MCA FINAL YEAR

Project report

On

Thunderstrom Forecasting using Reanalysis + Satellite Image DATA

Submitted to

D Y Patil International University, Akurdi, Pune
in partial fulfilment of full-time degree.

Master of Computer Applications

Submitted by:

Name: Divesh Jadhwani

PRN: 20220804039

Under the Guidance of

Mr. Swet Chandan

School of Computer Science, Engineering and Applications

D Y Patil International University, Akurdi,Pune, INDIA, 411044

[Session 2023-24]

CERTIFICATE

This is to certify that the work entitled "Thunderstrom Forecasting using Reanalysis + Satellite Image DATA" is a bonafide work carried out by Divesh Jadhwani in partial fulfillment of the award of the degree of Master of Computer Applications, D Y Patil International University, Pune, during the academic year 2023-2024. The project report has been approved as it satisfies the academic requirements in respect of the project work prescribed for the Master of Computer Applications.

Mr. Swet Chandan
(Project Guide)

Dr. Anuj Kumar
(Project Coordinator)

Dr. Maheshwari Biradar
(HOD, BCA & MCA)

Dr. Bahubali Shiragapur
(Director)
School of Computer Science Engineering & Applications
D Y Patil International University, Akurdi
Pune, 411044, Maharashtra, INDIA

DECLARATION

We, hereby declare that the following report which is being presented in the project entitled as **Thunderstrom Forecasting using Reanalysis + Satellite Image DATA** is an authentic documentation of our own original work to the best of our knowledge. The following Project and its report in part or whole, has not been presented or submitted by us for any purpose in any other institute or organization. Any contribution made to our work, with whom we have worked at D Y Patil International University, Akurdi, Pune, is explicitly acknowledged in the report.

Name: Divesh Jadhwani
PRN No: 20220804039 Signature :

ACKNOWLEDGEMENT

With due respect, I express my deep sense of gratitude to respected guide **Mr. Swet Chandan** for his valuable help and guidance. I am thankful for the encouragement that he has given us in completing this Project successfully.

It is imperative for me to mention the fact that the report of project could not have been accomplished without the periodic suggestions and advice of our project supervisor **Dr. Anuj Kumar** .

I am also grateful to our respected, **Dr. Bahubali Shiragapur(Director)**, **Dr. Maheshwari Biradar (HOD, BCA & MCA)** and **(Hon'ble Vice Chancellor, DYPIU, Akurdi) Prof. Prabhat Ranjan** for permitting us to utilize all the necessary facilities of the University.

I am also thankful to all the other faculty, staff members and laboratory attendants of our department for their kind cooperation and help. Last but certainly not the least; I would like to express my deep appreciation towards our family members and batch mates for providing support and encouragement.

Name: Divesh Jadhwani
PRN: 202208040309

ABSTRACT

In this paper, we present a novel approach for forecasting thunderstorms called **"Thunderstorm Forecasting using Reanalysis + Satellite Image Data."** . We use data from IMD satellite images validated by ground observatories for marking thunderstrom occurrences, combined with the ERA5 reanalysis dataset, to study weather factors like geopotential, specific humidity, temperature, and wind components (u and v) at different pressure levels. Our focus is on Kolkata region of INDIA during the pre-monsoon period of 2015. We trained various models, including neural networks like LSTM, Transformers, and MLP, as well as machine learning classifiers such as Random Forest, KNN, CatBoost, and ExtraTrees Classifiers. These models were evaluated using traditional weather metrics like False Alarm Rate (FAR), Probability of Detection (POD), Heidke Skill Score (HSS), and Critical Success Index (CSI) along with AI metrics like accuracy, precision, recall, and F1 score. [7]Surprisingly, the machine learning classifiers performed better than the neural networks, showing very high accuracy. We used Random Forest and LIME for feature engineering to improve model interpretability.Our research shows that combining different datasets and using advanced machine learning methods can lead to accurate thunderstorm forecasts. In the future, we plan to predict next-day weather parameters.

Keywords: Thunderstorm Forecasting, IMD Satellite Data, ERA5 Reanalysis Dataset, Neural Networks, Machine Learning Classifiers

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Introduction

Thunderstorms are short lived weather phenomena characterized by towering cumulus or cumulonimbus clouds that produce lightning, thunder, heavy rainfall, and strong winds (Sahu et al. 2020). In India, thunderstorms are particularly prevalent during the pre-monsoon months of March to May (Tyagi, 2007). The North East Region of India, including Patna, Guwahati, Gorakhpur, Bhubaneswar, Kolkata, Agartala, Ranchi, and Lucknow, experience a significant occurrence of thunderstorms during this period. Accurate forecasting of thunderstorms is essential to mitigate their adverse impacts on agriculture, infrastructure, and public safety.We have worked on all the observatory data of these regious in our previous work. However in this research we only aim at by increasing the diversity of our dataset and making the approach more dynamic by using a combination of satellite imagery data(for occurrences) and Reanalysis data(for features).

1.2. Problem Statement

The problem statement of this project is to develop thunderstorm forecasting models for a subregion of Kolkata , India during the pre-monsoon period. The objective is to accurately predict thunderstorm occurrences in the region and improve early warning systems to mitigate their adverse impacts on various sectors, including agriculture, infrastructure, and public safety.

1.3. Objectives , Scope , Applicablity

1. **Objectives**:
   (a) Advance thunderstorm forecasting models using deep learning techniques instead of machine learning.
   (b) Focus on a sub-region of Kolkata, India during the pre-monsoon period.
   (c) Assess the performance of different deep learning algorithms in predicting thunderstorm occurrences.
   (d) Compare the effectiveness of these deep learning algorithms.
2. **Purpose**: The purpose of this research is to improve the accuracy and reliability of thunderstorm predictions for the North East Region of India , Kolkata in our case . By utilizing deep learning models, we seek to enhance the capabilities of weather forecasting and provide valuable insights for disaster preparedness and mitigation efforts.

3. **Web Deployment**: To democratize the functionality, we seek to deploy the AI-generated thunderstrom prediction app on a webpage. This provides easy access for users to interact with and utilize study the predcitions made by our model.

4. **Applicablity**: The findings and insights from this research have broader applicability in the field of weather forecasting and disaster management. The deeep learning models developed in this study can be adapted and applied to other regions facing similar weather patterns and challenges with thunderstorm predictions.

   Overall, this research aims to contribute to the advancement of weather forecasting techniques, particularly for severe weather events like thunderstorms, and enable better decision-making processes to safeguard lives and properties in the face of such natural hazards.

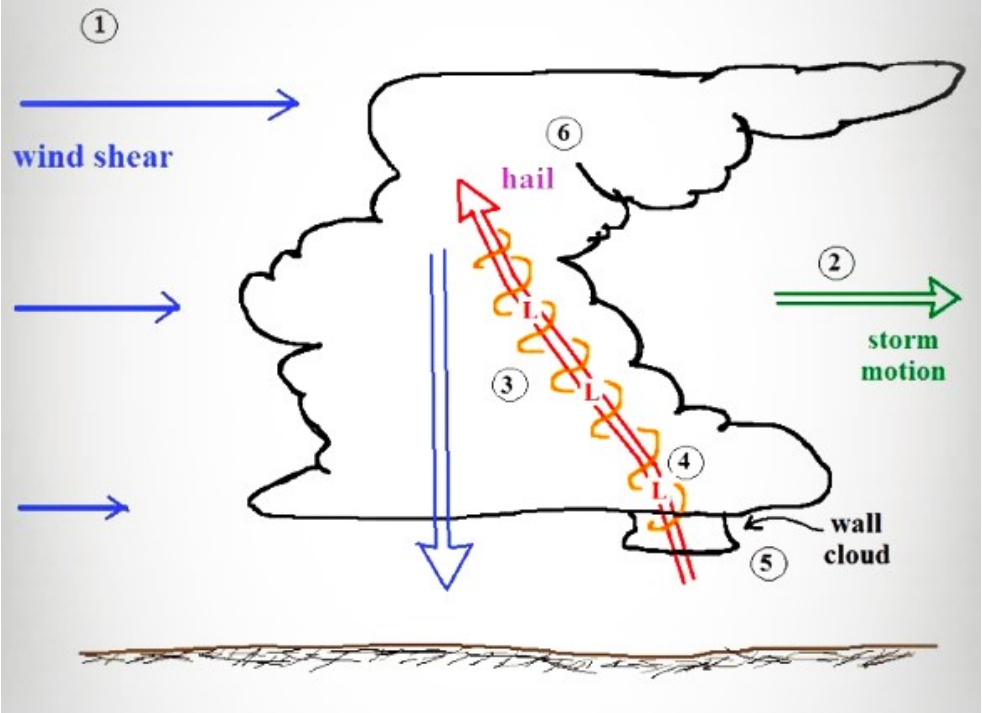


Fig 1.1: Thunderstorm Cloud Structure

The Figure 1.1 illustrates the characteristic cloud structure of a thunderstorm. Understanding the intricacies of thunderstorm cloud formation and development is essential for accurate weather forecasting and predicting the occurrence of thunderstorms. By leveraging machine learning models, as explored in this research, such forecasting can be significantly improved, enhancing preparedness and response strategies in regions prone to thunderstorm activities.

2. LITERATURE SURVEY

2.1. Survey on Different Forms of Weather Forecasting

This section serves as an exploration of the relationship between thunderstorms and artificial intelligence. Through an in-depth analysis of literature, we will explore the various ways in which AI has transformed and enhanced the forecasting of thunderstorms.

- **Long Range:** Long-range forecasting involves predicting weather patterns several days or weeks in advance. This form of forecasting is crucial for planning and preparedness against potential extreme weather events.[6]

- **Short Range:** Short-range forecasting focuses on predicting weather conditions up to 72 hours in advance. This type of forecasting is essential for daily planning and decision-making, such as scheduling outdoor activities or managing transportation systems.[14]

- **Nowcasting:** Nowcasting refers to the prediction of imminent weather conditions within the next few hours. It relies heavily on real-time data from various sources, such as radar and satellite imagery, to provide accurate and timely forecasts, especially for rapidly evolving weather phenomena like thunderstorms.[5]

- **Other Forms:** In addition to the mentioned forms of forecasting, there are specialized techniques such as ensemble forecasting, which involves running multiple forecast models to generate a range of possible outcomes and probabilistic forecasts. Additionally, advancements in machine learning and artificial intelligence have led to the development of innovative approaches for weather prediction, including the integration of deep learning algorithms with traditional forecasting models. These approaches show promising results in improving the accuracy and reliability of weather forecasts, particularly for extreme events like thunderstorms.[2]

2.2. Survey on Different Methods of Thunderstorm Forecasting

This section aims to provide a comprehensive overview of the various methods employed in thunderstorm forecasting, including Numerical Weather Prediction (NWP) and Artificial Intelligence (AI) techniques.

- **Numerical Weather Prediction (NWP):** NWP involves the use of mathematical models and computational algorithms to simulate and predict atmospheric conditions. These models utilize a wide range of meteorological data [11], such as temperature, humidity, wind speed, and pressure, to forecast weather phenomena, including thunderstorms. NWP models vary in complexity and resolution, with some capable of simulating fine-scale atmospheric processes that influence thunderstorm formation and behavior.

- **Artificial Intelligence (AI):** AI techniques, including machine learning and deep learning algorithms, have gained traction in thunderstorm forecasting due to their ability to process large volumes of data and extract complex patterns from atmospheric variables.[3] AI-based models can learn from historical weather data and improve their predictive accuracy over time. These models can be trained to recognize patterns associated with thunderstorm development, thereby enhancing forecast reliability.

- **Ensemble Forecasting:** Ensemble forecasting involves running multiple simulations using variations of initial conditions and model parameters to generate a range of possible weather outcomes. By considering the uncertainty inherent in atmospheric processes, ensemble forecasting provides probabilistic forecasts that account for different scenarios, including the likelihood of thunderstorm occurrence and intensity.[14]

- **Satellite and Radar Imaging:** Remote sensing technologies, such as satellite and radar imagery, play a crucial role in thunderstorm monitoring and forecasting.[10] These imaging techniques provide real-time data on cloud cover, precipitation patterns, and atmospheric dynamics, enabling meteorologists to track the evolution of thunderstorms and issue timely warnings to the public.

- **Data Assimilation Techniques:** Data assimilation involves integrating observational data from various sources, including ground-based weather stations, satellites, and aircraft, into numerical weather models to improve forecast accuracy.[1] By assimilating real-time observations into the model's initial conditions, data assimilation techniques help constrain uncertainties and enhance the fidelity of thunderstorm forecasts.

- **Hybrid Approaches:** Hybrid forecasting approaches combine NWP models with machine learning algorithms to leverage the strengths of both methodologies. By integrating physical principles with data-driven techniques, hybrid models aim to improve the representation of complex atmospheric processes associated with thunderstorm development, leading to more accurate and reliable forecasts.[7]

2.3. Survey on Different AI Models for Thunderstorm Forecasting

- **LSTM (Long Short-Term Memory):** LSTM is a type of recurrent neural network (RNN) architecture known for its ability to capture long-term dependencies in sequential data. LSTM can be used in thunderstorm forecasting by modeling the sequential patterns and dependencies in meteorological data.

- **Transformer:** Transformers are a type of neural network architecture introduced in the field of natural language processing (NLP) for sequence-to-sequence tasks. They have also shown promising results in handling sequential data in other domains, including time series forecasting. Transformers can capture long-range dependencies in input sequences and are capable of parallel processing, making them suitable for thunderstorm forecasting tasks.

- **Regression Models for Thunderstorm Forecasting:** Regression models are employed in thunderstorm forecasting to predict continuous variables, such as rainfall intensity, wind speed, and atmospheric pressure, based on various meteorological parameters. These models analyze historical weather data and other environmental factors to estimate the magnitude of thunderstorm-related phenomena. Regression models can provide valuable insights into the quantitative aspects of thunderstorms, aiding in disaster preparedness and risk assessment.

- **Classification Models for Thunderstorm Forecasting:** Classification models are commonly used in thunderstorm forecasting to categorize different weather patterns based on satellite images. These models aim to classify satellite images into distinct categories, such as clear skies, cloudy weather, thunderstorms, and other meteorological phenomena. Classification models can provide valuable insights into the spatial distribution and intensity of thunderstorms, enabling meteorologists to issue timely warnings and assess potential risks.

2.4. Gap Analysis

I conducted an extensive literature review to gain insights into the current state of Thunderstorm Forecasting research. This involved analyzing various papers and studies focusing on Different Forms of Weather Forecasting, Different Methods of Thunderstorm Forecasting, and Different AI Models for Thunderstorm Forecasting. By delving into the latest methodologies and technologies, I aimed to identify gaps in existing research and explore potential solutions. Through my literature survey, I discovered a common gap across multiple papers: the lack of a dual approach combining satellite imagery with reanalysis datasets for thunderstorm forecasting. While many studies focused on either satellite data or reanalysis datasets separately, none integrated both approaches for a comprehensive analysis. This gap sparked my interest in exploring the potential of a combined approach, leveraging satellite images to validate thunderstorm occurrences alongside reanalysis datasets. By adopting a classification approach in meteorology, I aimed to bridge this gap and provide a more holistic understanding of thunderstorm forecasting. This literature review not only deepened my understanding of the field but also inspired the unique approach taken in this research project. By leveraging insights from existing studies and addressing common gaps, I aimed to contribute to the advancement of Thunderstorm Forecasting techniques.

Tab 2.1: Gap Analysis Table

Title	**Methodology**	**Parameters Used**	**Gap Analysis**
Thunderstorm climatology over the Indian region	Multiscale Analysis of Satellite Images	Radiometric data processing, Clustering algorithms	Lack of comparison with other satellite-based classification techniques [8]
Simulation of location-specific severe thunderstorm events using high-resolution land data assimilation	Data Assimilation Techniques	Land data assimilation, Numerical Weather Prediction	Limited study on uncertainty quantification [12]
Prediction and Classification of Thunderstorms Using Artificial Neural Network	Artificial Neural Networks	Multilayer Perceptron, Backpropagation algorithm	Lack of investigation on ensemble models[3]
A machine learning approach to finding weather regimes and skillful predictor combinations for short-term storm forecasting	Weather Regime Identification	Clustering algorithms, Skillful predictor identification	Lack of exploration on unsupervised learning methods [16]

3. METHODOLOGY

3.1. Data set 1 with explaination

```
        'format': 'netcdf',
        'variable': [
            'geopotential', 'specific_humidity', 'temperature',
            'u_component_of_wind', 'v_component_of_wind',
        ],
        'pressure_level': [
            '100', '125', '150',
            '175', '200', '225',
            '250', '300', '350',
            '400', '450', '500',
            '550', '600', '650',
            '700', '750', '775',
            '800', '825', '850',
            '875', '900', '925',
            '950', '975', '1000',
        ],
        'year': '2015',
        'month': [
            '03', '04', '05',
            '06',
        ],
        'day': [
            '01', '02', '03',
            '04', '05', '06',
            '07', '08', '09',
            '10', '11', '12',
            '13', '14', '15',
            '16', '17', '18',
            '19', '20', '21',
            '22', '23', '24',
            '25', '26', '27',
            '28', '29', '30',
            '31',
        ],
        'time': '00:00',
        'area': [
            23.5, 87.5, 21.5,
            89.5,
```

Fig 3.1: .netcdf FILE of ERA5

The above image is the demonstration of Reanalysis Dataset which provides high-resolution atmospheric data in the netCDF format. The proposed project aims to leverage advanced data science and artificial intelligence techniques for thunderstorm prediction. Imagine ERA5 as a magical book that holds secrets about the weather-details about things like geopotential (imagine it as the height of the air), humidity (how much moisture is in the air), temperature, and winds. These details help us understand how thunderstorms come to life.

3.2. Data set 2 with explaination

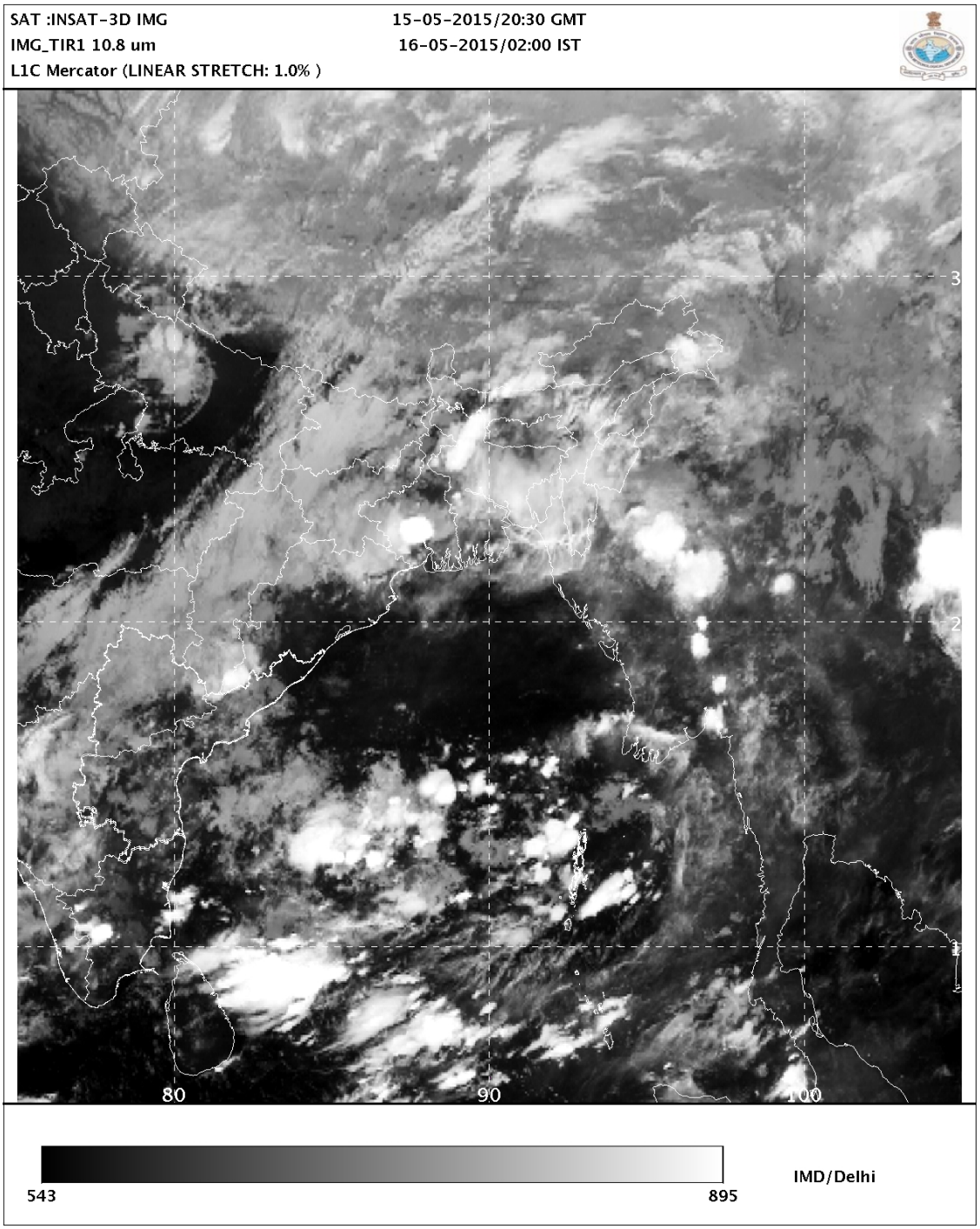


Fig 3.2: IMD Satellite Image example of Thunderstrom Occurrence

The above image is the demonstration of the ORIGINAL Satellite Image which is used by IMD for all types of forecasting . We are studying the Thermal Infrared Images for the region of kolkata on the co-ordinates same as era5 dataset for synchronization of the datasets . These images are used to mark the occurrences of the thunderstoms giving us a chance to validate ground observations and mark the missed ones and solve the drawback . If you focus on kolkata region you can spot a thunderstrom which is what exactly be segmenting and classfiying as 1 or 0.

3.3. Block Diagrams

The Block diagram illustrates the flow of actions within the Streamlit application. The ”User” interacts with the ”Streamlit” interface, where they can ”Enter weather params” within the constraints of a latitudes and longitudes, specifically between our selected region of kolkata, and ”Select Model” (either LSTM or ML or Transformers). The ”Predict Thunderstrom” component receives input from either of these models, allowing the user to predict thunderstrom with their chosen model.

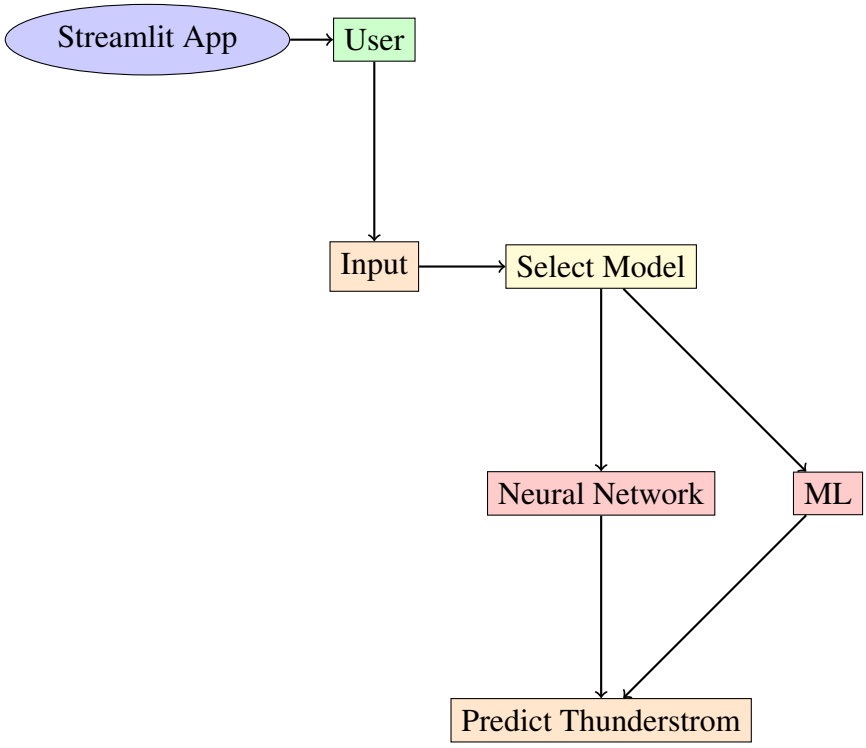


Fig 3.3: Example Deployment Diagram for the Streamlit Application

3.3.1. 0 level DFD

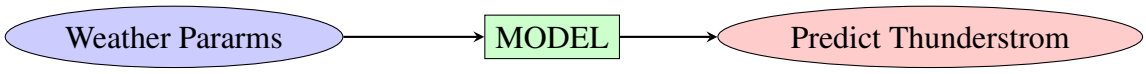


Fig 3.4: 0-level Data Flow Diagram for Thunderstom Prediciton

The 0-level Data Flow Diagram illustrates the core components of the Thunderstrom Predcition process, involving the user-provided weather Parameters, a pre-trained deep learning model , and the classification of the thunderstrom occurrence based on the weather params.

3.3.2. 1st level DFD

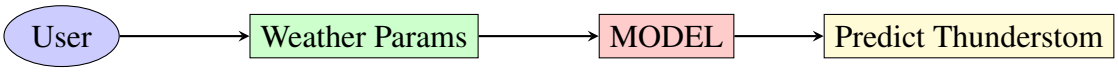


Fig 3.5: 1st-level Data Flow Diagram for Thunderstom Prediciton

This 1st-level Data Flow Diagram illustrates the flow of data in the Thunderstrom Predcition system. The user provides an Weather params, which is processed by the model to classify.

3.3.3. 2nd level DFD

Fig 3.6: 2nd-level Data Flow Diagram for Thunderstom Prediciton

This 2nd-level Data Flow Diagram presents a detailed view of theThunderstrom Predcition process. The User inputs a seed melody, selects a model, classifies, and can choose to print the generated results. Each step in the process is visually distinguished by different colors for clarity and comprehension.

3.4. Flowchart

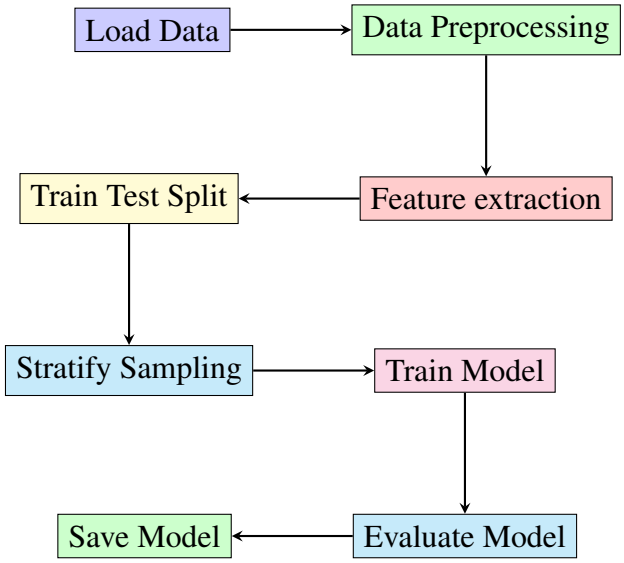


Fig 3.7: Model Life Cycle FLow CHART

- **Load Data:** Read the dataset into a DataFrame from a CSV file and opencv for images.
- **Data Preprocessing:** Clean and prepare the data by handling missing values, encoding categorical features, and normalizing numerical features.

- **Feature Extraction:** Select or engineer relevant features to improve model performance.
- **Train Model:** Fit a AI based model to the training data.
- **Evaluate Model::** Assess the model's performance using appropriate metrics on the validation or test data.
- **Save Model:** Persist the trained model to disk for future use.

3.5. Pseudo Code

Reanalysis Dataset Training **Input:** Normalized dataset, Feature Engineered, Training Weather parameter, model architecture, number of epochs, batch size. **Output:** Trained machine learning + neural network models.
1. Load the Normalzsed dataset, including all the features. 2. Feature extraction by random forest and LIME(XAI). 3. Split data into train test validation sets . 4. Build ML + neural network models, for instance, machine learning classifiers + LSTM , Transformer , MLP models with specified output units, hidden layers, loss function, and learning rate. 5. Train the model on the training data with the chosen number of epochs and batch size. 6. Save the trained model for future prediciton.

Satellite Image Dataset Training **Input:** Normalized dataset, Training Thunderstrom occurrence classes, model architecture, number of epochs, batch size. **Output:** Trained neural network model for visual thunderstrom occurrence classification
1. Load the dataset, including all the sperate classes. 2. Convert satellite images into gray scale , extract ROI , Apply binary threshold as preprocessing . 3. Split data into train test validation sets. 4. Build CNN model with specified output units, hidden layers, loss function, and learning rate. 5. Train the model on the training data with the chosen number of epochs and batch size. 6. Save the trained model for future segmentation of thunderstrom occurrence. 7. Run model on all satellite images and mark the date of occurrence and validate with ground observatories .

3.6. Performance Metrics

In this section, we present the performance metrics used to evaluate the thunderstorm forecasting models. These metrics provide different perspectives on the model's classification performance.

3.6.1. Classic Binary Classification Metrics

- **Accuracy**: Accuracy measures the overall correctness of the classification model.

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

- **Precision**: Precision measures the proportion of correctly predicted positive instances out of the total instances predicted as positive.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Recall (Sensitivity or True Positive Rate)**: Recall measures the proportion of correctly predicted positive instances out of the total actual positive instances.

$$\text{Recall} = \frac{TP}{TP + FN}$$

- **F1 Score**: The F1 score is the harmonic mean of precision and recall, providing a balanced measure between the two.

$$\text{F1 Score} = \frac{2 \cdot (\text{Precision} \cdot \text{Recall})}{\text{Precision} + \text{Recall}}$$

3.6.2. Additional Classification Metrics used in Weather Forecasting Domain :

- **Probability of Detection (POD)**: POD, also known as Sensitivity or True Positive Rate, measures the proportion of actual positive instances correctly predicted as positive.

$$\text{POD} = \frac{TP}{TP + FN}$$

- **False Alarm Rate (FAR)**: FAR measures the proportion of actual negative instances incorrectly predicted as positive.

$$\text{FAR} = \frac{FP}{FP + TN}$$

- **Heidke Skill Score (HSS)**: HSS evaluates the skill of a classification model compared to random chance. It considers the improvement of the model over the random model.

$$\text{HSS} = \frac{2 \cdot (TP \cdot TN - FP \cdot FN)}{(TP + FN) \cdot (FN + TN) + (TP + FP) \cdot (FP + TN)}$$

- **Critical Success index (CSI)**: CSI, also known as Threat Score or Gilbert Skill Score, measures the proportion of correctly predicted events (both positive and negative) out of the total events.

$$\text{CSI} = \frac{TP}{TP + FP + FN}$$

These metrics provide valuable insights into the performance of our thunderstorm forecasting models. We will now discuss and interpret the results in the following section.

4. PERFORMANCE ANALYSIS

Following are the TESTING results after model training

Machine Learning Classifiers

Classic Machine Learning Evaluation Testing Metrics :

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
KNeighborsClassifier	91.33	91.29	91.33	91.30
DecisionTreeClassifier	93.83	93.82	93.83	93.83
RandomForestClassifier	97.65	97.66	97.65	97.64
ExtraTreesClassifier	98.24	98.24	98.24	98.23
XGBClassifier	84.10	84.07	84.10	83.41

Tab 4.1: Classic ML Metrics in Percentage

Weather forecasting Evaluation Testing Metric :

Model	FAR (%)	POD (%)	CSI (%)	HSS (%)
KNeighborsClassifier	5.89	85.50	91.33	82.65
DecisionTreeClassifier	4.32	89.95	93.83	87.66
RandomForestClassifier	0.97	94.76	97.65	95.31
ExtraTreesClassifier	0.73	96.08	98.24	96.47
XGBClassifier	5.77	62.84	84.10	68.20

Tab 4.2: Weather Forecasting Metrics in Percentage

In our comprehensive evaluation of machine learning classifier models, we assessed a diverse range including Logistic Regression, Naive Bayes variants (Bernoulli, Categorical, Complement, Gaussian, and Multinomial), K-Nearest Neighbors, Decision Tree, Extra Tree, Random Forest, Extra Trees, XGBoost, LightGBM, and CatBoost. Among these, several models demonstrated remarkable performance, surpassing others in various metrics. Notably, **KNeighborsClassifier, DecisionTreeClassifier, RandomForestClassifier, ExtraTreesClassifier, and XGBClassifier** emerged as top performers, showcasing superior accuracy, precision, recall, and F1 scores. What's particularly intriguing is that these models not only excelled within their class but also outperformed neural network models, as evidenced in the subsequent section for comparison. This underscores the importance of exploring and leveraging diverse machine learning techniques to achieve optimal performance in various applications.

Deep Learning Models Testing Metrics

Classic Machine Learning Evaluation Testing Metrics :

Tab 4.3: Evaluation Metrics for Transformers, MLP, and LSTM

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Transformers	72.83	65.69	30.39	41.56
MLP	68.27	52.71	12.68	20.44
LSTM	70.19	61.11	17.06	26.67

Weather forecasting Evaluation Testing Metric :

Tab 4.4: Weather Forecasting Metrics for Transformers, MLP, and LSTM

Algorithm	POD (%)	FAR (%)	CSI (%)	HSS (%)
Transformers	60.39	40.95	72.83	45.66
MLP	26.68	72.388	11.38	28.11
LSTM	37.06	50.58	70.19	40.37

Table 4.3 and Table 4.4 show the evaluation of Neural network on our data . The results are not outperforming the machine learning model due to their complex architecture . Where our model required simple statistical approach.Sometimes we can achieve many things by applying simple approachs which is exactly what is demonstrated here .

Feature Importance

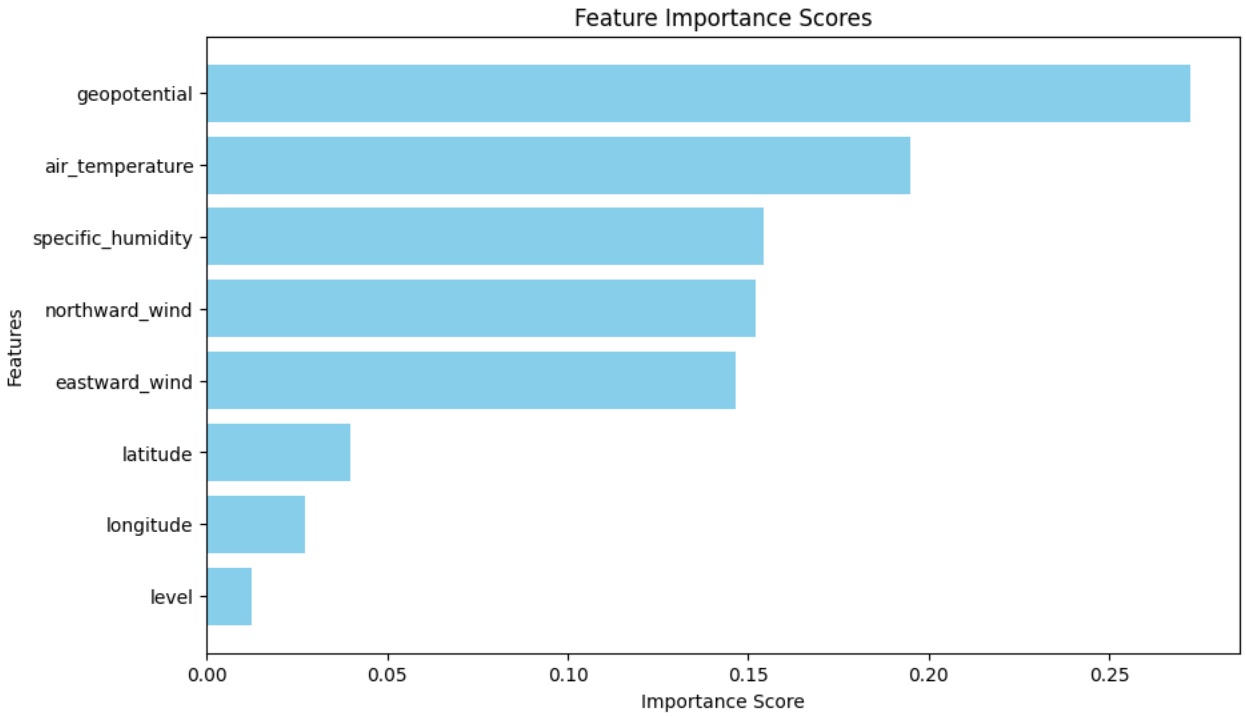


Fig 4.1: Feature Importance plot using Random Forest

Graphs

Line Chart

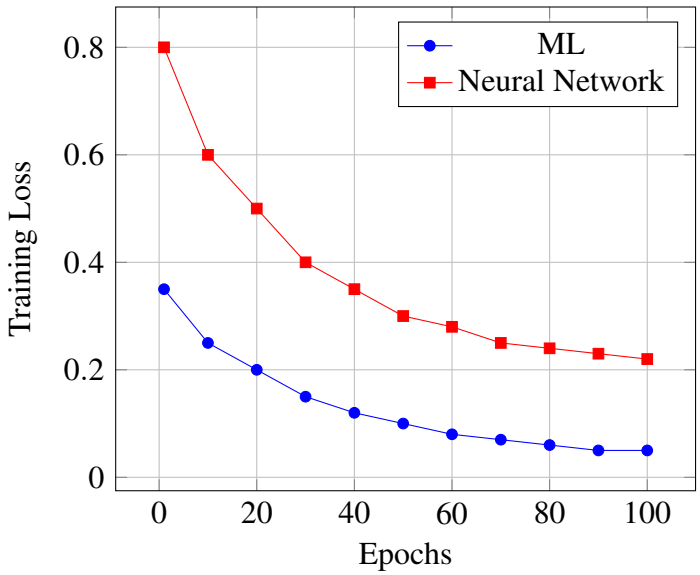
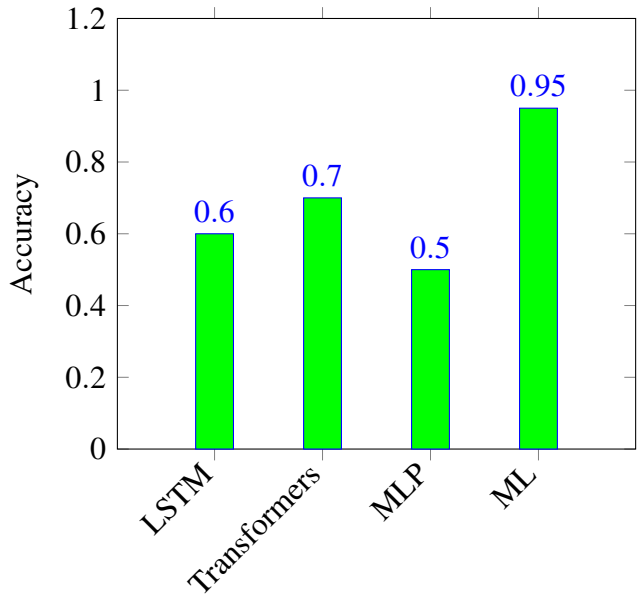


Fig 4.2: Training Loss Comparison between ML and Neural Networks

The graph is self explanatory where it demonstrates that ML have performed better than Neural Networks in terms of loss reduction

Bar Plot



The Bar plot illustrates the comparitive accuracy across different models . As we can see simple ML classifier has outperformed complex neural network models

Confusion Matrix

ERA5 DATA SET

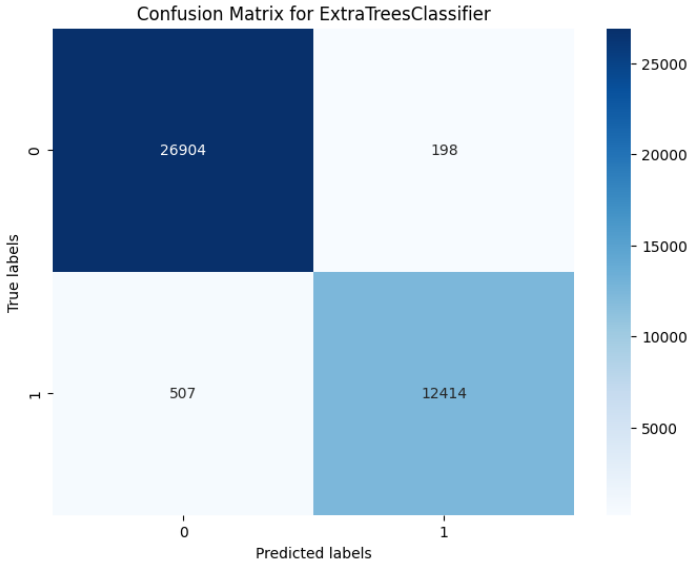


Fig 4.3: Extra Trees Classifier Best Performing ML Confusion Matrix

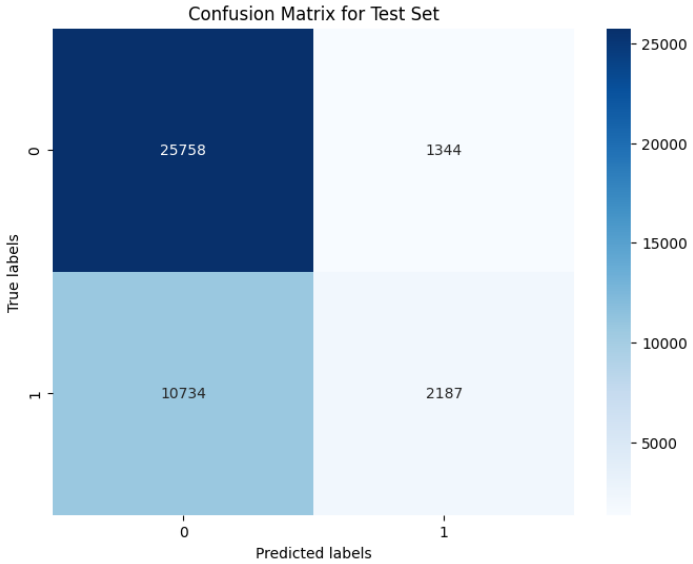


Fig 4.4: Transformers Confusion Matrix

Satellite Image DATA SET

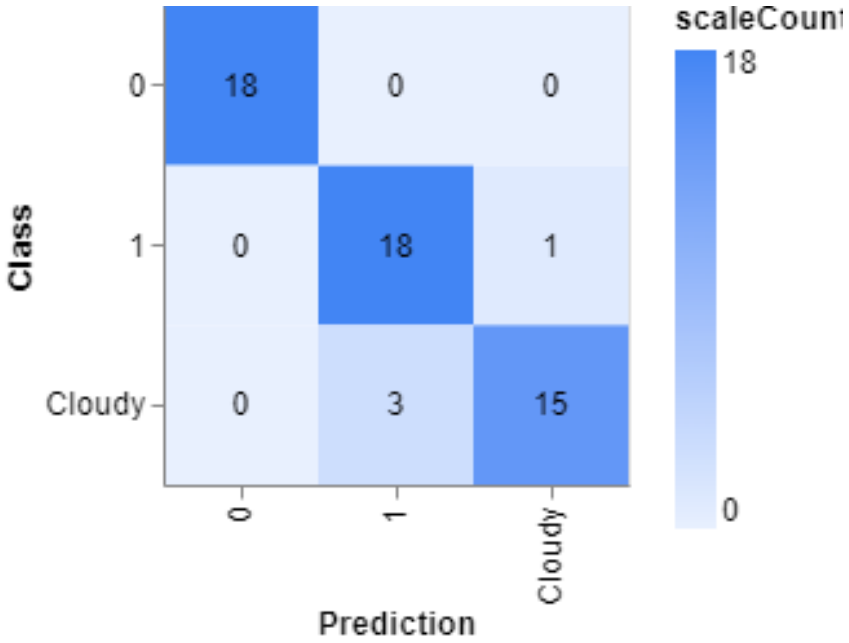


Fig 4.5: Satellite Image Training Confusion Matrix

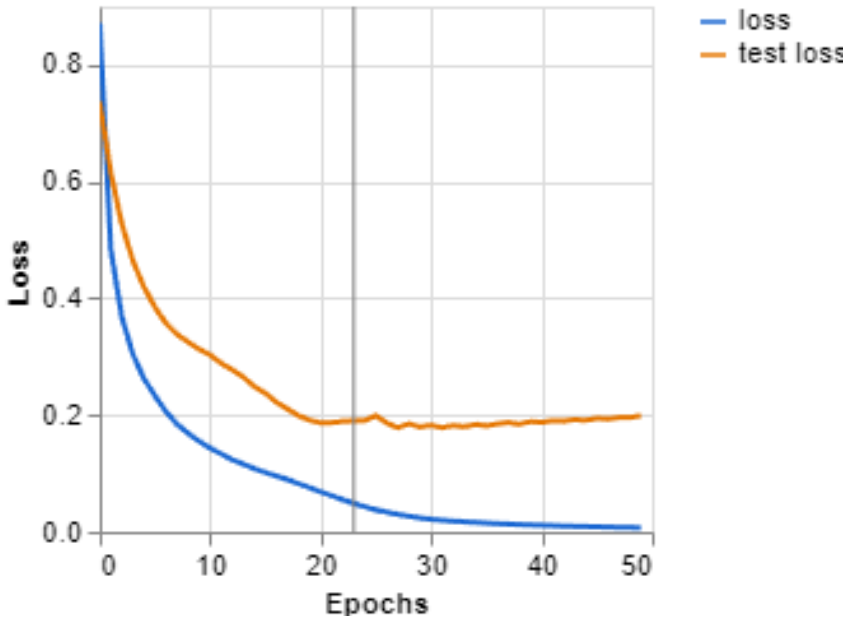


Fig 4.6: Satellite Image Training Loss and accuracy

5. CONCLUSION

5.1. Conclusion

In this study, we advanced our research by integrating IMD satellite data, ERA5 reanalysis, and ground observatory data to validate thunderstorm occurrences in Kolkata during the pre-monsoon period of 2015. Various machine learning classifiers outperformed neural network models in predicting thunderstorm events, achieving high accuracy. Feature engineering using Random Forest and LIME was employed. Future work will focus on regression to predict weather parameters for the next day.

5.2. Advantages and Strengths of Method

The strengths and advantages of the methods used in this project contribute to its effectiveness in delivering valuable insights for thunderstorm forecasting with AI. Some of the major advantages are as follows:

Strengths:

1. The dataset used is highly reliable, ensuring that the primary focus can be on training a good model without concerns about data quality.

2. The integration of satellite and ground observatory data with ERA5 reanalysis dataset provides a robust framework for accurate thunderstorm prediction.

3. The use of LIME for feature engineering enhances the interpretability of the models, making it easier to understand the impact of different features.

4. The superior performance of various machine learning classifiers demonstrates the versatility and robustness of our approach.

Weaknesses:

1. These models can be complex and tricky for users to interpret during preprocessing, potentially leading to suboptimal results.

2. Training sophisticated models like neural networks and ensemble classifiers require significant computational power and time.

3. The model's accuracy is highly dependent on the quality and comprehensiveness of the input data, necessitating continuous updates and validation.

5.3. Applications and Future Scope

The methods and models developed in this project have various practical applications:

- **Weather forecasting:** Enhancing the accuracy of thunderstorm predictions to improve early warning systems and reduce weather-related hazards.
- **Agriculture:** Providing precise weather forecasts to help farmers make informed decisions about planting, irrigation, and harvesting schedules.
- **Aviation:** Improving flight safety by offering better predictions of severe weather conditions that could impact flight operations.
- **Disaster management:** Assisting in the preparation and response strategies for natural disasters, potentially saving lives and reducing economic losses.
- **Climate research:** Contributing to the understanding of climate patterns and the impact of various meteorological parameters on severe weather events.

References

[1] Colby FP Jr (1984) Convective inhibition as a predictor of convection during AVESESAME II. Mon Weather Rev 112(11):2239–2252.

[2] George JJ (1960) Weather forecasting for aeronautics. Academic press, https://doi.org/10.1016/C2013-0-12567-6 (ISBN: 978–1–4832–3320–8)

[3] India Meteorological Department Standard Operation Procedure for Numerical Weather Prediction and Forecast Verification (2021) p 14. https://mausam.imd.gov.in/imd$_latest/contents/pdf/nwp_sop.pdf$

[4] Means LL (1952) Stability index computation graph for surface data. Unpublished manuscript available from F. Sanders, 9.

[5] Miller RC (1967) Notes on analysis and severe storm forecasting procedures of the Military Weather Warning Centre. AWS Tech Rep 200, USAF p170.

[6] Sahu RK, Dadich J, Tyagi B, Vissa NK (2020) Trends of thermodynamic indices thresholds over two tropical stations of north-east India during pre-monsoon thunderstorms, Journal of Atmospheric and Solar-Terrestrial Physics, 211, 105472, ISSN 1364-6826, https://doi.org/10.1016/j.jastp.2020.105472.

[7] Showalter AK (1953) A convective index as an indicator of cumulonimbus development. J Appl Meteorol 5:839–846.

[8] Tyagi A (2007) Thunderstorm climatology over the Indian region. Mausam 58(2):189–212.

[9] Williams E, Renno N (1993) An analysis of the conditional instability of the tropical atmosphere. Mon Weather Rev 121(1):21–36. https://doi.org/10.1175/1520-0493(1993)121%3c0021:AAOTCI%3e2.0.CO;2

[10] Kinza Amjad[1], Mubasher H. Malik[2*], Hamid Ghous[3], Aftab Hussain[4], Maryem Ismail[5]
1,4-5Institute of Southern Punjab, Multan, Pakistan
2Vision, Linguistics & Machine Intelligence Research Lab, Multan, Pakistan
3Australian Scientific & Engineering Solutions, Sydney, New South Wales, Australia
(2022).

[11] A. Sisodiya, S. Pattnaik, H. Baisya, G. S. Bhat, A. G. Turner
(2019). "Simulation of location-specific severe thunderstorm events using high-resolution land data assimilation." Dynamics of Atmospheres and Oceans, 87, 101098. doi: `https://doi.org/10.1016/j.dynatmoce.2019.101098`

[12] S. Mahesh Anad, Ansupa Dash, M. S. Jagadeesh Kumar, Amit Kesarkar
Vellore Institute of Technology University, Vellore, India; National Atmospheric Research Laboratory, Gadanki, India
(2023).

[13] Unashish Mondal, Anish Kumar, S. K. Panda, Devesh Sharma, Someshwar Das
(2021). "Comprehensive study of thunderstorm indices threshold favorable for thunderstorms during monsoon season using WRF–ARW model and ERA5 over India."

[14] Bartosz Czernecki, Mateusz Taszarek, Michał Marosz, Marek Półrolniczak, Leszek Kolendowicz, Andrzej Wyszogrodzki, Jan Szturc
(2020). "Application of machine learning to large hail prediction - The importance of radar reflectivity, lightning occurrence and convective parameters derived from ERA5."

[15] Kuai Dai, Xutao Li, Junying Fang, Yunming Ye, Demin Yu, Di Xian, Danyu Qin
(2022). "Four-hour thunderstorm nowcasting using deep diffusion models of satellite."

[16] John K. Williams, D. A. Ahijevych, C. J. Kessinger, T. R. Saxen, M. Steiner, S. Dettling
National Center for Atmospheric Research, Boulder, Colorado
(2021). "A machine learning approach to finding weather regimes and skillful predictor combinations for short-term storm forecasting."

[17] Liu Na, Xiong Anyuan, Zhang Qiang, Liu Yujia, Zhan Yunjian, Liu Yiming
(2024). "Development of Basic Dataset of Severe Convective Weather for Artificial Intelligence Training." DOI: https://doi.org/10.11898/1001-7313.20210502

[18] Agostino Manzato
(2025). "Sounding-derived indices for neural network based short-term thunderstorm and rainfall forecasts."

[19] Donald W. McCann
(1992). "A Neural Network Short-Term Forecast of Significant Thunderstorms." DOI: https://doi.org/10.1175/1520-0434(1992)007<0525:ANNSTF>2.0.CO;2