Predicting Weather Variations with Machine Learning for ClimateWins

January 2025

Ivonne Aspilcueta

Objectives and Overview

Core goals of ClimateWins:

Identify weather patterns outside the regional norm in Europe.

Determine if unusual weather patterns are increasing.

Generate possibilities for future weather conditions over the next 25 to 50 years.

Determine the safest regions for habitation in Europe over the next 25 to 50 years.

Summary of Three Thought Experiments:

1. Predicting Weather Anomalies with Random Forest: Focus on using ensemble models to analyze shifts in weather patterns based on historical data.

2. Deep Learning for Complex Weather Interactions: Utilize Convolutional Neural Networks (CNNs) to detect patterns in satellite images and radar data.

3. GANs for Synthetic Weather Projections: Generate possible future weather scenarios using GANs to simulate climate variations.

Data Requirements Beyond Historical Weather Data

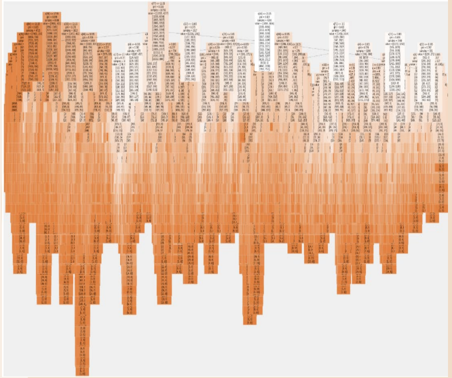
Additional Data Needs

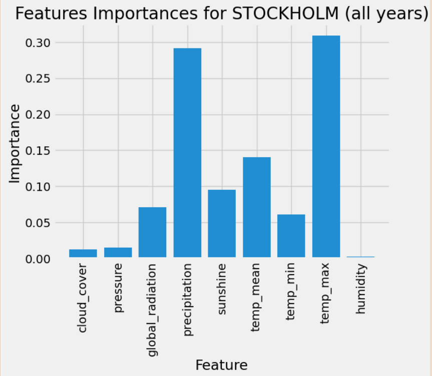
01 Satellite and Radar Imagery: To provide a visual context of changing weather patterns, enabling deep learning models like CNNs to extract meaningful patterns.

02 Real-time Weather Station Data: Allows for immediate updates to models, improving the accuracy of forecasts.

03 Topographical and Climate Zone Data: Essential for understanding how physical geography impacts regional weather patterns.

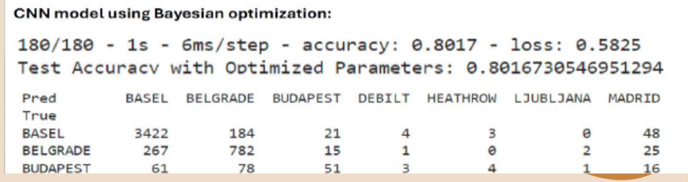
Overview of Machine Learning Approaches





Random Forest Model

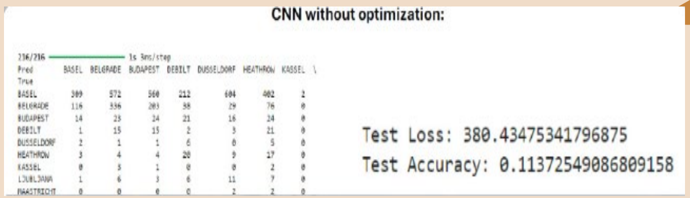
- **Description:** An ensemble learning method used to classify weather conditions, predict safe flight conditions, and analyze variable importance.
- **Results:** Achieved 73% accuracy with Randomized Search CV for multi-station data; 100% accuracy when focused on a single station like Maastricht.
- **Key Features**: Precipitation, temperature metrics, cloud cover, and sunshine.



CNN model using Bayesian optimization:

180/180 - 1s - 6ms/step - accuracy: 0.8017 - loss: 0.5825
Test Accuracy with Optimized Parameters: 0.8016730546951294

Pred / True	BASEL	BELGRADE	BUDAPEST	DEBILT	HEATHROW	LJUBLJANA	MADRID
BASEL	3422	184	21	4	3	0	48
BELGRADE	267	782	15	1	0	2	25
BUDAPEST	61	78	51	3	4	1	16



CNN without optimization:

236/236 ━━━━━━━━━━ 1s 3ms/step

Pred / True	BASEL	BELGRADE	BUDAPEST	DEBILT	DUSSELDORF	HEATHROW	KASSEL \
BASEL	309	572	560	212	604	402	2
BELGRADE	116	336	203	38	29	76	0
BUDAPEST	14	23	24	21	16	24	0
DEBILT	1	15	15	2	3	21	0
DUSSELDORF	2	1	1	6	0	5	0
HEATHROW	3	4	4	20	9	17	0
KASSEL	0	5	1	0	0	2	0
LJUBLJANA	1	6	3	6	11	7	0
MAASTRICHT	0	0	0	0	2	2	0

Test Loss: 380.43475341796875
Test Accuracy: 0.11372549086809158

Deep Learning with CNNs

- **Description:** CNNs were applied to classify weather conditions based on radar and satellite imagery.

- **Results:** Improved test accuracy to 80.17% through Bayesian optimization.

- **Key Features:** Enabled analysis of complex spatial patterns in weather images.

Predicting Anomalies with Random Forest

Hypothesis: A random forest model can identify abnormal weather trends based on historical patterns and station data.

Approach:

• Used RandomForestClassifier to analyze key features like precipitation and temperature across various European weather stations.

• Applied **RandomizedSearchCV** for hyperparameter tuning, optimizing parameters such as **n_estimators, max_depth, and min_samples_split to** improve model accuracy.

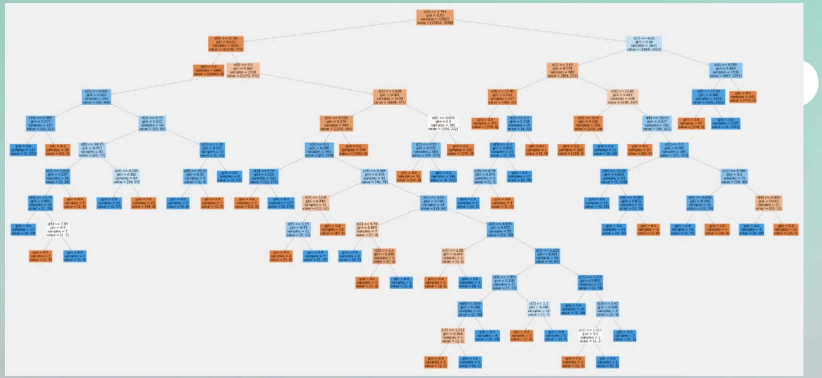
Model Used: RandomForestClassifier with optimized hyperparameters.

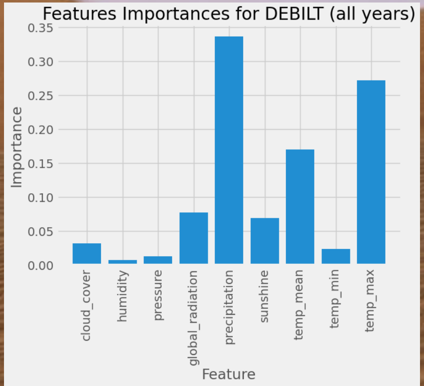
Result:

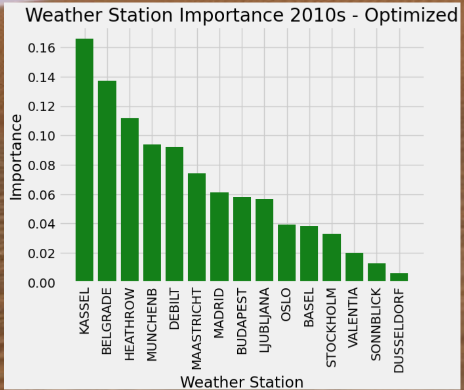
• **Accuracy:** Improved from an initial 71.2% to approximately 72% after hyperparameter optimization.

• **Feature Importance:** The most predictive features were Kassel, Belgrade and Heathrow stations, highlighting the areas with significant weather variation.

Conclusion: This approach successfully identified areas experiencing deviations from historical patterns, helping to detect increasing anomalies like shifts in precipitation and temperature extremes.







Deep Learning for ImageBased Weather Classification

Hypothesis: CNNs can better interpret radar and satellite imagery to classify weather conditions, improving the prediction of weather trends.

Approach:

• Developed a CNN model to classify radar images of various weather conditions (e.g., cloudy, rainy, sunny).

• Used Bayesian optimization to refine hyperparameters like the number of neurons, batch size, and learning rate for better accuracy.

Model Used: CNN with Bayesian optimization.

Result:

• Initial Accuracy: The unoptimized CNN achieved around 11% accuracy.

• Optimized Model Accuracy: After Bayesian optimization, accuracy improved significantly to 80%.

• Confusion Matrix: Showcases the model's ability to differentiate between weather conditions such as 'cloudy' and 'rainy,' highlighting areas where classification errors reduced after optimization.

Conclusion: This model demonstrated the potential for analyzing complex visual data, making it useful for predicting shifts in weather patterns.

CNN without optimization:

```
216/216 ━━━━━━━━━━ 1s 5ms/step
```

Pred / True	BASEL	BELGRADE	BUDAPEST	DEBILT	DUSSELDORF	HEATHROW	KASSEL \
BASEL	389	572	560	212	604	402	2
BELGRADE	116	336	203	38	29	76	0
BUDAPEST	14	23	24	21	16	24	0
DEBILT	1	15	15	2	3	21	0
DUSSELDORF	2	1	1	6	0	5	0
HEATHROW	3	4	4	20	9	17	0
KASSEL	0	3	1	0	0	2	0
LJUBLJANA	1	6	3	6	11	7	0
MAASTRICHT	0	0	0	0	2	2	0
MADRID	40	29	11	22	243	29	0
MUNCHENB	3	0	0	0	0	0	0
OSLO	0	0	0	0	0	0	0
STOCKHOLM	0	0	0	0	0	0	0
VALENTIA	0	0	0	1	1	1	0

```
Test Loss: 380.43475341796875
Test Accuracy: 0.11372549086809158
```

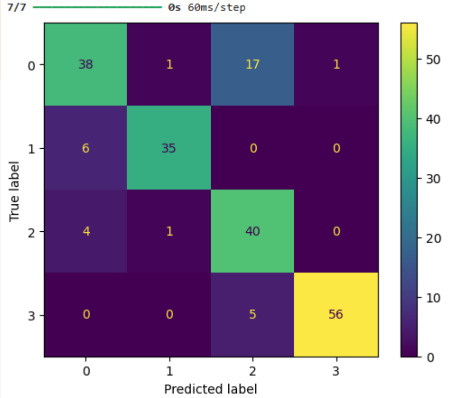
CNN model using Bayesian optimization:

```
180/180 - 1s - 6ms/step - accuracy: 0.8017 - loss: 0.5825
Test Accuracy with Optimized Parameters: 0.8016730546951294
```

Pred / True	BASEL	BELGRADE	BUDAPEST	DEBILT	HEATHROW	LJUBLJANA	MADRID
BASEL	3422	184	21	4	3	0	48
BELGRADE	267	782	15	1	0	2	25
BUDAPEST	61	78	51	3	4	1	16
DEBILT	37	16	10	12	2	0	5
DUSSELDORF	14	8	3	1	1	0	2
HEATHROW	29	10	7	1	12	1	22
KASSEL	6	3	2	0	0	0	0
LJUBLJANA	22	11	4	0	0	9	15
MAASTRICHT	9	0	0	0	0	0	0
MADRID	66	60	14	2	2	2	312
MUNCHENB	8	0	0	0	0	0	0
OSLO	0	1	2	0	0	0	2
STOCKHOLM	1	2	1	0	0	0	0
VALENTIA	1	0	0	0	0	0	0

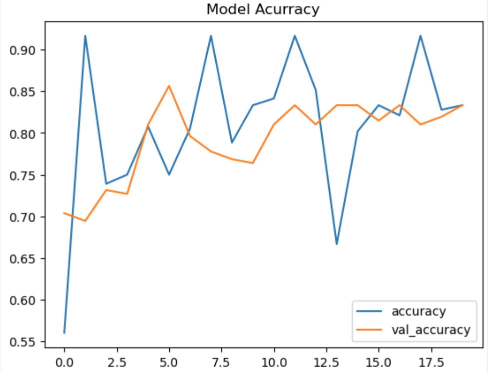
Model Performance Evaluation Using CNN

```
Accuracy: 0.8333333134651184, Val_Accuracy: 0.8333333134651184
Loss: 0.07480228692293167, Val_Loss: 0.05640774592757225
```



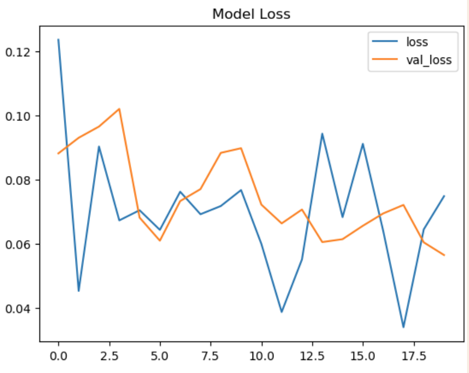
Confusion Matrix

Shows how well the model classifies weather types. Most predictions are accurate, but some classes are confused with others.



Accuracy Plot

Tracks accuracy over epochs. Higher values indicate better performance; fluctuations suggest varying model stability.



Loss Plot

Displays training and validation loss. Lower loss indicates better fit, with spikes showing areas for improvement.

GANs for Synthetic Weather Projections

Hypothesis:

GANs can generate new weather scenarios based on current trends, providing a range of possible futures.

Approach:

Use GANs to simulate various weather patterns and scenarios, such as temperature shifts or precipitation changes over time. Create synthetic weather maps to explore long-term predictions (e.g., 25 to 50 years).

Expected Outcome:

A tool that visualizes potential future climates, allowing stakeholders to plan for extreme scenarios.

Summary and Recommendations

Recommendation

Use Random Forest models for immediate analysis of feature importance and to identify key predictors of abnormal weather patterns.

Implement CNNs for analyzing satellite data and weather imagery to improve real-time classification of weather conditions.

Invest in developing GANs for longer-term scenario simulation, helping ClimateWins plan for potential future climates.

Next Steps

Validate model predictions with real-time data.

Collaborate with meteorological agencies for improved data access.

Explore scaling the models to include more complex climate data.

THANKS!

Do you have any questions?

iaspilcueta@gmail.com