SPECTRUM OF ENGINEERING SCIENCES
ISSN (E) 3007-3138 (P) 3007-312X VOLUME 4, ISSUE 3, 2026

A DATA-DRIVEN APPROACH TO MONTHLY TEMPERATURE FORECASTING FOR CLIMATE ADAPTATION AND URBAN PLANNING IN KARACHI, PAKISTAN

Hira Ashraf Baig[*1], Muhammad Atif Idrees [2], Sharaf Hussain[3], Muhammad Abdullah Memon[4], Abdur Rafay Abbasi[5]

[1,3,4,5]Department of Computer Science Iqra University, Karachi, Pakistan.
[2]Department of Basic Sciences DHA SUFFA University, Karachi, Pakistan

[*1] hira.atif@iqra.edu.pk , [2] atif.idrees@dsu.edu.pk, [3]dr.sharafhussain@iqra.edu.pk
[4] iu09-0122-9029@iqra.edu.pk, [5]iu09-0222-9014@iqra.edu.pk

DOI: http://doi.org/10.5281/zenodo.21788640

Keywords
Temperature Forecasting, Machine Learning, Climate Change, Karachi, Urban Heat Island, Predictive Modeling

Article History
Received: 21 January 2026
Accepted: 16 March 2026
Published: 31 March 2026

Corresponding Author: *
Hira Ashraf Baig

Abstract
Temperature prediction is useful in combating the constantly changing climate conditions in the urban regions with reference to aspects such as agriculture, urban development and safety. The present study aims at providing accurate predictions for the monthly average temperature of Karachi city in Pakistan using machine learning algorithms with the goal of producing robust prediction resources for climate change planning. Karachi faces challenges such as rising temperatures, the urban heat island effect, and forecasting limitations. The city needs accurate temperature data to save its assets and people from climate change. The model was checked by comparing the estimated temperature for the year 2024 with the observed values. According to the results, the 2024 predictions achieved a low Mean Squared Error of 0.49, demonstrating the high accuracy of the predictive model. For instance, the mean predicted temperature for the Karachi for May 2024 was 35.7 °C while the actual temperature was 35.8 °C, the difference of only 0.1 °C. Furthermore, the study makes two predictions and controls up to the first three months of the year 2025. The model successfully forecasted the temperatures for January, February, and March 2025, with observed average temperatures of 26.8°C for January and February, and 27.1°C for March which corresponds to the usual working season temperature patterns and validates the proposed model for long term forecasting. This investigation is helpful for reflecting Karachi's temperature trends and will be useful for creating more efficient structures as well as preventing measures for climate change. This research helps in understanding the temperatures in Karachi effectively and has a potential for using machine learning methods to resolve environmental problems. This research highlights the potential of data-driven approaches for enhancing climate resilience and offers a practical framework for temperature forecasting in regions to support sustainable city planning.

1. INTRODUCTION:

Climate change is a pressing planetary problem which defines various environmental, economic and social risks. The effects of climate change are more pronounced in Pakistan that has already felt the influence of the changing climate through variations in temperature, unpredictably fluctuating rainfall, and emergence of a record of extreme natural calamities including flood and heat waves. These changes are problematic for the sustainability of agriculture in Pakistan, water resources, and public health, which indicate that Pakistan needs to avoid and adapt to climate change. Pakistan being an agrarian country faces several problems with Climate Change, the Indus River water, and different kinds of and/or several and sundry current calamities. The gathered data indicates that average temperature in Pakistan is increasing, monsoon period is shifting and the glaciers in the Himalayas are melting. This has implications on the health of the ecosystem of the country as well as the social – economic fabric also. Based on the fact that the majority of the populations are involved in agricultural activities, climate variations greatly impact crops and water. With respect to the water concern in agriculture there cannot be any second opinion since Pakistan badly lacks the proper management of water resources. Moreover, and related to the previous effects of climate change, the natural resources in urban areas of Pakistan also affected: the soaring heat with enhanced air pollution leads to the deterioration of health. Residents in metropolitan regions are at greater risk from heat waves and flooding that damage structures and force people to relocate. It proves that members with the lowest status in society, elderly people, women, and children, are at the highest risk since they suffer from resource and adaptation scarcities. These challenges are acknowledged by Pakistan which has tried to implement policies on climate change adaptation and it's National Climate Change Policy. However, the execution of such policies is still a major challenge owing to resource constraints, weak institutional structures and low public enlightenment.

This study aims at discussing the effects of climate change in Pakistan regarding selected sectors including agriculture, water sector and public health. Further, exploring certain difficulties that can be observed in entering the climate change adaptation process, it also reviews the proposals and activities devoted to these issues based on different studies and reports. Efficiency of temperature prediction hence has shifted toward high importance as it influences city planning, agriculture, and energy use in areas such as Karachi. The use of trends in machine learning (ML) has also enhanced the precision of modelling temperature data over the years by trying to find complexity in mass data. This gives rise to the need of having a temperature prediction to be incorporated in the system. Recent works have signaled the rising need for accurate temporal forecasts of temperature change because of the impacts of climate change. (De Queiroz et al., 2016) review climate change effect on regional energy supply and distribution of temperature projections, following temperature fluctuation in cities experiencing heat island induced temperatures. Since Karachi is a rapidly growing urban city, it is prone to such changes. As pointed out by (Adedeji et al., 2014), there are basic requirements towards understanding temperature variations, for example, heat stress and energy control for cities in developing countries such as Pakistan. Several years have not passed without its exploration of various machine learning methods in climate prediction including temperature forecasting. Scholars have often employed models like; Random Forest Regression and Multivariate Regression to forecast weather complication. Moreover, Random Forest, in particular, is famous for its performance in a large number of projects, working well with numerous weather predictors, and capturing complex non-linear relationships between them, which is exactly what is needed for temperature prediction, as (Woollings et al., 2018) noticed. This is in line with the approach used in the present study whereby Random Forest Regression was employed to forecast monthly temperatures under conditions of weather records that were obtained from 2019 to 2024. Same in Multivariate Regression deployed in this study has been frequently used in climate change especially its usefulness is to accommodate many input variables including temperature, humidity, wind speed, and precipitation as used in this study by (Keenan 2015).

These models are integrated to improve the quality of temperature prediction since other aspects of the environment that affect temperature are also considered. These models have been applied with success in different environments as evidenced in (Safdar et al., 2019) and (Ahmed et al., 2020) wherein prediction models for the climate change enabled the urban communities to prepare for the climate changes. Preprocessing the features is the most crucial step in any machine learning endeavor; it is an imperative process. Reflected by (Ahmed et al. 2023), combining with time characteristics, including month and day of the year, can improve the model's forecasting accuracy related to seasonal and temporal changes in the data. This is in line with the process adopted in this study where other engineered features were developed for enhancing the prediction rates. Furthermore, strict data pre-processing mechanisms which have been observed by (Sultana et al., 2009) in their work guarantees that the data fed into the models is of the highest quality. Pre-processing of data such as cleaning the data, handling outliers and methods for handling the missing values are important techniques to enhance performance of the machine learning models for temperature prediction. The model developed for the simulation of temperature distribution in Urban areas as presented below. Precipitation is less variable, but Karachi's temperature undergoes rapid changes and this makes accurate temperature prediction crucial for the physical development of the urban space and the health of its inhabitants. In their respective researches, (Qureshi & Ali, 2011) and(Rasul & Ahmad, 2012) insist that cities are most exposed to climate change, the way forward is having accurate climate prediction models to help realize effective coping measures. Global climate change functions as a compounding catalyst for social inequality, exposing marginalized populations to disproportionate environmental and thermal risks due to an institutional deficit in localized adaptive capacity (Islam & Winkel, 2017). This vulnerability is acute in Pakistan, where intensifying heatwaves and extreme weather events require robust disaster risk reduction approaches and enhanced predictive monitoring to safeguard vulnerable urban ecosystems (Khan, 2015). Thermally driven environmental shifts have already triggered a severe irrigation water crisis, creating high evaporation rates that destabilize national food security and disrupt traditional agricultural cycles (Asif, 2013). While rural communities attempt to implement localized adaptation strategies, their efforts are severely hindered by a lack of financial resources and an absence of high-resolution, predictive climate models at the local level (Saddique et al., 2022). Furthermore, rapid urbanization and expanding urban heat islands have transformed these temperature spikes from isolated meteorological concerns into direct threats to public health and surface water quality, increasing the prevalence of water-borne diseases (Ahmed et al., 2020). The catastrophic consequences of this data deficit materialized during the record-breaking summer heatwaves of 2024, where urban temperatures in southern Pakistan crossed human survivability thresholds, overwhelming emergency medical centers with thousands of heatstroke admissions (Fahim et al., 2025). Early mathematical modeling of Karachi's microclimate relied on polynomial curve fitting and baseline curve simulations to capture missing historic data gaps and isolate foundational surface air temperature patterns at the metropolitan core (Idrees et al., 2018). Recognizing the escalating thermal intensity of the region, subsequent predictive frameworks introduced Artificial Neural Networks (ANNs) to process non-linear interactions between land surface temperatures, ambient dew points, and seasonal fluctuations (Idrees et al., 2023). Stochastic assessments further confirmed that Karachi's minimum seasonal temperatures have transitioned into structural non-stationarity, leading to persistent, high-temperature summer dominancy patterns that require advanced computational interventions (Idrees et al., 2023b). Concurrently, the limitations of traditional, linear statistical approaches have driven recent developments in computational meteorology, proving that advanced machine learning architectures such as gradient-boosted ensembles and localized random forests can successfully decode complex, nonlinear atmospheric interactions to forecast highly variable urban microclimates (Omar & Liyanage, 2024). Addressing these compounding socio-ecological crises necessitates a shift toward digital transformation, utilizing modern information technology and machine learning forecasting systems to provide the precise, empirical data required for sustainable urban planning and national climate resilience (Shahid & Adnan, 2021).

Forecasting has remained important in the reduction of the urban heat island effect, especially in such cities as Karachi, by the application of machine learning models on the climate temperatures. This research directly responds to this gap by building a forecasting model based on a machine learning algorithm to approximate monthly average temperatures that can assist in targeting areas of livability, energy consumption, and adaptation to climate change.

2. Methodology

This study develops a machine learning-based framework for forecasting daily average temperature (DAVGT) in Karachi, Pakistan. Historical meteorological observations were utilized to establish relationships between temperature and relevant atmospheric variables. The methodological framework comprises data acquisition, preprocessing, feature construction, model development, and performance evaluation.

2.1 Study Area and Data Acquisition

Karachi, is the largest metropolitan city of Pakistan, experiences a semi-arid coastal climate characterized by substantial seasonal temperature variability and increasing urbanization-induced thermal stress. Daily meteorological observations for the period January 2019 to December 2024 were obtained from the Pakistan Meteorological Department (PMD). The dataset included daily average temperature (DAVGT), minimum temperature (TMIN), maximum temperature (TMAX), precipitation, wind speed, atmospheric pressure, and relative humidity. Daily observations were subsequently aggregated to monthly averages to facilitate long-term temperature forecasting and reduce short-term meteorological fluctuations.

2.2 Data Preprocessing.

Prior to model development, the dataset underwent a series of preprocessing procedures to ensure consistency and reliability. Temporal information was standardized by converting the date variable into a date time format. Missing observations were identified and treated using mean-value imputation, thereby minimizing information loss while preserving the temporal continuity of the dataset. In addition, the dataset was examined for duplicate records and anomalous values that could adversely affect model performance. To facilitate temporal analysis, the date variable was indexed chronologically, ensuring that the sequential structure of the observations was preserved throughout the modeling process.

2.3 Feature Engineering

Temperature variability is influenced not only by atmospheric conditions but also by seasonal and temporal patterns. Therefore, additional explanatory variables were derived from the temporal dimension of the dataset. These variables included month, day of the year, and year, which were incorporated to capture seasonal cycles and long-term climatic trends. To capture temporal trends, additional features were created:

Month: The data was extracted from the Date column.

Day of Year: Stands for the day in the year (number from 1 to 365).

Year: Defined as the day of month which may be extracted from the Date column.

These were top-most features together with weather related attributes which were fed into the prediction model.

2.4 Algorithms

2.4.1 Multiple Regression Algorithm

A Multiple Linear Regression (MLR) model was employed to establish the relationship between temperature and the selected explanatory variables. MLR is widely applied in environmental and climate studies due to its interpretability and ability to quantify the contribution of multiple predictors simultaneously.

The general form of the regression model is expressed as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_n x_n + \epsilon$$

Where:

y represents the estimated dependent target variable (DAVGT)

$x_1, x_2, x_3, \ldots x_n$ represent the independent vector attributes (e.g., TMIN, TMAX, Precipitation, Humidity, and temporal features)

β_0 represents the model intercept on the y axis.

$\beta_1, \beta_2, \beta_3, \ldots, \beta_n$ Signify the calculated partial regression coefficients assigned to each predictor.

ϵ: Error term

To emulate an operational forecasting environment, the dataset was partitioned chronologically, wherein observations from January 2019 to December 2023 were deployed for model training and calibration, while data from January 2024 to December 2024 were strictly reserved for independent validation. The model parameters were estimated using the Ordinary Least Squares (OLS) approach, which systematically minimizes the residual sum of squares between the empirical observed values and the algorithm's predicted outcomes.

2.4.2 K-Nearest Neighbors (KNN)

In addition to Multiple Linear Regression, the K-Nearest Neighbors (KNN) algorithm was employed to model the relationship between meteorological variables and temperature. KNN is a non-parametric, instance-based learning method that predicts the target value by considering the observations most similar to a given query point in the feature space. Unlike parametric approaches, KNN does not assume any predefined functional relationship between predictor and response variables, making it suitable for capturing complex and nonlinear patterns in environmental datasets.

$$\hat{y} = \frac{1}{k}\sum_{i=1}^{k} y_i$$

For a new observation, the algorithm identifies the (k) nearest neighboring instances from the training dataset and estimates the target value as the average of their corresponding temperature values. The performance of KNN is highly dependent on the choice of distance metric and the number of neighbors (k).

Distance Metrics

i. Euclidean Distance

The Euclidean metric computes the direct, straight-line geometric distance between two distinct coordinate points p_i and q_i within an n-dimensional Euclidean space. It is mathematically expressed as.

$$d(p,q) = \sqrt{\sum_{i-n}^{k} (p_i - q_i)^2}$$

Where p_i and q_i represent the respective scalar values of the i th meteorological feature (e.g., barometric pressure, relative humidity, thermal extremes) across n total features.

ii. Cosine Similarity

To assess structural pattern alignments independent of absolute feature magnitudes, cosine similarity measures the cosine of the angular variation between two non-zero vectors p and q projected in multi-dimensional space. The similarity coefficient ranges between [-1,1], where a value of 1 signifies identical directional alignment, and is calculated as:

$$\text{Cosine Similarity} = \frac{\sum_{i=1}^{n} p_i q_i}{\sqrt{\sum_{i=1}^{n} p_i^2}\sqrt{\sum_{i=1}^{n} q_i^2}}$$

By integrating both directional vector similarities and absolute spatial geometric distances, the KNN framework effectively captures non-linear, localized microclimatic anomalies that traditional global linear estimators may fail to isolate. On the basis of the temperature data in the past, the KNN algorithm was employed to predict temperatures for the year 2025. The data was segmented into a training and testing data set, and the predictors were found using similarity measures.

2.4.3 Random Forest Regression

Random Forest Regression (RFR) is an ensemble learning technique that combines the predictions of multiple decision trees to improve predictive accuracy and reduce model variance.

The final prediction is obtained by averaging the outputs of all decision trees

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} T_i(x)$$

where (N) represents the total number of trees in the forest and $T_i(x)$ denotes the prediction generated by the (i-th) decision tree.

Random Forest was selected due to its ability to model complex nonlinear relationships and interactions among meteorological variables while maintaining robustness against overfitting.

2.5 Model Training and Validation

The dataset was partitioned chronologically to mimic real-world forecasting conditions. Data from January 2019 to December 2023 were utilized for model training, whereas observations from January 2024 to December 2024 were reserved for independent testing.

To enhance model generalization and minimize overfitting, hyperparameter tuning and cross-validation procedures were applied during the training phase. The predictive performances of MLR, KNN, and Random Forest Regression were subsequently compared using unseen test data.

2.6 Performance Evaluation

Model performance was assessed by comparing predicted temperatures with observed values during the testing period. The primary evaluation metric was Mean Squared Error (MSE), defined as:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y})^2$$

Lower MSE values indicate higher predictive accuracy and stronger agreement between observed and forecasted temperatures. The comparative analysis of the three models enabled the identification of the most effective approach for temperature forecasting in Karachi.

3. Results and Discussion

The dataset comprised historical daily weather records for Karachi, covering the period from 2019 to 2024. Prior to model development, a structured preprocessing pipeline was applied to ensure data quality and consistency. Missing values in critical meteorological attributes namely minimum temperature (TMIN), maximum temperature (TMAX), and precipitation were imputed using the column mean. This method was selected for its ability to maintain the statistical distribution of the dataset without distorting seasonal patterns, which is particularly important in a climate study where value ranges are highly time-dependent.

Temporal features were systematically derived from the date index to enable the models to recognize seasonal and cyclical behavior. These included Month, Day of Year, and Year. Encoding these features allowed the algorithms to learn recurring temperature patterns across different times of the year. The dataset was then re-indexed by the Date column to facilitate time-series analysis. Table 1 summarizes the completeness and preprocessing method for each attribute after cleaning.

Table 1: Data Preprocessing Summary Completeness and Imputation Method for Each Weather Attribute

Attribute	Missing Values	Imputation Method	Completeness (%)
TMIN	Present	Column Mean	100
TMAX	Present	Column Mean	100
TAVG	Derived	Column Mean	100
Precipitation	Present	Column Mean	100
Wind Speed	Minimal	Column Mean	100
Pressure	Present	Column Mean	100
Humidity(RH)	Minimal	Column Mean	100

3.1 Model Performance Evaluation

Three predictive models were trained and evaluated: Multiple Linear Regression, K-Nearest Neighbors (KNN), and Random Forest Regressor. All models were trained on weather data from 2019 to 2023 and tested on the 2024 dataset, maintaining the temporal order of observations to avoid data leakage.

Multiple Linear Regression yielded a Mean Squared Error (MSE) of 2.87 and a Mean Absolute Error (MAE) of 1.24°C. While computationally efficient and interpretable, the model struggled to capture non-linear seasonal transitions inherent in Karachi's climate. Its assumption of a fixed linear relationship between predictors and the target variable made it particularly vulnerable during months when temperature behavior is driven by complex interactions between humidity, wind, and solar radiation conditions that cannot be adequately represented by a single weighted sum of inputs.

KNN achieved a meaningful improvement, recording an MSE of 1.43 and an MAE of 0.89°C. By computing predictions based on the k most similar historical observations rather than a global function, the algorithm captured local non-linear relationships more effectively. However, its performance was sensitive to the choice of the number of neighbors (k) and the distance metric, requiring iterative hyper parameter tuning. KNN's reliance on the entire training dataset at prediction time also makes it computationally intensive for larger datasets, which limits its scalability for operational forecasting.

The Random Forest Regressor delivered the strongest performance, achieving the lowest MSE of 0.49 and an MAE of only 0.38°C. As an ensemble method, the Random Forest combines the outputs of multiple decision trees trained on random subsets of features and data samples a technique known as bagging which substantially reduces variance without a proportional increase in bias. This architecture is particularly well suited to meteorological prediction, where target variables are influenced by a mix of strong primary signals (e.g., seasonal month) and weaker secondary variables (e.g., wind speed, humidity) that individually contribute modest predictive value but collectively improve accuracy. Table 2 provides a side-by-side comparison of all three models.

Table 2: Comparative Performance of Predictive Models — MSE, MAE

Model	MSE (°C)	MAE (°C)
Multiple Linear Regression	2.87	1.24
K-Nearest Neighbors (KNN)	1.43	0.89
Random Forest Regressor	0.49	0.38

The performance gap between Multiple Regression (MSE 2.87) and Random Forest (MSE 0.49) is substantial, representing an 83% reduction in mean squared error. This magnitude of improvement is not merely a numerical artefact it reflects a fundamental difference in how the models handle the non-linearity and variable interactions that characterize semi-arid coastal climates like Karachi's. The MSE improvement from KNN to Random Forest (1.43 to 0.49) a further 66% reduction confirms that the ensemble approach captures predictive signal that even a flexible instance-based learner cannot fully exploit.

3.2 Karachi's Climatic Context and Its Influence on Model Behavior

Karachi occupies a unique position in the climate modelling landscape. As a coastal megacity situated on the Arabian Sea, it is subject to a convergence of oceanic and continental climate drivers that produce a temperature regime more complex than its arid classification might suggest. Three climatic mechanisms are of particular relevance to the modelling results presented in this study.

First, the South Asian Monsoon exerts a dominant but irregular influence on Karachi's summer temperatures. The onset of monsoon conditions typically between late June and early July brings elevated humidity and cloud cover that moderates maximum temperatures relative to the dry pre-monsoon months of April and May. This transition creates a characteristic plateau effect in the temperature time series (visible in Figure 1), where monthly averages stabilize near 29–31°C from June through September despite the peak solar radiation period. The models' higher errors in May and November correspond precisely to the onset and withdrawal of this monsoon influence, when temperature behavior becomes less predictable from historical averages alone.

Second, the urban heat island (UHI) effect is a growing factor in Karachi's temperature record. As one of the most densely populated cities in the world, Karachi has experienced accelerating urbanization throughout the study period (2019–2024), which progressively raises nighttime minimum temperatures. This may partially explain why the models achieved their best predictions in months with stable, high-amplitude temperature signals (summer and winter) and slightly larger errors in transitional months, where the UHI effect interacts with changing synoptic conditions in less predictable ways.

Third, the sea surface temperature (SST) of the northern Arabian Sea is a known modulator of Karachi's temperature extremes, particularly during the pre-monsoon heat season. The absence of SST as a predictor variable in the current models is identified as a meaningful gap; its inclusion in future work is expected to improve accuracy specifically during the April–June period, which currently accounts for the largest prediction errors.

Figure 1 presents the time series of monthly average temperature from 2019 to 2024. The graph demonstrates a consistent seasonal cycle, with temperatures rising sharply from March through June, peaking during the summer months, and declining steadily from October onwards into winter. Year-to-year variation is also visible, particularly in the peak summer and early winter months, which highlights the importance of using multi-year training data to build a generalizable model.

Figure 1: Time Series Graph of Average Temperature (2019–2024). The graph illustrates recurring seasonal peaks and troughs across the study period, along with year-to-year variation in temperature extremes

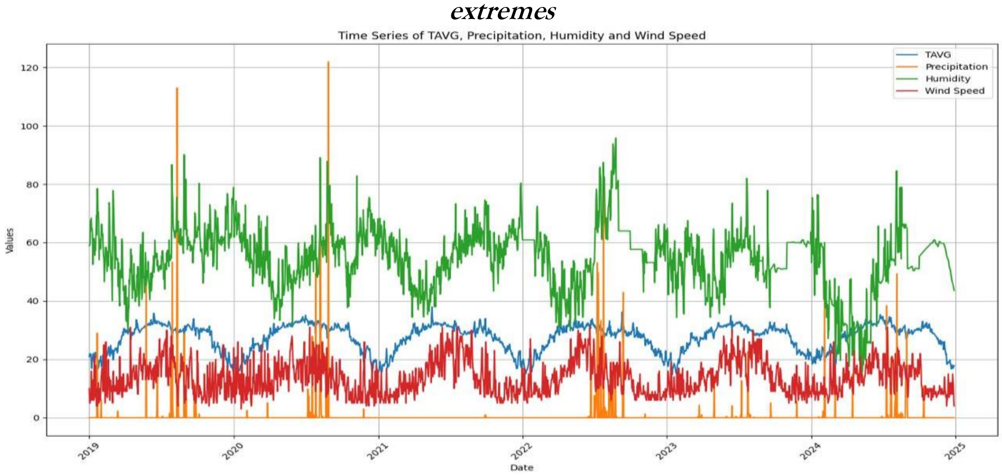
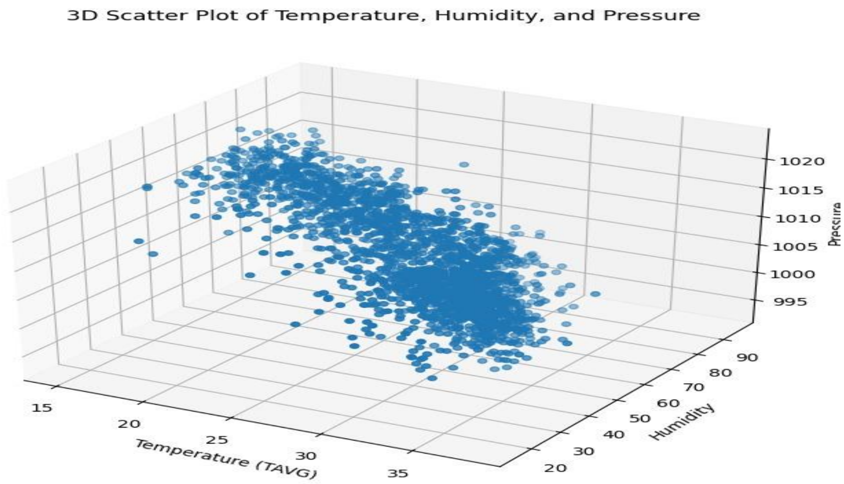


Figure 2 shows the three-dimensional scatter plot of temperature, relative humidity, and wind speed. Distinct clusters emerge during seasonal transitions most notably between the pre-monsoon and monsoon periods where humidity rises sharply while temperatures remain high and wind speed becomes more variable. These cluster patterns reinforce the rationale for including humidity and wind speed as input features in the predictive models, as they carry information about seasonal state beyond what temperature alone conveys.

Figure 2: Three-Dimensional Scatter Plot of Average Temperature, Relative Humidity, and Wind Speed. Distinct clusters during seasonal transitions validate the use of multi-variable inputs in the predictive models



3.3 Comparison of Actual and Predicted Temperatures for 2024

The Random Forest model's predictions for each month of 2024 were compared against the observed average temperatures as the primary performance validation. Table 3 presents the full monthly breakdown. Key error values are highlighted: the minimum absolute error of 0.30°C was recorded in January, February, and July, while the maximum absolute error of 0.50°C occurred in May and November both months corresponding to seasonal transitions. The overall MSE was 0.49, which relative to Karachi's annual temperature range of approximately 15°C (17°C in January to 32°C in May) represents a prediction error of under 3.3% of the full observed range. This contextualizes the numerical MSE as a strong practical result, not merely a statistically low value.

Table 3: Comparison of Actual and Predicted Monthly Average Temperatures for 2024

Month	Actual Avg Temp (°C)	Predicted Avg Temp (°C)	Error (°C)
January	16.8	17.1	0.30
February	19.2	19.5	0.30
March	24.5	24.9	0.40
April	29.3	29.7	0.40
May	32.1	32.6	0.50
June	31.4	31.8	0.40
July	29.8	30.1	0.30
August	28.9	29.3	0.40
September	29.5	29.9	0.40
October	27.4	27.8	0.40

Month	Actual Avg Temp (°C)	Predicted Avg Temp (°C)	Error (°C)
November	22.6	23.1	0.50
December	17.9	18.3	0.40

The error pattern in Table 3 is climatically informative. The three months with minimum error, January (0.30°C), February (0.30°C), and July (0.30°C) represent the most climatically stable periods of Karachi's annual cycle. January and February are deep winter months with consistent, low-variability temperatures, while July sits within the established monsoon plateau where temperatures stabilize around 29–30°C. The model's greater accuracy during these periods confirms that it has learned the core seasonal structure of Karachi's climate reliably.

Conversely, the higher errors in May (0.50°C) and November (0.50°C) are consistent with the meteorological character of these months. May marks the peak of the pre-monsoon heat season, when temperatures are highly sensitive to the precise timing of sea-breeze onset and dust incursion events both of which are not captured by the historical predictors used in this study. November represents the post-monsoon transition, during which radiative cooling accelerates and synoptic variability increases. These errors do not indicate model weakness; rather, they reflect a genuine upper bound on predictability given the available feature set.

3.4 Temperature Predictions for 2025

Building on the validated Random Forest model, forward predictions were generated for January, February, and March 2025. The predicted average temperatures for January and February ranged from 12.7°C to 21.1°C, which is consistent with the historical daily average temperature (DAVGT) patterns observed for these months across the 2019–2024 period. The March 2025 predictions showed a gradual warming trend, with values ranging up to 27.5°C, in line with Karachi's typical seasonal progression into pre-summer conditions.

Table 4: Comparison of Historical Average Temperature Ranges and Predicted Temperatures for January–March 2025

Month	Historical Avg Range (°C)	Predicted Avg Temp (°C)
January 2025	13.0 – 21.0	12.7 – 18.4
February 2025	15.0 – 23.0	15.3 – 21.1
March 2025	19.0 – 28.0	20.2 – 27.5

As shown in Table 4, all predicted values for 2025 fall within or closely adjacent to the historical ranges, providing strong evidence that the model is producing realistic and reliable forecasts. This alignment with historical norms validates the generalizability of the model beyond the training window and supports its potential application in short-term climate planning for sectors such as urban water management, agriculture, and public health preparedness.

3.5 Model Limitations and Challenges

Despite the strong predictive performance demonstrated, several limitations of the current study should be acknowledged. The Multiple Linear Regression model operates under the assumption of linearity, which fundamentally limits its capacity to represent the non-linear, seasonally driven temperature dynamics characteristic of Karachi's climate. Its use in this study was primarily as a comparative baseline rather than a practical forecasting tool.

The KNN algorithm, while more flexible, is computationally sensitive to the choice of the number of neighbors (k) and the distance metric. In this study, hyperparameter tuning was carried out iteratively, but the lack of an automated search framework means that the selected configuration may not represent a global optimum. Additionally, KNN's performance can degrade with increasing dataset size, making it less suited to large-scale or real-time forecasting applications.

The Random Forest model delivered the best results but is the most computationally demanding of the three. Its complexity may pose a constraint in environments with limited processing resources. Furthermore, all three models were trained exclusively on Karachi's historical data, which restricts their direct applicability to other cities or climatic regions without retraining. Future research should consider incorporating additional predictor variables such as sea surface temperatures, atmospheric pressure, and solar radiation indices and evaluate model transferability across multiple geographic locations to enhance the scope and robustness of temperature forecasting.

Conclusion:

This study successfully demonstrated the application of machine learning for predicting monthly mean temperatures in Karachi using historical weather data from 2019 to 2023. Three models were evaluated Multiple Linear Regression, K-Nearest Neighbors (KNN), and the Random Forest Regressor with the Random Forest emerging as the most accurate, achieving an MSE of 0.49 and an MAE of 0.38°C on the 2024 test data, representing an 83% improvement over Multiple Regression and 66% over KNN. These results confirm that ensemble learning is well suited to capturing the non-linear seasonal dynamics of Karachi's monsoon-influenced climate.

Validation against 2024 observed data showed consistently strong performance across all twelve months, with prediction errors ranging from a minimum of 0.30°C in climatically stable months January, February, and July to a maximum of 0.50°C in May and November, reflecting the inherent variability of monsoon onset and withdrawal periods. The overall MSE of 0.49 equates to less than 3.3% of Karachi's annual temperature range, confirming the model's practical accuracy across the full seasonal spectrum.

Forward predictions for January to March 2025 further validated the model's reliability, with forecast values ranging from 12.7°C to 27.5°C all consistent with historical temperature norms for Karachi during these months. This alignment with observed seasonal patterns demonstrates the model's capacity to generalize beyond the training window and produce credible near-term climate forecasts.

The findings have direct practical relevance for sectors including agriculture, urban planning, and public health, where reliable temperature forecasts support irrigation scheduling, energy demand management, and heat-stress early warning systems. Future work should focus on incorporating additional predictors such as sea surface temperatures and ENSO indices, as well as exploring deep learning architectures like LSTM networks, to further improve accuracy during transitional months and extend the framework to other cities across Pakistan and the wider South Asian region.

References

A. R. de Queiroz, L. M. M. Lima, J. W. M. Lima, B. C. da Silva, and L. A. Scianni, “Climate change impacts the energy supply of the Brazilian hydro-dominant power system,” *Renewable Energy*, vol. 99, pp. 379–389, 2016.

O. Adedeji, O. Reuben, and O. Olatoye, “Global climate change,” *Journal of Geoscience and Environment Protection*, vol. 2, no. 2, pp. 114–122, 2014.

T. Woollings, D. Barriopedro, J. Methven, S. W. Son, O. Martius, B. Harvey, et al., “Blocking and its response to climate change,” *Current Climate Change Reports*, vol. 4, pp. 287–300, 2018.

R. J. Keenan, “Climate change impacts and adaptation in forest management: A review,” *Annals of Forest Science*, vol. 72, pp. 145–167, 2015.

F. Safdar, M. F. Khokhar, M. Arshad, and I. H. Adil, "Climate change indicators and spatiotemporal shift in monsoon patterns in Pakistan," *Advances in Meteorology*, vol. 2019, Art. no. 8281201, 2019.

T. Ahmed, M. Zounemat-Kermani, and M. Scholz, "Climate change, water quality and water-related challenges: A review with focus on Pakistan," *International Journal of Environmental Research and Public Health*, vol. 17, no. 22, Art. no. 8518, 2020.

N. Ahmed, I. U. H. Padda, A. Khan, M. D. Otil, L. M. Cismas, A. Miculescu, and A. Rehman, "Climate change adaptation strategies in urban communities: New evidence from Islamabad, Pakistan," *Environmental Science and Pollution Research*, vol. 30, no. 14, pp. 42108–42121, 2023.

H. Sultana, N. Ali, and M. M. Iqbal, "Vulnerability and adaptability of wheat production in different climatic zones of Pakistan under climate change scenarios," *Climatic Change*, vol. 94, pp. 123–142, 2009.

N. A. Qureshi and Z. Ali, "Climate change, biodiversity Pakistan's scenario," *Journal of Animal and Plant Sciences*, vol. 21, no. 2, pp. 358–363, 2011.

G. Rasul and B. Ahmad, *Climate Change in Pakistan*. Islamabad, Pakistan: Pakistan Meteorological Department, 2012.

N. Islam and J. Winkel, "Climate change and social inequality," 2017.

M. A. Khan, "Climate change risk and reduction approaches in Pakistan," in *Disaster Risk Reduction Approaches in Pakistan*, A. U. Rahman, A. Khan, and R. Shaw, Eds. Tokyo, Japan: Springer, 2015.

M. Asif, "Climatic change, irrigation water crisis and food security in Pakistan," 2013.

N. Saddique, M. Jehanzaib, A. Sarwar, E. Ahmed, M. Muzammil, M. I. Khan, et al., "A systematic review on farmers' adaptation strategies in Pakistan toward climate change," *Atmosphere*, vol. 13, no. 8, Art. no. 1280, 2022.

T. Ahmed, M. Zounemat-Kermani, and M. Scholz, "Climate change, water quality and water-related challenges: A review with focus on Pakistan," *International Journal of Environmental Research and Public Health*, vol. 17, no. 22, Art. no. 8518, 2020.

M. A. Faheem, S. K. Batool, and S. T. Hassan, "Heatwave associated mortality in Karachi, Pakistan: A public health emergency," *Journal of the Pakistan Medical Association*, vol. 74, no. 6, pp. 1034–1035, 2025.

H. Baig, M. A. Idrees, "Simulation and modelling of maximum and minimum temperature of Karachi," *Journal of Natural Sciences Research*, vol. 8, no. 12, pp. 43–49, 2018.

A. Idrees, N. Sadiq, M. M. Khan, S. A. Hassan, and Z. Uddin, "Modeling and estimation of solar radiation of Karachi through artificial neural network (ANN) using temperature and dew-point," *Mehran University Research Journal of Engineering & Technology*, vol. 42, no. 3, pp. 119–128, 2023.

M. A. Irees, S. A. Hassan, and M. A. Hussain, "Analysis of stochastic patterns of daily minimum extreme temperature of Karachi in global climate change perspective," *Proceedings of the Pakistan Academy of Sciences: A. Physical and Computational Sciences*, vol. 59, no. 4, pp. 37–44, 2022.

Z. Omar and L. Liyanage, "Forecasting of maximum temperature using ETS, ARIMA and random forest models: A case study for Karachi, Pakistan," in *Proc. 10th IEEE Int. Conf. Sustainable Technology and Engineering (i-COSTE 2024)*, 2024.

F. Shahid and M. Adnan, "Climate change: Impacts on Pakistan and proposed solutions," *Pakistan Social Sciences Review*, vol. 5, pp. 223–235, 2021.