# Understanding Climate Variability in North India: A Machine Learning Perspective

**Presented By**


**Aayushi Tandon**

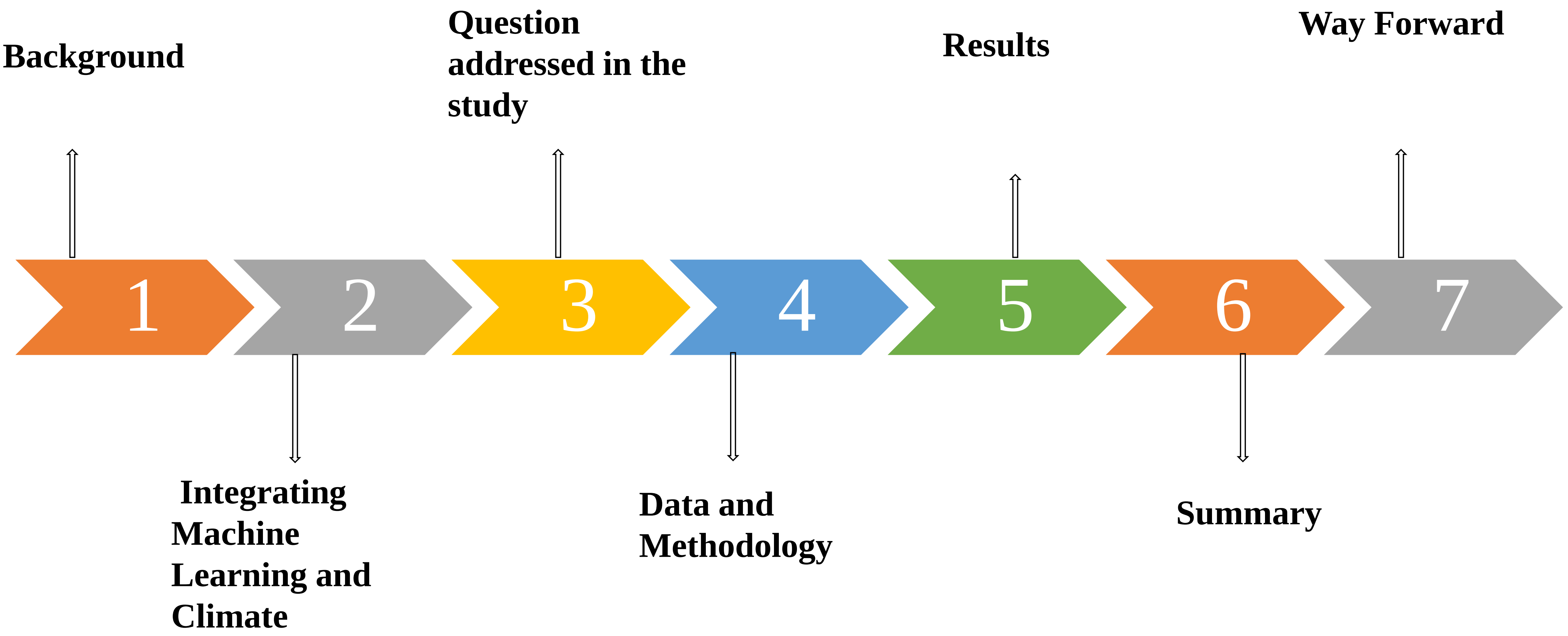
*Research Scholar*
*University of Petroleum and Energy Studies*
*Dehradun, India*

*Contributing Author*
**Dr. Amit Awasthi & Dr. Kanhu Charan Pattnayak**



*National Symposium "Understanding the science of heatwaves under the warming scenario and challenges ahead, (18th- 19th March 2024)*

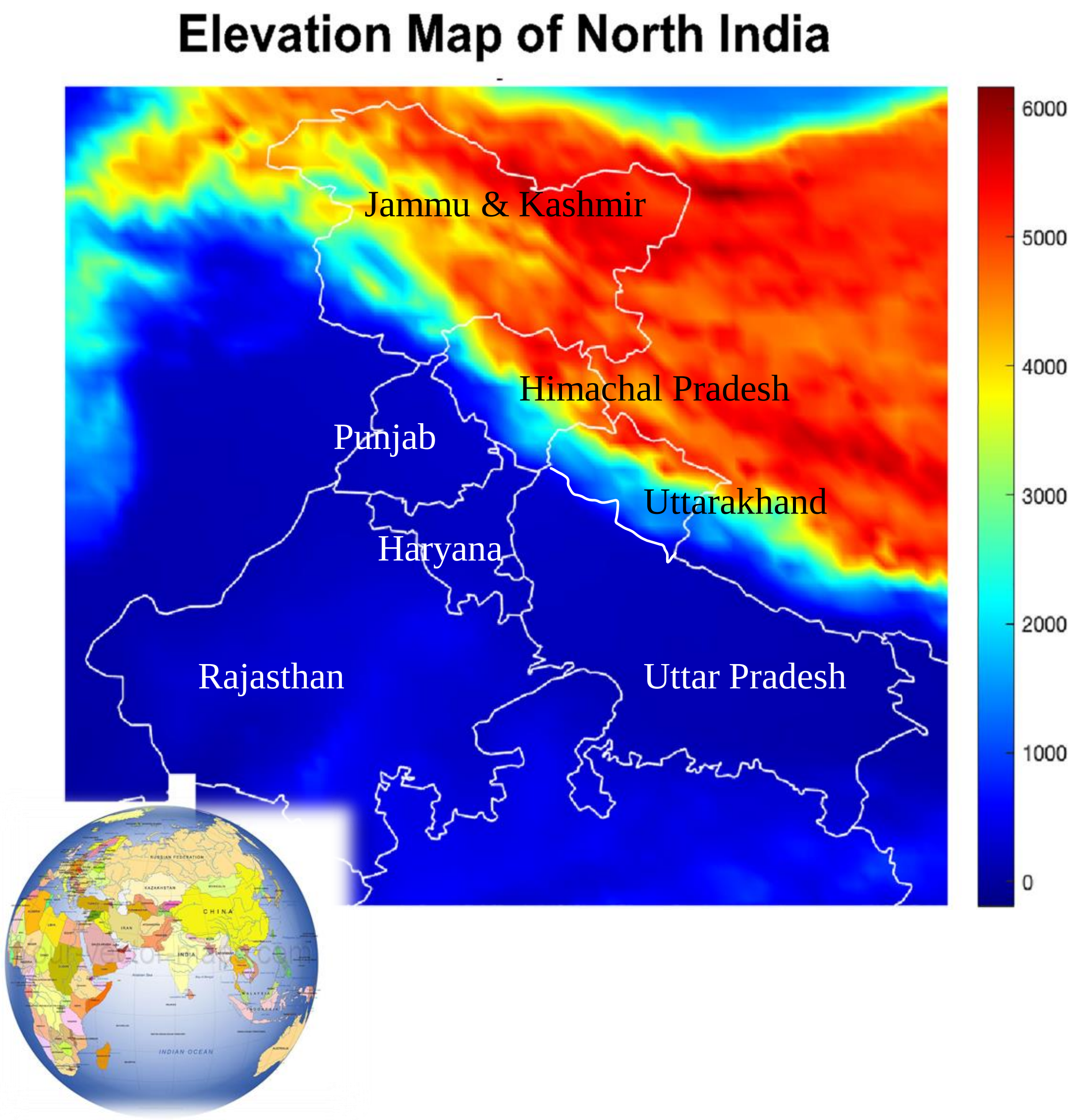
# Outline



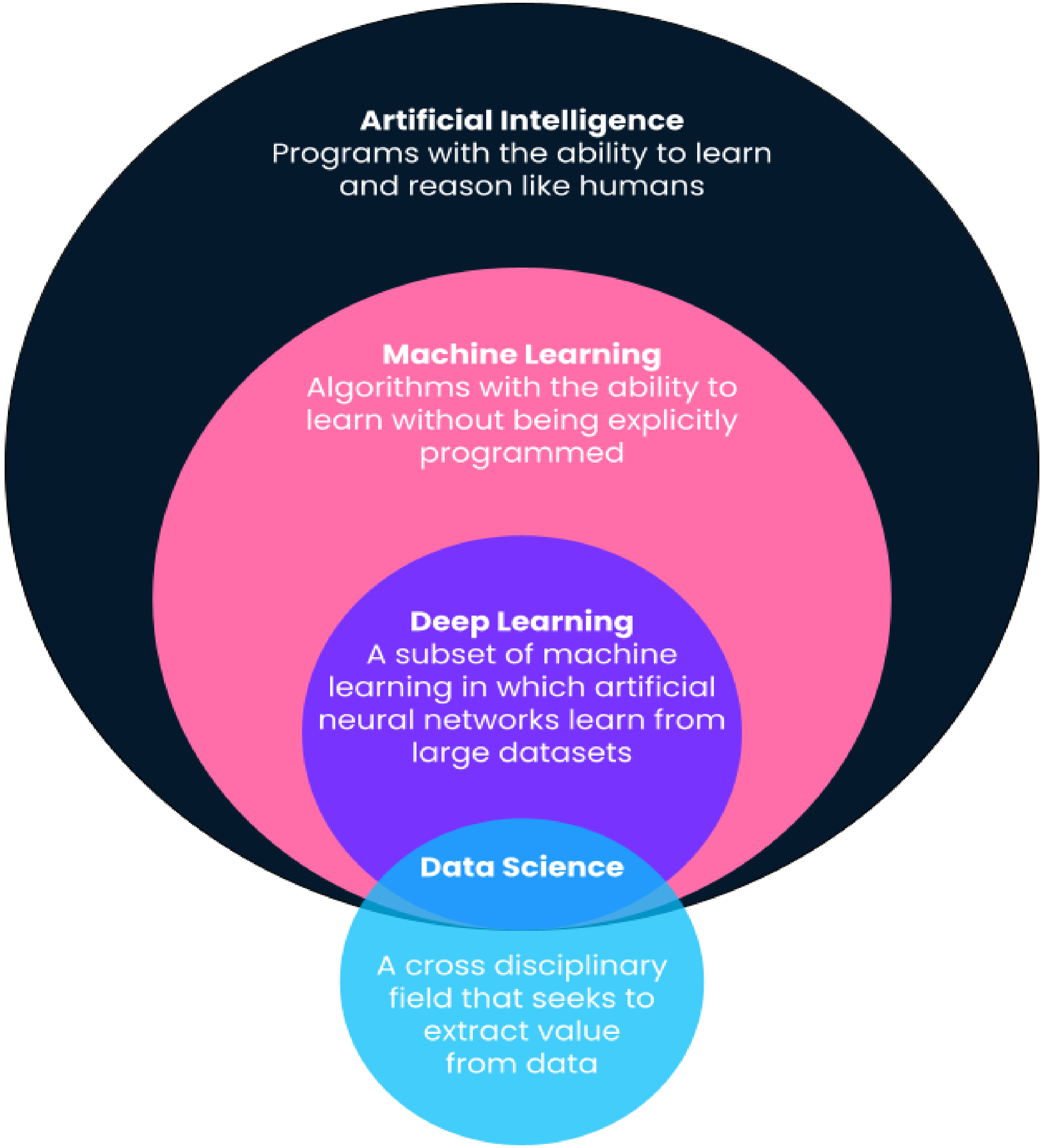
1
2
3
4
5
6
7
Background
Integrating Machine Learning and Climate
Question addressed in the study
Data and Methodology
Results
Summary
Way Forward

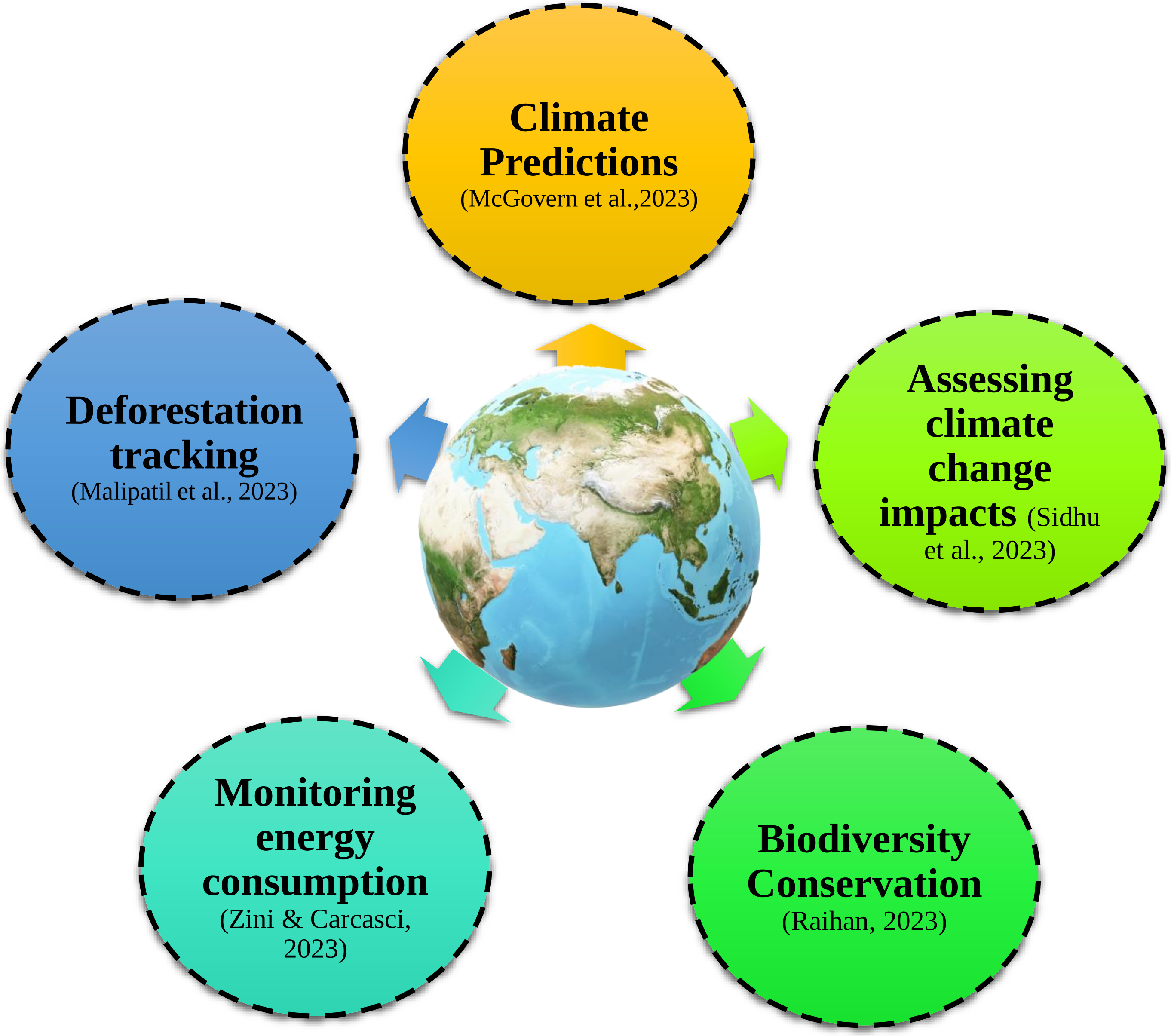
# 1. Background

- The geographical features and elevation of North India pose challenges for scientific exploration.
- Climate change is altering the variability and intensity of extreme weather events (IPCC, 2012).
- While temperature extremes and precipitation extremes have been studied separately, their inter-relationship has been relatively understudied in the literature ((Awasthi et al., 2023; Pattnayak et al., 2023)
- Artificial intelligence (AI) and machine learning (ML) can be harnessed to combat climate change in India (Srivastava & Maity, 2023).
- ML offers promising avenues to analyze climate data and predict future trends, enabling India to better prepare for climate change (Sahil et al., 2023).



# 2. Integrating Machine Learning and Climate



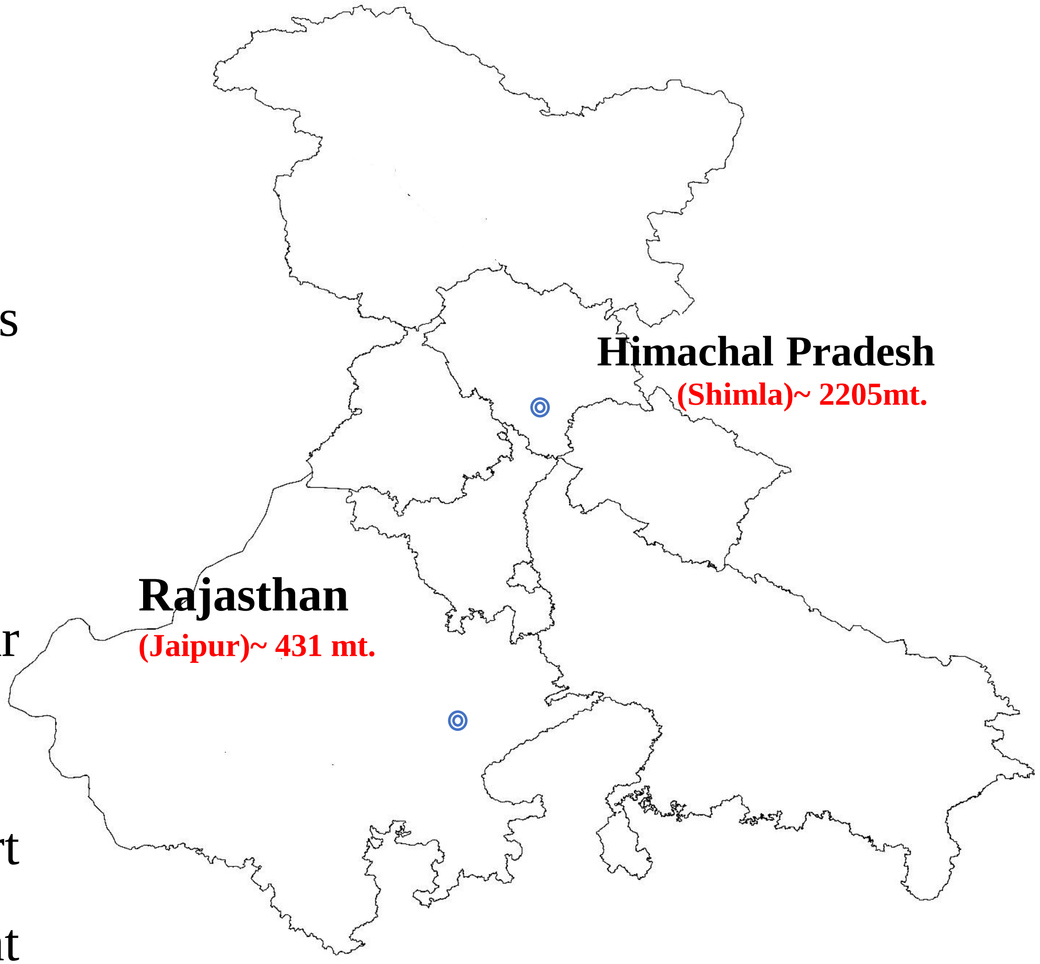



# 3. Questions addressed in this study

- Is machine learning adept at identifying relationships among diverse atmospheric variables?
- What is the achievable predictive accuracy of machine learning algorithms in climate predictions?
- Do different machine learning algorithms produce consistent results, and if not, what factors contribute to variability?
- How reliable and robust is the skill score of the model?
- Can this methodology be applied to predict extreme climatic events?

# 4. Data and Methodology

- Dataset: MERRA 2
- Study Domain: Rajasthan and Himachal Pradesh
- Time span: 1984-2023
- Spatial resolution: 0.5 degrees longitude by 0.6 degrees latitude
- Temporal resolution: Daily
- Parameters: Temperature, Precipitation, Wind speed, Solar irradiance, Surface pressure, Dew point
- Machine Learning Methods- Random Forest, Support Vector Machine, K Nearest Neighbour, Xtreme Gradient Boost.

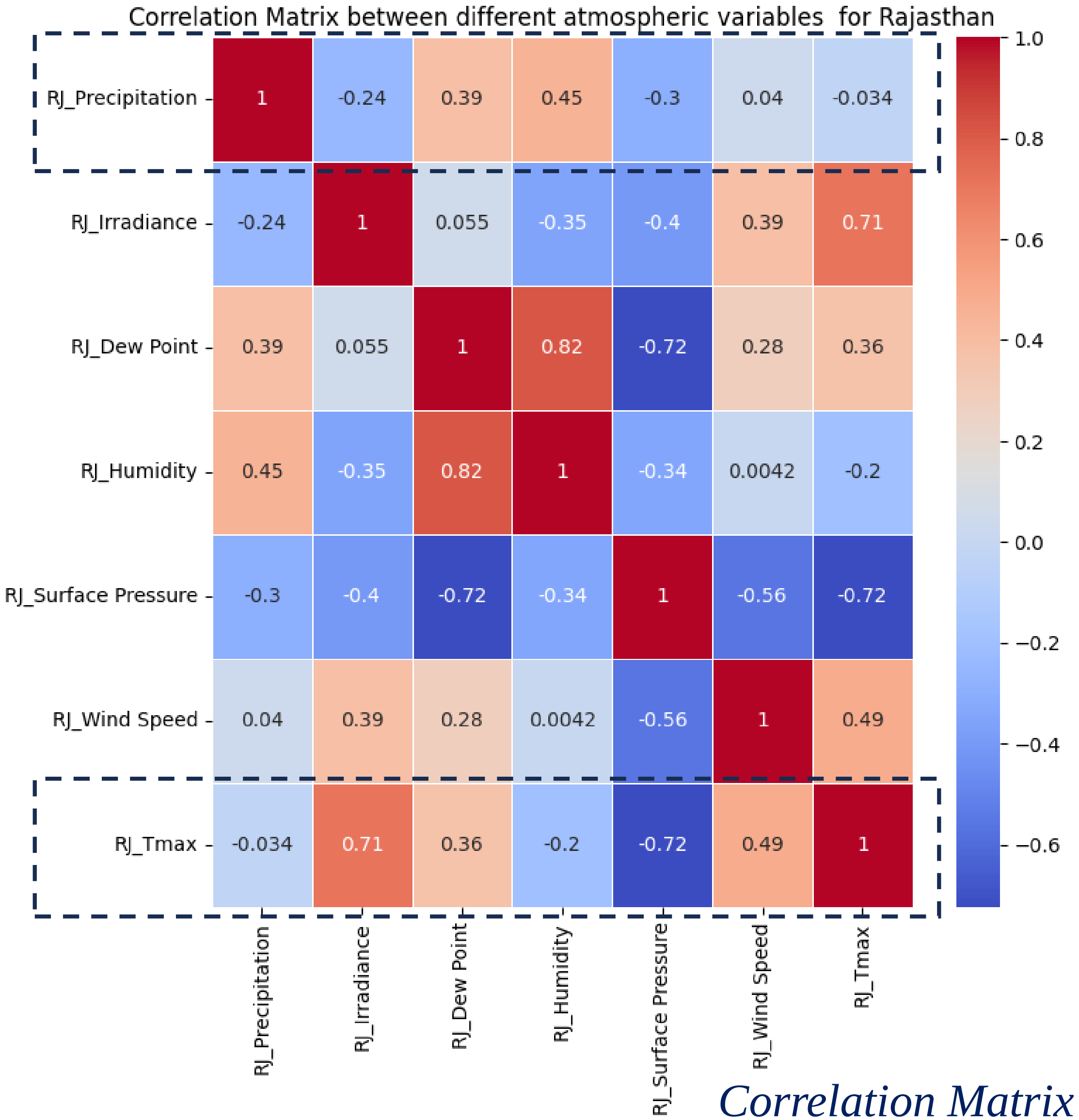



*Study Area*

# 5. Results

## Correlations between different atmospheric variables

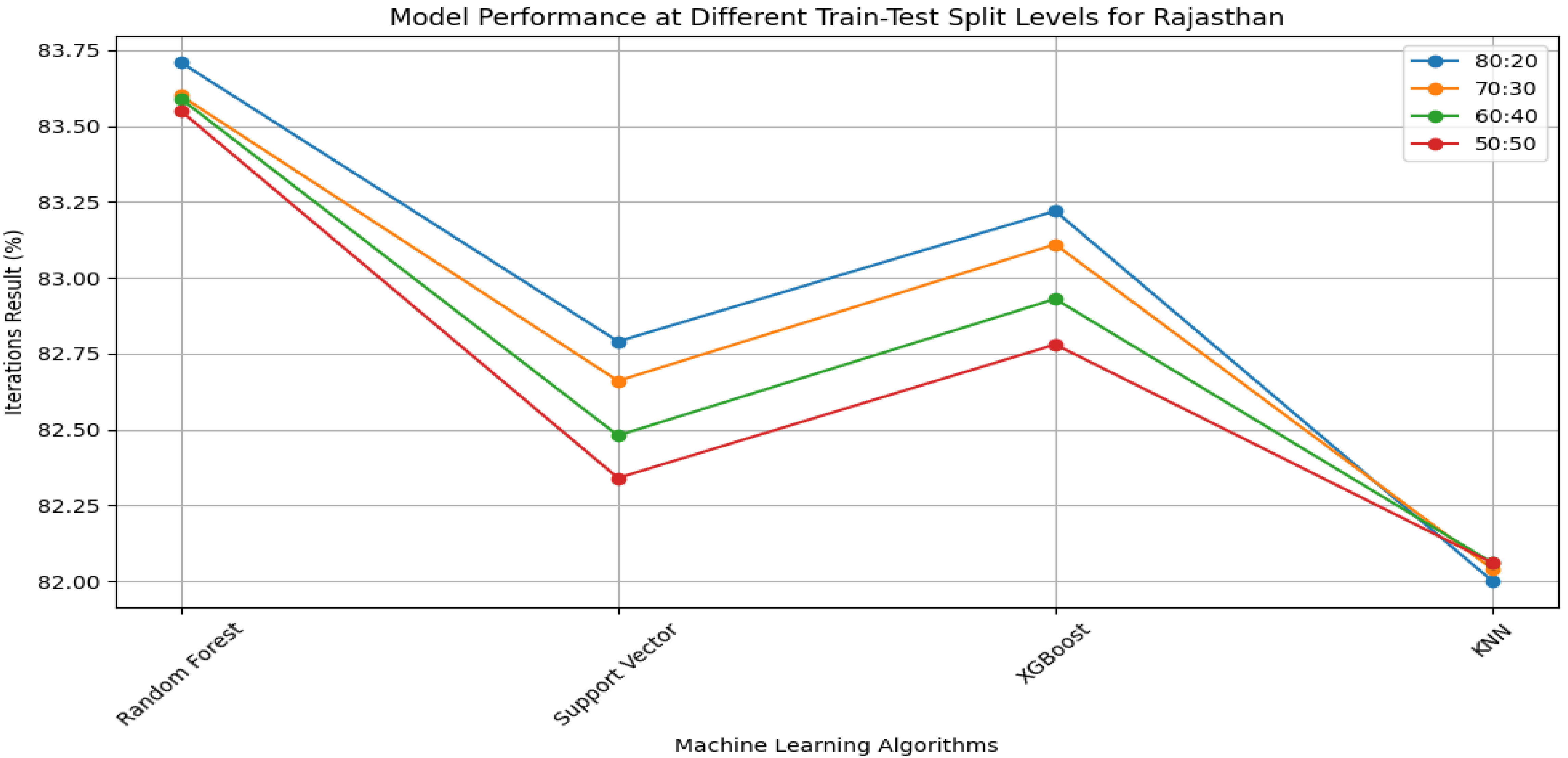
| Parameter | Precipitation | Temperature |
|---|---|---|
| Solar Irradiance | Weak Negative (-0.24) | Strong positive (0.71) |
| Dew Point | Strong positive (0.39) | Strong Positive (0.36) |
| Humidity | Strong positive (0.45) | Weak negative (-0.2) |
| Surface Pressure | Moderate negative (-0.3) | Strong negative (-0.72) |
| Wind Speed | Weak positive (0.04) | Strong Positive (0.49) |
| Precipitation | - | Weak negative (-0.034) |
| Temperature | Weak Negative (-0.034) | - |



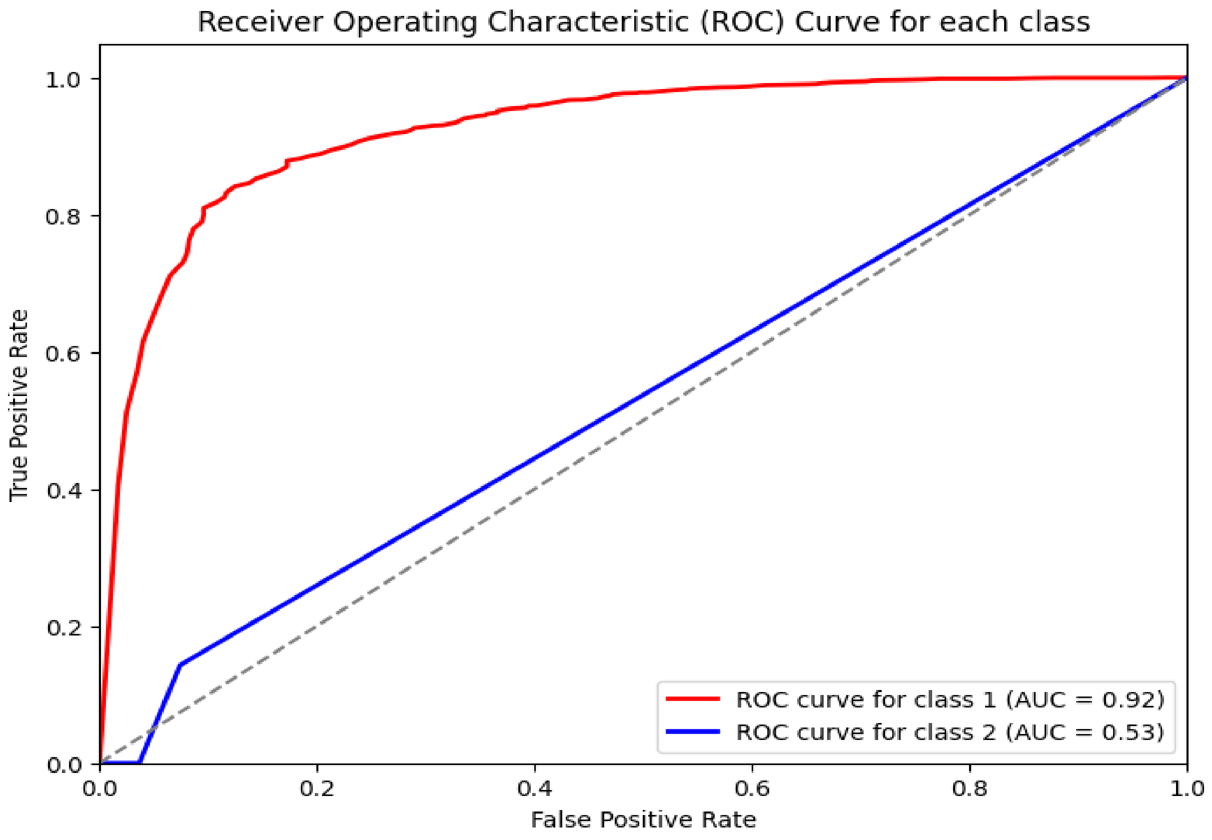

*Correlation Matrix*

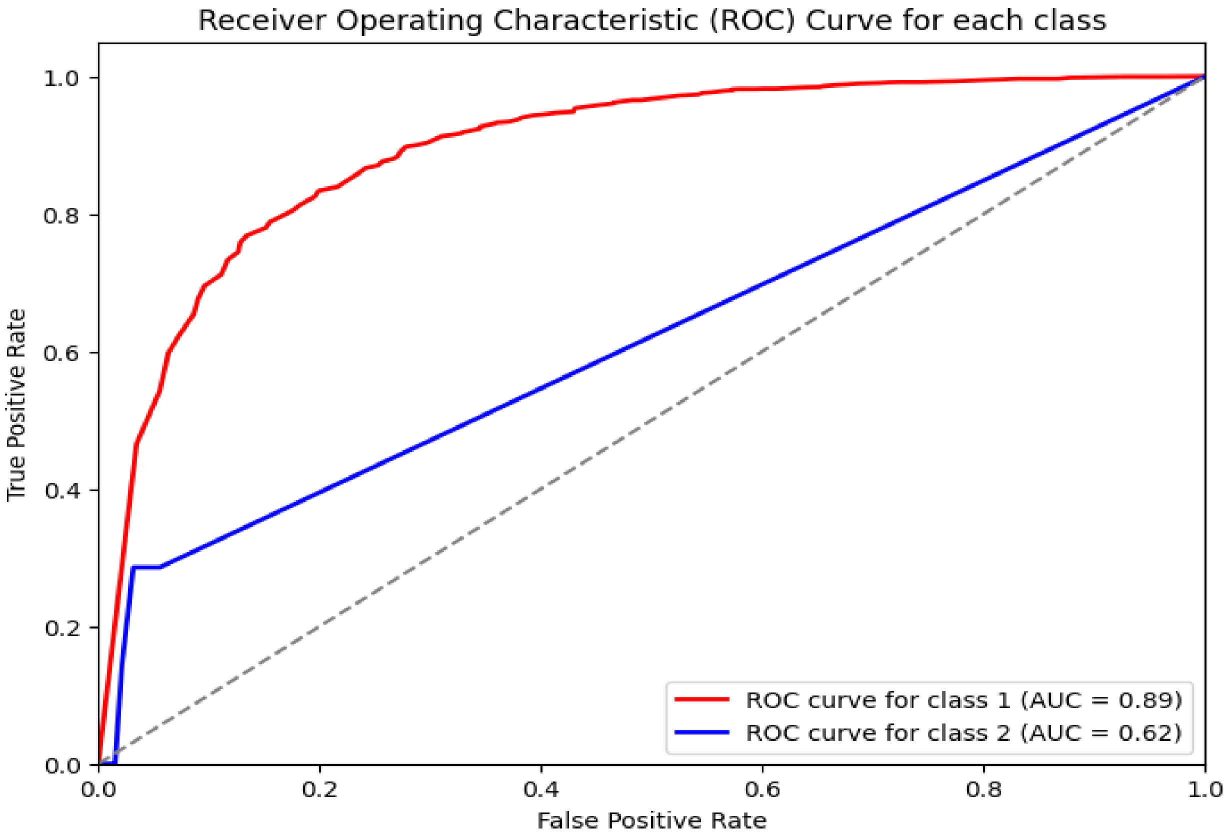
Continued…

# Comparison of Accuracy Scores Across Various Machine Learning Algorithms at Different Train-Test Split Levels



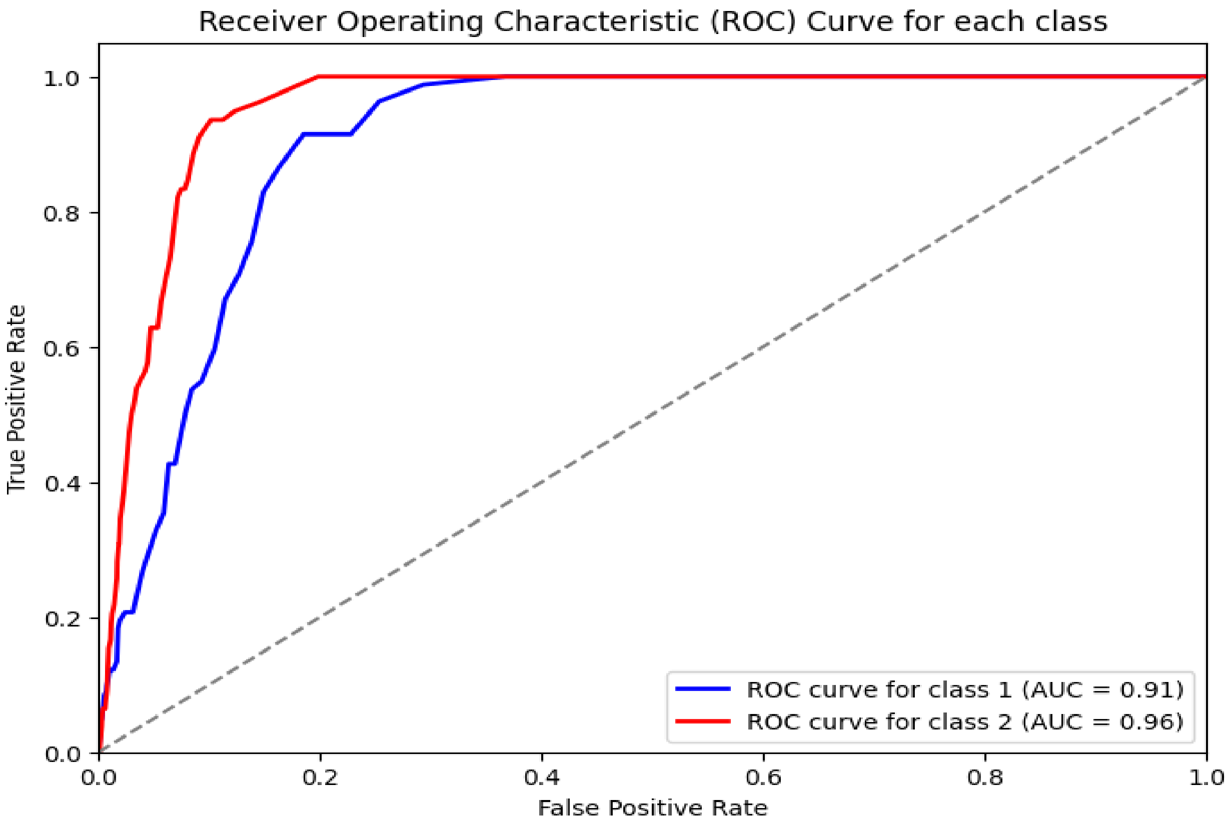
# Precipitation Extremes

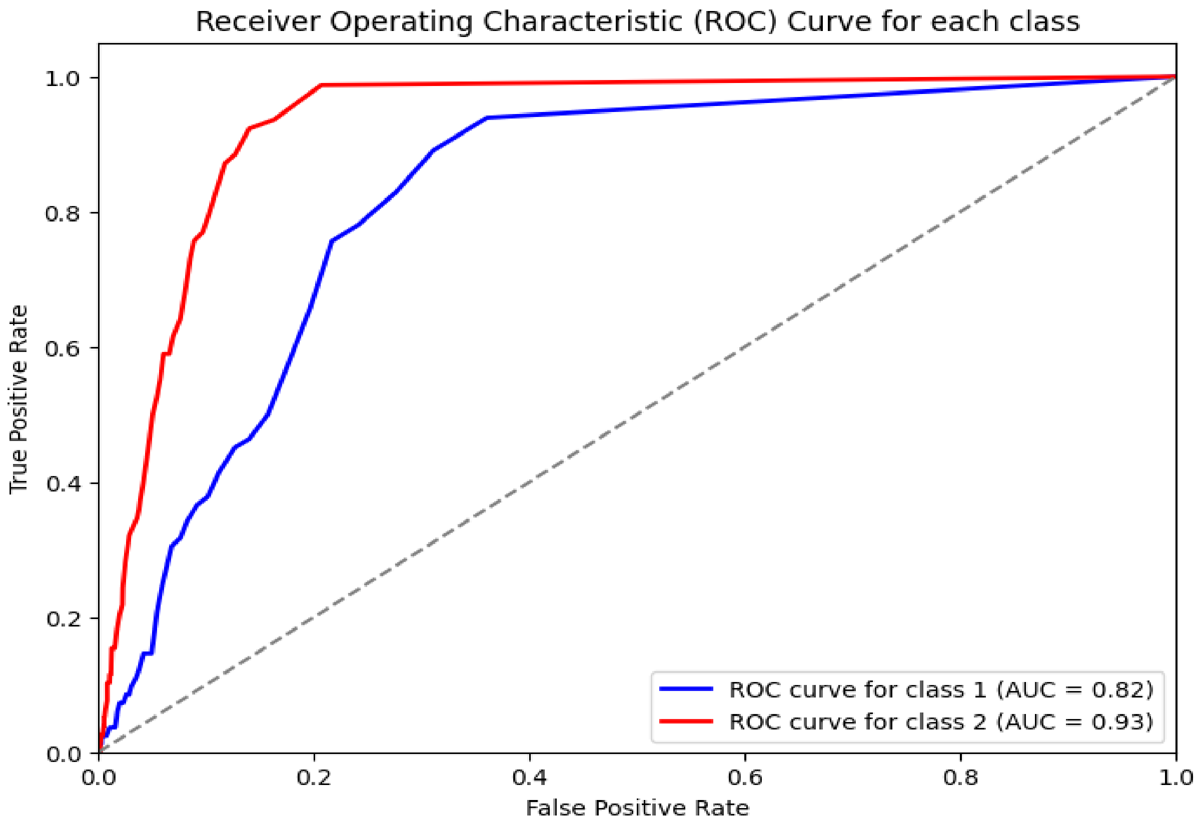



- Thresholds for extreme rain events: 10th percentile approximately 0.30 mm, 95th percentile approximately 18mm.
- High performance of the machine learning model for events below the 10th percentile .
- Model performance is poor for extreme events, specifically those exceeding the 95th percentile.
- Enhanced model performance by focusing on statistically significant variables for extreme rain events, particularly those above the 95th percentile.

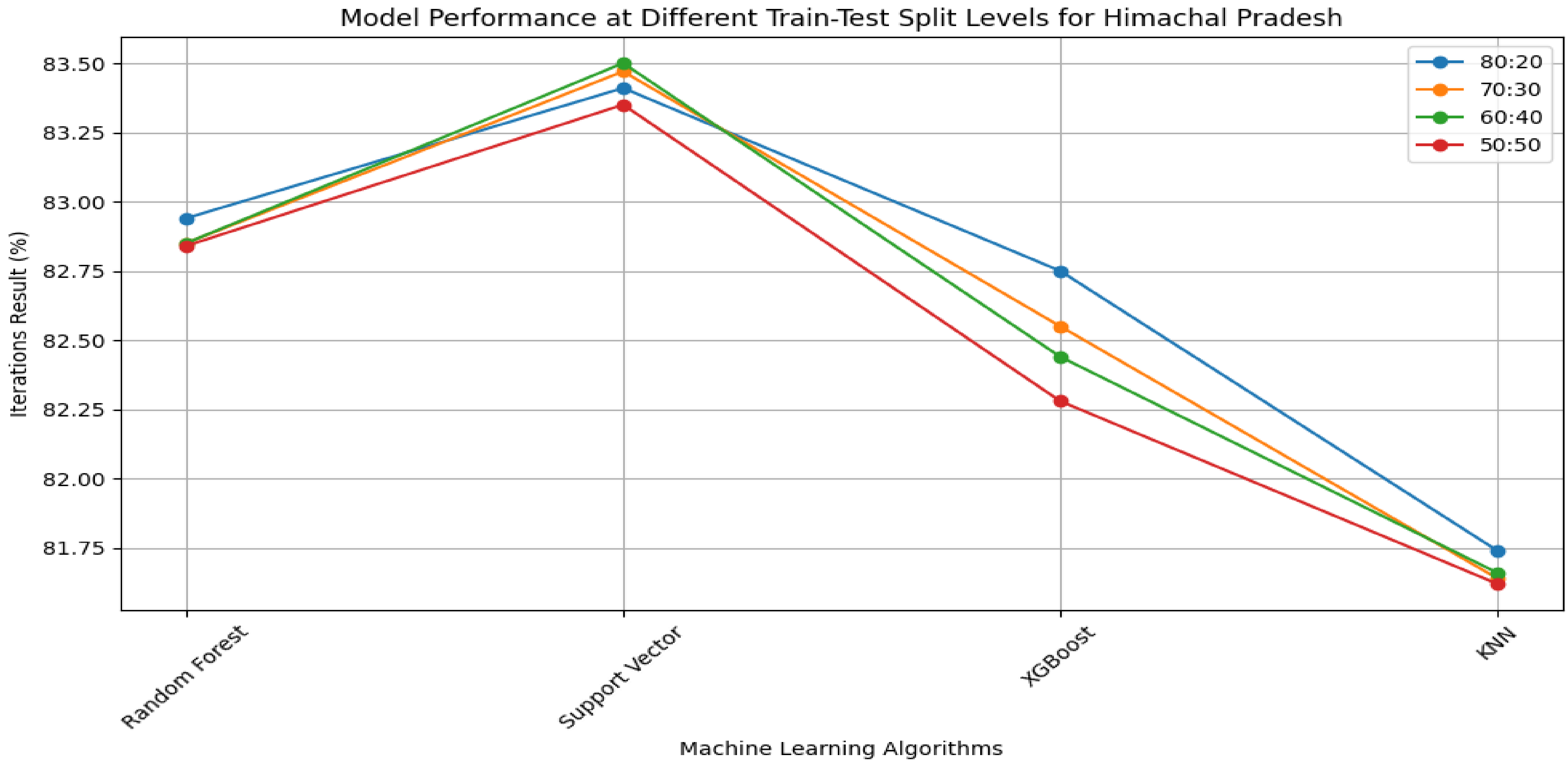
# Temperature Extremes



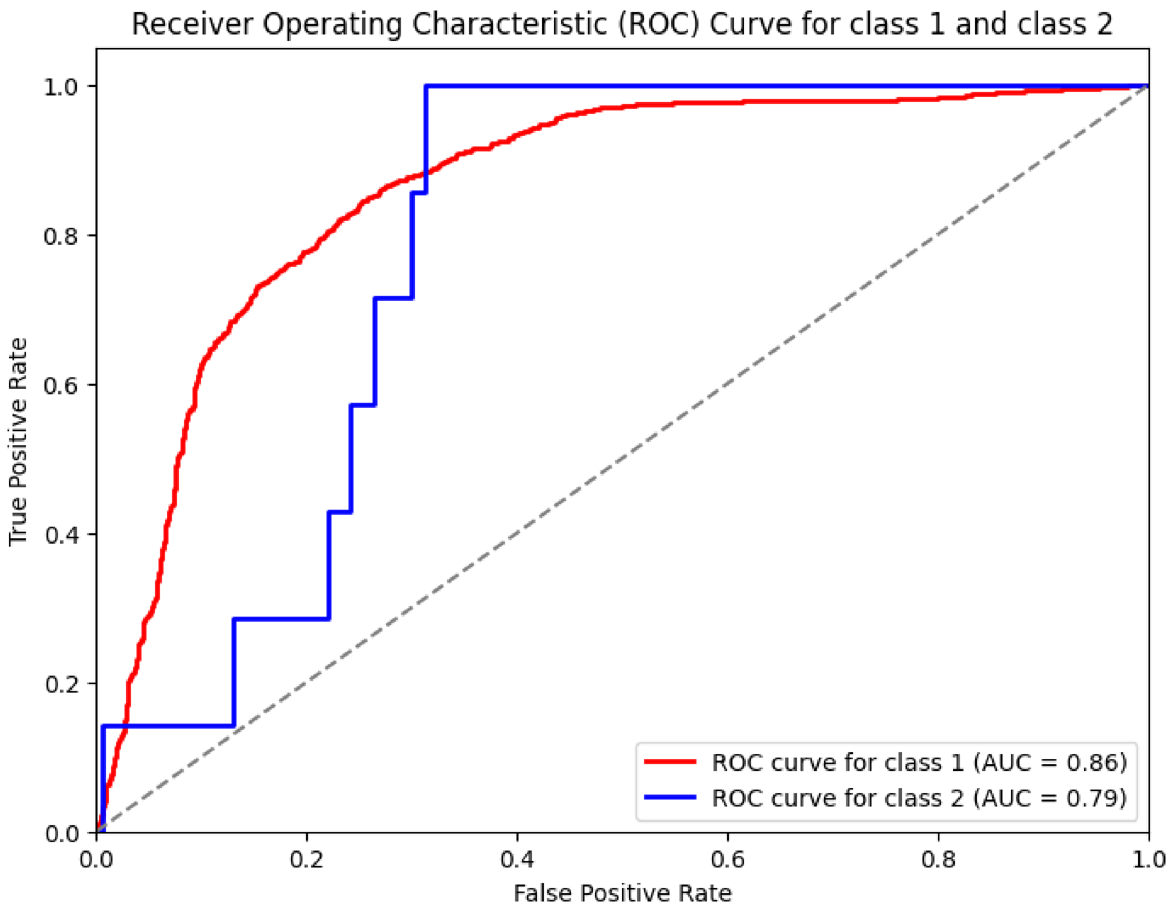



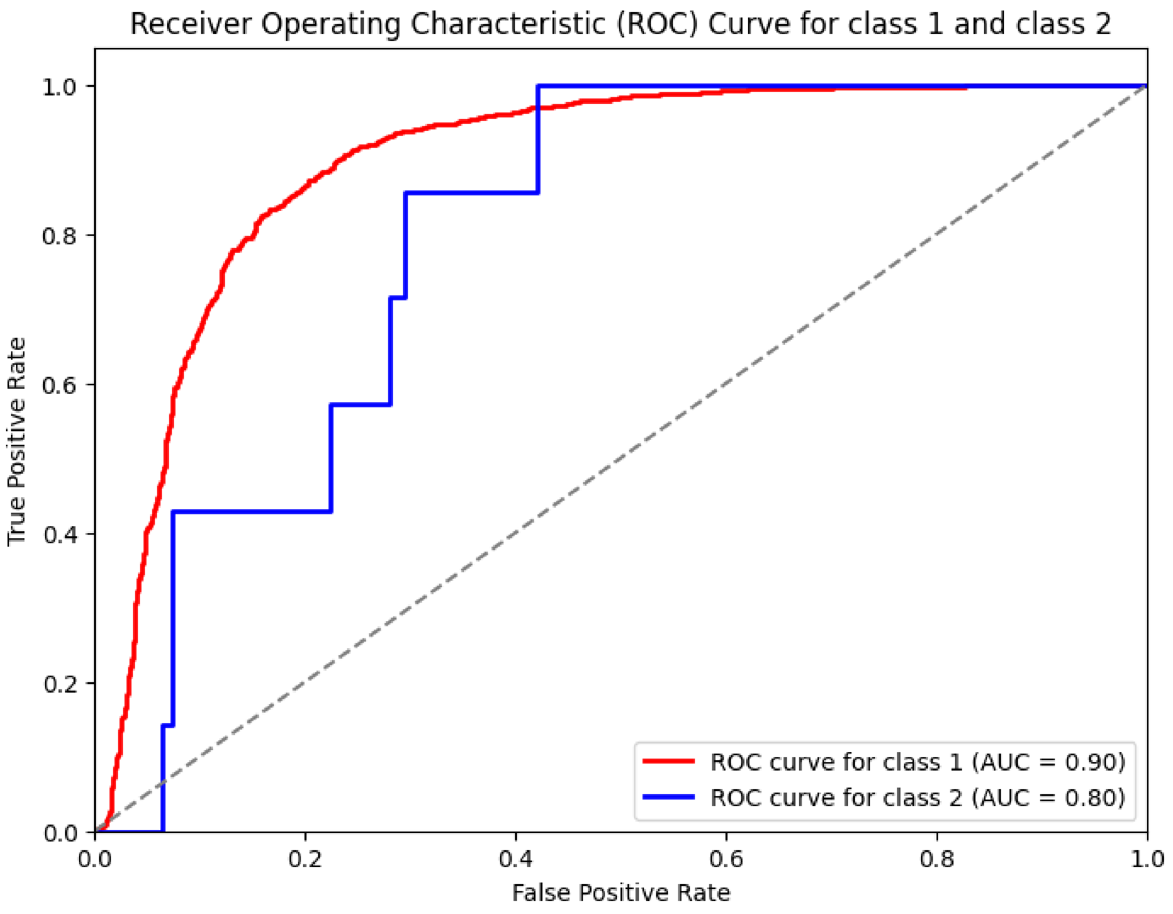

- Thresholds for extreme temperature events: 10th percentile approximately 24 degrees Celsius, 95th percentile approximately 43 degrees Celsius.
- The machine learning model demonstrates overall high performance in predicting temperature extreme values.
- However, performance decreases when only statistically significant variables are included in the model.

# Comparison of Accuracy Scores Across Various Machine Learning Algorithms at Different Train-Test Split Levels



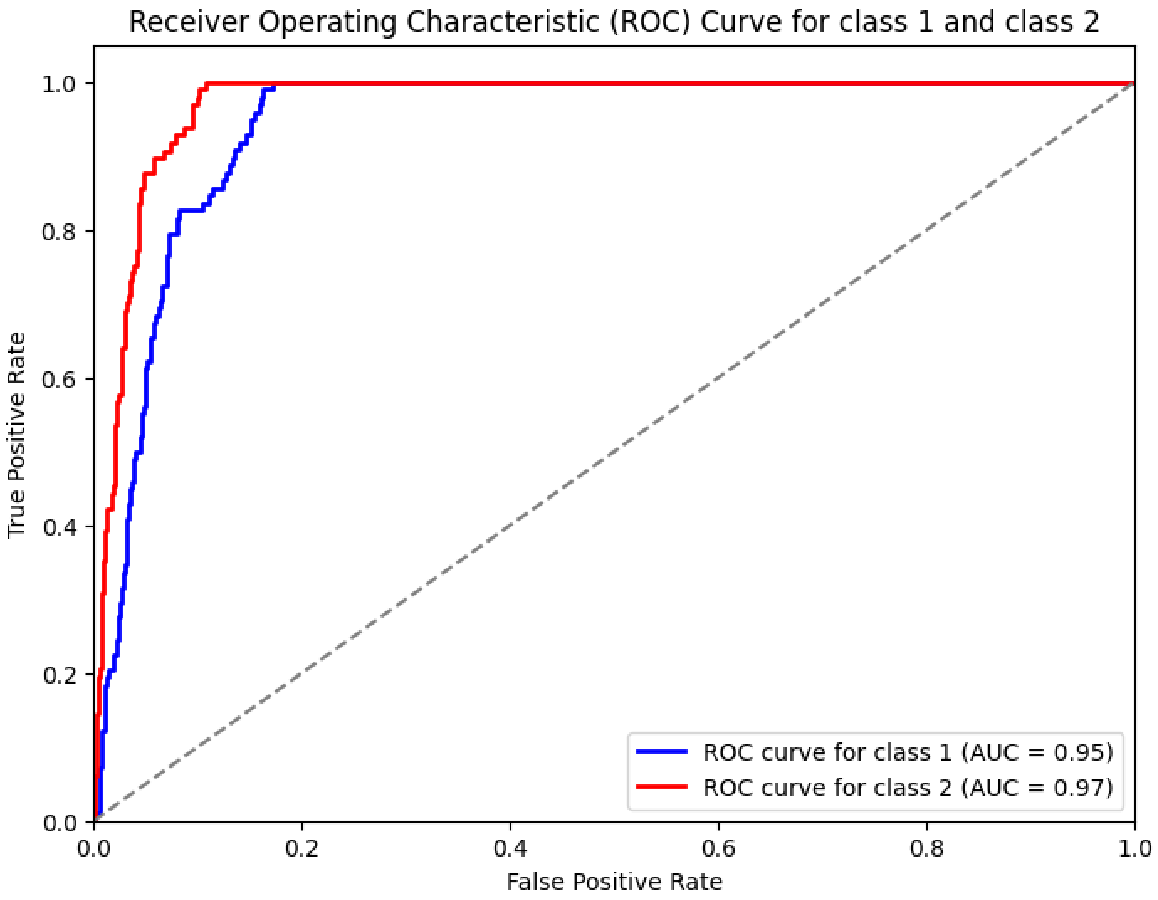
# Precipitation Extremes

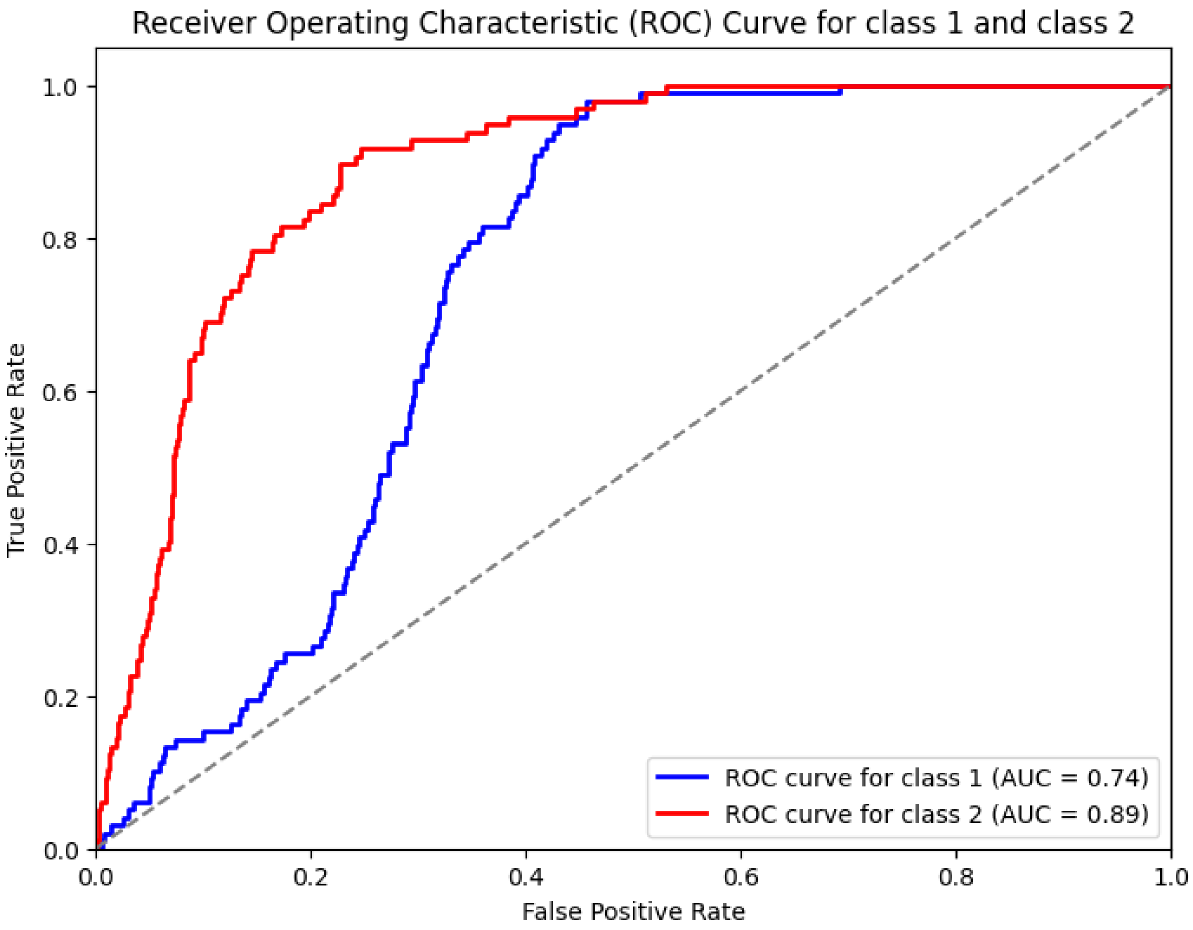



- Thresholds for extreme rain events: 10th percentile approximately 0.30 mm, 95th percentile approximately 23.78 mm.
- Machine learning model performed well for events below the 10th percentile .
- Enhanced model performance is observed by focusing on statistically significant variables for extreme rain events.

# Temperature Extremes




•Thresholds for extreme temperature events: 10th percentile approximately 19.01 degrees Celsius, 95th percentile approximately 38.38 degrees Celsius.

•The machine learning model demonstrates overall high performance in predicting temperature extreme values when all variables are considered.

•However, performance decreases when only statistically significant variables are included in the model.

Continued…

# Robustness of the model in predicting different events

| **Brier Score** | Rain (all variables) | Rain (only significant variables) | Temperature (all variables) | Temperature (only significant variables) |
|---|---|---|---|---|
| **Rajasthan** | 0.49 | 0.56 | 0.32 | 0.35 |
| **Himachal Pradesh** | 0.80 | 0.95 | 0.41 | 0.41 |

- The Brier Scores for both the states, especially in temperature prediction, are relatively close to 0, indicating robust and accurate models.
- Himachal Pradesh's higher Brier Scores, particularly in rainfall prediction, suggest less accurate models and higher uncertainty in forecasts compared to Rajasthan.
- Further refinement and improvement of predictive models, especially in Himachal Pradesh, may be necessary to enhance accuracy, particularly in predicting rainfall events

# 6. Summary

- **Is machine learning adept at identifying relationships among diverse atmospheric variables?**

Yes, machine learning demonstrates efficacy in identifying such relationships.

- **What is the achievable predictive accuracy of machine learning algorithms in future data forecasting?**

The predictive accuracy ranges between 82 to 84 percent across different machine learning algorithms

- **Do different machine learning algorithms produce consistent results, and if not, what factors contribute to variability?**

The variability in outcomes is influenced by factors such as the inherent complexity of algorithms, data quality and quantity, feature selection, hyperparameter tuning, and specific problem characteristics. Additionally, choice of evaluation metrics and performance measures also play a major role.

- **How reliable and robust is the skill score of the model?**

Current models are robust for some conditions such a for temperature extremes, further refinements are needed to draw stronger conclusions specifically in case of rainfall extreme events.

- **Can this methodology be applied to predict extreme climatic events ?**

While it holds potential, it remains a prospect for future exploration rather than a current certainty.

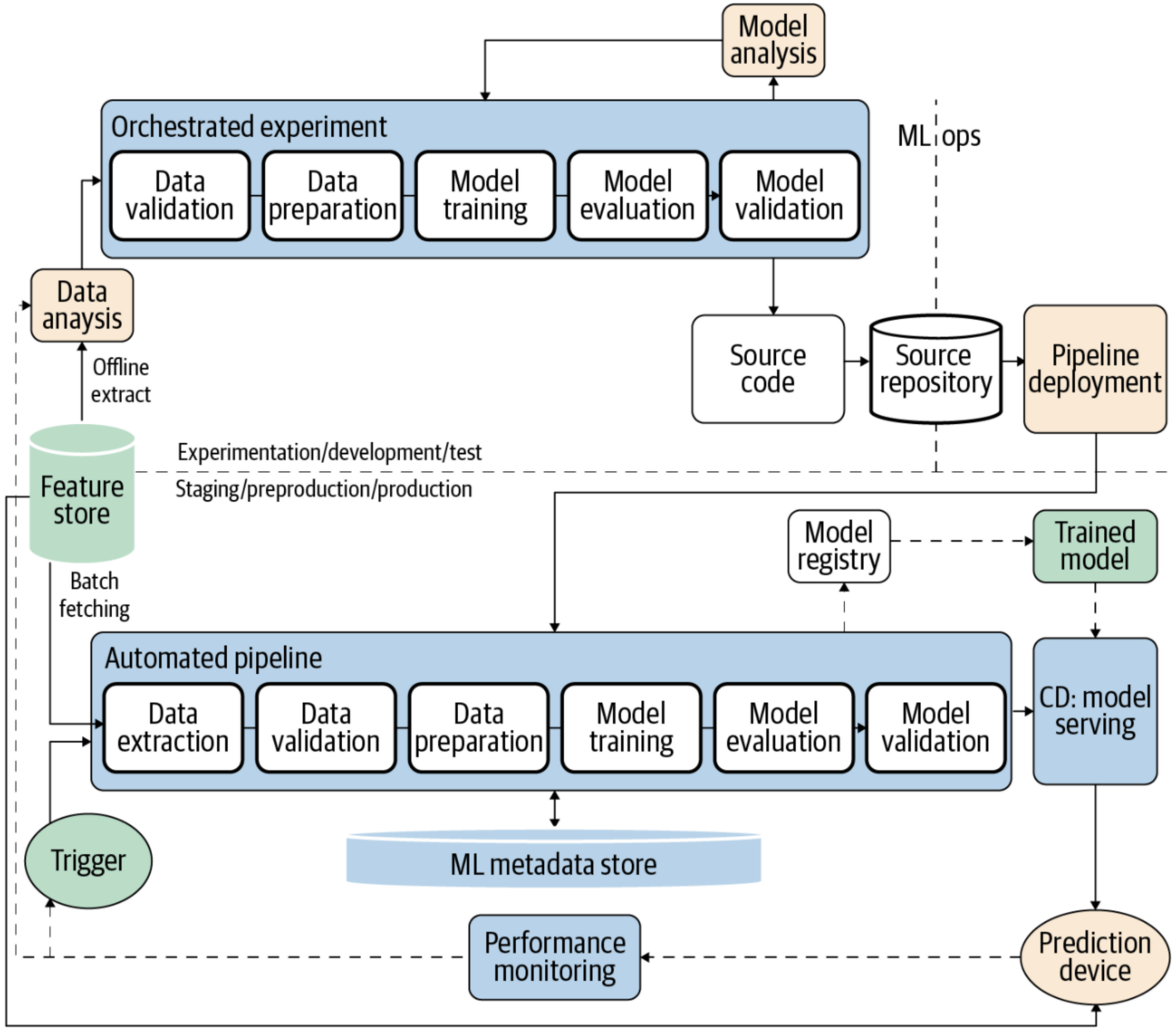
# 7. Way Forward

1. Enhance algorithm accuracy through advanced machine learning techniques and feature engineering.

2. Develop a real-time prediction framework for meteorological extremes, including droughts and heatwaves.

3. Deploy the framework on scalable cloud infrastructure and automate the model update and deployment process.

4. Extend the framework's application to evaluate and forecast various meteorological extremes.

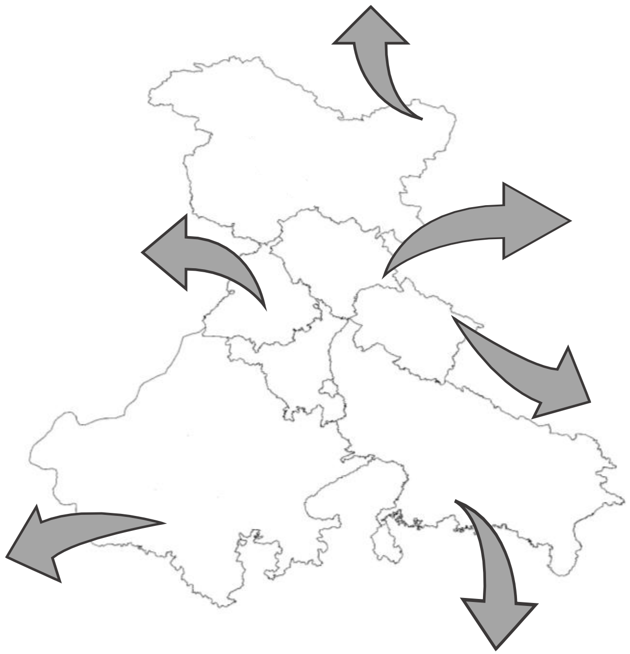


# Thank You!

# Supplementary

# ML algorithm for North Indian States

| Machine Learning Algorithms | Iterations result at different train – test split levels for Jammu and Kashmir | | | |
|---|---|---|---|---|
| | 80:20 | 70:30 | 60:40 | 50:50 |
| Random Forest Classifier | 79.59 | 79.72 | 79.61 | 79.54 |
| Support Vector Classifier | 79.77 | 79.82 | 79.80 | 79.79 |
| Xtreme Gradient Boost | 79.05 | 79.02 | 78.87 | 78.80 |
| K nearest neighbour | 77.71 | 77.68 | 77.63 | 77.54 |



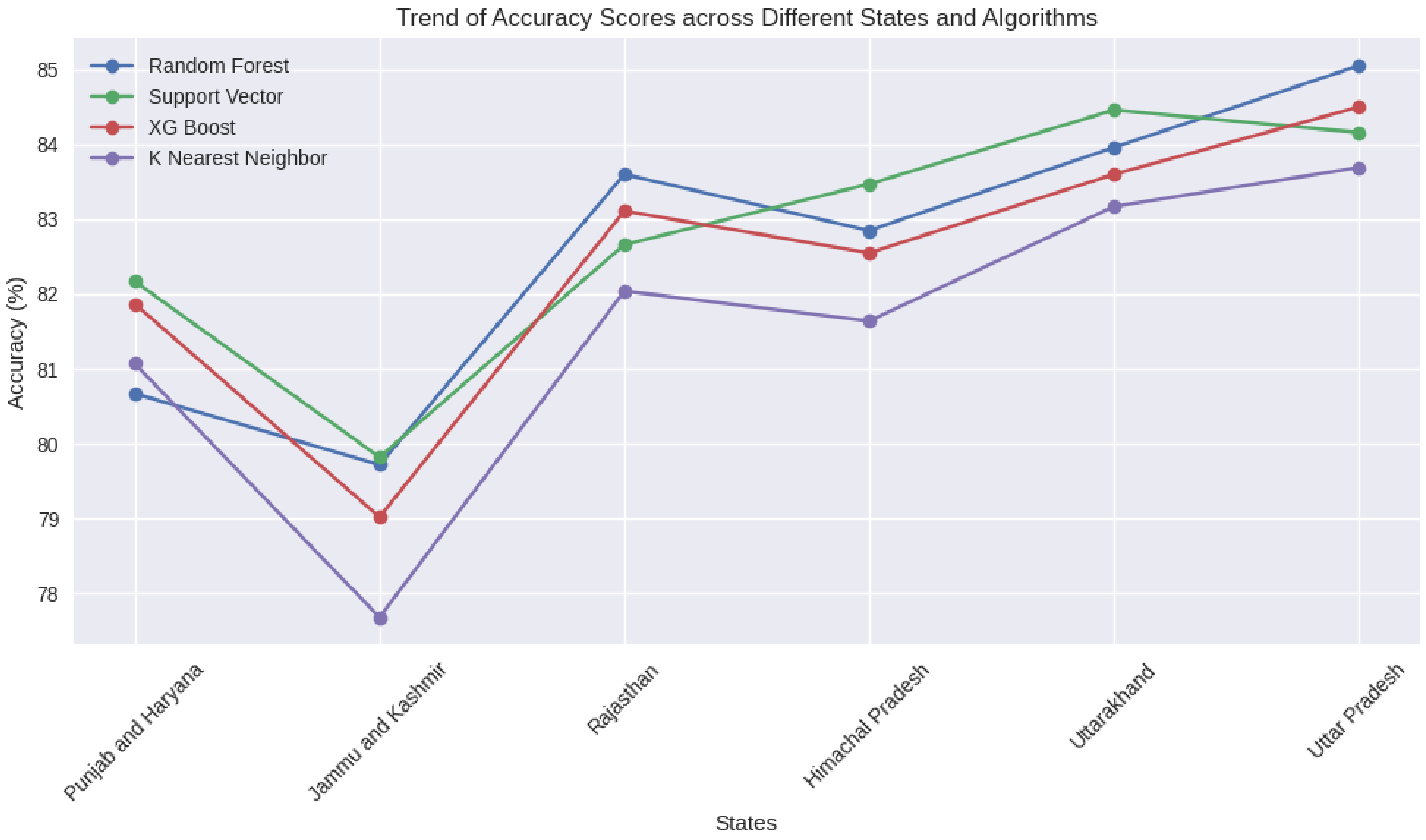
| Machine Learning Algorithms | Iterations result at different train – test split levels for Punjab and Haryana | | | |
|---|---|---|---|---|
| | 80:20 | 70:30 | 60:40 | 50:50 |
| Random Forest Classifier | 80.65 | 80.67 | 80.62 | 80.67 |
| Support Vector Classifier | 82.25 | 82.17 | 81.98 | 81.85 |
| Xtreme Gradient Boost | 81.93 | 81.87 | 81.75 | 81.54 |
| K nearest neighbour | 81.13 | 81.07 | 80.92 | 80.79 |

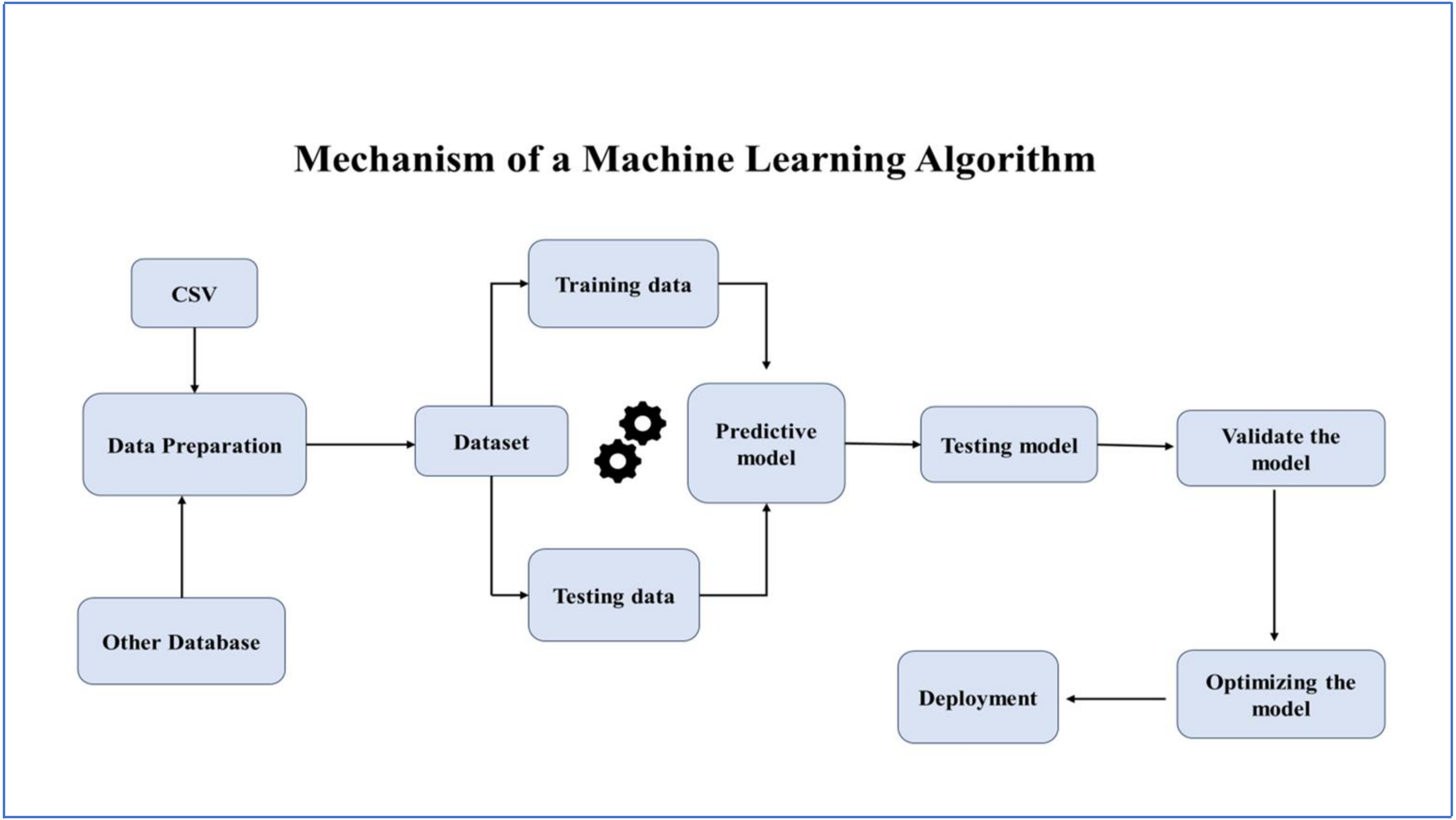
| Machine Learning Algorithms | Iterations result at different train – test split levels for Himachal Pradesh | | | |
|---|---|---|---|---|
| | 80:20 | 70:30 | 60:40 | 50:50 |
| Random Forest Classifier | 82.94 | 82.85 | 82.85 | 82.84 |
| Support Vector Classifier | 83.41 | 83.47 | 83.50 | 83.35 |
| Xtreme Gradient Boost | 82.75 | 82.55 | 82.44 | 82.28 |
| K nearest neighbour | 81.74 | 81.64 | 81.66 | 81.62 |

| Machine Learning Algorithms | Iterations result at different train – test split levels for Rajasthan | | | |
|---|---|---|---|---|
| | 80:20 | 70:30 | 60:40 | 50:50 |
| Random Forest Classifier | 83.71 | 83.60 | 83.59 | 83.55 |
| Support Vector Classifier | 82.79 | 82.66 | 82.48 | 82.34 |
| Xtreme Gradient Boost | 83.22 | 83.11 | 82.93 | 82.78 |
| K nearest neighbour | 82.00 | 82.04 | 82.06 | 82.06 |

| Machine Learning Algorithms | Iterations result at different train – test split levels for Uttarakhand | | | |
|---|---|---|---|---|
| | 80:20 | 70:30 | 60:40 | 50:50 |
| Random Forest Classifier | 83.96 | 83.96 | 83.90 | 83.84 |
| Support Vector Classifier | 84.47 | 84.46 | 84.36 | 84.16 |
| Xtreme Gradient Boost | 83.76 | 83.60 | 83.51 | 83.33 |
| K nearest neighbour | 83.25 | 83.17 | 83.12 | 82.93 |

| Machine Learning Algorithms | Iterations result at different train – test split levels for Uttar Pradesh | | | |
|---|---|---|---|---|
| | 80:20 | 70:30 | 60:40 | 50:50 |
| Random Forest Classifier | 85.05 | 84.91 | 84.83 | 84.70 |
| Support Vector Classifier | 84.16 | 84.01 | 83.86 | 83.78 |
| Xtreme Gradient Boost | 84.50 | 84.24 | 84.11 | 83.93 |
| K nearest neighbour | 83.73 | 83.69 | 83.58 | 83.57 |


Trend of Accuracy Scores across Different States and Algorithms
Random Forest
Support Vector
XG Boost
K Nearest Neighbor
Accuracy (%)
85
84
83
82
81
80
79
78
Punjab and Haryana
Jammu and Kashmir
Rajasthan
Himachal Pradesh
Uttarakhand
Uttar Pradesh
States


Mechanism of a Machine Learning Algorithm
CSV
Data Preparation
Other Database
Dataset
Training data
Testing data
Predictive model
Testing model
Validate the model
Optimizing the model
Deployment