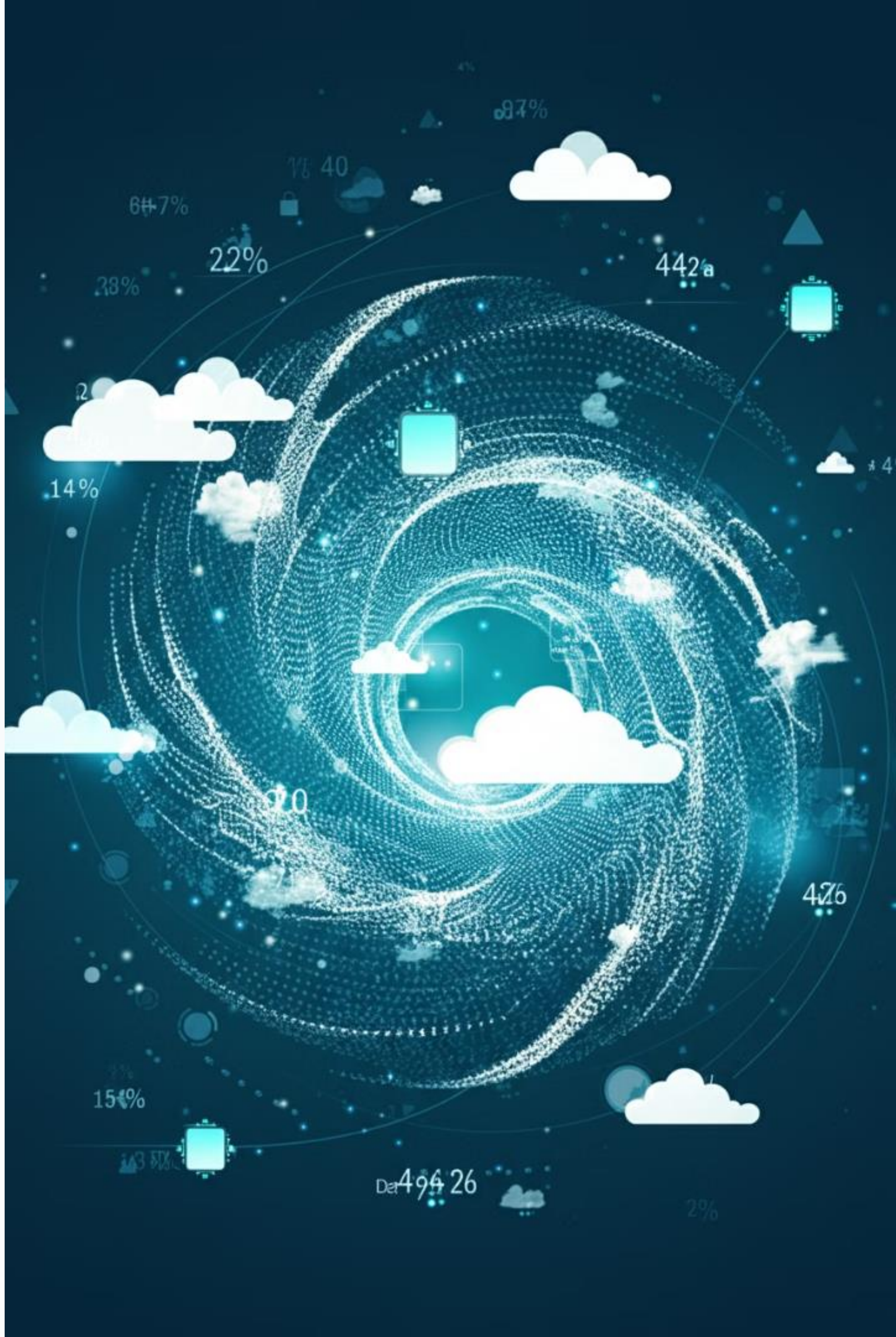
Machine Learning for Weather Prediction

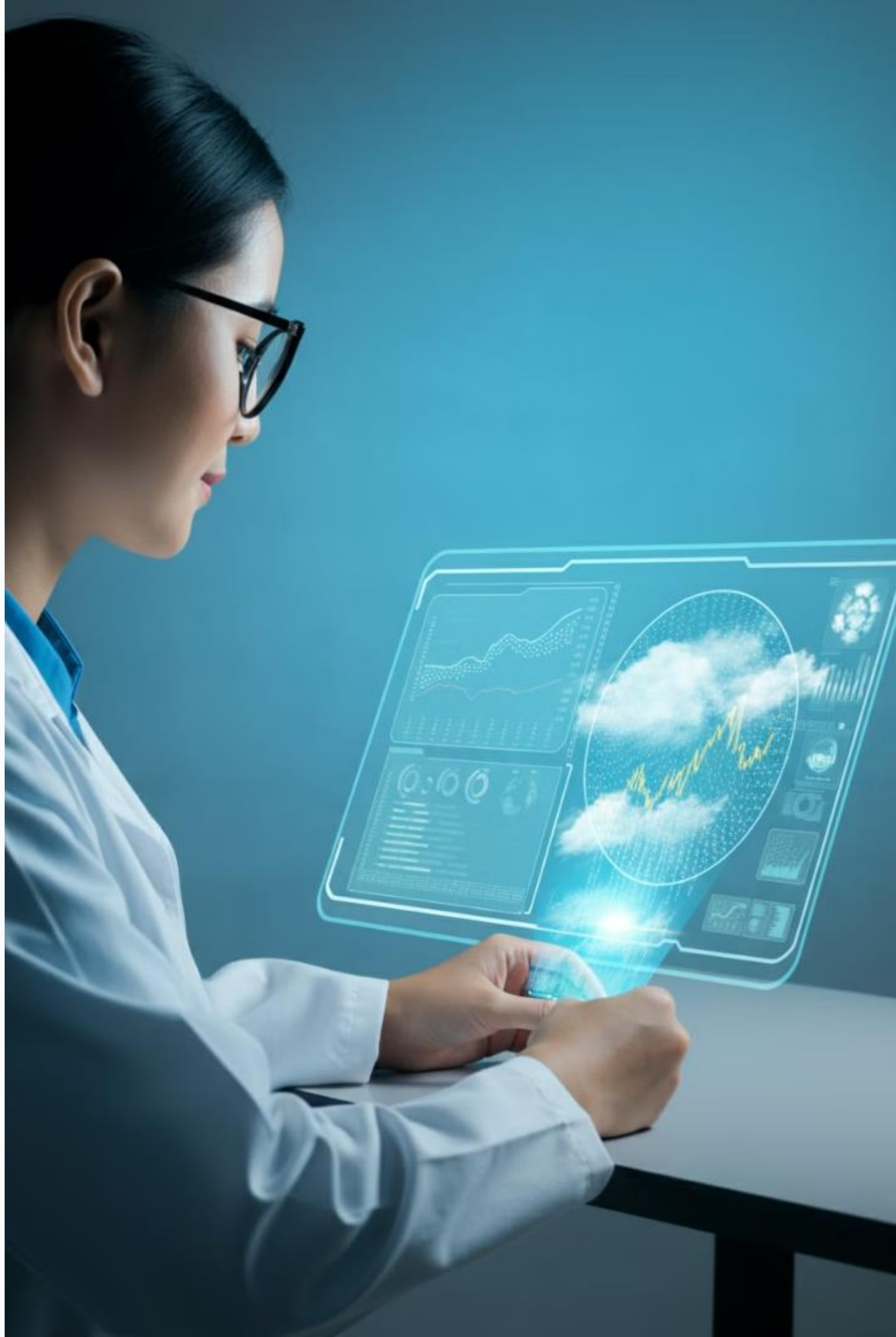
A proposal for ClimateWins to predict and analyze weather patterns in Europe using advanced machine learning techniques.

by Rose Alappat Joy



Project Objectives

1. Find new patterns in weather changes over the last 60 years
2. Identify unusual weather patterns in Europe
3. Determine if unusual patterns are increasing
4. Generate future weather predictions for next 25-50 years



Data and Methodology

Data Source

Weather observations from 18 European stations, late 1800s to 2022

Key Variables

Temperature, wind speed, snow, global radiation, and more

Approach

Unsupervised and supervised machine learning, deep learning models

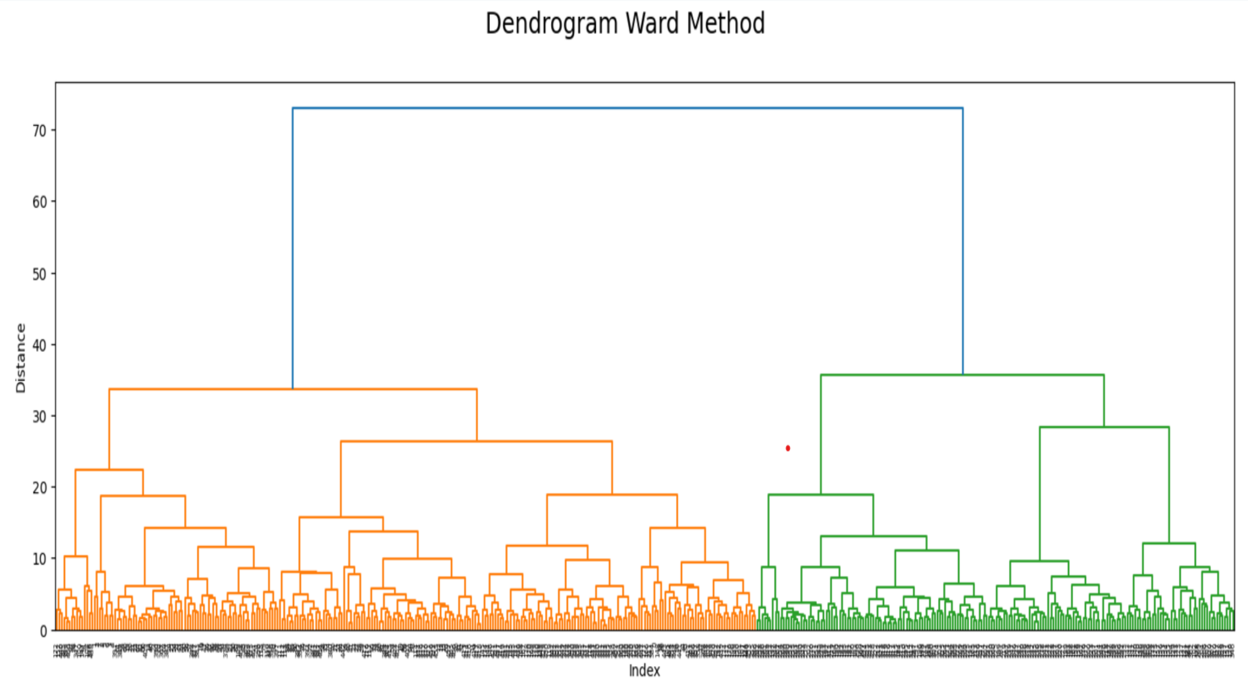
Unsupervised Learning Insights

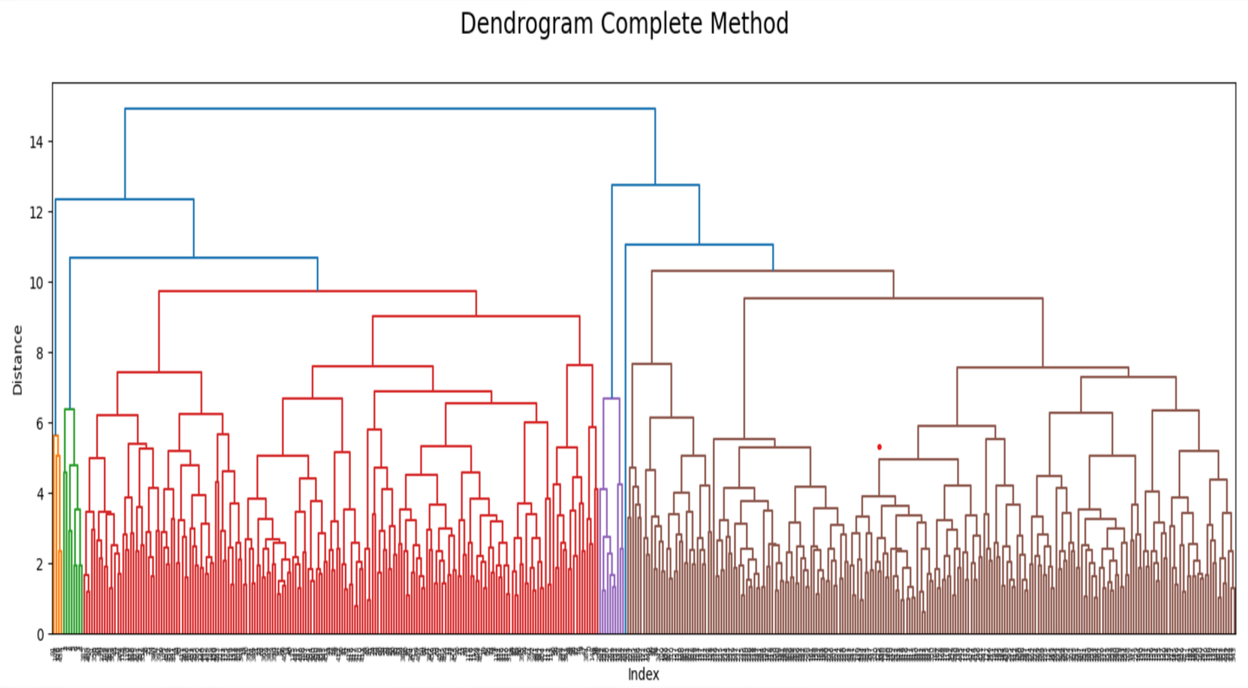
Clustering Methods

Single Linkage, Complete Linkage, Average Linkage, and Ward methods applied to compare weather patterns between stations and years.

Key Findings

Complete Linkage and Ward methods showed clearest distinctions between climate patterns, revealing significant differences between regions like Madrid and Munich.





Supervised Learning Models

1. **CNN Model**

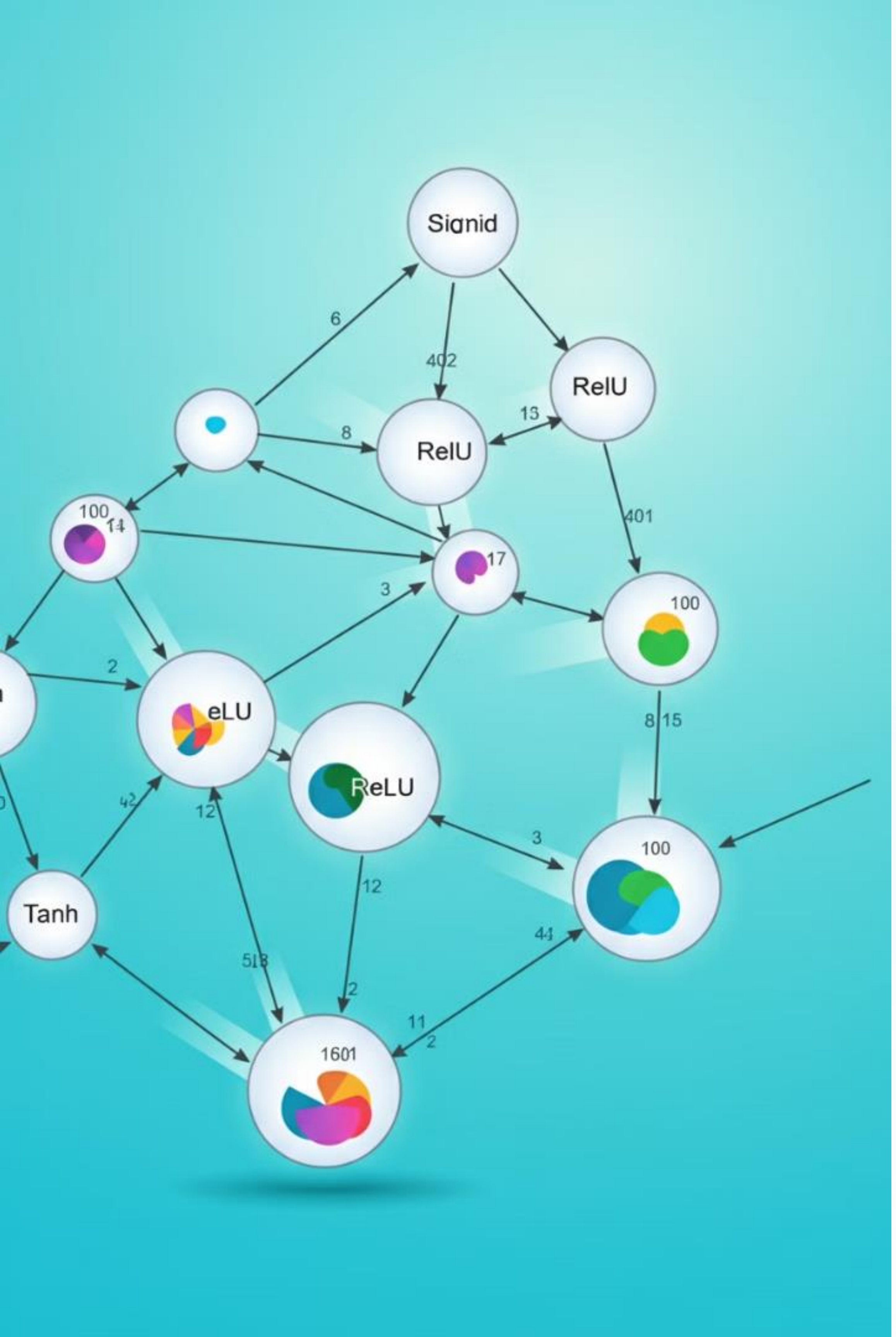
   Initial accuracy: 11.92%. Optimized accuracy: 92.2%

2. **RNN Model**

   Faced challenges in learning data patterns effectively

3. **Random Forest**

   Achieved 100% accuracy for individual stations like Stockholm



CNN Model Performance

51.6%

Tanh Accuracy

Moderate performance, identifying patterns in 5 out of 15 classes

12%

Sigmoid Accuracy

Poor performance, likely due to vanishing gradients

64.4%

ReLU Accuracy

Higher accuracy but unstable loss, identifying only 1 class

```
Epoch 1/30
1076/1076 - 22s - 20ms/step - accuracy: 0.0782 - loss: 24.5252
Epoch 2/30
1076/1076 - 13s - 12ms/step - accuracy: 0.0658 - loss: 24.4837
Epoch 3/30
1076/1076 - 13s - 12ms/step - accuracy: 0.0881 - loss: 25.0586
Epoch 4/30
1076/1076 - 13s - 12ms/step - accuracy: 0.0736 - loss: 24.7445
Epoch 5/30
1076/1076 - 13s - 12ms/step - accuracy: 0.0511 - loss: 25.1092
Epoch 6/30
1076/1076 - 13s - 12ms/step - accuracy: 0.0510 - loss: 24.8289
Epoch 7/30
1076/1076 - 12s - 11ms/step - accuracy: 0.0267 - loss: 24.5668
Epoch 8/30
1076/1076 - 13s - 12ms/step - accuracy: 0.0280 - loss: 24.8879
Epoch 9/30
1076/1076 - 12s - 11ms/step - accuracy: 0.0130 - loss: 24.8415
Epoch 10/30
1076/1076 - 13s - 12ms/step - accuracy: 0.0107 - loss: 25.0344
Epoch 11/30
1076/1076 - 13s - 12ms/step - accuracy: 0.0078 - loss: 24.7568
Epoch 12/30
1076/1076 - 21s - 19ms/step - accuracy: 0.0098 - loss: 24.5851
Epoch 13/30
1076/1076 - 13s - 12ms/step - accuracy: 0.0150 - loss: 24.6720
Epoch 14/30
1076/1076 - 13s - 12ms/step - accuracy: 0.0187 - loss: 24.2592
Epoch 15/30
1076/1076 - 21s - 20ms/step - accuracy: 0.0148 - loss: 24.5139
Epoch 16/30
1076/1076 - 20s - 19ms/step - accuracy: 0.0279 - loss: 24.3615
Epoch 17/30
1076/1076 - 13s - 12ms/step - accuracy: 0.0264 - loss: 24.2377
Epoch 18/30
1076/1076 - 13s - 12ms/step - accuracy: 0.0455 - loss: 24.1651
Epoch 19/30
1076/1076 - 13s - 12ms/step - accuracy: 0.0626 - loss: 24.3312
Epoch 20/30
1076/1076 - 13s - 12ms/step - accuracy: 0.0562 - loss: 25.0605
Epoch 21/30
1076/1076 - 12s - 12ms/step - accuracy: 0.0667 - loss: 25.0434
Epoch 22/30
1076/1076 - 13s - 12ms/step - accuracy: 0.0742 - loss: 24.9951
Epoch 23/30
1076/1076 - 12s - 11ms/step - accuracy: 0.0770 - loss: 24.2285
Epoch 24/30
1076/1076 - 13s - 12ms/step - accuracy: 0.0464 - loss: 24.2724
Epoch 25/30
1076/1076 - 13s - 12ms/step - accuracy: 0.0407 - loss: 24.3492
Epoch 26/30
1076/1076 - 13s - 12ms/step - accuracy: 0.0536 - loss: 24.5829
Epoch 27/30
1076/1076 - 21s - 19ms/step - accuracy: 0.0234 - loss: 24.3111
Epoch 28/30
1076/1076 - 20s - 19ms/step - accuracy: 0.0252 - loss: 24.4421
Epoch 29/30
1076/1076 - 13s - 12ms/step - accuracy: 0.0363 - loss: 24.4558
Epoch 30/30
1076/1076 - 12s - 12ms/step - accuracy: 0.0495 - loss: 24.4351
```

RNN/LSTM Model Results

1. **Initial Performance**

   Lower accuracy compared to CNN, with significant challenges in learning data patterns

2. **Training Progress**

   Loss decreases over epochs but fluctuates, showing irregular convergence

3. **Final Outcome**

   Despite lower loss compared to CNN, accuracy remains poor, suggesting underfitting or difficulty handling specific data categories

Random Forest Model Insights

City/Category	Key Indicators for Weather	Observations
Madrid & Budapest	`temp_max`, `temp_mean`, `temp_min`	Temperature observations are critical for predicting pleasant or unpleasant weather.
Basel	`temp_max`, `precipitation`, `sunshine`	A mix of temperature, precipitation, and sunshine determines pleasant or unpleasant weather.
General Observations	`temp_max` consistently ranks in the top three indicators across weather stations	Rising mean temperatures reflect global climate change, highlighting its importance in forecasting.

Key Findings: Feature Importance

Temperature

Maximum temperature consistently ranked as the most crucial indicator

Precipitation

Second most important feature for weather prediction

Sunshine

Third most significant factor in weather forecasting

Hyperparameter Optimization Results

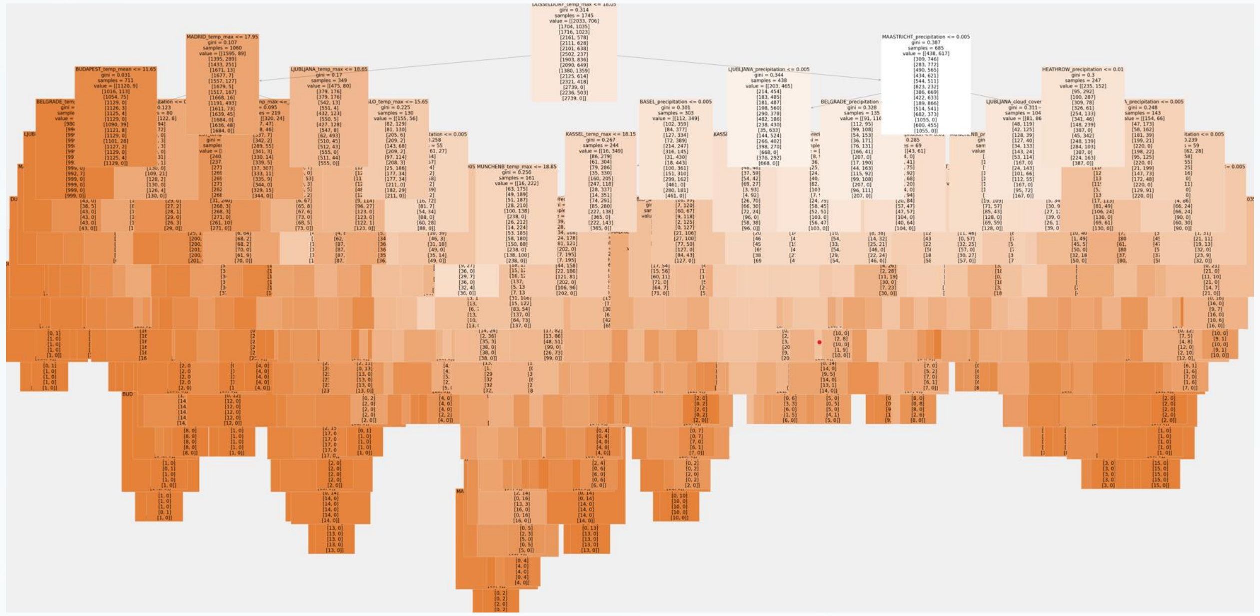
All Weather Stations

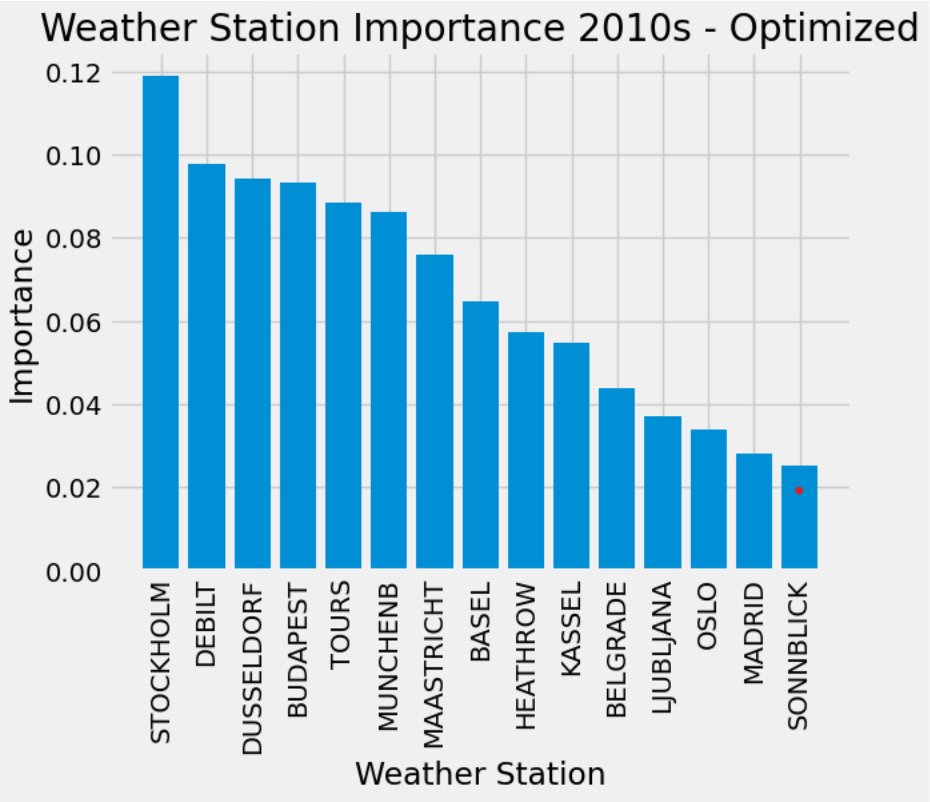
Accuracy improved from 58.7% to 65.60% after optimization

Stockholm Station

Accuracy increased from 98.78% to 100% after optimization

All Weather Stations After Optimization





Deep Learning CNN Optimization

11.92%

Before

Initial accuracy with Exponential loss growth

92.2%

After

Optimized accuracy with loss reduced to 0.2319

Pred True	BASEL	BELGRADE	BUDAPEST	DEBILT	DUSSELDORF	HEATHROW	KASSEL
BASEL	3538	69	10	3	11	6	1
BELGRADE	81	988	15	1	0	4	0
BUDAPEST	23	8	169	13	0	1	0
DEBILT	14	1	9	56	2	0	0
DUSSELDORF	4	0	1	3	14	6	0
HEATHROW	5	1	0	2	7	67	0
KASSEL	0	0	1	1	1	0	4
LJUBLJANA	10	4	1	0	0	3	1
MAASTRICHT	5	0	0	0	0	2	0
MADRID	16	9	14	1	6	17	0
MUNCHENB	7	1	0	0	0	0	0
OSLO	0	0	0	0	1	0	0
STOCKHOLM	1	0	0	0	0	0	0
VALENTIA	1	0	0	0	0	0	0

Pred True	LJUBLJANA	MAASTRICHT	MADRID	MUNCHENB	OSLO	STOCKHOLM
BASEL	4	2	38	0	0	0
BELGRADE	0	0	3	0	0	0
BUDAPEST	0	0	0	0	0	0
DEBILT	0	0	0	0	0	0
DUSSELDORF	1	0	0	0	0	0
HEATHROW	0	0	0	0	0	0
KASSEL	2	0	1	1	0	0
LJUBLJANA	41	0	1	0	0	0
MAASTRICHT	0	2	0	0	0	0

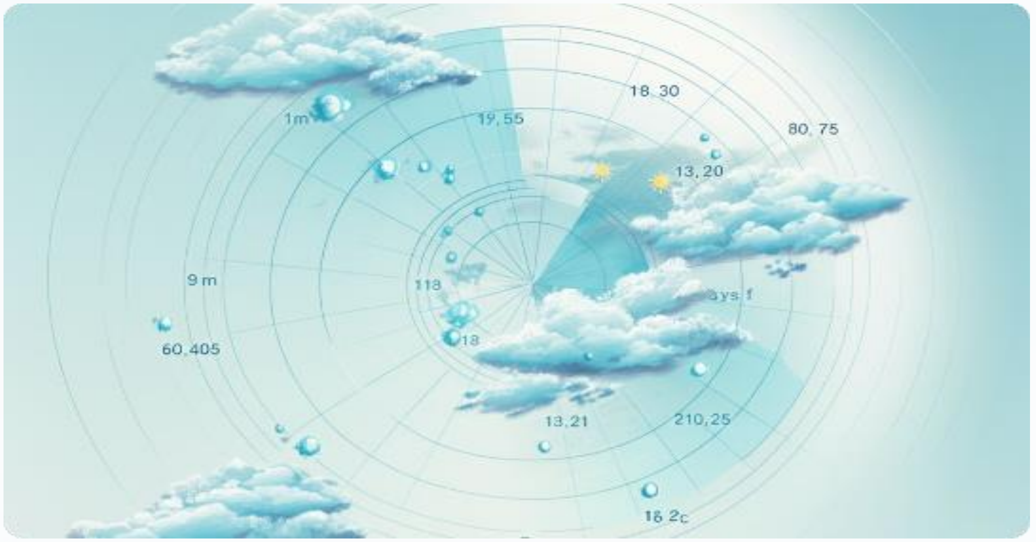
Visual Applications in Weather Prediction

Handwritten Recognition

Model achieved 40% accuracy on handwritten data, showing room for improvement

Radar Recognition

75% training accuracy and 74% validation accuracy, with low loss values indicating good generalization



GAN Applications in Weather Prediction

Satellite Imagery

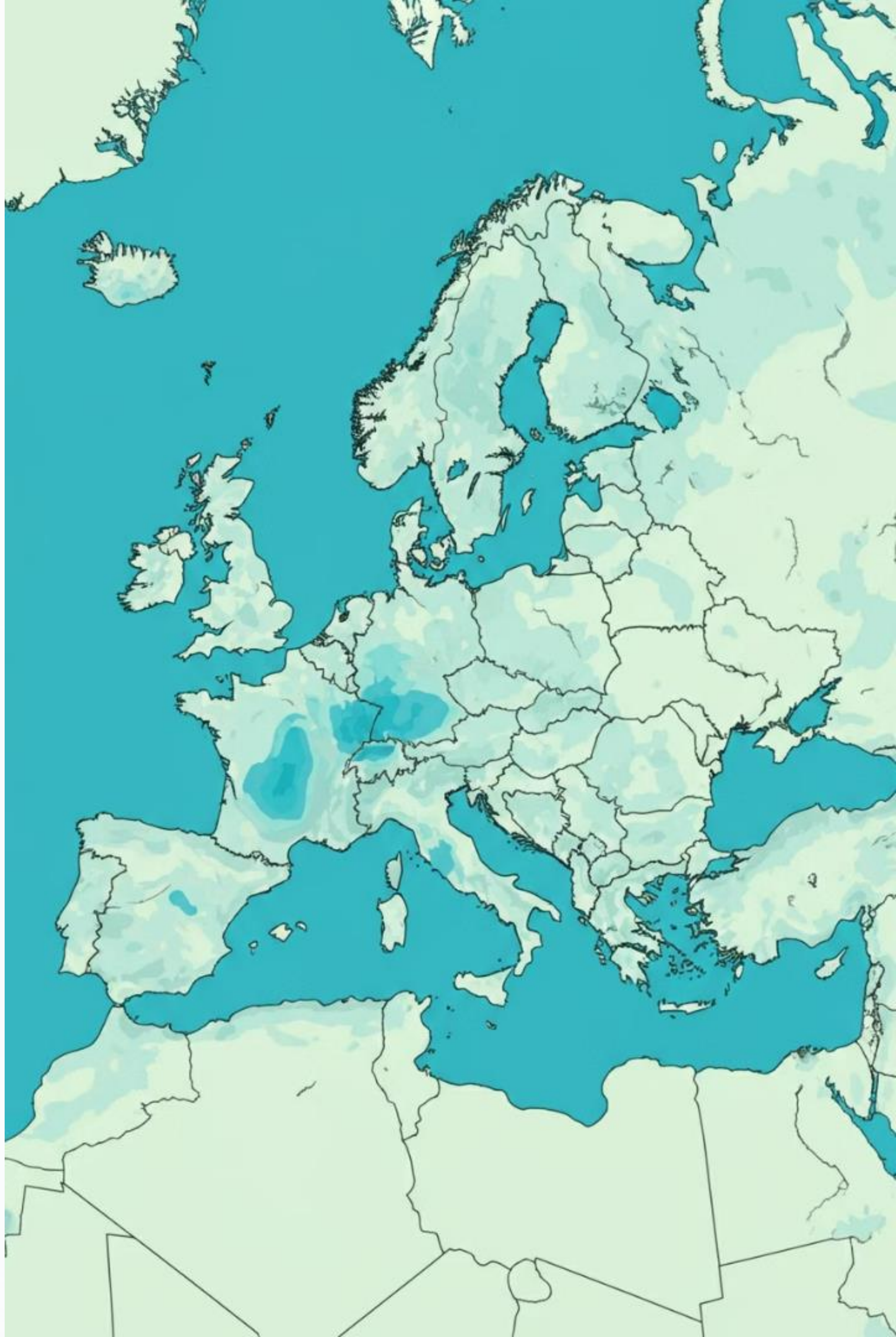
Monitor storms and precipitation patterns

Climate Simulation

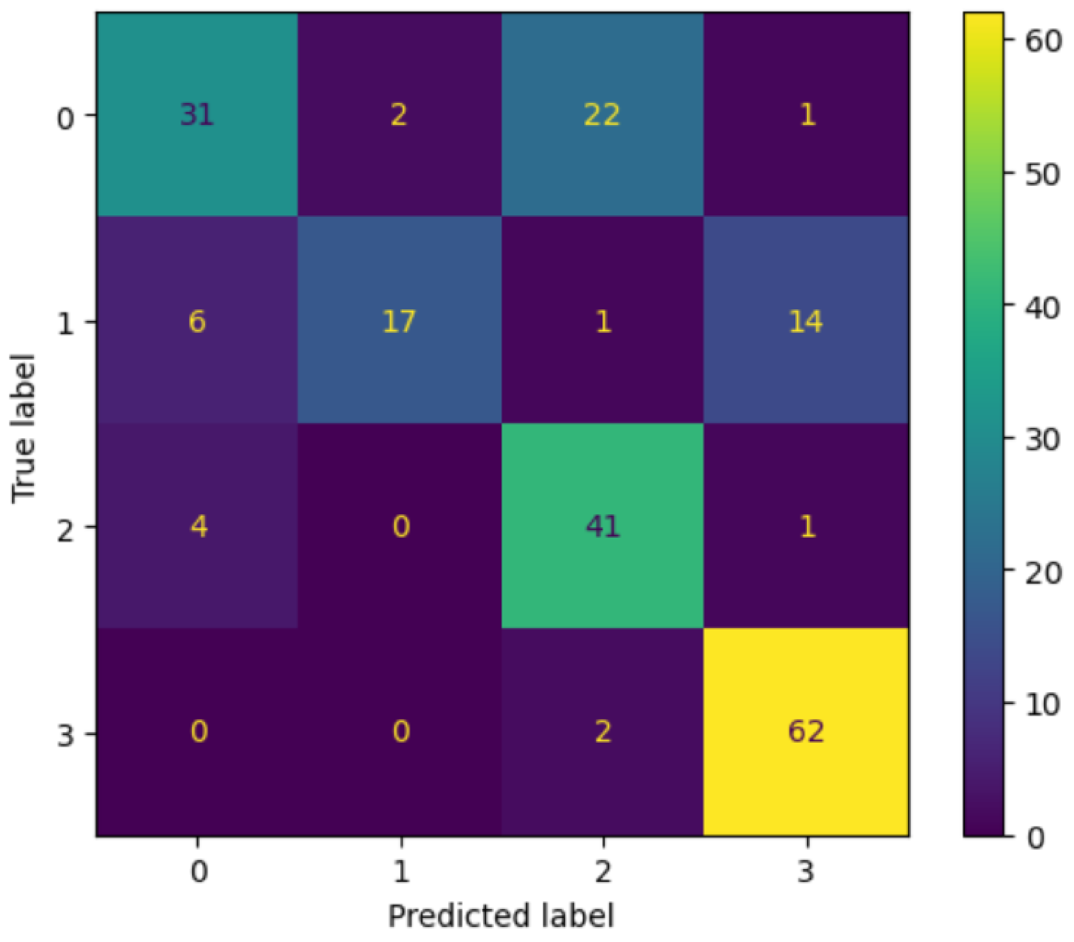
Generate future climate scenarios

Radar Processing

Automatic segmentation of meteorological phenomena



Evaluating Model Performance for Weather Event Classification



- A confusion matrix evaluates how well CNNs and Random Forests classify weather events by comparing predictions with actual outcomes.
- The confusion matrix highlights frequent misclassifications of "cloudy" as "shine" (22 times)
- Difficulties with "cloudy" vs. "rain." However, "shine" and "sunrise" are classified with high accuracy

Recommendations for ClimateWins

1 Segmented Analysis

Divide dataset by locations, time intervals, or weather features for focused predictions

2 Model Selection

Start with optimized random forest for interpretability, then introduce CNN for complex patterns

3 Key Variables

Prioritize temperature (especially maximum) and precipitation in weather modeling

4 Continuous Optimization

Regularly refine models with new data and emerging machine learning techniques

Next Steps for Implementation

1. **Data Integration**

   Combine historical and real-time data sources

2. **Model Development**

   Implement and fine-tune recommended ML models

3. **Validation**

   Rigorously test predictions against new data

4. **Deployment**

   Integrate predictive systems into ClimateWins operations

Thank You

Rose Alappt Joy
Mail Id : roseaj1510@gmail.com
www.linkedin.com/in/rose-alappat-joy-a9274896