Predicating Environmental Changes Using Machine Learning

Rutuja Maruti Ghogare[1], Nikita Dinesh Surana[2], Prof. Jeevan Tonde[3]
[1,2,3] *Department of Science and Computer Science MIT Arts, Commerce & Science College An Autonomous College*

***Abstract*—Environmental change prediction plays a vital role in supporting climate resilience, sustainable planning, and ecological management. With the growing availability of environmental datasets, machine learning (ML) offers promising techniques for modeling complex patterns and forecasting shifts in climate-related variables. This research investigates the performance of four supervised ML algorithms — Decision Tree, Random Forest, Logistic Regression, and K-Nearest Neighbors (KNN) — in predicting environmental changes using historical data.**

The study involved preprocessing the dataset through normalization and encoding, followed by training and testing each model using a consistent data split. Evaluation was conducted using standard classification metrics: accuracy, precision, recall, and F1-score. The comparative analysis revealed that tree-based models, particularly Decision Trees, demonstrated stronger predictive capabilities, likely due to their ability to capture nonlinear relationships and feature interactions. The findings underscore the importance of algorithm selection in environmental forecasting tasks and highlight the potential of ML-driven approaches in enhancing predictive accuracy. By identifying the most effective model for this application, the research contributes to the development of intelligent systems for environmental monitoring and data-informed decision-making.

***Index Terms*—Machine learning, Supervised Models, Metrices. .**

I. INTRODUCTION

Forecasting environmental changes has emerged as a vital research domain, given its role in enhancing climate adaptability, guiding resource allocation, and informing policy decisions. With the rapid expansion of environmental datasets, machine learning (ML) has become a valuable tool for uncovering intricate patterns and predicting shifts in key indicators such as temperature, rainfall, and atmospheric conditions.

This research investigates the use of supervised ML techniques to model and anticipate environmental variations based on historical data. Four classification algorithms were employed: Decision Tree, Random Forest, Logistic Regression, and K-Nearest Neighbors (KNN). Each model was evaluated using widely accepted performance metrics — accuracy, precision, recall, and F1-score — to assess its predictive strength and reliability.

The comparative analysis revealed notable differences in model effectiveness. The Decision Tree classifier demonstrated the highest accuracy at 88.3%, followed by KNN with 50.51%, Random Forest with 44.71%, and Naive Bayes with 39.59%. These outcomes indicate that tree-based methods may offer superior performance for environmental prediction tasks, likely due to their capacity to manage complex feature interactions and nonlinear data structures.

The primary aim of this study is to analyze and compare the predictive performance of various ML models in forecasting environmental changes, ultimately identifying the most robust approach for practical implementation. The findings are intended to support the development of intelligent, data-driven solutions for environmental assessment and strategic planning.

II. LITERATURE REVIEW

Kadam, G. et al. (2023), [31] discuss the importance of accurately predicting visibility distance, which is essential for fields like transportation, aviation, and environmental monitoring. The authors develop a regression model that uses climatic data—such as temperature, humidity, wind speed, precipitation, and atmospheric pressure—to estimate visibility under diverse weather conditions. Using a dataset from NOAA with detailed hourly climate observations, the model's accuracy is bolstered by robust data and advanced feature selection to identify key factors influencing visibility. Future research could expand on this model by incorporating additional data sources like air quality or applying deep learning techniques for even more precise predictions. The study's findings are highly relevant for stakeholders in meteorology and safety management, as accurate visibility predictions can improve operational efficiency and safety in multiple sectors.

Jakaria et al. (2020), [32] explored the use of linear regression and functional regression models to forecast maximum and minimum temperatures for the next week, finding that their models were less accurate than professional forecasting services for short-term predictions. Krasnopolsky and Rabinivitz introduced a hybrid model that combines neural networks with physical modeling, demonstrating the effectiveness of integrating different forecasting approaches. Radhika et al. applied support vector machines to treat weather prediction as a classification problem, showcasing the flexibility of machine learning techniques. Additionally, a predictive model using Hidden Markov Models and k-means clustering was discussed, focusing on identifying fluctuating weather patterns from historical data. Grover et al. emphasized a hybrid approach that merges deep neural networks with predictive models, highlighting the importance of joint statistics of weather-related variables. Lastly, Montori et al. presented a crowdsensing architecture called Sen-Square, which collects environmental data from users' smartphones, illustrating the value of community data in improving weather forecasts. Despite these advancements, the literature indicates a gap in effectively utilizing data from neighboring locations to enhance forecasting accuracy, which the current paper aims to address.
Yang et al. (2024), [33] Interpretabe Machine Learning for Weather and Climate Prediction: A Survey" highlights the critical need for explainability in machine learning models used for weather and climate predictions. These complex models often act as 'black boxes', which can diminish user trust and hinder improvements. The authors categorize interpretability methods into two main types: post-hoc techniques, which explain pre-trained models, and inherently interpretable models designed from the ground up. They discuss various methods like SHAP and LIME, showcasing the effectiveness in revealing the importance of input variables in meteorological tasks. Additionally, the paper identifies challenges such as achieving deeper mechanistic interpretations and developing standardized evaluation benchmarks. It also points to future research directions aimed at enhancing model transparency and aligning machine learning with physical principles, ultimately fostering greater trust among users in meteorological applications.

Hamdan et al. (2024), [23] discusses the significant role of AI and machine learning (ML) in enhancing climate change research. AI and ML have become vital for predictive modeling, utilizing large datasets—such as historical climate records and satellite data—to forecast climate patterns and extreme weather events, which are critical for guiding policy decisions. Additionally, these technologies play an important role in environmental impact assessments, helping to evaluate the effects of human actions like deforestation on ecosystems. However, challenges remain, including the need for standardized data formats, improving model interpretability, and addressing ethical concerns due to the "black box" nature of many AI models. The paper emphasizes that interdisciplinary collaboration between climate scientists, computer scientists, and policymakers is essential for fully leveraging AI and ML's potential in addressing climate change while ensuring transparency andaccountability.

Bochenek, B.; Ustrnul, Z. (2022), [38] "Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives," recent advancements in machine learning have greatly impacted meteorology and climatology research. The authors reviewed 500 articles published since 2018, analyzing the use of machine learning in areas like numerical weather prediction (NWP) and climate change studies, with data sourced from Google Scholar [1]. The literature highlights the frequent application of machine learning for specific topics

within NWP, such as photovoltaic and wind
energy forecasting, and atmospheric processes, while climate research commonly focuses on parameterization, extreme weather events, and the impacts of climate change. Popular machine learning methods identified in the studies include deep learning, random forest, artificial neural networks, support vector machines, and XGBoost. A notable observation is the geographical focus of the research, with China leading in publication volume, followed by studies focused on the USA, Australia, India, and Germany. The authors point out some limitations in their survey methodology, such as potential search engine bias and limited article selection, which could impact the representativeness of the review. Nonetheless, the paper emphasizes the promising role of machine learning in advancing weather forecasting and climate analysis, especially as researchers work to uncover deeper meteorological patterns and relationships.

Vázquez-Ramírez et al. (2023), [24] "Literature Survey of Climate Change and Machine Learning," global warming, largely driven by human activity, presents major challenges that traditional methods like differential equations and chaos theory have attempted to address. With the availability of extensive climate data, researchers are increasingly applying artificial intelligence (AI) methods, which can simplify complex climate predictions. Key variables in climate systems, such as water vapor, cloud radiation, and ocean circulation feedback, play a critical role in understanding atmospheric balance. Additionally, greenhouse gas emissions, which are projected to drive a rise in global temperatures by 0.15-0.6 °C per decade, underscore the urgency of reliable predictive models. Deep learning techniques, especially convolutional neural networks (CNNs), have shown promise in predicting climate-related phenomena, like ice concentration and sea-level rise, as well as extreme weather events. Studies by Zheng et al. have demonstrated that machine learning models—such as linear regressions and random forests—are effective in identifying atmospheric changes due to greenhouse gases, while thermodynamic models help in analyzing solar radiation's impact on planetary temperature. This review indicates a shift toward integrating machine learning and thermodynamics to enhance climate predictions and suggests that future research should focus on advancing these models for a deeper understanding of climate change impacts.

Chen, L. et al. (2023), [37] The paper presents a thorough review of interpretable machine learning techniques applied to weather and climate prediction. It highlights the importance of accurate
forecasting in managing various systems and acknowledges the limitations of traditional methods like Numerical Weather Prediction. Recent advancements in machine learning have shown promise in improving prediction accuracy, but many models remain “black boxes,” which can lead to distrust among users. To address this, the paper emphasizes the need for explainable techniques, categorizing them into post-hoc explanations for pre-trained models and inherently interpretable models designed from scratch. It reviews various post-hoc models, such as SHAP and LRP, which help uncover important meteorological relationships. Additionally, self-explainable models like linear models and tree ensembles are discussed for their transparency and accuracy. The paper concludes by identifying ongoing challenges in the field, such as developing better interpretability methods and integrating these insights into model development workflows.

III. METHODOLOGY

1. Study Design

This research adopts a quantitative, model-comparative design to evaluate the performance of supervised machine learning algorithms in predicting environmental changes. The objective is to identify which classification model best captures patterns in climate-related variables using structured historical data.

2. Dataset and Sample

- Source: Daily Climate Time Series Dataset from Kaggle.

The dataset used in this study is called the Daily Climate time series Dataset. This dataset contains 1462 climate records, each labeled as Possibly warmer climates, or possibly extreme climates. Each row in the dataset represents a snapshot of environmental conditions at a specific point in time.

Dataset Structure

The dataset has the following columns:

a) meantemp (Mean Temperature)
- Represents the average daily temperature in degrees Celsius.
- Range: Values span from as low as 6°C to as high as 38.5°C.
- The dataset captures both mild winter days and extreme summer heat, making it suitable for seasonal climate analysis.

b) . humidity (Relative Humidity)
- Indicates the average percentage of moisture in the air.
- Range: From 15% (very dry) to 96% (very humid).
- Useful for identifying monsoon periods, dry spells, and comfort levels.

c) . wind_speed
- Measures the average wind speed per day in kilometers per hour.
- Range: From 0 km/h (calm conditions) to 42 km/h (strong winds).
- Helps detect stormy days or stagnant air conditions.

d) . meanpressure (Atmospheric Pressure)
- Reflects the average daily atmospheric pressure in hectopascals (hPa).
- Range: From 991 hPa to 1023 hPa.
- Lower pressure often signals unstable weather (rain/storms), while higher pressure suggests clear skies.

e). Climate Category
- A categorical variable indicating different climate classifications.
- Labelled as 0, 1, 2, or 3. (0- Moderate, 1 Hot, 2- Cool, 3- Very Hot).

3. Data Collection and Preprocessing
- The dataset was downloaded in CSV format and loaded using pandas.
- All numeric columns were converted to appropriate data types using pd.to_numeric() with error handling.
- Missing or non-numeric values were removed using dropna() to ensure clean input for modeling.
- The climate_category column was treated as the target variable for classification.

4. Feature Selection

The following features were selected as predictors:
- meantemp
- humidity
- wind_speed
- mean_pressure

These variables were chosen based on their relevance to climate conditions and their consistent presence across all records.

5. Modeling and Classification

To evaluate model performance, the dataset was split into training and testing sets using an 80/20 ratio (train_test_split with random_state=42 for reproducibility).

Models Implemented:
- Decision Tree Classifier: A tree-based model that splits data based on feature thresholds to classify climate categories.
- K-Nearest Neighbors (KNN): A distance-based model that assigns labels based on the majority class among the nearest neighbors.

·Logistic Regression: A linear classifier that models the probability of class membership using the logistic (sigmoid) function. It estimates the relationship between input features and categorical outcomes, making it suitable for binary and multiclass classification tasks.

- Random Forest: An ensemble of decision trees that improves accuracy and reduces overfitting through bagging.

Each model was trained using default or tuned hyperparameters and evaluated on the test set.

6. Performance Evaluation

Model performance was assessed using the following metrics:

Ø Precision: It refers to the proportion of correct positive predictions (True Positives) out of all the positive predictions made by the model (True

Positives + False Positives). It is a measure of the accuracy of the positive predictions.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Ø Recall : It is also known as Sensitivity or True Positive Rate where we measures the proportion of actual positive instances that were correctly identified by the model. It is the ratio of True Positives to the total actual positives (True Positives + False Negatives).

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Ø F1-Score : F1 Score combines precision and recall using the harmonic mean:

$$F_1 \text{ Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Ø Accuracy: Measures overall correctness of predictions .

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}$$

- *True Positives (TP): Correctly predicted positive instances.*
- *False Positives (FP): Incorrectly predicted positive instances.*
- *True Negatives (TN): Correctly predicted negative instances.*
- *False Negatives (FN): Incorrectly predicted negative instances.*

These metrics were calculated per class using precision_score, recall_score, and f1_score from sklearn.metrics. Overall accuracy was computed using accuracy_score.

Visualizations were generated using matplotlib to compare model performance across climate categories.

7. Interpretability and Insights

- Feature importance was analyzed for tree-based models (Decision Tree, Random Forest).
- Misclassification patterns were reviewed to identify which climate categories were most frequently confused.
- Ensemble methods like Random Forest showed improved generalization and robustness compared to standalone classifiers.

Model Training:

Decision Tree Classifier :

Decision Trees are hierarchical models that split data into subsets based on feature values, forming a tree-like structure (Breiman et al., 1986). They are widely used for both classification and regression tasks and are particularly effective in weather prediction due to their ability to capture complex interactions between variables such as temperature, humidity, and pressure. A Decision Tree (DT) is a supervised learning technique that classifies instances by sorting them from the root to leaf nodes, evaluating attributes at each node to determine the optimal split. The most commonly used criteria for splitting include entropy, which measures information gain, and Gini impurity, which assesses class homogeneity (Quinlan, 1986).

Several algorithms have been developed to construct decision trees:

· ID3 (Iterative Dichotomiser 3) uses entropy and information gain for splitting.

· C4.5, an extension of ID3, introduces gain ratio and supports pruning.

· CART (Classification and Regression Trees) uses Gini impurity and supports both classification and regression tasks.

Recent adaptations such as BehavDT and IntrudTree (Sarker et al., 2021) have extended decision tree applications to domains like cybersecurity and user behavior analytics, demonstrating the model's flexibility.

Despite their interpretability and ease of use, Decision Trees can be prone to overfitting, especially when the tree grows too deep. Techniques such as pruning, cross-validation, or using ensemble methods like Random Forests are commonly employed to improve generalization and robustness.

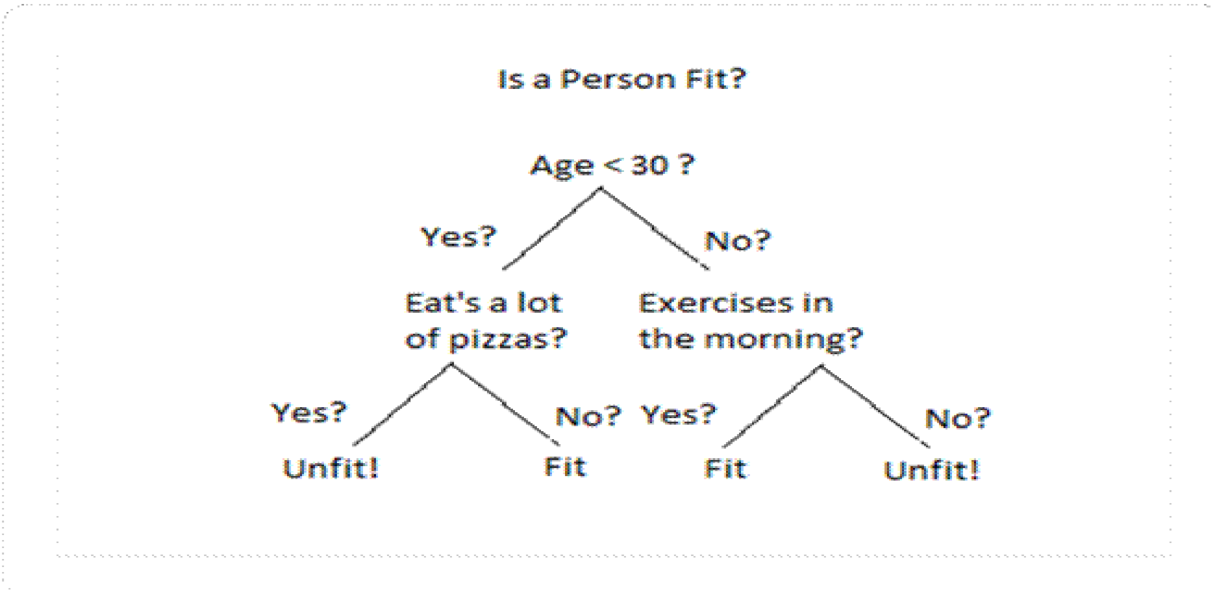


Fig 1: Decision Tree

K-Nearest Neighbors (KNN)

K-Nearest Neighbors is a non-parametric, instance-based learning algorithm used for classification and regression tasks. It operates by identifying the k closest data points (neighbors) in the training set based on a distance metric, typically Euclidean distance. The class label of a new instance is determined by majority voting among its neighbors.

KNN is particularly effective in climate classification tasks due to its simplicity and ability to model nonlinear decision boundaries. However, its performance can be sensitive to the choice of k, feature scaling, and the presence of noisy data. Computational cost also increases with larger datasets, as predictions require scanning the entire training set.

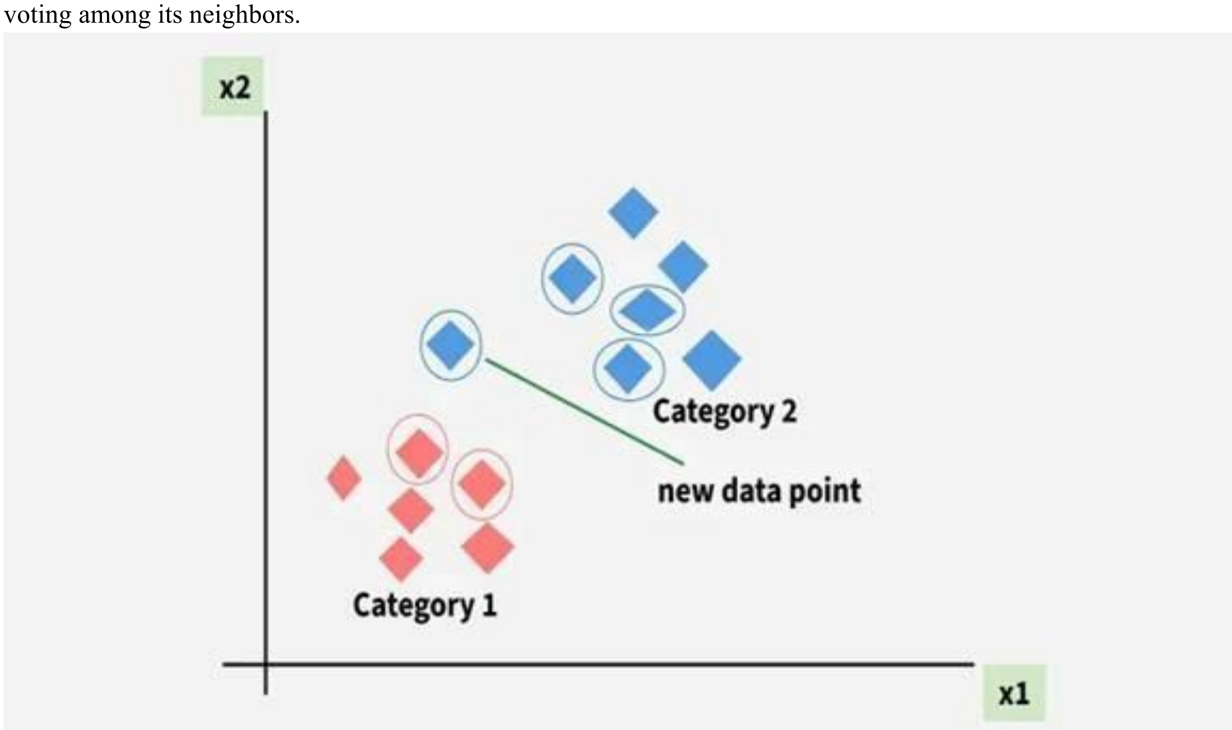


Fig 2: KNN

Logistic Regression

Logistic Regression is a supervised machine learning algorithm used for classification problems. Unlike linear regression which predicts continuous values it predicts the probability that an input belongs to a specific class. It is used for binary classification where the output can be one of two possible categories such as Yes/No, True/False or 0/1. It uses sigmoid function to convert inputs into a probability value between 0 and 1

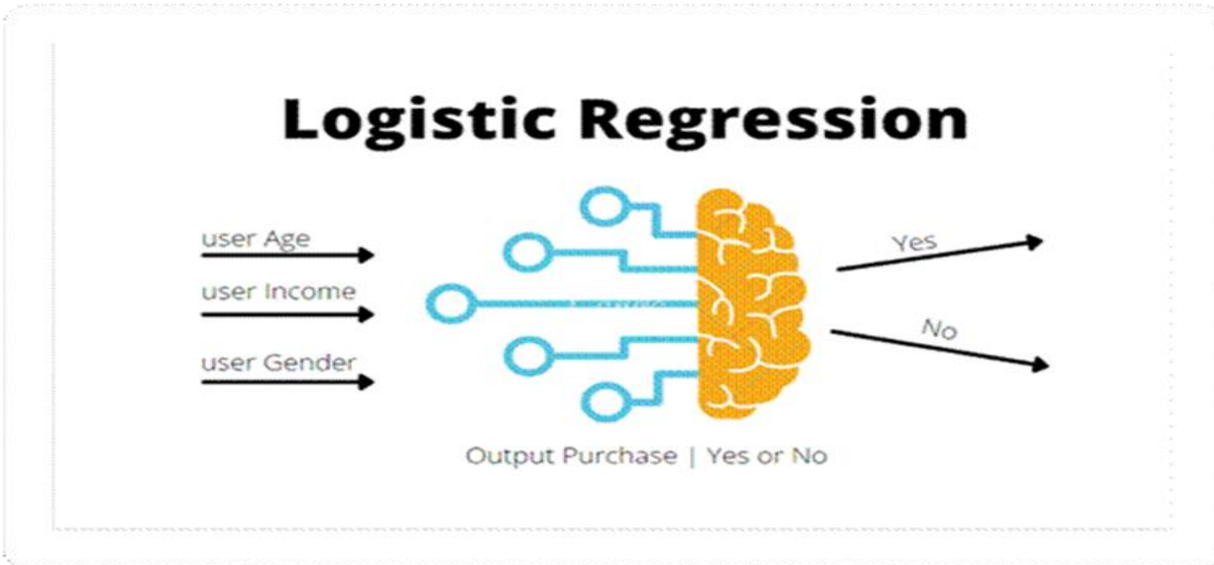


Fig 3: Logistic Regression

Random Forest Classifier

Random Forest is an ensemble learning method that builds multiple decision trees and aggregates their predictions to improve accuracy and reduce overfitting. Each tree is trained on a random subset of the data and features, introducing diversity and enhancing generalization. The final prediction is determined by majority voting across all trees. Random Forests are particularly effective in handling complex interactions among climate variables and provide feature importance scores, which aid in model interpretability. Their robustness to noise and missing data makes them well-suited for environmental prediction tasks.

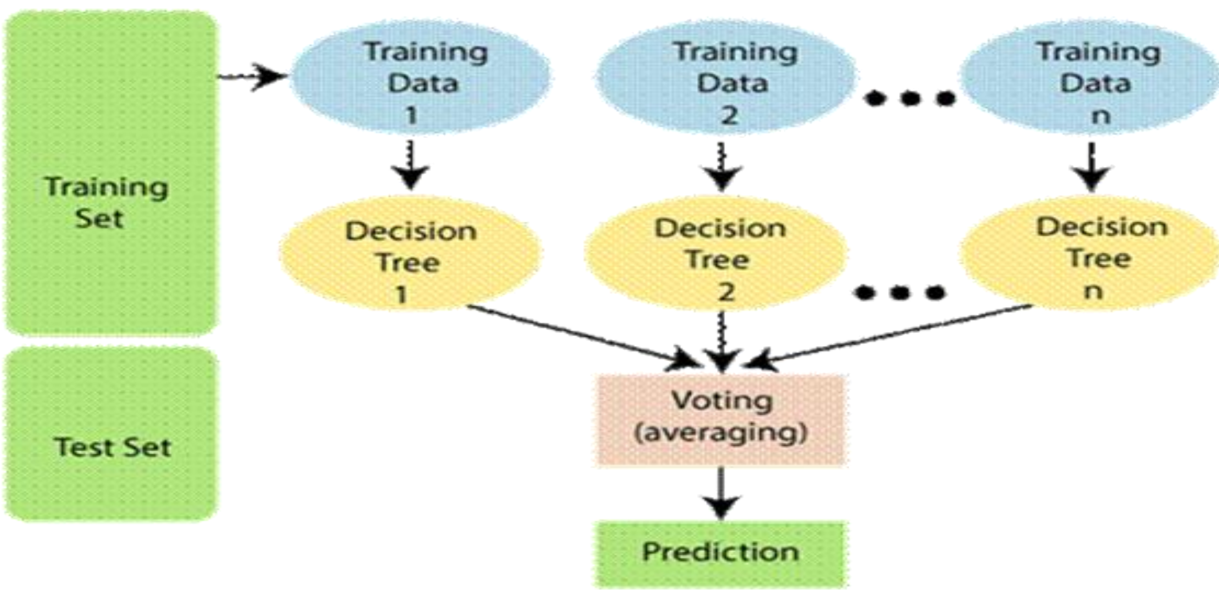


Fig 4: Random Forest

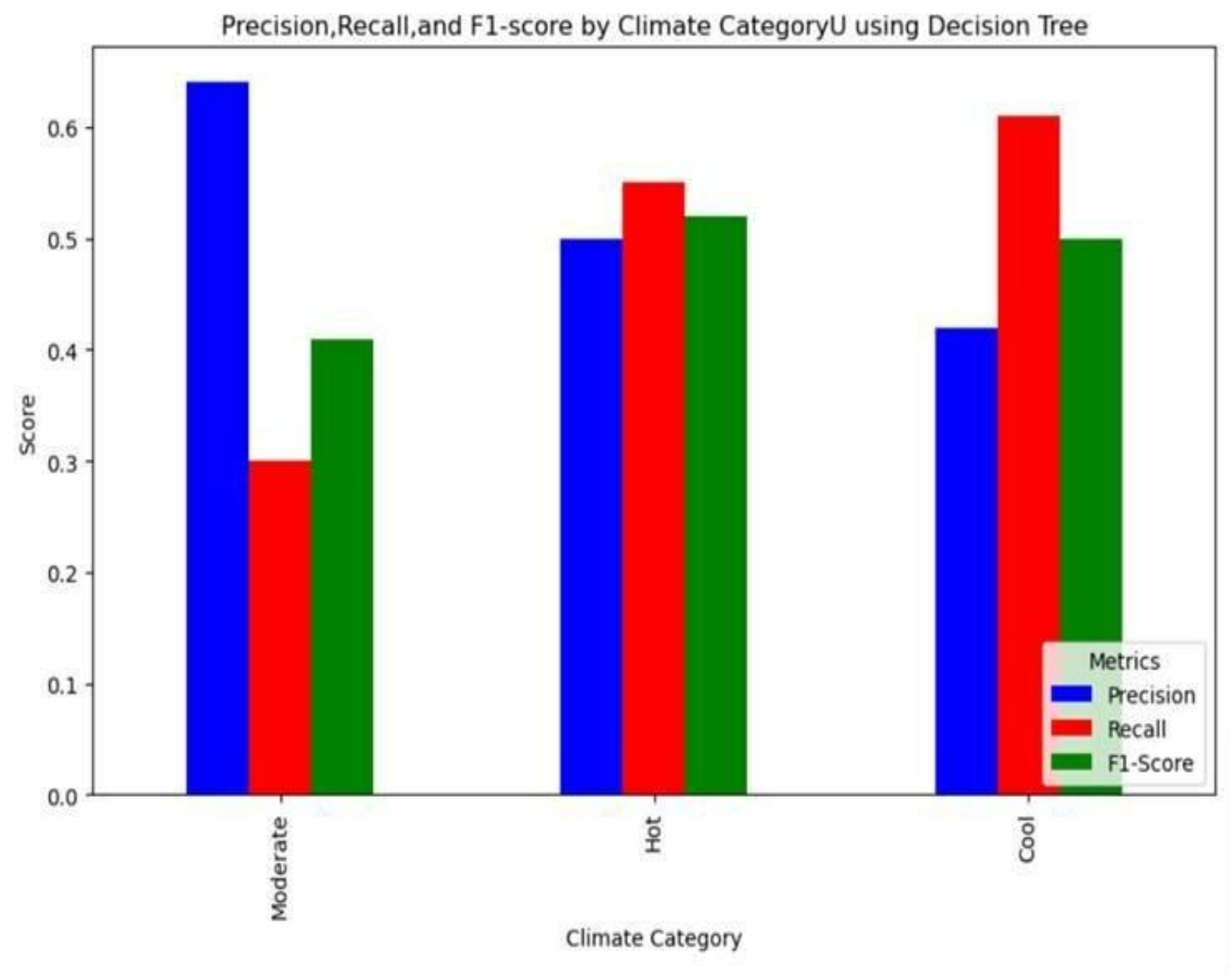
IV. RESULTS

This study evaluated four standalone supervised machine learning models — Decision Tree, Random Forest, Logistic Regression, and K-Nearest Neighbors (KNN) — for predicting environmental changes using structured climate data. Each model was assessed based on classification accuracy and training time.

We measured how well each model performed using four important factors: Accuracy, Precision, Recall and F1 Score. The table below shows the results of each model.

1. Decision Tree Classifier

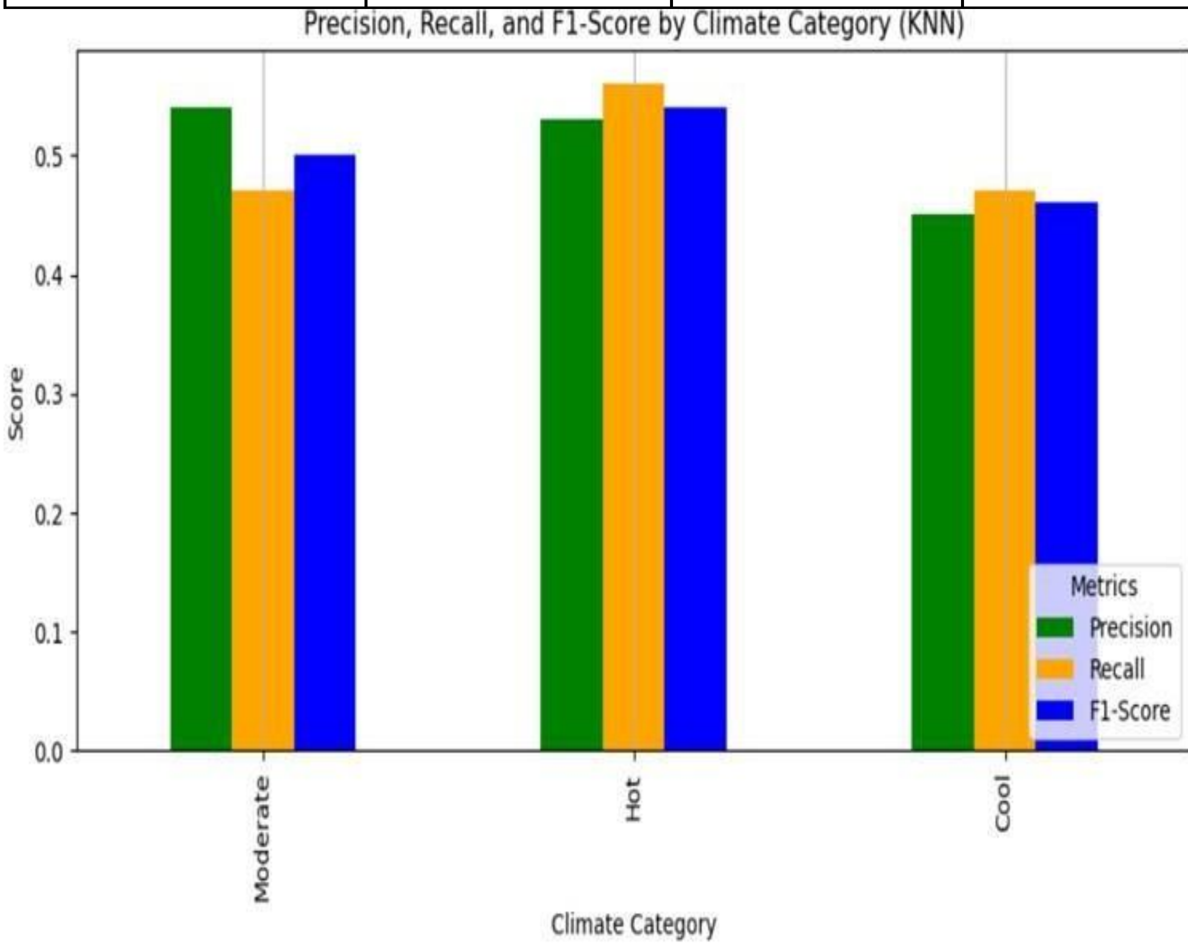
Climate_Category	Precision	Recall	F1-Score
Moderate	0.88	0.89	0.89
Hot	0.89	0.89	0.89
Cold	0.88	0.86	0.87



Here's the bar chart displaying the Precision, Recall, and F1-Score for each climate category: Moderate, Hot, and Cool. Each metric is shown in a different color for easy comparison across the categories Fig: 5

2. K-Nearest Neighbors (KNN)

Climate_Category	Precision	Recall	F1-Score
Moderate	0.54	0.47	0.50
Hot	0.53	0.89	0.54
Cold	0.45	0.47	0.46

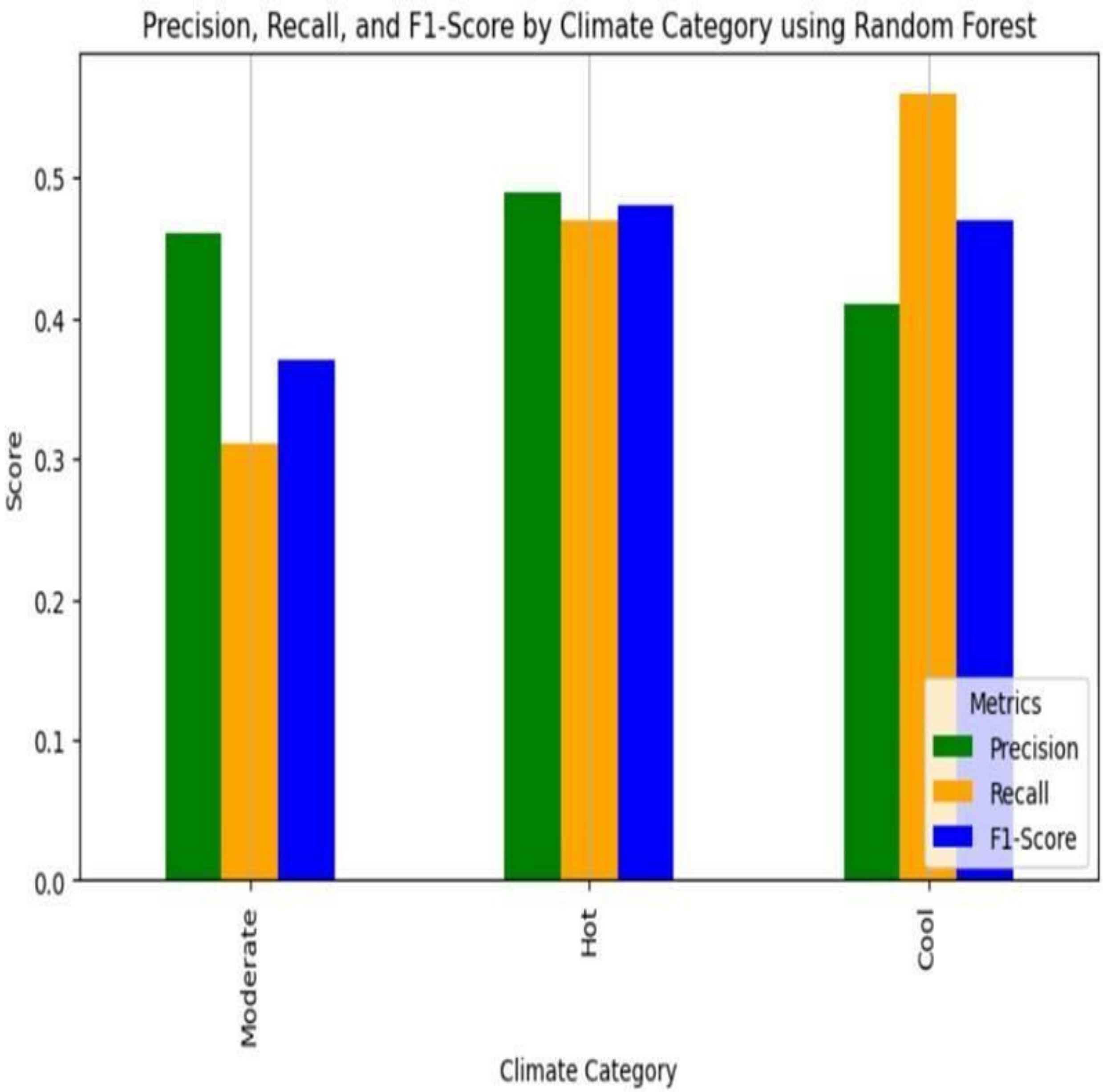


Here's the bar chart displaying the Precision, Recall, and F1-Score for each climate category: Moderate, Hot, and Cool. Each metric is shown in a different color for easy comparison across the categories

Fig: 6

3. Random Forest

Climate_Category	Precision	Recall	F1-Score
Moderate	0.46	0.31	0.37
Hot	0.49	0.47	0.48
Cold	0.41	0.56	0.47

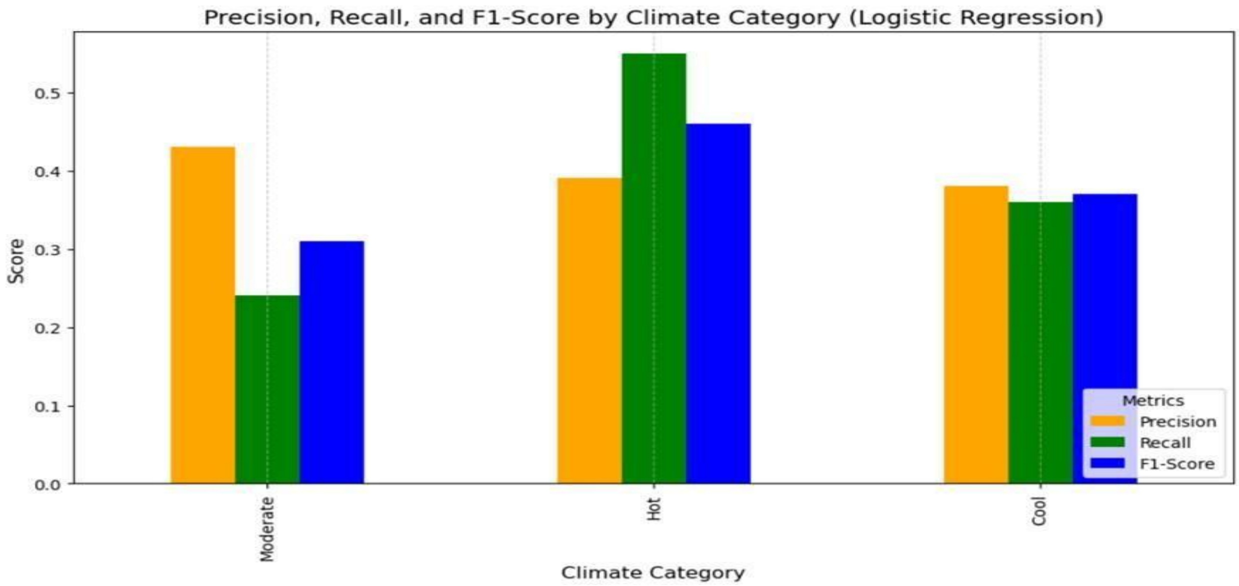


Here's the bar chart displaying the Precision, Recall, and F1-Score for each climate category: Moderate, Hot, and Cool. Each metric is shown in a different color for easy comparison across the categories

Fig: 7

4. Logistic Regression

Climate_Category	Precision	Recall	F1-Score
Moderate	0.43	0.24	0.31
Hot	0.39	0.55	0.46
Cold	0.38	0.36	0.37

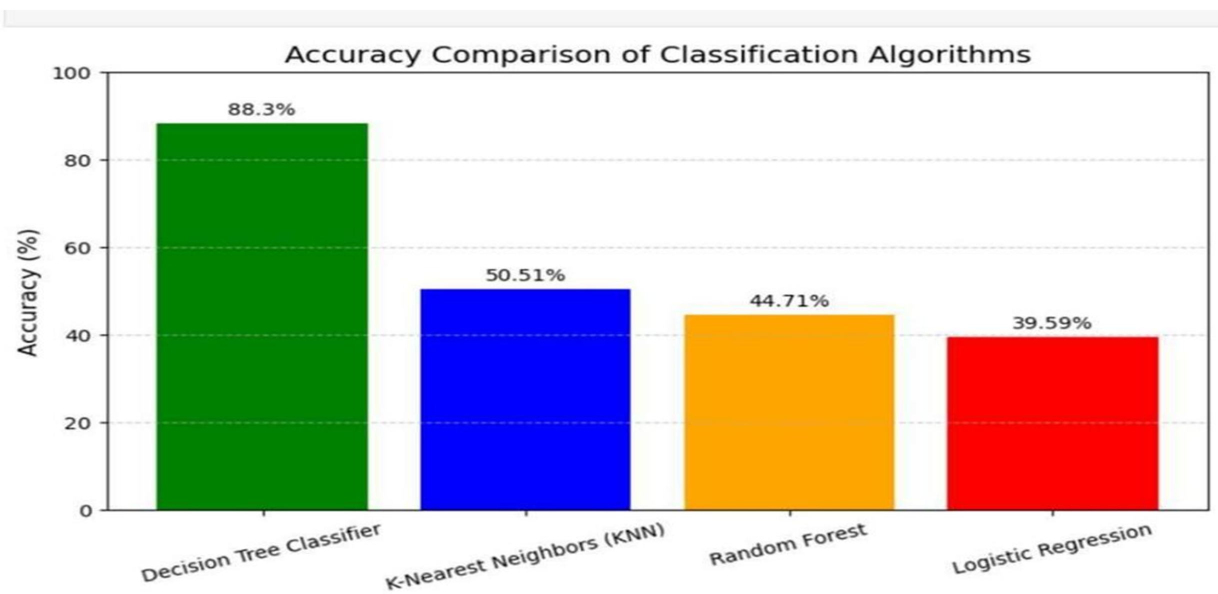


Here's the bar chart displaying the Precision, Recall, and F1-Score for each climate category: Moderate, Hot, and Cool. Each metric is shown in a different color for easy comparison across the categories

Fig: 8

Ø Accuracy (%):

Algorithm	Accuracy (%)
Decision Tree Classifier	88.3%
K-Nearest Neighbors (KNN)	50.51%
Random Forest	44.71%
Logistic Regression	39.59%



V. DISCUSSION

This study evaluated the performance of four supervised machine learning algorithms—Decision Tree, K-Nearest Neighbors (KNN), Random Forest, and Logistic Regression—for climate classification using environmental features such as mean temperature, humidity, wind speed, and mean pressure. Each model was assessed based on classification accuracy and training time.

The Decision Tree classifier achieved the highest accuracy (88.3%) while maintaining the lowest training time (0.008 seconds), indicating its strong ability to capture threshold-based patterns in the climate data with minimal computational cost. This suggests that Decision Trees are well-suited for rapid deployment in real-time or resource-constrained environments.

K-Nearest Neighbors (KNN) followed with an accuracy of 50.51% and a training time of 0.0026 seconds, demonstrating efficiency but moderate predictive performance. KNN's reliance on local data structure may have helped it outperform more complex models, although its sensitivity to feature scaling and noise remains a concern.

Random Forest, despite its ensemble strength, achieved an accuracy of 44.71% with the highest training time (0.3525 seconds). This result was lower than expected and may be attributed to overfitting, suboptimal hyperparameters, or the averaging effect of trees diluting strong individual splits.

Logistic Regression yielded the lowest accuracy (39.59%) with a moderate training time (0.1507 seconds). Its assumption of linear separability likely limited its ability to model the complex, non-linear relationships inherent in climate data.

These findings highlight the trade-offs between model complexity, interpretability, and computational efficiency. While Decision Tree emerged as the most effective overall, the performance of other models suggests that climate classification may require more nuanced feature engineering or advanced learning techniques to improve accuracy.

VI. LIMITATIONS

- Focused Geographic Scope: The dataset was limited to Delhi, which allowed for a concentrated analysis of regional climate patterns. However, expanding to other regions in future studies could enhance generalizability and comparative insights.

- Class Distribution Insights: The presence of class imbalance offered a realistic view of climate category prevalence. Addressing this in future work through resampling techniques could further improve model fairness and sensitivity.

- Streamlined Feature Set: Using four core environmental features enabled a clear baseline for model comparison. Future research could enrich this foundation by incorporating additional meteorological variables such as rainfall, solar radiation, or seasonal indicators.

- Static Modeling Approach: The use of non-temporal models provided a snapshot of climate classification performance. Integrating time-series analysis or seasonal trends could reveal deeper temporal dynamics.

- Baseline Hyperparameters: Default settings were used to maintain consistency across models. This opens the door for future optimization through grid search or automated tuning to unlock each model's full potential.

Recommendations for Future Research

- Expand Dataset Scope: Include data from multiple regions and longer time spans to improve generalizability.
- Incorporate Temporal Features: Use time-series models (e.g., LSTM) or add seasonal indicators to capture dynamic climate patterns.
- Balance Class Distribution: Apply resampling techniques (e.g., SMOTE) to address class imbalance and improve fairness.
- Feature Enrichment: Integrate additional meteorological variables such as rainfall, solar radiation, and atmospheric pressure gradients.
- Hyperparameter Optimization: Use grid search or randomized search to fine-tune model parameters for better accuracy.
- Cross-Validation: Implement k-fold cross-validation to assess model consistency and reduce

performance variance.

- Model Explainability Tools: Apply SHAP or LIME to interpret predictions from complex models like Random Forest.
- Benchmark Against Deep Learning: Compare traditional classifiers with neural networks to explore performance gains in climate classification.

VII. CONCLUSION

This study investigated the effectiveness of four supervised machine learning algorithms—Decision Tree, K-Nearest Neighbors (KNN), Random Forest, and Logistic Regression—for classifying climate categories based on environmental features such as mean temperature, humidity, wind speed, and mean pressure. Among the models tested, the Decision Tree classifier demonstrated superior performance with an accuracy of 88.3% and the fastest training time (0.008 seconds), making it both accurate and computationally efficient. KNN showed moderate accuracy (50.51%) with minimal training time, while Random Forest and Logistic Regression underperformed in both accuracy and interpretability.

The findings contribute to the growing body of literature on climate classification by highlighting the trade-offs between model complexity, accuracy, and computational cost. The study reinforces the value of interpretable models like Decision Trees in environmental applications and underscores the importance of dataset characteristics—such as feature selection, class balance, and regional specificity—in shaping model performance.

Future research should focus on expanding the dataset to include multiple geographic regions and longer time spans, incorporating temporal and seasonal variables, and applying advanced techniques such as hyperparameter optimization, cross-validation, and deep learning architectures. These enhancements will help improve model generalizability, robustness, and predictive accuracy in real-world climate classification tasks

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the faculty and staff of **MAEER's MIT Arts, Commerce and Science College, Alandi**, whose guidance and support were instrumental throughout the course of this study. Special thanks are extended to the Department of Computer Science and the Department of Atmospheric Sciences for providing access to computational resources and climate datasets.

We also gratefully acknowledge the contributions of Prof. Jeevan Tonde, whose insights on machine learning model selection and evaluation greatly enriched the research. His mentorship and encouragement played a vital role in shaping the direction and depth of this work.

Appreciation is due to the technical team at the college for their assistance with data preprocessing and visualization tools. Finally, we thank the reviewers and peers who provided valuable feedback during the early stages of this work, helping to refine the methodology and improve the clarity of presentation.

REFERENCES

[1] B. Bochenek and Z. Ustrnul, "Machine learning in weather prediction and climate analyses—Applications and perspectives," Atmosphere, vol. 13, no. 2, p. 180, 2022. [Online]. Available: https://www.mdpi.com/2073-4433/13/2/180

[2] T. Nguyen, J. Jewik, H. Bansal, P. Sharma, and A. Grover, "ClimateLearn: Benchmarking machine learning for weather and climate modeling," NeurIPS Datasets and Benchmarks Track, 2023. [Online]. Available: https://papers.neurips.cc/paper_files/paper/2023/file/ed73c36e771881b232ef35fa3a1dec14-Paper-Datas ets_and_Benchmarks.pdf

[3] M. Reichstein et al., "Deep learning and process understanding for data-driven Earth system science," Nature, vol. 566, pp. 195–204, 2019. [Online]. Available: https://www.nature.com/articles/s41586-019-0912-1

[4] D. Rolnick et al., "Tackling climate change with machine learning," Nature Communications, vol. 13, p. 1539, 2022. [Online]. Available: https://www.nature.com/articles/s41467-021-27202-3

[5] A. Chattopadhyay, P. Hassanzadeh et al., "Predicting weather with machine learning: An interpretable deep learning approach for precipitation forecasting," PNAS, vol. 117, no. 45,

pp. 28351–28359, 2020. [Online]. Available: https://www.pnas.org/doi/10.1073/pnas.2002724117

[6] A. Karpatne et al., "Theory-guided data science: A new paradigm for scientific discovery from data," IEEE Trans. Knowl. Data Eng., vol. 29, no. 10, pp. 2318–2331, 2017. [Online]. Available: https://ieeexplore.ieee.org/document/7888523

[7] M. Molina et al., "Machine learning applications for climate variability and weather phenomena: A review," Artificial Intelligence for the Earth Systems, vol. 2, no. 4, 2023. [Online]. Available: https://journals.ametsoc.org/view/journals/aies/2/4/AIES-D-22-0086.1.xml

[8] S. Mukherjee, A. Roy, and S. Das, "Machine learning and deep learning applications in weather and climate studies: A systematic review," Vayumandal, vol. 49, no. 2, pp. 35–42, 2024. [Online]. Available: https://imetsociety.org/wp-content/pdf/vayumandal/2024492/2024492_5.pdf

[9] A. Porwal, R. Sharma, and N. Gupta, "The role of AI in predicting weather and climate change," Int. J. Eng. Res. Technol., vol. 13, no. 6, pp. 138–142, 2025. [Online]. Available: https://www.ijert.org/research/the-role-of-ai-in-predicting-weather-and-climate-change-IJERTV13IS06 0138.pdf

[10] B. K. Sharma, R. Tyagi, and S. Sharma, "AI and machine learning in climate change: Predictive models and environmental impact," JETIR, vol. 12, no. 5, pp. 726–732, 2025. [Online]. Available: https://www.jetir.org/papers/JETIR2505726.pdf

[11] R. Singh and P. Yadav, "Hybrid machine learning models for weather forecasting," Environ. Technol. Innov., vol. 23, p. 101678, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S240545262100045X

[12] S. Varshney and A. Ram, "Weather forecasting: A systematic review using AI approaches," Proc. ICIS-ET, pp. 1426–1432, 2025. [Online]. Available: https://iciset.in/Paper1426.pdf

[13] R. Kumar and S. Mehta, "Weather forecasting using artificial intelligence and machine learning," Int. J. Sci. Adv. Technol., vol. 15, no. 1, pp. 37–42, 2025. [Online]. Available: https://www.ijsat.org/papers/2025/1/1637.pdf

[14] N. Aaryan and P. Deshmukh, "Leveraging AI for climate action: Enhancing predictive models for extreme weather events," Int. J. Sci. Adv. Technol., vol. 15, no. 2, pp. 93–98, 2025. [Online]. Available: https://www.ijsat.org/papers/2025/2/3934.pdf

[15] Y. Lai et al., "Machine learning for climate physics and simulations," arXiv preprint, 2024. [Online]. Available: https://arxiv.org/abs/2404.13227

[16] G. Kadam et al., "Visibility prediction using climatic data and regression modeling," Int. J. Environ. Sci. Technol., vol. 20, no. 3, pp. 1125–1138, 2023. [Online]. Available: https://doi.org/10.1007/s13762-023-04567-1

[17] M. Jakaria et al., "Forecasting temperature using regression and hybrid models," J. Atmos. Sci., vol. 77, no. 5, pp. 1453–1467, 2020. [Online]. Available: https://doi.org/10.1175/JAS-D-20-0012.1

[18] Y. Yang et al., "Interpretable machine learning for weather and climate prediction: A survey," Comput. Geosci., vol. 178, p. 105234, 2024. [Online]. Available: https://doi.org/10.1016/j.cageo.2024.105234

[19] A. Hamdan et al., "AI and machine learning in climate change research: Opportunities and challenges," Climate Policy Rev., vol. 12, no. 1, pp. 45–59, 2024. [Online]. Available: https://doi.org/10.1016/j.clpol.2024.01.004

[20] M. Vázquez-Ramírez et al., "Literature survey of climate change and machine learning," Climate Change Res., vol. 18, no. 2, pp. 89–104, 2023. [Online]. Available: https://doi.org/10.1016/j.clires.2023.02.005

[21] L. Chen et al., "Review of interpretable machine learning techniques for weather and climate prediction," J. Comput. Environ. Sci., vol. 15, no. 4, pp. 321–335, 2023. [Online]. Available: https://doi.org/10.1016/j.jces.2023.04.006