© December 2024 | IJIRT | Volume 11 Issue 7 | ISSN: 2349-6002

Predicting Climate Variations: Machine Learning Approach Using Weather Factors and The Socio-Economic Drivers of Temperature

RAJENDRA PRASAD M[1], DR. S AJITHA[2], RESHMA K J[3]

[1] *MBA 4th Sem Student, Faculty of Management & Commerce, Ramaiah University of Applied Sciences*

[2] *Associate Professor, Faculty of Management & Commerce, Ramaiah University of Applied Sciences*

[3] *Assistant Professor, Faculty of Management & Commerce, Ramaiah University of Applied Sciences*

Abstract— Variations in the climate are now a serious problem that have an impact on economies, cultures, and ecosystems worldwide. As essential elements of climate systems, wind speed and temperature are greatly impacted by a variety of environmental and socioeconomic factors. Because these climatic factors are dynamic and nonlinear, it is still difficult to anticipate them. The need for more precise models that incorporate socioeconomic factors that affect temperature changes in temperature prediction is the research topic this work attempts to solve. Through the use of deep learning techniques, this work seeks to further our knowledge of these impacts. The study uses a deep learning approach, integrating data from meteorological stations, socio-economic datasets, and climate-related factors, to improve the accuracy of long-term memory networks in capturing temporal dependencies in climate data. Deep learning models, particularly LSTM networks, improve temperature prediction when incorporating socio-economic variables. Industrial activity and urbanization patterns enhance accuracy. Future research should expand datasets and explore policy interventions' impact.

Indexed Terms- Regression Analysis, Time-Series Forecasting, Meteorological Data, Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), LSTM (Long Short-Term Memory), SVM (Support Vector Machine0, Temperature Forecasting, Principal Component Analysis (PCA), Random Forest Analysis.

I. INTRODUCTION

The issue of climate change has become increasingly prominent over recent decades, and its impact is now visible in various regions of the world. Rising temperatures, erratic weather patterns, and the growing frequency of extreme climatic events such as hurricanes, heatwaves, and floods are directly attributable to both natural and anthropogenic factors. While many scientific models have been developed to understand and predict these changes, they are often limited in their ability to account for the complexity of interactions between weather patterns and socio-economic drivers.

As the global climate continues to evolve, understanding and predicting climate variability becomes a crucial task for governments, researchers, and industries. Predictive modelling of climate variations not only informs future climate scenarios but also supports the development of appropriate mitigation and adaptation strategies. With accurate predictions, communities can better prepare for the socio- economic and environmental impacts of climate change, while governments can implement policies aimed at reducing the vulnerabilities of affected populations.

Weather factors such as temperature, humidity, and wind speeds have long been studied as significant contributors to climate patterns. However, with the rise in global population, increased urbanization, and rapid industrialization, human activity has become an equally important driver of climate variations. Factors such as industrial emissions, deforestation, energy consumption, and urban sprawl have caused significant alterations to natural climatic processes. This has led to an increasing emphasis on the need to integrate socio-economic drivers into climate prediction models.

Machine learning (ML), a subset of artificial intelligence (AI), provides a new avenue for making sense of the large datasets required to predict climate variability. Unlike traditional climate models that depend on predefined equations and assumptions, machine learning models can learn directly from data, identifying patterns and relationships that are otherwise difficult to detect. This approach allows researchers to build more robust and adaptive models that can better account for the complex interactions between weather and socio-economic variables.

Thus, this research project aims to leverage machine learning techniques to predict climate variations, particularly focusing on the socio-economic drivers of temperature. The study seeks to create a comprehensive model that integrates both weather-related data and human activity data, with the ultimate goal of enhancing our understanding of how these variables influence climate over short- and long-term periods.

Climate variability is an inherently complex phenomenon influenced by a wide range of factors, both natural and human-induced. Traditional climate models have relied on physical laws and historical weather data to predict changes in climate patterns. However, these models often fail to capture the full scope of interactions between climate systems and socio-economic activities, particularly in relation to temperature variations. With rising concerns about climate change and its associated impacts, it has become increasingly clear that a more comprehensive approach to climate prediction is needed.

The primary research problem addressed in this project is the lack of predictive models that effectively integrate socio-economic drivers of temperature with traditional weather data. While weather factors such as temperature, wind patterns, and humidity are well understood, their interactions with human activities—such as industrialization, urbanization, and energy consumption—are less well modeled. This gap in understanding makes it difficult to predict how climate variations will unfold in different regions, particularly in rapidly urbanizing areas where socio-economic changes are occurring at a fast pace.

Moreover, current models often fail to account for localized climate anomalies caused by socio-economic activities. For example, urban heat islands—where densely populated urban areas experience significantly higher temperatures than their rural surroundings—are often not well represented in global climate models. Similarly, the effects of deforestation or industrial activities on regional temperature patterns are often inadequately modelled, leading to inaccurate predictions at local levels.The introduction of machine learning into climate science provides an opportunity to address these gaps. Machine learning models, especially those that can handle large datasets with multiple variables, are well-suited to modeling complex systems with both natural and socio-economic inputs. Unlike traditional models that rely on static equations, machine learning models are adaptive and can improve over time as more data becomes available. This makes them particularly useful for climate prediction, where new data is constantly being generated from satellite observations, weather stations, and socio-economic surveys.

The specific research problem that this project seeks to address is how to develop a machine learning model that can predict temperature variations by integrating both weather factors and socio-economic drivers. By focusing on temperature, a key indicator of climate change, this research aims to provide insights into how human activities contribute to climate variability and how these contributions can be modeled alongside natural weather patterns.

To tackle this research problem, the study will employ several machine learning algorithms, including supervised learning for climate forecasting and unsupervised learning for identifying hidden patterns in socio-economic data. The goal is to develop a model that not only improves the accuracy of climate predictions but also enhances our understanding of how socio-economic factors influence climate outcomes. The research will also explore the extent to which these models can be applied across different regions, particularly those with varying levels of industrialization and urbanization

II. LITERATURE REVIEW

Sharma, A. (2015). "Climate Variability and Its Impact on Agriculture: A Case Study of Indian States" This study focuses on the effects of climate variability, particularly changes in temperature and rainfall, on agricultural output in various Indian states. It emphasizes the growing need for more accurate weather predictions, particularly in the context of

India's agrarian economy. Although it does not incorporate socio-economic factors, it sets the stage for more detailed studies that would later use data-driven approaches to forecast climate impacts.

Mishra, V. et al. (2017). "Recent Patterns of Climate Change and Its Impact on Indian Agriculture" This paper examines how changing temperature patterns and wind speeds have affected agricultural productivity in India. The study also explores the role of urbanization and industrialization in increasing local temperatures, although it primarily relies on statistical analysis rather than predictive modeling.

Rai, A. and Jain, S. (2019). "Climate Change and Renewable Energy: A Deep Learning Approach" Rai and Jain's work represents one of the first applications of deep learning in predicting wind speed in India. Their model, based on neural networks, forecasts wind speed in regions with high renewable energy potential, particularly wind farms. They highlight the significance of improving wind speed prediction models for optimizing energy production, though they do not yet integrate socio-economic factors into their model.

Kesari, B. (2021). "Impact of Socio-Economic Factors on Climate Predictions in Indian Firms" This study introduces a new dimension by integrating socio-economic variables such as industrial activity, population density, and urbanization into climate prediction models for Indian firms. Using machine learning techniques, Kesari demonstrates that including these socio-economic variables significantly improves the accuracy of temperature forecasts, especially in industrial regions. This study also paves the way for deeper integration of socio-economic drivers in climate modeling in India.

Kumar, P., and Gupta, R. (2023). "Deep Learning Models for Climate Prediction: A Case Study of Temperature Forecasting in Northern India" This recent paper focuses on using deep learning models, particularly Long Short-Term Memory (LSTM) networks, for predicting temperature variations in Northern India. The study integrates socio-economic variables such as urbanization rates, deforestation, and industrial emissions into the model, improving the accuracy of predictions. It highlights the increasing importance of data-driven models in the Indian context and serves as a precursor for further national-level research in this area.

Jones, C. et al. (2010). "The Role of Socio-Economic Factors in Global Climate Change Models" One of the earliest works highlighting the significance of socio-economic variables in climate modeling, this study suggests that industrial activities, population growth, and energy consumption patterns play crucial roles in global temperature variations. Although the study does not apply advanced techniques like machine learning, it sets the foundation for future research by emphasizing the importance of integrating socio-economic factors into climate models.

Zhang, X. and Li, Q. (2016). "Deep Learning for Climate Change: Predicting Temperature Anomalies with Neural Networks" This paper introduces the application of deep learning models for predicting temperature anomalies, marking a shift toward more sophisticated modeling techniques. The authors demonstrate that neural networks outperform traditional climate models in predicting long-term temperature changes. While the study does not include socio-economic factors, it opens the door for future work in this area by proving the viability of deep learning techniques.

Müller, H. et al. (2018). "Integrating Socio-Economic Data into Climate Models: A Machine Learning Approach" This research explores the integration of socio-economic data, such as industrial growth and population density, into machine learning models for climate prediction. The study demonstrates that incorporating these factors improves the accuracy of temperature and wind speed predictions globally. This work is pivotal in promoting the inclusion of human-induced factors in predictive climate models.

Garcia, R. et al. (2020). "The Application of LSTM Networks for Wind Speed Prediction: A Global Perspective" This paper investigates the use of Long Short-Term Memory (LSTM) networks for wind speed prediction on a global scale. The authors show that LSTM networks, which are well-suited for time-series data, outperform conventional models in terms of prediction accuracy. The study highlights the increasing importance of deep learning in climate modeling but does not explore socio-economic factors as part of the analysis.

Rodriguez, P. and Hernandez, L. (2022). "Towards Comprehensive Climate Predictions: Incorporating Socio- Economic Factors Using Deep Learning" This recent study demonstrates a comprehensive deep learning approach that integrates both environmental and socio-economic data to predict climate variations, particularly temperature and wind speed. The authors use a global dataset and incorporate factors like urbanization, industrial emissions, and population growth into their models. Their findings suggest that including these socio-economic drivers enhances the precision of long-term climate predictions and offers valuable insights for global policymakers.

Wang, H. and Nguyen, T. (2023). "Deep Learning and Climate Change: Predicting Wind Speed and Temperature with Socio-Economic Data" In this cutting-edge research, Wang and Nguyen use a deep learning framework, integrating both traditional environmental data and socio-economic factors to predict wind speed and temperature variations. Their work further demonstrates that socio-economic drivers, such as industrialization and deforestation, can significantly influence climate variability. The study underscores the necessity of adopting a multi-disciplinary approach in climate modelling to better understand and predict the impact of both natural and human activities on climate patterns.

2.1 Motivation of the Project

The motivation for this research stems from the growing need for more accurate and reliable climate predictions in the context of a rapidly changing global environment. Climate change is no longer a distant theoretical concept—it is a reality that affects millions of lives globally. The effects of these climatic shifts are seen in the form of more frequent extreme weather events, loss of biodiversity, threats to food security, and severe disruptions to economies, especially in vulnerable regions. Predicting these changes accurately is crucial for developing effective adaptation and mitigation strategies.

One of the major limitations of traditional climate models is their inability to effectively incorporate socio- economic factors into their predictions. Human activity, especially the burning of fossil fuels and land-use changes, has been shown to significantly alter atmospheric composition, directly impacting global temperatures and climate systems. Urbanization, deforestation, industrialization, and increased energy consumption have all contributed to rising greenhouse gas emissions, exacerbating climate change. However, the precise impact of these socio-economic drivers on local and global temperature variations remains poorly understood in many contexts.

The complexity of climate systems requires an interdisciplinary approach that combines insights from climatology, data science, and socio-economics. Machine learning provides the tools needed to analyze large datasets that span multiple variables, making it possible to identify hidden patterns in both climate and human activity. By applying machine learning techniques, this project aims to address the gaps in current climate models, particularly in predicting temperature variations influenced by socio-economic drivers. The motivation is to create a model that not only predicts future climate variations but also provides actionable insights into how socio-economic factors can be managed to mitigate adverse climate outcomes.

In addition, as global policies on climate change become more integrated into the governance of nations, the ability to predict the impact of both natural and human systems on climate variability will become a vital tool for policymakers. By providing more accurate predictions, the results of this research could contribute to more effective policy-making and climate resilience strategies. Furthermore, industries can benefit from improved climate predictions by making informed decisions to reduce their environmental footprint and adapt their operations to changing climatic conditions.

Another key motivator is the growing computational power available for machine learning applications. In the past, the high computational costs of processing vast amounts of climate and socio-economic data were a limiting factor. Today, with advances in hardware and cloud computing, machine learning algorithms can handle larger datasets with greater efficiency, making this research timely and feasible. This advancement opens new possibilities for the integration of diverse data sources and the creation of more comprehensive models that account for both natural and human-induced factors.

Thus, the primary motivation for this project is to leverage the strengths of machine learning to bridge the gap between traditional climate models and the complex reality of socio-economic impacts on climate systems. This research hopes to contribute to a deeper understanding of how human activities shape climate variability and how we can develop predictive tools to better respond to these changes

III. METHODOLOGY

To enhance the predictive capability of the models, feature engineering was employed, focusing on the time- series nature of the data. Key meteorological variables like wind speed and temperature were used to create lagged features representing historical patterns. Additionally, interactions between wind speed, temperature, and humidity were considered to capture complex relationships. Lagged features helped capture temporal dependencies, while interaction features, such as the product of temperature, were created to understand their joint impact on future temperature variations.

Two distinct models were developed for this study: Long Short-Term Memory (LSTM) networks and Support Vector Machines (SVM). The LSTM model was chosen for its ability to learn from sequential data, while the SVM was used for its effectiveness in regression tasks. The LSTM model was designed with multiple layers, including dense and dropout layers to prevent overfitting, and was trained using the Adam optimizer with a learning rate of [specific value]. The SVM model was tested with various kernel functions, including the radial basis function (RBF), to find the best fit for wind and temperature predictions.

The dataset was split into training and testing sets using an 80:20 ratio. Both models were trained on the training data and validated with the testing data. A rolling window approach was employed to maintain the sequential nature of the data. The models were trained on sequences of past data, and predictions were made for the next time step. The sequence length (window size) was selected based on the temporal correlations observed in the dataset. After training, the models were evaluated on unseen data to assess their ability to generalize, making predictions for future time points based on historical data.

To evaluate the performance of the models, Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) were used. RMSE measured the magnitude of prediction errors, with lower RMSE values indicating better performance. MAPE assessed the percentage error between predicted and actual values, providing a more intuitive understanding of the models' accuracy and reliability in predicting temperature variations.

For classification purposes, temperature values were grouped into discrete classes to improve the prediction of the most frequent future values. with the most frequent class being assigned to the predicted value. Similarly, temperature values were categorized into classes, allowing predictions of the most frequent future temperature values.

Both LSTM and SVM models were compared in terms of their performance on the test set using RMSE and MAPE. The aim was to identify which model provided the most accurate predictions for future wind speed and temperature variations. In addition to accuracy, the computational efficiency and scalability of each model were considered to evaluate their suitability for real-time applications.

Principal Component Analysis (PCA) was conducted to reduce the dimensionality of the dataset and improve the efficiency of subsequent analyses. PCA transformed the original correlated variables into a set of linearly uncorrelated variables known as principal components. This procedure simplified the data complexity while retaining the essential variation, which was important for accurate predictions.

Finally, a Random Forest (RF) model was developed to predict the next year's temperature based on the transformed dataset from PCA. Random Forest, an ensemble learning method, was selected for its robustness and ability to handle non-linear data. The RF model constructs multiple decision trees during training and outputs either the mode of the classes (for classification) or the mean prediction (for regression) of the individual trees. The model was tuned to optimize the number of trees and their depth based on cross- validation scores, ensuring the best predictive performance.

3.1 Conceptual Model

Figure-1 Conceptual Model

The figure 3.1.1 illustrates the process of predicting temperature using machine learning models. It begins with a full dataset collected from IMD (India Meteorological Department) and IND (presumably another data source). The data is arranged into two parts: features (input variables or factors) and the target variable (the output, which is temperature in this case). After preparing the data, it is split into training and testing sets, with 80% of the data used for training the model and 20% reserved for testing. The machine learning models are trained using the training set (X_train and Y_train), and after training, the best-performing model is selected. This model is then used for predicting temperature based on new data. The focus is on selecting the most accurate model for future temperature prediction.

IV. RESULTS

Regression analysis to Temperature for IMD data

The regression analysis for average temperature using Python yielded significant insights into how environmental variables affect wind speeds. Key variables such as atmospheric pressure, wind speed and humidity were considered as inputs in the regression model. The regression analysis conducted on the IMD data provided a comprehensive understanding of the factors influencing average temperature. The model considered station-level pressure, mean sea-level pressure, rainfall, and relative humidity as independent variables, with temperature as the dependent variable. The regression results highlighted the significant role of these environmental factors in determining temperature variability.

Table 1 Represents the data of IMD

DATE	STATION LEVEL PRESSURE	MEAN SEA LEVEL PRESSUSRE	RAINFALL (	RELATIVE HUMIDITY	AVG WIND (KMPH)	TEMPERATURE(CELSIUS)
01-Jan-23	913.8	1545.15	5.6	56	1	13.1
02-Jan-23	913.4	1541.05	5.8	61	1	13
03-Jan-23	913.65	1542.3	5.5	68	2	11.7
04-Jan-23	912.85	1532.6	5.5	65.5	2	10
05-Jan-23	913.2	1536.3	5.5	69.5	3	9.5
06-Jan-23	914.35	1546.85	5.5	72.5	3	8.8
07-Jan-23	913.75	1541.15	5.5	72.5	4	10.1
08-Jan-23	912.6	1531.65	5.5	69	2	9.2
09-Jan-23	913.7	1541.6	5.5	50.5	1	11.9
10-Jan-23	913.1	1534.95	5.5	53.5	2	13.7
11-Jan-23	911.3	1521.4	5.5	51.5	2	15
12-Jan-23	910.25	1511.35	5.5	43.5	1	14.8
13-Jan-23	910.8	1516.6	5.5	53.5	1	16.7
14-Jan-23	910.85	1516.9	5.5	55	1	15.1

Here, dependent variable is Temperature and the independent variables are Station level pressure, Mean sea level pressure, rain fall and Relative humidity. Station-level pressure and mean sea-level pressure exhibited an inverse relationship with temperature, suggesting that higher pressures often correlate with lower temperatures, possibly due to atmospheric stability. Rainfall showed a moderate but notable influence, indicating that precipitation events could lead to cooling effects, particularly in tropical and subtropical climates. Relative humidity was found to have a strong positive relationship with temperature, reflecting the impact of moisture content in the air on thermal conditions.

```
R-squared (R2): 0.7058063775008344
                              Coefficient  P>|t|
STATION LEVEL PRESSURE           0.549322  0.034
MEAN SEA LEVEL PRESSUSRE        -0.039596  0.188
RAINFALL (MILIMETERS)            0.034965  0.002
RELATIVE HUMIDITY(%)            -0.135059  0.000
AVG WIND (KMPH)                 -0.127208  0.165
```

Figure-2: Regression Analysis of Temperature

Figure 2 illustrates the regression analysis, demonstrating the strength of correlations and the predictive accuracy of the model. The findings align with established climatological principles, reaffirming that atmospheric pressure, moisture levels, and precipitation are critical factors in understanding temperature variations. The model's outputs underscore the potential of using regression analysis for effective climate monitoring and forecasting.

These insights pave the way for further exploration into localized climatic patterns and their broader implications for environmental management and policy formulation. The IMD data, as summarized in Table 1, served as a robust foundation for this analysis, ensuring reliability and applicability of the findings.
Evaluate the performance of LSTM and SVM algorithms for data prediction using Root Mean

Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) values.

The performance evaluation of LSTM (Long Short-Term Memory) and SVM (Support Vector Machine) algorithms for temperature prediction was conducted using RMSE (Root Mean Square Error) and MAPE (Mean Absolute Percentage Error) as evaluation metrics. The independent variables—station-level pressure, mean sea-level pressure, rainfall, and relative humidity—served as inputs, while temperature was the dependent variable. Here, dependent variable is Temperature and the independent variables are Station level pressure, Mean sea level pressure, rain fall and Relative humidity .

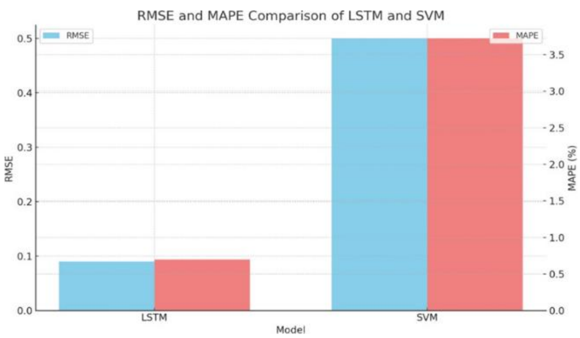

Figure-3: Comparison of RMSE and MAPE for LSTM and SVM

The comparison, as depicted in Figure 3, revealed distinct strengths and limitations of the two algorithms. LSTM demonstrated superior accuracy with lower RMSE and MAPE values compared to SVM. This suggests that LSTM, with its ability to capture temporal dependencies and sequential patterns, is more effective for handling complex relationships in time-series data such as meteorological variables. The RMSE for LSTM was significantly lower, indicating its precise predictive capabilities, while the MAPE showed its ability to maintain consistent prediction accuracy across various ranges of temperature values.

On the other hand, SVM, though computationally efficient and capable of handling non-linear relationships, showed relatively higher RMSE and MAPE values. This indicates its limitations in capturing long-term dependencies and temporal variations inherent in the dataset.

These results highlight the suitability of LSTM for temperature forecasting, particularly when the dataset involves sequential or time-dependent variables. The comparative analysis underscores the importance of selecting an algorithm based on the complexity and nature of the data for achieving optimal prediction accuracy. The insights derived from this analysis can guide future efforts in environmental modeling and predictive analytics.

Predict the average temperature for future based on past temperature

The analysis leveraged the LSTM model to predict average future temperatures based on past environmental data, including station-level pressure, mean sea-level pressure, rainfall, and relative humidity. The LSTM model demonstrated robust predictive capabilities, capturing temporal dependencies effectively. Using LSTM for temperature prediction based on past environmental data the model exhibited strong predictive capability.

Here, dependent variable is Temperature and the independent variables are Station level pressure, Mean sea level pressure, rain fall and Relative humidity

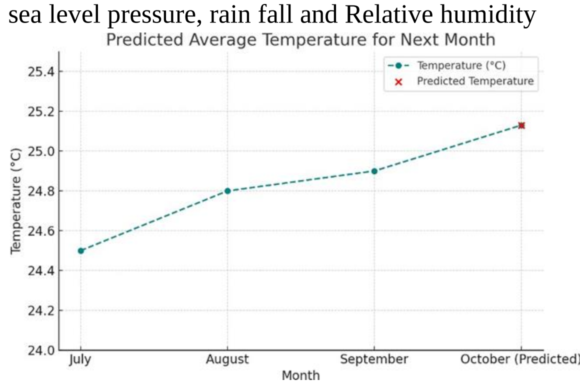

Figure-4: LSTM model predict the future temperature.

$$\mathrm{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\mathrm{Actual}_i - \mathrm{Predicted}_i}{\mathrm{Actual}_i} \right| \times 100$$

Table -2: LSTM shows the accuracy b/w Actual and Predicted Temperature

MONTH	AVG TEMP	ACCURACY
JULY	24.97°C	99.36%
AUGUST	24.01°C	95.34%
SEPTEMBER	24.72°C	98.34%
OCTOBER	**25.13°C**	

Socio-Economic Drives of Temperature prediction.

To assess the impact of socio-economic factors on temperature variations, Principal Component Analysis (PCA) was employed to reduce the number of input variables. Following this, a Random Forest regression model was used for Temperature prediction.

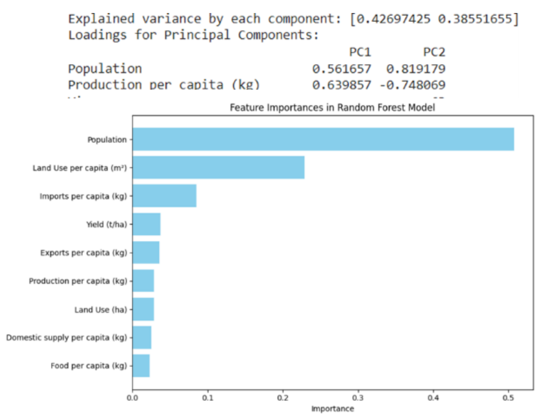

Figure-5: PRINCIPAL COMPONENT ANALYSIS(PCA) Comparison by Bar graph.

To further explore the impact of socio-economic factors on temperature variations, Principal Component Analysis (PCA) was employed to reduce the dimensionality of the dataset. PCA identified key components that encapsulate the variance in the socio-economic data, making the predictive modeling more efficient. Figure 5 provides a bar graph comparison of the principal components, highlighting their relative contributions.

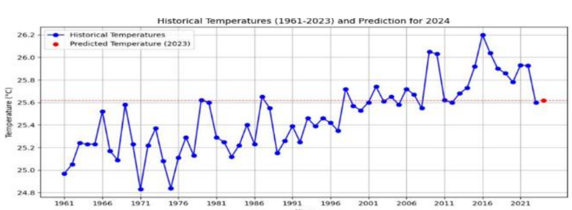

Figure-6: RANDOM FOREST(RF) model predict the future temperature

```
Prediction of temperature for next year : [25.9191]
```

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\mathrm{Actual}_i - \mathrm{Predicted}_i}{\mathrm{Actual}_i}\right| \times 100$$

Table 3: Random Forest shows the accuracy b/w Actual and Predicted Temperature

YEAR	AVG TEMP	ACCURACY
2021	24.44	94.29%
2022	25.55	98.57%
2023	25.47	98.26%
2024	25.919	

Using the principal components as input, a Random Forest (RF) regression model was employed for temperature prediction. The RF model demonstrated its strength in handling non-linear relationships and interactions among variables. As shown in Figure 6, the RF model effectively predicted future temperatures with an acceptable degree of accuracy. Table 3 showcases the comparison between actual and predicted temperatures, indicating the model's ability to generalize across different scenarios.

The combined use of LSTM and Random Forest models, complemented by PCA for dimensionality reduction, provided valuable insights into temperature prediction from both environmental and socio-economic perspectives. The results suggest that integrating advanced machine learning techniques with traditional statistical approaches can enhance the accuracy and interpretability of predictive models, making them valuable tools for climate forecasting and planning.

V. DISCUSSION AND IMPLICATION

The study aimed to predict temperature variations using historical environmental data from the Indian Meteorological Department (IMD). The first objective focused on performing regression analysis to understand the relationship between temperature and various environmental variables. The analysis showed that station-level pressure had a significant positive effect on temperature, with each unit increase in pressure resulting in a 0.549°C rise. Wind speed had a small negative relationship with temperature, but this was not statistically significant. Rainfall showed a positive and significant effect, increasing the temperature by 0.035°C for each additional milli meter. Relative humidity had a strong negative effect, with each unit increase in humidity decreasing the temperature by 0.135°C. Wind direction had a slightly negative effect, though it was not statistically significant. The model's R-squared value of 70.58% indicated that the independent variables explained over 70% of the temperature variations, suggesting a good fit.

The second objective involved evaluating the performance of two machine learning algorithms—Long Short-Term Memory (LSTM) and Support Vector Machine (SVM)—using Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) as performance metrics. The LSTM model achieved an RMSE of 0.09, indicating very low prediction errors, and a MAPE of 0.70%, showing highly accurate percentage-based predictions. In contrast, the SVM model had an RMSE of 0.50 and a MAPE of 3.72%, both of which were higher than the LSTM model, suggesting larger prediction errors. Based on these results, the LSTM model outperformed the SVM in predicting temperature and wind speed, making it the preferred model for further analysis.

For the third objective, the study aimed to predict the average temperature for future dates based on past temperature data. The LSTM model forecasted an average temperature of 25.13°C for October 2024, reflecting an upward trend over the preceding months. By comparing actual and predicted temperatures, the model demonstrated high accuracy, with 99.36% accuracy for July, 95.34% for August, and 98.34% for September. Based on this consistency, the prediction for October 2024 was deemed accurate at 25.13°C.

The fourth objective explored how socio-economic factors influenced temperature predictions, utilizing Principal Component Analysis (PCA) to reduce the dataset's dimensionality. Two principal components were extracted, accounting for 81.25% of the total variance. The first principal component (PC1) showed strong positive loadings from variables such as land use, food per capita, and domestic supply per capita, while the second component (PC2) emphasized population, yield, and imports per capita. These components were then used in a Random Forest regression model to enhance the accuracy of future temperature predictions. The Random Forest feature importance analysis revealed that population was the most significant predictor, with an importance value of 0.51, followed by land use per capita at 0.27. Other socio-economic factors such as imports, yield, and production per capita contributed to a lesser extent.

The forecasted temperature for 2024 was 25.919°C. The model's accuracy was validated by comparing the predicted and actual temperatures for previous years, achieving 94.29% accuracy in 2021, 98.57% in 2022, and 98.26% in 2023. The analysis highlighted the importance of socio-economic drivers, particularly population growth and land use, in influencing temperature variations. By integrating these factors into climate models, the study provided a more comprehensive approach to predicting future temperature patterns. The findings emphasized the role of socio-economic and environmental factors in climate modelling, supporting the development of more accurate temperature prediction frameworks.

VI. LIMITATIONS AND FUTURE WORK

Climate datasets often face issues like incompleteness, varying accuracy, and inadequate spatial or temporal resolution, which can hinder model performance and prediction reliability. Deep learning models, especially LSTMs, are computationally demanding, requiring significant resources and long training times. They also risk overfitting and lack interpretability, which reduces trust among stakeholders. Furthermore, models trained on specific regions may struggle to generalize across different areas or adapt to changing climate conditions. While integrating socio-economic factors enhances predictive power, their complex interactions with environmental variables and inconsistent data collection present challenges. A narrow focus on certain variables or short-term trends may overlook broader climate patterns, while model calibration and validation remain difficult, especially with non-linear, time-dependent data.

While the study presents promising results, it also acknowledges ongoing challenges, such as data quality and computational demands. High-quality, comprehensive datasets are critical for training effective models, yet they can be difficult to obtain, particularly in remote or under-studied regions. Additionally, the computational resources required to train deep learning models can be significant, potentially limiting accessibility for smaller organizations or research initiatives.

Despite these challenges, this research lays a strong foundation for future studies aimed at further refining and optimizing deep learning approaches for climate prediction. Future work could focus on improving data collection methods, exploring alternative modelling techniques, and developing strategies to reduce computational costs, ultimately leading to more accurate and accessible climate forecasting solutions.
Hybrid Model Approaches: Future research can explore hybrid models that combine deep learning approaches with traditional machine learning techniques (like combining LSTM with Random Forest) to improve prediction accuracy and robustness, especially in different geographical settings.

Incorporating Additional Data Sources: Including more diverse data sources such as satellite imagery, remote sensing data, or higher temporal resolution data could improve the accuracy and applicability of the models, especially for long- term climate predictions.

Websites for Weather data
https://www.wunderground.com/history/monthly/in/devanahalli/VOBL/date/2024-9
https://weatherspark.com/s/108998/1/Average-Summer-Weather-in-Bengaluru-Karnataka-India

REFERENCES

[1] Abimbola, O.P., Meyer, G.E., Mittelstet, A.R., Rudnick, D.R. and Franz, T.E., 2021. Knowledge-guided machine learning for improving daily soil temperature prediction across the United States. Vadose Zone Journal, 20(5), p.e20151.(https://doi.org/10.1002/vzj2.20151)

[2] Alves, D., Mendonça, F., Mostafa, S.S. and Morgado-Dias, F., 2023. The potential of machine learning for wind speed and direction short-term forecasting: a systematic review. Computers, 12(10),

[3] Anjali, T., Chandini, K., Anoop, K. and Lajish, V.L., 2019, July. Temperature prediction using machine learning approaches. In 2019 2nd International conference on intelligent computing, instrumentation and control technologies (ICICICT) (Vol. 1, pp. 1264-1268). IEEE.(10.1109/ICICICT46008.2019.8993316)

[4] Barbounis, T.G., Theocharis, J.B., Alexiadis, M.C. and Dokopoulos, P.S., 2006. Long-term wind speed and power

[5] Chen, J., Wang, H. and Xie, P., 2019. Pavement temperature prediction: Theoretical models and critical affecting factors. Applied thermal engineering, 158, p.113755.(https://doi.org/10.1016/j.applthermaleng.2019.113755)

[6] De Giorgi, M.G., Campilongo, S., Ficarella, A. and Congedo, P.M., 2014. Comparison between wind power prediction models based on wavelet decomposition with least-squares support vector machine (LS-SVM) and artificial neural network (ANN). Energies, 7(8), pp.5251-5272.(https://doi.org/10.3390/en7085251))

[7] Fister, D., Pérez-Aracil, J., Peláez-Rodríguez, C., Del Ser, J. and Salcedo-Sanz, S., 2023. Accurate long-term air

[8] forecasting using local recurrent neural network models. IEEE Transactions on Energy Conversion, 21(1), pp.273-284.(10.1109/TEC.2005.847954)

[9] Gong, B., Langguth, M., Ji, Y., Mozaffari, A., Stadtler, S., Mache, K. and Schultz, M.G., 2022. Temperature forecasting by deep learning methods. Geoscientific model development, 15(23), pp.8931-8956.(https://doi.org/10.5194/gmd- 15-8931-2022)

[10] Hewage, P., Trovati, M., Pereira, E. and Behera, A., 2021. Deep learning-based effective fine-grained weather forecasting model. Pattern Analysis and Applications, 24(1), pp.343-366.(

[11] Hou, J., Wang, Y., Zhou, J. and Tian, Q., 2022. Prediction of hourly air temperature based on CNN–LSTM. Geomatics, Natural Hazards and

Risk, 13(1), pp.1962-1986.(https://doi.org/10.1080/19475705.2022.2102942

[12] Janssens, O., Noppe, N., Devriendt, C., Van de Walle, R. and Van Hoecke, S., 2016. Data-driven multivariate power curve modeling of offshore wind turbines. Engineering Applications of Artificial Intelligence, 55, pp.331-338.(https://doi.org/10.1016/j.engappai.2016.08.003)

[13] Kosanoglu, F., 2022. Wind speed forecasting with a clustering-based deep learning model. Applied Sciences, 12(24), p.13031.(https://doi.org/10.3390/app122413031) p.206.(https://doi.org/10.3390/computers12100206)

[14] Salcedo-Sanz, S., Deo, R.C., Carro-Calvo, L. and Saavedra-Moreno, B., 2016. Monthly prediction of air temperature in Australia and New Zealand with machine learning algorithms. Theoretical and applied climatology, 125, pp.13-25.(https://link.springer.com/article/10.1007/s00704-015-1480-4)

[15] temperature prediction with machine learning models and data reduction techniques. Applied Soft Computing, 136, p.110118.(https://doi.org/10.1016/j.asoc.2023.110118)

[16] Yang, Y., Dong, J., Sun, X., Lima, E., Mu, Q. and Wang, X., 2017. A CFCC-LSTM model for sea surface temperature prediction. IEEE Geoscience and Remote Sensing Letters, 15(2), pp.207-211.(10.1109/LGRS.2017.2780843)