_leaf	0.1, . . . , 0.9
	Max_features	“Auto,” “log2,” “sqrt,” none
	Max_leaf_nodes	None, 10, 20, 30, 40, 50, 60, 70, 80,90
SVR	Kernel	“linear,” “poly,” “rbf,” “sigmoid”
	C	1, 5, 10
	coef0	0.001, 0.01, 0.4, 0.5
	Gamma	“auto,” “scale”
ANN	Hidden_layer_sizes	(50, 50), (100, 50), (100, 1)
	Activation	“Relu,” “tanh,” “logistic,” “identity”
	Alpha	0.0001, 0.001, 0.01, 0.05, 0.1
	Learning_rate	“Constant,” “adaptive”
	Solver	“Adam,” “sgd,” “lbfgs”

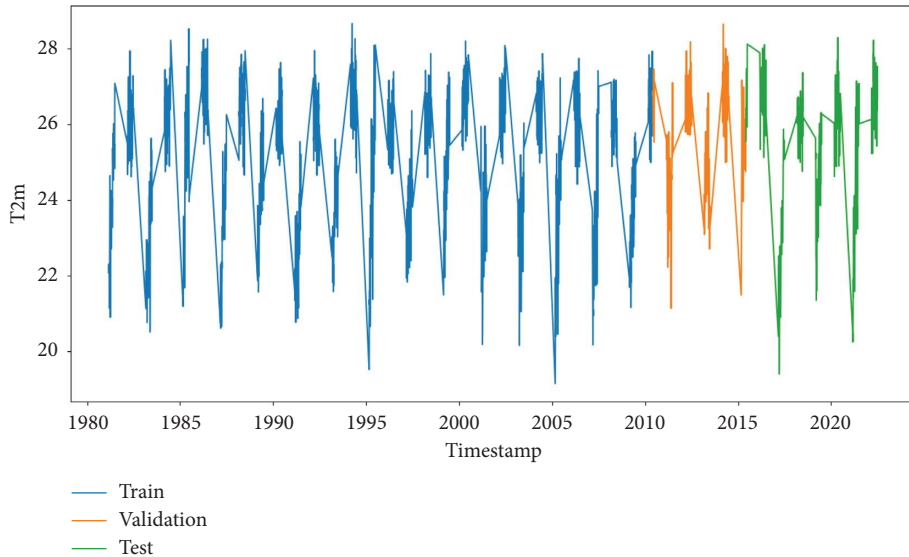


FIGURE 8: 2 m air temperature time series, including the training, the validation, and the test sets.

TABLE 3: Summary of hyperparameters used for both LSTM and GRU models in this study.

Model hyperparameters	Values
LSTM	4
GRU	4
Activation function (hidden layer)	Tanh
Activation function (dense layer)	Linear, ReLU
Neurons	50, 100
Optimizer	Adam
Epochs	50, 100
Regularization	L2
Batch size	60
Loss function	MSE
Validation metrics	MAE

variance of the observed t2m values. The models with the highest NSE and the lowest {MAPE, RMSE} show the best performance. These statistical performances are defined by the following equations:

$$\begin{aligned}
 R^2 &= \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y}_t)}{\sum_{t=1}^n (y_t - \bar{y}_t)^2}, \\
 MAE &= \frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{n}, \\
 RMSE &= \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}}, \\
 MAPE &= \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|, \\
 \text{Pearsonr} &= \frac{\sum_{t=1}^N [(\hat{y}_t - \bar{y}_t)(y_t - \bar{y}_t)]}{\sqrt{\left(\sum_{t=1}^N (\hat{y}_t - \bar{y}_t)^2\right) \times \sqrt{\left(\sum_{t=1}^N (y_t - \bar{y}_t)^2\right)}}, \\
 \text{NSE} &= 1 - \frac{\sum_{t=1}^T (\hat{y}_t - y_t)^2}{\sum_{t=1}^T (y_t - \bar{y}_t)^2},
 \end{aligned} \tag{15}$$

where n is the number of samples used in the test set, y_t are the observed t2m values, \hat{y}_t are the predicted t2m values, and \bar{y}_t is the mean value of the observed t2m values.

Table 4 outlines the criteria for the previously mentioned statistical methods used in the search for a better prediction model.

4. Experiments and Results

This section describes the experiments carried out to confirm the ability of machine learning (ML) models, especially neural network models, to learn. The results for each method are investigated more in detail.

4.1. Feature Correlations. Before the learning algorithm can be applied to the dataset, specific characterizations of the data must be confirmed. Because the current investigation is based on information from the samples, the idea of presenting the correlation that exists among the variables arose. Therefore, an examination of the visualization and analysis of the characteristic variables is carried out on the basis of the Pearson correlation coefficient (r) in the range between -1 and 1 [66]. The information describing the relationship between features is then stored in Pearson's correlation matrix (Figure 9a), where $0 < r \leq 1$ means positive correlation, $r = 0$ no correlation, and $-1 \leq r < 0$ negative correlation.

Figure 9(a) reveals that the features are correlated between themselves and with the target. Some variables have a high correlation, whereas others have a low correlation. This figure shows a (high) positive correlation between the features (rh1000, td, and z500) and the target. The t500 and

t850 are also correlated and remain within the accepted range ($r > 0.3$). However, other variables have a lower correlation with the target. The five most correlated variables (rh1000, td, t850, t850, and z500) are selected for multi-variate model development in the entire study region.

4.2. Assessment of Machine Learning Models. In this part of the study, models are developed to predict subseasonal temperatures over West Africa (Senegal) for the period March to June (MAMJ) using ERA-5 data and applied machine learning (ML) methods to gain a better understanding of the predicted results. There are six different ML models, including linear regression (LR), decision tree (DT), support vector regression (SVR), artificial neural networks (ANN), long short-term memory (LSTM), and gated recurrent unit (GRU). The developed ML and deep neural network models are used to predict the air temperature (t2m) pattern in Senegal using the provided dataset. The dataset contains 2749026 records of daily t2m, dew point, temperature, relative humidity, and geopotential from 1981 to 2022. The training set (1981–2015) is split into two parts: 80% for training and 20% for validation. The data from 1/06/2015–29/06/2022 are used for testing to decide on the best model. Approximately, the past three weeks were used to predict the next two weeks. Figure 9(b) shows the distribution of the data. From this histogram, it appears that the data follow the Gaussian distribution.

The Python packages Sklearn and Keras were used to train the ML and neural network models, respectively. Firstly, the data are split into training, validation, and test sets. Secondly, the training and validation data are ingested into the models, and the training process starts. Different model configurations were tested in the training using the error in order to obtain the best training configuration (best model). Finally, after obtaining the best training result, the test set is introduced to the trained models, and the results are then compared to the observed values.

The following statistical indicators were used to evaluate the relationship between the observed and predicted values: the coefficient of determination (R^2), the root mean squared error (RMSE), the mean absolute percentage error (MAPE), the mean absolute error (MAE), Pearson's correlation (Pearson's r), and the Nash–Sutcliffe efficiency (NSE). The detailed results for each model are given in Table 5, which shows the implementation and effectiveness of the developed models in the summer air temperature (t2m) estimation for West Africa (Senegal) with n_shift equal to 25 and two time steps ahead (y_{t+1} and y_{t+2}), i.e., the next 2 weeks. The optimal hyperparameters values for ML algorithms are obtained using hyperparameters tuning with GridSearchCV.

The first ML approach used to predict t2m values is LR with l2 regularization. The optimal value for alpha is 7, and the solver used is lsqr. Figure 10(a) shows the correlation (regression line) between the observation and the prediction of the LR method. The plot shows a slightly better fit, but the accuracy (NSE and R^2) of the t2m prediction is lower (Table 5).

TABLE 4: General performance ratings of the statistical indices.

Performance rating	RMSE	MAE	MAPE	NSE	Pearson's r
Very good	<0.009	0	0	0.75~1	0.8~1
Good	0.009~0.08	Not defined	Not defined	0.65~0.75	0.6~0.79
Satisfactory	0.08~0.5	Not defined	Not defined	0.5~0.65	0.3~0.59
Unsatisfactory	>0.5	1	∞	<0.5	<0.29

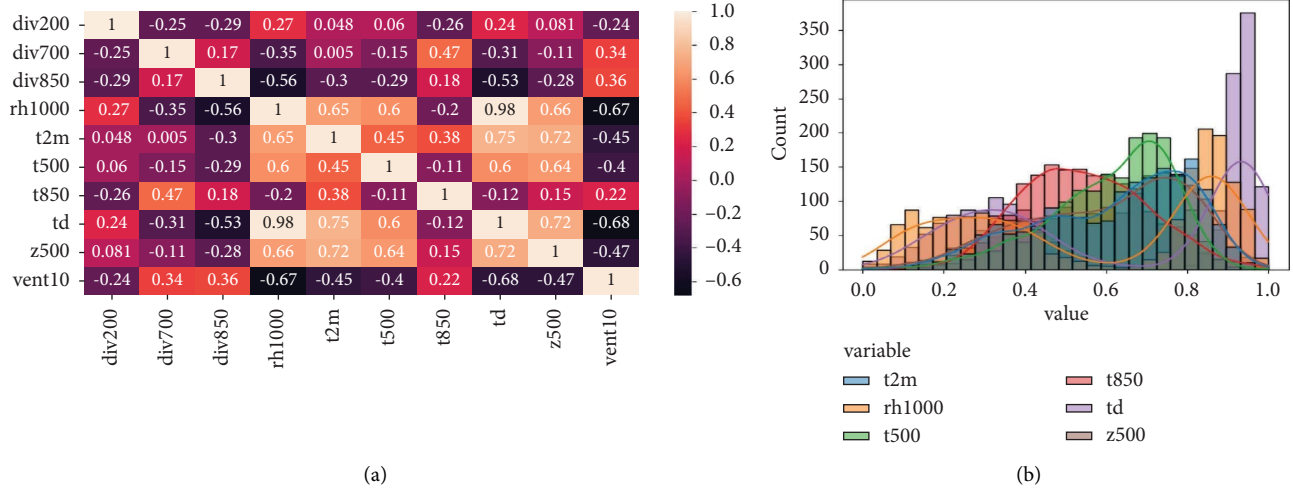


FIGURE 9: Pearson correlation of all variables and KDE of selected variables. (a) Correlation matrix, displaying the correlation between all variables. (b) Data distribution with kernel density estimate (KDE) for selected variables from 1981–2010.

TABLE 5: Statistic performance on the test set: RMSE, MAE, R^2 , Pearson's r , and NSE values, as well as MAPE reported for all the machine learning models.

	Method	MSE	RMSE	MAE	MAPE	R^2	Pearson's r	NSE
LR	Alpa: 7	0.0190	0.1381	0.1	0.27	0.4493	0.6752	0.4493
DT	Max_depth: 5	0.0203	0.1425	0.105	0.2907	0.4132	0.6492	0.414
SVR	Kernel: rbf	0.0202	0.1422	0.0984	0.2683	0.416	0.6545	0.416
ANN	A: ReLU, solver: Adam	0.018	0.134	0.0948	0.2561	0.4814	0.6968	0.516
LSTM	A: ReLU, E: 50, N: 50	0.0098	0.099	0.0676	0.0257	0.6852	0.8626	0.7294
	A: linear, E: 100, N: 50	0.0100	0.1	0.0706	0.0267	0.6319	0.8607	0.716
	A: ReLU, E: 100, N: 50	0.0103	0.1015	0.0717	0.0271	0.6622	0.8648	0.714
	A: ReLU, E: 50, N: 100	0.0108	0.1039	0.0753	0.0283	0.5511	0.854	0.7005
GRU	A: ReLU, E: 50, N: 50	0.0092	0.0959	0.0653	0.0251	0.6954	0.8707	0.7468
	A: ReLU, E: 100, N: 50	0.0096	0.098	0.0679	0.0259	0.6885	0.8605	0.7353
	A: linear, E: 100, N: 50	0.0096	0.098	0.0673	0.0258	0.6804	0.866	0.7371
	A: ReLU, E: 50, N: 100	0.0098	0.099	0.0683	0.0261	0.658	0.8556	0.7304

The best performance is marked in bold. Here, A = activation, E = epoch, and N = number of units.

The second ML method applied is DT. The following values are used for this study: max_depth = 5, max_features = "auto", max_leaf_nodes = 10, max_leaf_nodes = 10, and splitter = "best". Figure 10(b) shows the regression line between the observed values and the test results of the DT method. Table 5 and Figure 10(b) indicate that the DT model provided the worst accuracy than the LR model. As can be seen from the picture, DT has problems predicting the t2m pattern.

Another ML method used is SVR. The kernel function, the radial basis function (RBF), the kernel coefficient (gamma = auto), the optimization tolerance (tol = 0.001), and the regularization parameter ($C = 1.0$)

were the optimal hyperparameters selected for this study. Figure 10(c) indicates the correlation between the observed values and the predicted results of the SVR model in the test set.

The ANN method used in this study is a multilayer perceptron (MLP) with backpropagation, which is commonly used for the prediction of temperatures [38, 54]. The ReLU activation function, the Adam optimizer, the alpha value of 0.1, and the hidden size of 50 were used to obtain the best test results. Table 5 and Figure 10(d) show the correlation between the observed values and the predicted values of the ANN model in the test set.

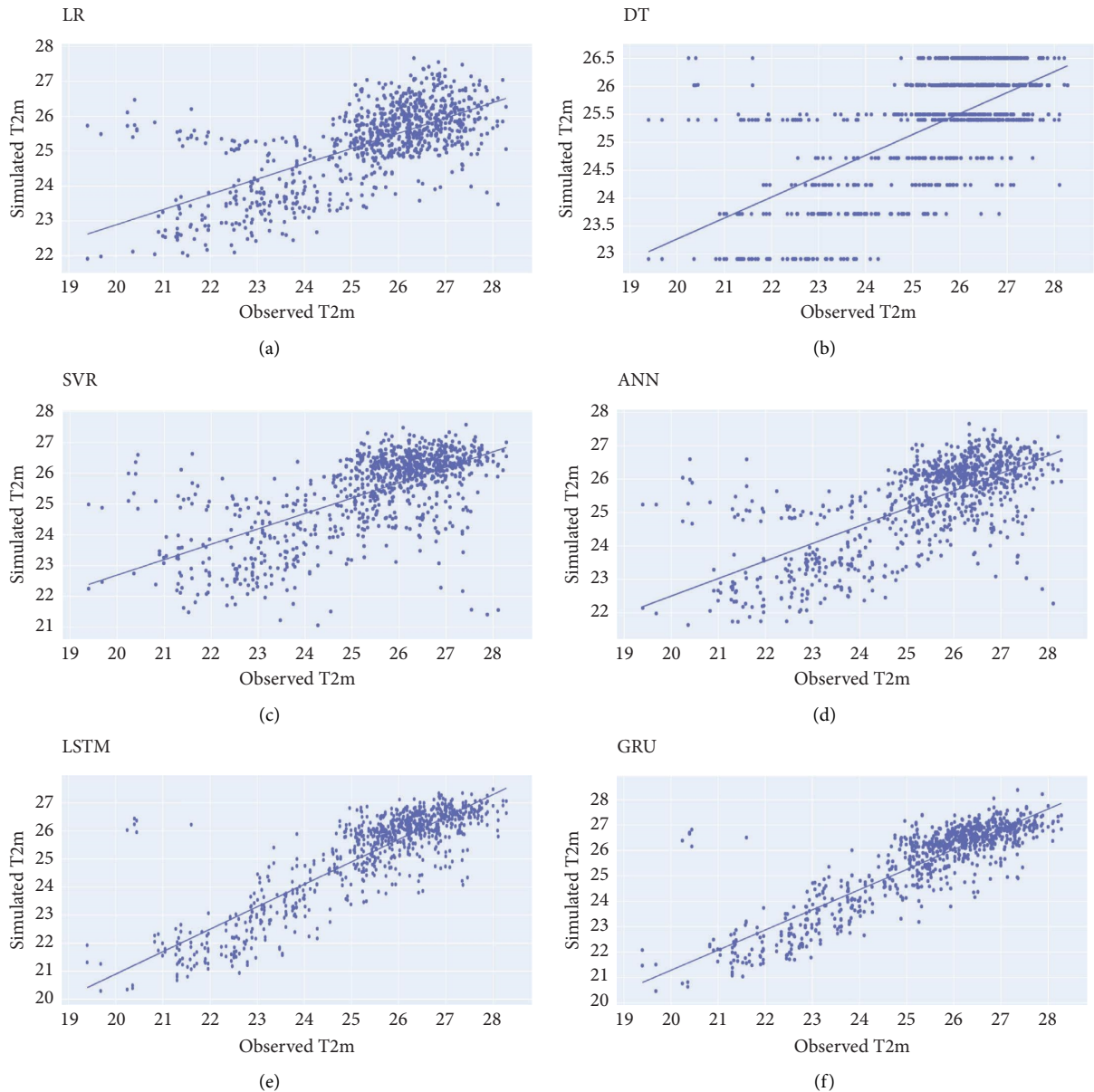


FIGURE 10: Correlation between the observed and the simulated output on the test set for (a) LR, (b) DT, (c) SVR, (d) ANN, (e) LSTM, and (f) GRU.

The deep learning models used in this study are LSTM and GRU. Different model configurations are tested. The number of epochs, the neurons, and different activation functions (Figure 5) were tested through these configurations to obtain the best result for the test set. All LSTM and GRU model configurations were trained with the Adam optimizer. So, the best test results were obtained for the model configuration with 50 epochs, 50 neurons, tanh activation function in the hidden layers, and ReLU activation function in the dense layer. Table 5 and Figures 10(e) and 10(f) show the correlation between the observed and predicted t2m values of the LSTM and GRU models on test data. The plots show a better fit, and the

results are highly accurate. These models performed approximately well, with NSE values of 72.74% and 74.68%, respectively.

However, the GRU model (ReLU activation function, 50 epochs, and 50 neurons) shows the best predictions of t2m values (NSE = 0.7468, MAPE = 0.0251, Pearson's $r = 0.8707$) compared to the LSTM model (ReLU activation function, 50 epochs, and 50 neurons) with $R^2 = 0.6852$, NSE = 0.7294, MAPE = 0.0257, and Pearson's $r = 0.8626$. Overall, the GRU model with low variance is the best performing deep learning algorithm for $n_{\text{sift}} = 25$ (≈ 3 weeks) in this study of summer temperature prediction (next 2 weeks) over West Africa, especially Senegal.

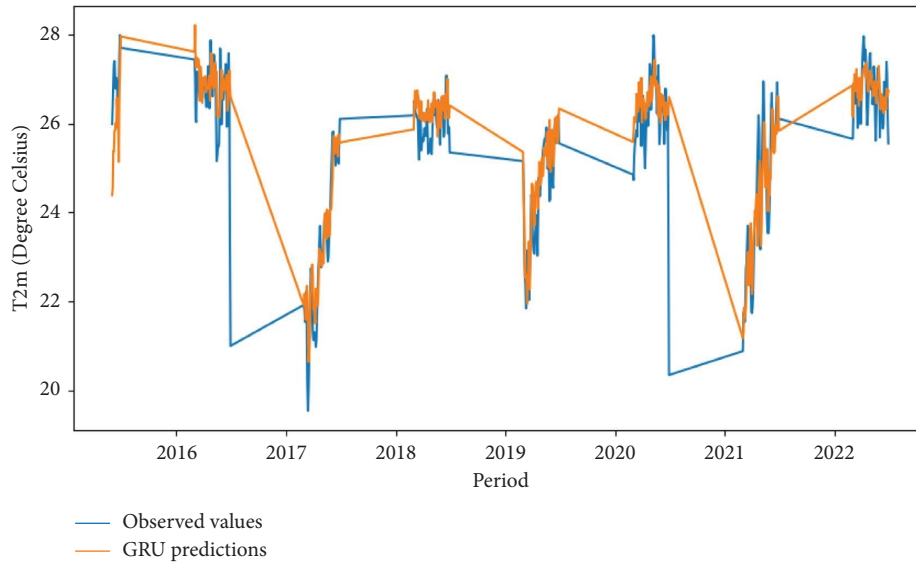


FIGURE 11: Comparison of the GRU results with the observed values.

4.3. Visualization of Predictions: Comparison of Observed with Predicted Values. As can be seen from Table 5 and Figure 10, deep learning methods such as long short-term memory (LSTM) and gated recurrent units (GRU) are well trained and show acceptable test performance compared to other ML methods. The ANN method performs relatively better than LR, DT, and SVR, but it still underperforms LSTM and GRU. Among all the methods, GRU (MAPE = 0.0251, NSE = 0.7468, and Pearsons' $r=0.8707$) and DT (MAPE = 0.2907, NSE = 0.414, and Pearsons' $r=0.6492$) show, respectively, the best and the worst performance. The GRU model is capable of predicting temperatures over 20 degrees Celsius, in contrast to the DT model.

Figure 11 shows the comparison of the predicted t2m values with the observed values for the GRU model. The figure reveals how the GRU model performs better in estimating summer air temperatures. Although the performance of most LSTM models was pretty good and accurate, the GRU model (with ReLu activation, 50 epochs, and 50 units) provides the best results for temperature predictions (MAPE = 2.51%, NSE = 74.68%) and is able to capture long-term dependencies and provide predicted t2m values that are consistent with the observed values than the other models. Table 6 shows the ranking of the models based on the performance of the prediction steps. Overall, the GRU model ranks first among all models, followed by the LSTM model.

Thus, as mentioned in the previous section, the model with the GRU algorithm provides the best prediction for the t2m value compared to the observed values, using the past 25 days (≈ 3 weeks) to predict the next 1-2 weeks, with a low RMSE, a low MAPE, and a higher NSE according to the performance indices shown in Table 5. Overall, the GRU model showed a better ability to predict the t2m than the other models and can therefore be considered a useful method for modeling subseasonal summer temperatures and keeping the population safe from the risks associated with natural phenomena caused by high temperatures.

The results of this study show that, compared to other modern machine learning methods, deep learning algorithms are one of the best, most efficient, and most powerful tools that can be used for subseasonal summer temperature prediction on large datasets [51, 54]. However, these analyses show that when using different variables such as relative humidity, dew point, temperature, and geopotential data, there is a better possibility for t2m values with high accuracy.

5. Discussion

It is indeed possible to validate the effectiveness of deep networks for predicting time series data based on state-of-the-art studies. Thanks to the development of new models and advances in computing power, the field of prediction using deep neural networks for forecasting has grown rapidly. Machine learning (ML) approaches such as linear regression (LR), decision tree (DT), support vector machine (SVM), and artificial neural network (ANN) are commonly used for climate modeling, especially in temperature modeling. Deep neural networks (LSTM and GRU) models, two of the best subseasonal forecasting models, are growing in popularity, although most of the top models have more traditional architectures.

The GRU network, a modified and improved version of LSTM, is capable of supporting a higher resampling rate and can be trained on smaller and larger datasets. Therefore, our experiments carried out on more than 1.5 million training sets have indicated that

- (i) GRU neural network, which is easier and faster to train, is the best model that provided good results
- (ii) A deep neural network GRU with encoder and decoder layers has high performance and the ability to forecast subseasonal temperatures (next 1-2 weeks) with 2.51% MAPE precision

Based on the recent work in the literature, GRU showed promising results and was considered a good predictor [92–94]. However, we recall that these studies were performed

TABLE 6: ML and DL methods ranking based on their performance.

Rank	RMSE	MAE	MAPE	R^2	Pearson's r	NSE	Overall
1	GRU	GRU	GRU	GRU	GRU	GRU	GRU
2	LSTM	LSTM	ANN	LSTM	LSTM	LSTM	LSTM
3	ANN	ANN	LSTM	ANN	ANN	ANN	ANN
4	SVR	SVR	LR	LR	LR	LR	LR
5	SVR	LR	LR	SVR	SVR	SVR	SVR
6	DT	DT	DT	DT	DT	DT	DT

using a univariate approach and not multivariate approaches. Furthermore, Park et al. [95] and Du et al. [96] have proposed the LSTM model with encoder and decoder layers but not compared to the GRU model. Benet et al. [51], Van Straaten et al. [50], Sharaff and Roy [53], and dos Santos [55] have also proposed machine learning models (LR and RF) for sub-seasonal time series forecasting but not deep neural networks. Therefore, the present investigation compares LR, DT, SVR, ANN, LSTM, and GRU methods for subseasonal forecasting based on a multivariate approach to predict summer temperatures for the next 2 weeks. This multivariate approach for multistep subseasonal forecasting has not been investigated yet in recent work for West African regions.

Our experiments performed with the Adam optimizer, linear and tanh activation functions in the encoder and decoder layers, 50 epochs, and 50 units have shown the GRU neural network to be the best predictor with smaller prediction errors (MAPE = 2.51%, MAE = 0.0653, RMSE = 0.0959) than those indicated by the other ML methods (MAE = 1.60 and RMSE = 2.03) [53, 55]. Because the GRU neural network is more resistant to exploding or vanishing gradient problems, it can learn in a larger variety of configurations or long-range dependencies. The DT and SVR models demonstrated more unsatisfactory results and were therefore considered the worst models.

6. Conclusion

Nowadays, in many fields such as agriculture, meteorology, hydrology, and water resource management, the forecasting of subseasonal temperatures is of crucial importance [67]. Therefore, this study investigated the ability of state-of-the-art machine learning methods to estimate the subseasonal temperatures (1 to 2 weeks) for the summer (MAMJ) season.

The major contribution of this study is to compare the ability of different machine learning (ML) techniques to forecast subseasonal temperatures in West Africa, especially in the Senegal region. In addition, this study adds to the existing body of knowledge by providing insights into model performance, comparing different methods, considering region-specific factors, assessing subseasonal forecasting skills, and potentially introducing methodological advancements [97].

In this paper, six ML methods are used, including LR, DT, SVR, ANN, long short-term memory (LSTM), and gated recurrent unit (GRU) [62]. This study used data from 02/03/1891 to 29/06/2022. Data from 02/03/1981 to 31/05/2015 were selected for training, and 20% of the training set was used for validation. The remaining data from 01/06/2015

to 29/06/2022 were selected to test the ML models. The test results show that the deep neural network (LSTM and GRU) models were better trained and accurate than the other ML models (LR, DT, SVR, and ANN). The evaluation metrics, including RMSE, MAE, MAPE, R^2 , Pearson's r , and NSE, help select the best model. However, these performance indicators reveal that deep neural network (LSTM and GRU) methods are the most suitable. Overall, the experiment results show that the GRU model is one of the most powerful deep learning methods for subseasonal time series forecasting in West Africa, with a precision of approximately 2.51% MAPE and 0.0959 RMSE, 87.07% Pearson's r , and 74.68% of the NSE score.

In summary, ML methods, especially deep neural networks, a promising direction for future research in sub-seasonal forecasting, are shown to help extend the prediction time of summer temperature to subseasonal. Therefore, these contributions enhance our understanding of sub-seasonal climate forecasting, especially in the context of predicting summer temperatures in West Africa (Senegal).

Although deep neural networks have shown great promise, there are still more studies to be done in this area. Due to limited resources, the author of this article was unable to investigate studies in different regions of West Africa. For further research, one might plan to

- (i) Investigate the previous model over very long time scales or long-horizon forecasting (up to 6 weeks lead time)
- (ii) Build a global model and expand the study to other areas in West and Central Africa, especially in the Sahel regions
- (iii) Improve the accuracy of subseasonal temperature predictions during the rainy seasons
- (iv) Investigate the attention-based models and compare their performances with the previous model

However, these methods have limitations when it comes to subseasonal temperature forecasting, especially in the context of climate. This is because the climate is constantly changing and models may not have the capacity to capture the full extent of climate change, which can often lead to extreme phenomena occurring under previously stable conditions [98].

Data Availability

The data supporting the current study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

All authors contributed to the preparation of this manuscript. Annine Duclaire Kenne proposed the methodology, analyzed data, designed the models, performed the experiments, and wrote, reviewed, and edited the manuscript. Mory Toure designed the study, visualized and analyzed climate patterns, supervised the study, and wrote, reviewed, and edited the manuscript. Lema Logamou Seknewna performed statistical analysis and investigated, reviewed, and supervised the study. Herve Landry Ketsemen performed the LR method and investigated, reviewed, and edited the manuscript.

Acknowledgments

The authors would like to express their sincere gratitude to the African Institute for Mathematical Sciences (AIMS) and African Master of Machine Intelligence (AMMI) programs, teaching staff, tutors, and advisors, especially Mr. Mory Toure. Their valuable suggestions and meticulous review have undoubtedly strengthened the quality and clarity of the content. The authors are truly appreciative of their commitment to the peer review process, which has significantly enriched the overall depth and coherence of the manuscript.

References

- [1] A. Costello, M. Abbas, A. Allen et al., "Managing the health effects of climate change: lancet and university college london institute for global health commission," *The Lancet*, vol. 373, no. 9676, pp. 1693–1733, 2009.
- [2] S. Thober, R. Kumar, N. Wanders et al., "Multi-model ensemble projections of european river floods and high flows at 1.5, 2, and 3 degrees global warming," *Environmental Research Letters*, vol. 13, no. 1, Article ID 014003, 2018.
- [3] K. Mohammed, A. K. M. S. Islam, G. M. T. Islam et al., "Future Floods in Bangladesh under 1.5°C, 2°C, and 4°C Global Warming Scenarios," *Journal of Hydrologic Engineering*, vol. 23, no. 12, Article ID 04018050, 2018.
- [4] K. Azizi and C. Meier, "Improving the characterization of urban flash floods through application of local knowledge," *Authorea Preprints*, 2022.
- [5] S. Ebrahimi and M. Khorram, "Variability effect of hydrological regime on river quality pattern and its uncertainties: case study of zarjoob river in Iran," *Journal of Hydroinformatics*, vol. 23, no. 5, pp. 1146–1164, 2021.
- [6] M. Zhang, S. Tan, C. Zhang, S. Han, S. Zou, and E. Chen, "Assessing the impact of fractional vegetation cover on urban thermal environment: a case study of hangzhou, China," *Sustainable Cities and Society*, vol. 96, Article ID 104663, 2023.
- [7] J. Zhang, M. Khoshbakht, J. Liu, Z. Gou, J. Xiong, and M. Jiang, "A clustering review of vegetation-indicating parameters in urban thermal environment studies towards various factors," *Journal of Thermal Biology*, vol. 110, Article ID 103340, 2022.
- [8] M. Zhang, S. Tan, Z. Pan, D. Hao, X. Zhang, and Z. Chen, "The spatial spillover effect and nonlinear relationship analysis between land resource misallocation and environmental pollution: evidence from China," *Journal of Environmental Management*, vol. 321, Article ID 115873, 2022.
- [9] Solutions basées sur la nature pour la ville contemporaine/re-naturer la ville/réflexions sur les paysages urbains, "les services écosystémiques et les solutions basées sur la nature dans les villes/infrastructures vertes multifonctionnelles et adaptation au changement climatique: le verdissement des friches industrielles comme stratégie d'adaptation pour les communautés vulnérables?/delivering green infrastructure through planning: insights from practice in fingal, Ireland/planning for biophilic cities: from theory to practice," *Théorie de la planification and Pratique*, vol. 17, pp. 267–300.
- [10] WHO, "Heat-health action plans: guidance," 2008, <https://www.who.int/publications/i/item/9789289071918>.
- [11] F. C. Bull, S. S. Al-Ansari, S. Biddle et al., "World health organization 2020 guidelines on physical activity and sedentary behaviour," *British Journal of Sports Medicine*, vol. 54, no. 24, pp. 1451–1462, 2020.
- [12] M. L. Parry, O. Canziani, P. Jean, V. D. L. Paul, and C. Hanson, *Climate Change 2007-impacts, Adaptation and Vulnerability: Working Group II Contribution to the Fourth Assessment Report of the IPCC*, Cambridge University Press, Cambridge, UK, 2007.
- [13] W. L. Filho, A. L. Balogun, O. E. Olayide et al., "Assessing the impacts of climate change in cities and their adaptive capacity: towards transformative approaches to climate change adaptation and poverty reduction in urban areas in a set of developing countries," *Science of the Total Environment*, vol. 692, pp. 1175–1190, 2019.
- [14] IPCC, "Framing and context," 2021, <https://www.ipcc.ch/sr15/chapter/chapter-1/>.
- [15] T. R. Karl and K. E. Trenberth, "Modern global climate change," *Science*, vol. 302, no. 5651, pp. 1719–1723, 2003.
- [16] A. J. Hobday, L. V. Alexander, S. E. Perkins et al., "A hierarchical approach to defining marine heatwaves," *Progress in Oceanography*, vol. 141, pp. 227–238, 2016.
- [17] N. Minot, "Contract farming and its effect on small farmers in less developed countries," Technical report, Agecon Search, Research in Agricultural & Applied Economics, East Lansing, MI, USA, 1986.
- [18] I. Sy, B. Cisse, B. Ndao et al., "Heat waves and health risks in the northern part of Senegal: analysing the distribution of temperature-related diseases and associated risk factors," *Environmental Science and Pollution Research*, vol. 29, no. 55, pp. 83365–83377, 2022.
- [19] I. Diouf, A. M. Adeola, J. A. Gbenga et al., "Impact of future climate change on malaria in west africa," *Theoretical and Applied Climatology*, vol. 147, pp. 1–13, 2022.
- [20] N. Nicholls and L. Alexander, "Has the climate become more variable or extreme? progress 1992–2006," *Progress in Physical Geography: Earth and Environment*, vol. 31, no. 1, pp. 77–87, 2007.
- [21] M. A. McGeehin and M. Mirabelli, "The potential impacts of climate variability and change on temperature-related morbidity and mortality in the United States," *Environmental Health Perspectives*, vol. 109, no. suppl 2, pp. 185–189, 2001.
- [22] D. G. Miralles, P. Gentile, S. I. Seneviratne, and A. J. Teuling, "Land-atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges," *Annals of*

- the New York Academy of Sciences*, vol. 1436, no. 1, pp. 19–35, 2019.
- [23] C. J. White, H. Carlsen, A. W. Robertson et al., “Potential applications of subseasonal-to-seasonal (s2s) predictions,” *Meteorological Applications*, vol. 24, no. 3, pp. 315–325, 2017.
- [24] R. Lowe, M. García-Díez, J. Ballester et al., “Evaluation of an early-warning system for heat wave-related mortality in europe: implications for sub-seasonal to seasonal forecasting and climate services,” *International Journal of Environmental Research and Public Health*, vol. 13, no. 2, p. 206, 2016.
- [25] S.-M. Chen and J.-R. Hwang, “Temperature prediction using fuzzy time series,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 30, no. 2, pp. 263–275, 2000.
- [26] M. Zhang, A.-A. Kafy, P. Xiao et al., “Impact of urban expansion on land surface temperature and carbon emissions using machine learning algorithms in wuhan, China,” *Urban Climate*, vol. 47, Article ID 101347, 2023.
- [27] S. R. Moosavia, D. A. Woodb, and S. A. Samadanic, “Modeling performance of foam-co2 reservoir flooding with hybrid machine-learning models combining a radial basis function and evolutionary algorithms,” *Methods*, vol. 4, no. 5, 2020.
- [28] M. Alizamir, O. Kisi, A. N. Ahmed et al., “Advanced machine learning model for better prediction accuracy of soil temperature at different depths,” *Public Library of Science One*, vol. 15, no. 4, Article ID e0231055, 2020.
- [29] C. Zhang and M. Zhang, “Wavelet-based neural network with genetic algorithm optimization for generation prediction of pv plants,” *Energy Reports*, vol. 8, pp. 10976–10990, 2022.
- [30] M. Moeini, A. Shojaezadeh, and M. Geza, “Supervised machine learning for estimation of total suspended solids in urban watersheds,” *Water*, vol. 13, no. 2, p. 147, 2021.
- [31] K. A. Marill, “Advanced statistics: linear regression, part ii: multiple linear regression,” *Academic Emergency Medicine*, vol. 11, no. 1, pp. 94–102, 2004.
- [32] B. Azari, K. Hassan, J. Pierce, and S. Ebrahimi, “Evaluation of machine learning methods application in temperature prediction,” *Computational Research Progress in Applied Science and Engineering*, vol. 8, no. 1, pp. 1–12, 2022.
- [33] M. Apaydın, M. Yumuş, A. Değirmenci, and Ö. Karal, “Evaluation of air temperature with machine learning regression methods using seoul city meteorological data,” *Pamukkale University Journal of Engineering Sciences*, vol. 28, no. 5, pp. 737–747, 2022.
- [34] R. Kumar, “Decision tree for the weather forecasting,” *International Journal of Computer Application*, vol. 76, no. 2, pp. 31–34, 2013.
- [35] Y. Radhika and M. Shashi, “Atmospheric temperature prediction using support vector machines,” *International journal of computer theory and engineering*, vol. 1, no. 1, pp. 55–58, 2009.
- [36] X. Wang, L. Ma, and X. Wang, “Apply semi-supervised support vector regression for remote sensing water quality retrieving,” in *Proceedings of the 2010 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2757–2760, IEEE, Honolulu, HI, USA, July 2010.
- [37] C. Qi, W. Wang, and S. Wang, “Application of indoor temperature prediction based on svm and bpnn,” in *Proceedings of the 27th Chinese Control and Decision Conference (2015 CCDC)*, pp. 2883–2887, IEEE, Qingdao, China, May 2015.
- [38] D. B. Shank, G. Hoogenboom, and R. W. McClendon, “Dewpoint temperature prediction using artificial neural networks,” *Journal of Applied Meteorology and Climatology*, vol. 47, no. 6, pp. 1757–1769, 2008.
- [39] A. Najah, A. El-Shafie, O. A. Karim, and A. H. El-Shafie, “Application of artificial neural networks for water quality prediction,” *Neural Computing and Applications*, vol. 22, no. S1, pp. 187–201, 2013.
- [40] K. P. Singh, A. Basant, A. Malik, and G. Jain, “Artificial neural network modeling of the river water quality—a case study,” *Ecological Modelling*, vol. 220, no. 6, pp. 888–895, 2009.
- [41] A. Wang, M. Zhang, A.-A. Kafy, B. Tong, D. Hao, and Y. Feng, “Predicting the impacts of urban land change on lst and carbon storage using invest, ca-ann and woa-lstm models in guangzhou, China,” *Earth Science Informatics*, vol. 16, no. 1, pp. 437–454, 2023.
- [42] P. Lara-Benítez, M. Carranza-García, J. M. Luna-Romera, and J. C. Riquelme, “Temporal convolutional networks applied to energy-related time series forecasting,” *Applied Sciences*, vol. 10, no. 7, p. 2322, 2020.
- [43] J. Yeomans, S. Thwaites, W. S. P. Robertson, D. Booth, B. Ng, and D. Thewlis, “Simulating time-series data for improved deep neural network performance,” *IEEE Access*, vol. 7, pp. 131248–131255, 2019.
- [44] M. Zhang, C. Zhang, A.-A. Kafy, and S. Tan, “Simulating the relationship between land use/cover change and urban thermal environment using machine learning algorithms in wuhan city, China,” *Land*, vol. 11, no. 1, p. 14, 2021.
- [45] A. Dikshit, B. Pradhan, and M. Santosh, “Artificial neural networks in drought prediction in the 21st century—a scientometric analysis,” *Applied Soft Computing*, vol. 114, Article ID 108080, 2022.
- [46] K. H. Guigma, D. MacLeod, M. Todd, and Y. Wang, “Prediction skill of sahelian heatwaves out to subseasonal lead times and importance of atmospheric tropical modes of variability,” *Climate Dynamics*, vol. 57, no. 1–2, pp. 537–556, 2021.
- [47] D. I. V. Domeisen, E. A. B. Eltahir, E. M. Fischer et al., “Prediction and projection of heatwaves,” *Nature Reviews Earth and Environment*, vol. 4, no. 1, pp. 36–50, 2022.
- [48] B. F. Zaitchik, “Madden-Julian oscillation impacts on tropical african precipitation,” *Atmospheric Research*, vol. 184, pp. 88–102, 2017.
- [49] G. Brunet, M. Shapiro, B. Hoskins et al., “Collaboration of the weather and climate communities to advance subseasonal-to-seasonal prediction,” *Bulletin of the American Meteorological Society*, vol. 91, no. 10, pp. 1397–1406, 2010.
- [50] C. Van Straaten, K. Whan, D. Coumou, B. Van den Hurk, and M. Schmeits, “Using explainable machine learning forecasts to discover subseasonal drivers of high summer temperatures in western and central europe,” *Monthly Weather Review*, vol. 150, no. 5, pp. 1115–1134, 2022.
- [51] E. W. Benet, M. Pyrina, B. J. Esteve, E. Fraenkel, J. Cohen, and D. I. V. Domeisen, “Sub-seasonal prediction of central european summer heatwaves with linear and random forest machine learning models,” *Artificial Intelligence for the Earth Systems*, vol. 2, 2022.
- [52] E. J. Alfaro, A. Gershunov, and D. Cayan, “Prediction of summer maximum and minimum temperature over the central and western United States: the roles of soil moisture and sea surface temperature,” *Journal of Climate*, vol. 19, no. 8, pp. 1407–1421, 2006.
- [53] A. Sharaff and S. R. Roy, “Comparative analysis of temperature prediction using regression methods and back propagation neural network,” in *Proceedings of the 2018 2nd International Conference on Trends in Electronics and*

- Informatics (ICOEI)*, pp. 739–742, IEEE, Tirunelveli, India, May 2018.
- [54] T. Anjali, K. Chandini, K. Anoop, and V. L. Lajish, “Temperature prediction using machine learning approaches,” in *Proceedings of the 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, pp. 1264–1268, IEEE, Kannur, India, July 2019.
- [55] R. S. dos Santos, “Estimating spatio-temporal air temperature in london (UK) using machine learning and earth observation satellite data,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 88, Article ID 102066, 2020.
- [56] R. Z. Zhang, X. J. Jia, and Q. F. Qian, “Seasonal forecasts of eurasian summer heat wave frequency,” *Environmental Research Communications*, vol. 4, no. 2, Article ID 025007, 2022.
- [57] G. A. Meehl, J. H. Richter, H. Teng et al., “Initialized earth system prediction from subseasonal to decadal timescales,” *Nature Reviews Earth and Environment*, vol. 2, no. 5, pp. 340–357, 2021.
- [58] C. Lu, D. Huang, Y. Shen, and F. Xin, “Data-driven global subseasonal forecast model (gsfm v1. 0) for intraseasonal oscillation components,” *Geoscientific Model Development Discussions*, pp. 1–18, 2022.
- [59] J. A. Weyn, D. R. Durran, R. Caruana, and N. Cresswell-Clay, “Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models,” *Journal of Advances in Modeling Earth Systems*, vol. 13, no. 7, Article ID e2021MS002502, 2021.
- [60] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [61] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, “On the properties of neural machine translation: encoder-decoder approaches,” 2014, <https://arxiv.org/abs/1409.1259>.
- [62] J. Chung, C. Gulcehre, K.H. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” 2014, <https://arxiv.org/abs/1412.3555>.
- [63] Senegal, “Senegal,” 1980, <https://en.wikipedia.org/wiki/Senegal>.
- [64] UNICEF, *Facts for Life*, Unicef, New York, NY, USA, 2010.
- [65] T. N. Palmer, C. Brankovic, F. Molteni et al., “The european centre for medium-range weather forecasts (ecmwf) program on extended-range prediction,” *Bulletin of the American Meteorological Society*, vol. 71, no. 9, pp. 1317–1330, 1990.
- [66] K. Pearson and L. N. G. Filon, “Vii. mathematical contributions to the theory of evolution.—iv. on the probable errors of frequency constants and on the influence of random selection on variation and correlation,” *Philosophical Transactions of the Royal Society of London- Series A: Containing Papers of a Mathematical or Physical Character*, vol. 191, pp. 229–311, 1898.
- [67] S. He, X. Li, T. DelSole, P. Ravikumar, and A. Banerjee, “Sub-seasonal climate forecasting via machine learning: challenges, analysis, and advances,” *Proceedings of the Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence*, vol. 35, no. 1, pp. 169–177, 2021.
- [68] S. Ali, M. Geza, J. McCray, and S. H. Terri, “Site-scale integrated decision support tool (i-dstss) for stormwater management,” *Water*, vol. 11, no. 10, p. 2022, 2019.
- [69] J. R. Quinlan, “Induction of decision trees,” *Machine Learning*, vol. 1, pp. 81–106, 1986.
- [70] M. Xu, P. Watanachaturaporn, P. Varshney, and M. K. Arora, “Decision tree regression for soft classification of remote sensing data,” *Remote Sensing of Environment*, vol. 97, no. 3, pp. 322–336, 2005.
- [71] T. Akbar, A. G. Putrada, and F. Dawani, “Decision tree regression with adaboost ensemble learning for water temperature forecasting in aquaponic ecosystem,” in *Proceedings of the 2020 6th International Conference on Interactive Digital Media (ICIDM)*, pp. 1–5, IEEE, Bandung, Indonesia, December 2020.
- [72] B. Schölkopf and J. S. Alexander, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and beyond*, MIT press, Cambridge, MA, USA, 2002.
- [73] I. N. Daliakopoulos, P. Coulibaly, and I. K. Tsanis, “Groundwater level forecasting using artificial neural networks,” *Journal of Hydrology*, vol. 309, no. 1-4, pp. 229–240, 2005.
- [74] M. Hayati and Z. Mohebi, “Application of artificial neural networks for temperature forecasting,” *International Journal of Electrical and Computer Engineering*, vol. 1, no. 4, pp. 662–666, 2007.
- [75] S. Samantaray, A. Sahoo, and D. Prakash Sathpathy, “Temperature prediction using hybrid mlp-Goa algorithm in keonjhar, odisha: a case study,” in *Proceedings of the Smart Intelligent Computing and Applications, Volume 1: Proceedings of Fifth International Conference on Smart Computing and Informatics (SCI 2021)*, pp. 319–330, Springer, Singapore, April 2022.
- [76] C. W. Dawson and R. L. Wilby, “Hydrological modelling using artificial neural networks,” *Progress in Physical Geography*, vol. 25, no. 1, pp. 80–108, 2001.
- [77] D. P. Kingma and B. Jimmy, “Adam: a method for stochastic optimization,” 2014, <https://arxiv.org/abs/1412.6980>.
- [78] M. I. Jordan, “Serial order: a parallel distributed processing approach,” *Advances in Psychology*, vol. 121, pp. 471–495, 1997.
- [79] P. Ren, Y. Xiao, X. Chang et al., “A survey of deep active learning,” *Association for Computing Machinery Computing Surveys*, vol. 54, no. 9, pp. 1–40, 2021.
- [80] Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult,” *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.
- [81] S. Hochreiter, “The vanishing gradient problem during learning recurrent neural nets and problem solutions,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 06, no. 02, pp. 107–116, 1998.
- [82] A. A. Nay, J. El Khoury, and C. Mansour, “Trip-based prediction of hybrid electric vehicles velocity using artificial neural networks,” in *Proceedings of the 2021 IEEE 3rd International Multidisciplinary Conference on Engineering Technology (IMCET)*, pp. 60–65, IEEE, Beirut, Lebanon, December 2021.
- [83] J. V. Tembhurne, M. M. Almin, and T. Diwan, “Mc-dnn: fake news detection using multi-channel deep neural networks,” *International Journal on Semantic Web and Information Systems*, vol. 18, no. 1, pp. 1–20, 2022.
- [84] Y. Li and Y. Lu, “Lstm-ba: ddos detection approach combining lstm and bayes,” in *Proceedings of the 2019 Seventh International Conference on Advanced Cloud and Big Data (CBD)*, pp. 180–185, IEEE, Suzhou, China, September 2019.

- [85] A. S. Santra and J.-L. Lin, "Integrating long short-term memory and genetic algorithm for short-term load forecasting," *Energies*, vol. 12, no. 11, p. 2040, 2019.
- [86] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, <https://arxiv.org/abs/1409.0473>.
- [87] W. Li, T. Logenthiran, and W. L. Woo, "Multi-GRU prediction system for electricity generation's planning and operation," *Institution of Engineering and Technology Generation, Transmission and Distribution*, vol. 13, no. 9, pp. 1630–1637, 2019.
- [88] H. M. Lynn, S. B. Pan, and P. Kim, "A deep bidirectional gru network model for biometric electrocardiogram classification based on recurrent neural networks," *IEEE Access*, vol. 7, pp. 145395–145405, 2019.
- [89] C. F. Gaitán, "Machine learning applications for agricultural impacts under extreme events," in *Climate Extremes and Their Implications for Impact and Risk Assessment*, pp. 119–138, Elsevier, Amsterdam, Netherlands, 2020.
- [90] M. Abadi, B. Paul, J. Chen et al., "Tensorflow: a system for large-scale machine learning," in *Proceedings of the 12th USENIX symposium on operating systems design and implementation (OSDI 16)*, pp. 265–283, Savannah, GA, USA, April 2016.
- [91] N. Ketkar and N. Ketkar, "Introduction to keras," *Deep learning with python: a hands-on introduction*, pp. 97–111, 2017.
- [92] F. Shahid, A. Zameer, and M. Muneeb, "Predictions for covid-19 with deep learning models of lstm, gru and bi-lstm," *Chaos, Solitons and Fractals*, vol. 140, Article ID 110212, 2020.
- [93] R. Fu, Z. Zhang, and L. Li, "Using lstm and gru neural network methods for traffic flow prediction," in *Proceedings of the 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pp. 324–328, IEEE, Wuhan, China, November 2016.
- [94] S. Kumar, L. Hussain, S. Banarjee, and M. Reza, "Energy load forecasting using deep learning approach-lstm and gru in spark cluster," in *Proceedings of the 2018 fifth international conference on emerging applications of information technology (EAIT)*, pp. 1–4, IEEE, Kolkata, India, January 2018.
- [95] S. H. Park, B. D. Kim, M. K. Chang, C. C. Chung, and J. W. Choi, "Sequence-to-sequence prediction of vehicle trajectory via lstm encoder-decoder architecture," in *Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1672–1678, IEEE, Changshu, China, June 2018.
- [96] S. Du, T. Li, Y. Yang, and S. J. Horng, "Multivariate time series forecasting via attention-based encoder-decoder framework," *Neurocomputing*, vol. 388, pp. 269–279, 2020.
- [97] G. Brunet, M. Shapiro, M. Moncrieff et al., "Toward a seamless process for the prediction of weather and climate: the advancement of sub-seasonal to seasonal prediction," *Bulletin of the American Meteorological Society*, 2009.
- [98] M. Scheffer, S. Carpenter, J. A. Foley, C. Folke, and B. Walker, "Catastrophic shifts in ecosystems," *Nature*, vol. 413, no. 6856, pp. 591–596, 2001.