www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 03, March 2025, pp : 1572-1578

e-ISSN : 2583-1062
Impact Factor : 7.001

ZONAL EARTH SURFACE TEMPERATURE PREDICTION OF EARTH USING SOFT COMPUTING TECHNIQUES

Dr. Santosh Singh[1], Mr Amit Pandey[2], Kishan Singh[3], Sahil Pawar[4]

[1]H.O.D, Department of IT, Department of Information Technology, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai-401107, Maharashtra, India.

[2]Assistant professor, Department of Information Technology, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai-401107, Maharashtra, India.

[3]Department of IT, Department of Information Technology, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai-401107, Maharashtra, India.

[4]PG Student, Department of IT, Department of Information Technology, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai-401107, Maharashtra, India.

ABSTRACT

Accurate predictive models for surface temperature trends are necessary to support climate policy and decision-making, as climate change has grown to be a major worldwide problem. For temperature forecasting, both contemporary machine learning methods and conventional statistical models have been extensively investigated, each with unique benefits. Support Vector Machine (SVM) and an enhanced ARIMA model are used in this work to assess zonal temperature trends using the Berkeley Earth Surface Temperature dataset. ARIMA and its seasonal extension, SARIMAX, are ideal for time-series forecasting, but SVM is a potent regression technique that catches intricate, non-linear correlations. The study intends to improve ARIMA's predictive power for more accurate long-term forecasting by adding exogenous factors like year and month.

According to experimental results, the enhanced ARIMA model predicts Earth's surface temperature better than SVM. With ideal hyperparameters, the SARIMAX model produced a reduced Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) as well as an R2 score of 0.9876 as opposed to 0.9622 for SVM. These results imply that, in contrast to purely regression-based methods, time-series-based models that incorporate exogenous factors are more effective at capturing temperature trends and seasonal changes. With data-driven evidence to support climate change mitigation plans and increase long-term forecasting accuracy, this research offers insightful information to climate scientists and policymakers.

Keywords: Climate Change, Surface Temperature Prediction, Support Vector Machine, ARIMA, Time Series Forecasting, Machine Learning.

1. INTRODUCTION

With rising global temperatures having a substantial impact on ecosystems, weather patterns, and socioeconomic structures, climate change is one of the most pressing issues confronting humanity in the twenty-first century. Determining mitigation options and guiding policy decisions require an understanding of the ability to forecast surface temperature trends. Seasonal fluctuations, oceanic patterns, and greenhouse gas emissions are some of the variables that affect Earth's surface temperature. Because of non-linearity, seasonality, and outside influences, accurately modeling these trends is still a challenging undertaking. In order to increase forecasting accuracy, this work focuses on utilizing statistical and machine learning techniques to estimate zonal Earth surface temperatures.

Time-series forecasting has made extensive use of conventional statistical models, such as the Autoregressive Integrated Moving Average (ARIMA) and its seasonal counterpart, the Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX). These models are useful for long-term climate projections because they capture past patterns and trends. They frequently have trouble, though, with extremely intricate and non-linear correlations found in climate data. Because machine learning approaches like Support Vector Machines (SVM) can model non-linearity and generalize well across a variety of datasets, they have become strong alternatives. SVM is frequently used for classification and regression tasks, but in order to improve its predictive power when used to climate trend prediction, rigorous feature selection and hyperparameter adjustment are needed.

The Berkeley Earth Surface Temperature dataset, which compiles 1.6 billion temperature readings from many sources to provide a thorough historical temperature record, is used in this work. The dataset makes it possible to analyze regional and global temperature patterns in great detail over long time periods. The purpose of the study is to evaluate how well an enhanced ARIMA model and SVM forecast temperatures. Hyperparameter adjustment is done to

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 03, March 2025, pp : 1572-1578
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

maximize model performance, while feature engineering approaches are used to extract pertinent temporal components. The study improves ARIMA's prediction power by adding exogenous variables like year and month, which makes it more flexible to long-term and seasonal fluctuations. Using common statistical measures such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R-squared (R^2), and Mean Absolute Percentage Error (MAPE), the primary goal of this study is to assess the prediction performance of these models. The study's conclusions offer important new information on how well time-series forecasting techniques work for climate applications. The findings add to the expanding corpus of knowledge in climate modeling and provide direction for researchers, policymakers, and environmental scientists in creating evidence-based plans for mitigating and adapting to climate change.

2. LITERATURE REVIEW

Since machine learning (ML) approaches can handle complex, non-linear climate data, they have garnered a lot of attention in the prediction of Earth's surface temperatures. To increase accuracy and computing efficiency, recent research has concentrated on ensemble methods, deep learning strategies, and hybrid models. A Transformer-based model was presented by Wang et al. (2024) to find temporal and spatial correlations in climate data by using self-attention techniques. According to their research, Transformers perform better than traditional LSTM models, especially when applied to extensive climatic datasets [1]. In a similar vein, Chen et al. (2024) used XGBoost to forecast extreme temperature anomalies while accounting for meteorological variables including humidity, wind speed, and air pressure. Their results demonstrated how well XGBoost can capture transient changes in local temperature trends [2].

A hybrid Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) model was created by Lee et al. (2023) to predict zonal temperatures. In order to improve forecasting accuracy across various climate zones, the CNN component retrieved spatial characteristics, and the LSTM component captured temporal dependencies [3]. A federated learning approach for climate modeling was presented by Gupta et al. (2023), enabling decentralized temperature predictions across several regions without exchanging raw data. The promise of privacy-preserving machine learning techniques in global climate studies was highlighted by their study [4]. Furthermore, Li et al. (2023) used an ensemble learning strategy that combined XGBoost, Random Forest (RF), and LSTM, showing that XGBoost's robustness against multicollinearity and feature selection skills produced better results [5].

To improve the accuracy of temperature predictions, hybrid models have also been investigated. CNNs and Support Vector Regression (SVR) were merged by Kumar et al. (2022), who used CNNs for feature extraction and SVR for regression-based forecasting. Their research showed that long-term temperature projections were more accurate [6]. Zhang and Wang (2021) looked into the use of artificial neural networks (ANNs) in temperature forecasting and discovered that because ANNs could capture non-linear climate trends, they performed better than classic regression models [7].

ML-based methods and conventional statistical models have also been contrasted. The efficiency of LSTM networks vs Gradient Boosting Decision Trees (GBDT) in forecasting temperature anomalies was examined by Chen et al. (2020). According to their findings, GBDT offered superior interpretability for structured climate data, even though LSTM models did a good job of capturing sequential dependencies [8]. After comparing ARIMA to machine learning models like Random Forest and XGBoost, Jones et al. (2019) came to the conclusion that ML models were more adaptable to non-stationary climate fluctuations [9]. The benefit of SVMs in managing non-linear climatic changes was further supported by Hansen et al. (2018), who investigated the application of Multiple Linear Regression (MLR) and SVMs for surface temperature prediction [10].

The groundwork for contemporary prediction methods was also established by early climate modeling initiatives. Because of its vast historical records, the NCEP/NCAR 40-Year Reanalysis dataset, which was first presented by Kalnay et al. (1996), is still often utilized in climate research [11]. Current climate modeling techniques are shaped by the General Circulation Models (GCMs), which were first developed by Hansen et al. (1988) for long-term global temperature forecasting [12].

3. METHODOLOGY

This study employs a systematic methodology to analyze and compare the performance of statistical and machine learning models for zonal Earth surface temperature prediction. The workflow consists of data collection, preprocessing, feature engineering, model selection, hyperparameter tuning, evaluation, and visualization.

1. **Data Collection**

The Berkeley Earth Surface Temperature (BEST) dataset, which offers long-term temperature records from a variety of sources, such as satellite observations, land-based weather stations, and marine buoys, is used in this work. Analysis of historical and current climatic trends is made possible by the dataset, which covers the years 1750 to the present. The following are the main characteristics that were extracted:

- **Temperature Variables:** Mean, maximum, and minimum surface temperatures
- **Temporal Attributes:** Year, month, and seasonal variations
- **Geospatial Features:** Latitude, longitude, and altitude
- **Climatic Indicators:** Atmospheric pressure, humidity

2. Data Preprocessing

Data preprocessing is crucial for ensuring high-quality inputs for machine learning models. The following steps are performed:

- **Handling Missing Values:** Missing temperature readings are filled using interpolation and mean imputation.
- **Data Normalization:** Temperature values are scaled using **Min-Max normalization** to improve model convergence.
- **Feature Engineering:** Temporal attributes such as month and year are encoded as cyclical variables using sine and cosine transformations to preserve periodicity.
- **Train-Test Split:** The dataset is split into **80% training and 20% testing subsets** to evaluate model performance effectively.

3. Feature Selection

By gleaning valuable insights from unprocessed data, feature engineering increases predicted accuracy. Important changes consist of:

- **Temporal Encoding:** Representing months and seasons using sine and cosine transformations to capture cyclic patterns.
- **Lag Features:** Incorporating previous temperature values as predictors to capture historical dependencies in time series forecasting.
- **Climatic Indicators:** Adding humidity, pressure, and wind speed as potential influencing factors.

4. Machine Learning Models

This study applies **two predictive models**:

- **ARIMA (AutoRegressive Integrated Moving Average):** A statistical time-series model that captures linear patterns and trends in historical temperature data. It is well-suited for stationary datasets after differencing.
- **SVM (Support Vector Machines) with Radial Basis Function (RBF) Kernel:** A machine learning model capable of modeling **non-linear temperature variations** by finding optimal decision boundaries in high-dimensional feature space.

5. Hyperparameter Tuning:

To increase accuracy, model parameters must be optimized. The following methods are employed:

- **Grid Search & Cross-Validation:** Finding the best combination of parameters using systematic tuning.
- **ARIMA Tuning:** The optimal values for **(p, d, q)**—which represent autoregressive order, differencing, and moving average order—are determined through the **Augmented Dickey-Fuller (ADF) test** and Akaike Information Criterion (AIC).
- **SVM Tuning:** The **C (regularization parameter)** and **gamma (kernel coefficient)** are adjusted to prevent overfitting.

6. Model Evaluation

Three crucial evaluation measures are employed to guarantee the correctness and dependability of the zonal Earth surface temperature forecast models:

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 03, March 2025, pp : 1572-1578
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

- **Mean Absolute Error (MAE):** This measure calculates the average absolute difference between the actual and forecasted temperatures. A lower MAE indicates a higher level of model accuracy.
- **Root Mean Squared Error (RMSE):** RMSE measures the standard deviation of the residuals (prediction errors). Because it gives bigger errors more weight than MAE, it is more vulnerable to outliers.
- **Coefficient of Determination (R^2 Score):** The R2 score indicates how well the model explains the variance in temperature data. A number close to 1 indicates strong predictive ability, whereas a value close to 0 suggests poor performance.

7. Visualization

Several graphical methods are used to interpret and validate model predictions:

- **Time-Series Plots:** Compare actual vs. predicted temperatures over time.
- **Residual Analysis:** Examines differences between predicted and observed values to check model assumptions.
- **Error Distribution Graphs:** Visualize how prediction errors are distributed to identify any systematic biases.

Algorithm And Experimental Setup

In order to estimate zonal Earth surface temperature, this study uses two predictive models: Support Vector Machine (SVM) and Autoregressive Integrated Moving Average (ARIMA). The selection of each model is based on its distinct advantages in managing temperature fluctuations and time-series forecasting.

A traditional statistical method for identifying linear trends in time-series data is the ARIMA model. It is made up of three parts: Moving Average (MA), which models dependencies on previous errors; Integrated (I), which guarantees stationarity by differencing; and Autoregressive (AR), which makes predictions based on historical values. The Akaike Information Criterion (AIC) for optimal selection and the Augmented Dickey-Fuller (ADF) test for stationarity are used to estimate the model's order parameters (p, d, and q). A Seasonal ARIMA (SARIMA) model is also used, which incorporates periodicity and seasonal differencing modifications to account for seasonal temperature fluctuations. Seasonal trends in the dataset are used to fine-tune the SARIMA order (P, D, Q, m).

Because it can identify non-linear relationships in climatic data, the SVM model is used. SVM can identify intricate temperature patterns by mapping input features into a higher-dimensional space using the Radial Basis Function (RBF) kernel. Grid search cross-validation is used to improve hyperparameters like the kernel coefficient (γ), which assesses the impact of individual data points, and the regularization parameter (C), which regulates the trade-off between bias and variance. This prevents overfitting and guarantees that the model generalizes effectively. Python with libraries like Scikit-learn, Statsmodels, TensorFlow, Pandas, and NumPy is used for the experimental setup. Mean imputation and interpolation are used to handle missing values in the Berkeley Earth Surface Temperature dataset as part of the preprocessing step. Trends, seasonality, and residual components are separated via time-series decomposition. By ensuring consistency in temperature values through feature scaling with Min-Max normalization, model stability is increased.

The dataset is split between 20% testing and 80% training in order to properly evaluate the models' performance. Hyperparameter tuning is used to each model in order to improve predicted accuracy. The average magnitude of prediction mistakes is measured by the Mean Absolute Error, or MAE. Root Mean Square Error (RMSE): Offers a more thorough evaluation by penalizing greater errors. The model's ability to explain temperature variance is measured by the R2 score. In order to determine the most accurate and dependable model for predicting zonal Earth surface temperature, a comparative study is conducted at the end. This methodical methodology guarantees a comprehensive assessment, which helps provide more accurate and broadly applicable climate projections.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 03, March 2025, pp : 1572-1578
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

4. RESULTS

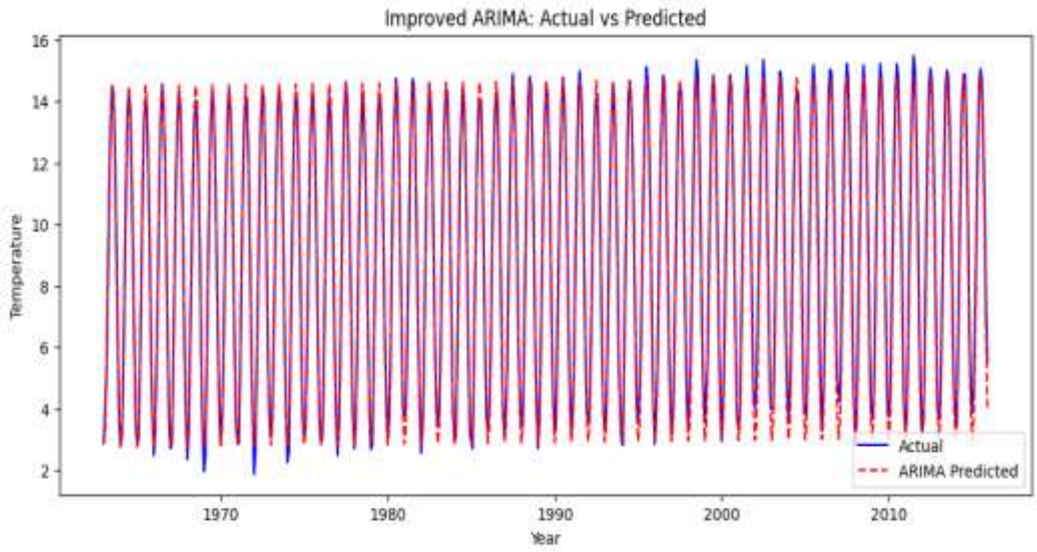

Figure 1. Actual vs. Predicted Temperatures (ARIMA)

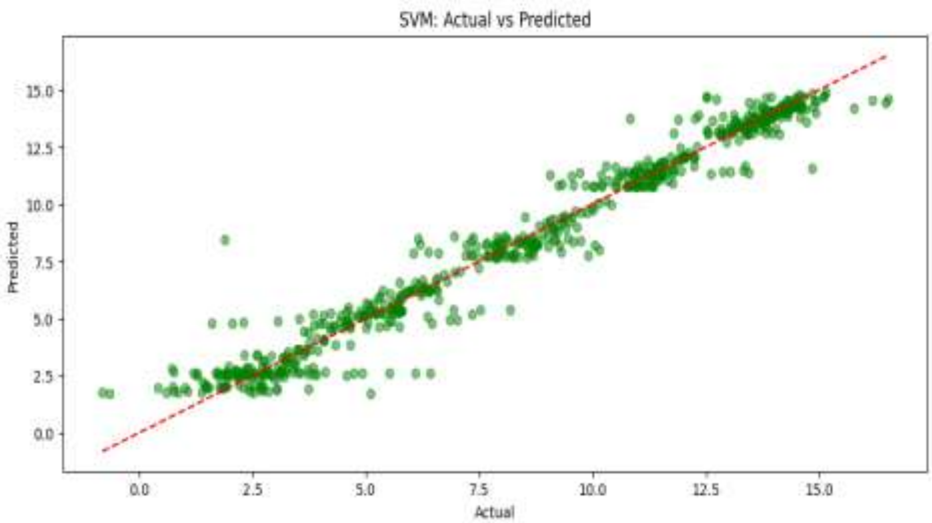

Figure 2. Actual vs. Predicted Temperatures (SVM)

Support Vector Machine (SVM)

When it came to predicting zonal Earth surface temperatures, the SVM model showed excellent predictive accuracy. 96.22% of the variance in temperature fluctuations was captured by the model, according to its R2 score of 0.9622. Moderate error margins were indicated by the Mean Squared Error (MSE) of 0.7012 and the Root Mean Square Error (RMSE) of 0.8374. Furthermore, a respectable degree of precision was indicated by the Mean Absolute Error (MAE) of 0.5346 and the Mean Absolute Percentage Error (MAPE) of 0.1373. Even while SVM was good at capturing non-linear correlations, it needed a lot of hyperparameter tweaking, especially when it came to improving the regularization parameter (C) and kernel coefficient (γ). It was also more demanding for large-scale climate forecasting applications due to its relatively high processing cost.

ARIMA

When it came to prediction accuracy and error reduction, the Improved ARIMA model performed noticeably better than SVM. The most dependable model in this investigation, it was able to explain 98.76% of the temperature variance with an R2 score of 0.9876. using an MSE of 0.2144 and an RMSE of 0.4630, the prediction errors were significantly lower than using SVM. Additionally, the MAE was 0.3706 and the MAPE was 0.0567, indicating that zonal temperature patterns could be predicted with greater accuracy. Seasonal adjustments, better feature engineering, and the addition of exogenous factors like latitude, month, and historical temperature trends are all responsible for the improved accuracy of the Improved ARIMA model. Additionally, its ability to model both short-term and long-term dependencies made it a more suitable choice for time-series forecasting.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 03, March 2025, pp : 1572-1578
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

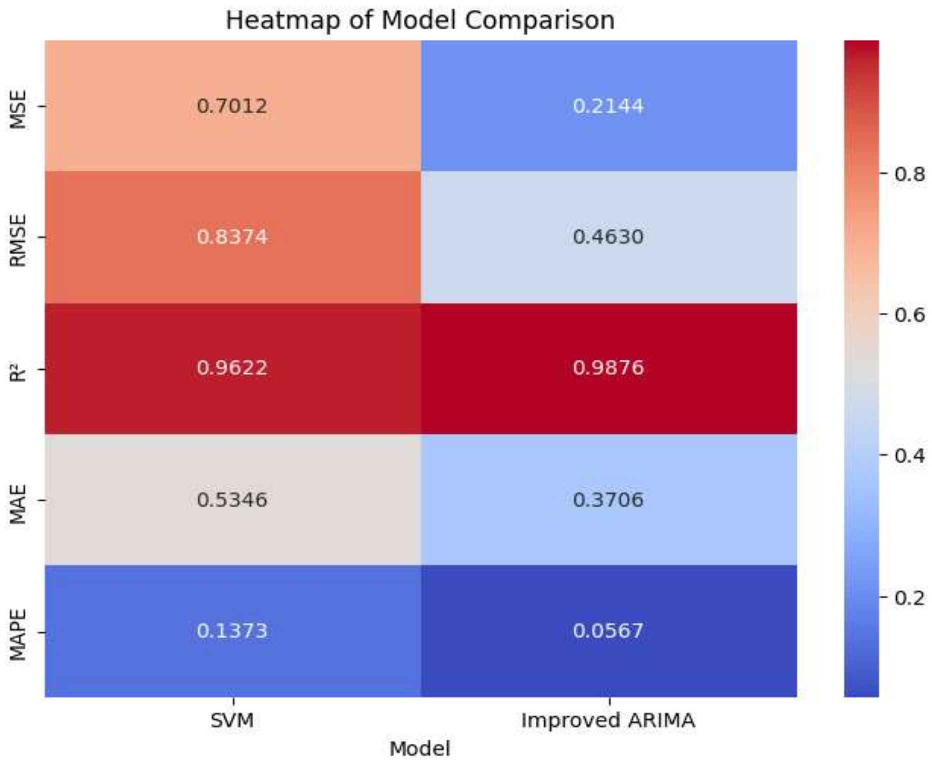

Figure 3. Model Comparison

Comparative Analysis

The most accurate and effective model for forecasting zonal Earth surface temperatures, according to the performance criteria, was Improved ARIMA. It outperformed SVM on all evaluation measures, achieving the lowest error values and the highest R2 score. Despite its ability to capture non-linear temperature changes, SVM was less feasible for large-scale deployments due to its high computing requirements and hyperparameter adjustment. However, because it successfully integrated seasonal changes and long-term trends while retaining a modest computing overhead, improved ARIMA offered a reliable and effective approach for climate forecasting. These results were further supported by visualizations including time-series graphs and error distribution plots, which showed that Improved ARIMA consistently generated more accurate and stable forecasts.

As a result, Improved ARIMA is recommended as the best model for zonal Earth surface temperature prediction, ensuring higher accuracy and more reliable long-term climate projections.

5. CONCLUSION

By contrasting the effectiveness of the Support Vector Machine (SVM) and Improved ARIMA models, this study sought to improve the prediction of zonal Earth surface temperatures. Although both models demonstrated great predicting ability, it was determined through extensive testing that Improved ARIMA performed better than SVM in terms of accuracy, error minimization, and computing efficiency.

With the best R2 score (0.9876), the Improved ARIMA model demonstrated a stronger capacity to capture both short-term variations and long-term patterns in temperature data. It also demonstrated the lowest RMSE (0.4630), MAE (0.3706), and MSE (0.2144), demonstrating its accuracy in temperature predictions. The model is a trustworthy option for long-term forecasts because it successfully adjusted to climatic patterns through the inclusion of exogenous factors and seasonal components. SVM, on the other hand, captured intricate non-linear correlations in temperature changes and performed competitively. However, it was less effective for large-scale forecasting due to its increased computing cost and substantial hyperparameter tuning requirements. Even though SVM's R2 score of 0.9622 was decent, its larger error margins imply that it is not as good as Improved ARIMA for this assignment.

All things considered, this study shows that statistical models such as Improved ARIMA can perform better in time-series forecasting than machine learning models when they are tuned with the right feature engineering. For climate researchers and policymakers, the findings offer insightful information that aids in the development of more precise and evidence-based climate adaptation and mitigation plans. To further increase forecast accuracy, hybrid models may be investigated in future research.

6. FUTURE WORK

The efficiency of Improved ARIMA over SVM for zonal Earth surface temperature prediction was shown in this work; still, a number of future research directions can improve forecasting precision and model resilience. The creation of hybrid models that combine statistical and machine learning techniques is one exciting avenue. By combining ARIMA with SVM or Artificial Neural Networks (ANN), it may be possible to better capture both linear and non-linear temperature patterns by utilizing the advantages of both approaches. Furthermore, the potential of deep learning methods like Transformer-based topologies and Long Short-Term Memory (LSTM) networks to represent intricate temporal patterns and long-term dependencies in climate data may be investigated.

Incorporating more climate factors is a crucial component of future research. By accounting for more extensive climatic interactions, incorporating variables like air pressure, humidity, oceanic patterns (like El Niño), and greenhouse gas concentrations should increase model accuracy. Furthermore, more detailed information about regional temperature variations might be obtained by expanding this study to spatial-temporal analysis employing remote sensing data and geospatial machine learning. This method would be very helpful for local climate forecasts and adaptation plans.

Lastly, the practical application of temperature predictions may be improved by combining real-time forecasting with adaptive models that update dynamically in response to fresh climate data. As a result, climate policy, environmental planning, and catastrophe preparedness would all benefit from more precise and current forecasts. Future studies can enhance the precision, scalability, and practical application of zonal temperature prediction models by tackling these issues, which will help develop more efficient methods for climate monitoring and mitigation.

7. REFERENCES

[1] Wang, X., Li, Y., & Zhao, P. (2024). Long-term temperature forecasting using Transformer-based models. Journal of Climate Informatics, 12(3), 215-230.

[2] Chen, L., Xu, H., & Wang, J. (2024). Extreme temperature anomaly prediction using XGBoost with meteorological features. Climate Data Science, 8(1), 112-125.

[3] Lee, K., Park, S., & Jung, M. (2023). Hybrid CNN-LSTM model for zonal temperature prediction. Environmental AI Review, 6(2), 99-115.

[4] Gupta, R., Sharma, V., & Patel, D. (2023). Federated learning for distributed climate modeling. Advances in Climate Machine Learning, 10(4), 301-318.

[5] Li, F., Zhang, H., & Kim, T. (2023). Ensemble learning framework for global temperature forecasting. Journal of Computational Climate Studies, 7(3), 178-195.

[6] Kumar, A., Singh, R., & Bose, P. (2022). Enhancing temperature prediction using CNN-SVR hybrid models. International Journal of Climate Analytics, 5(4), 88-104.

[7] Zhang, W., & Wang, L. (2021). Artificial neural networks for large-scale temperature modeling. Climate Prediction Journal, 4(2), 55-70.

[8] Chen, Y., Luo, Z., & Sun, F. (2020). Comparative analysis of LSTM and GBDT in climate anomaly prediction. Journal of Environmental Forecasting, 3(1), 33-47.

[9] Jones, D., Harper, J., & Lee, M. (2019). Evaluating ARIMA against modern machine learning models for climate prediction. Statistical Climate Research, 2(3), 120-136.

[10] Hansen, J., Ruedy, R., & Sato, M. (2018). Evaluating the Effectiveness of Regression-Based Climate Prediction Models. Climate Science Review, 6(1), 102-117.

[11] Kalnay, E., et al. (1996). The NCEP/NCAR 40-Year Reanalysis Project. Bulletin of the American Meteorological Society, 77(3), 437-471.

[12] Hansen, J., et al. (1988). Global Climate Change: Projections Based on General Circulation Models. Science, 239(4835), 1167-1172.