Uses of Machine Learning in Predicting Climate Change

Deepa Sury

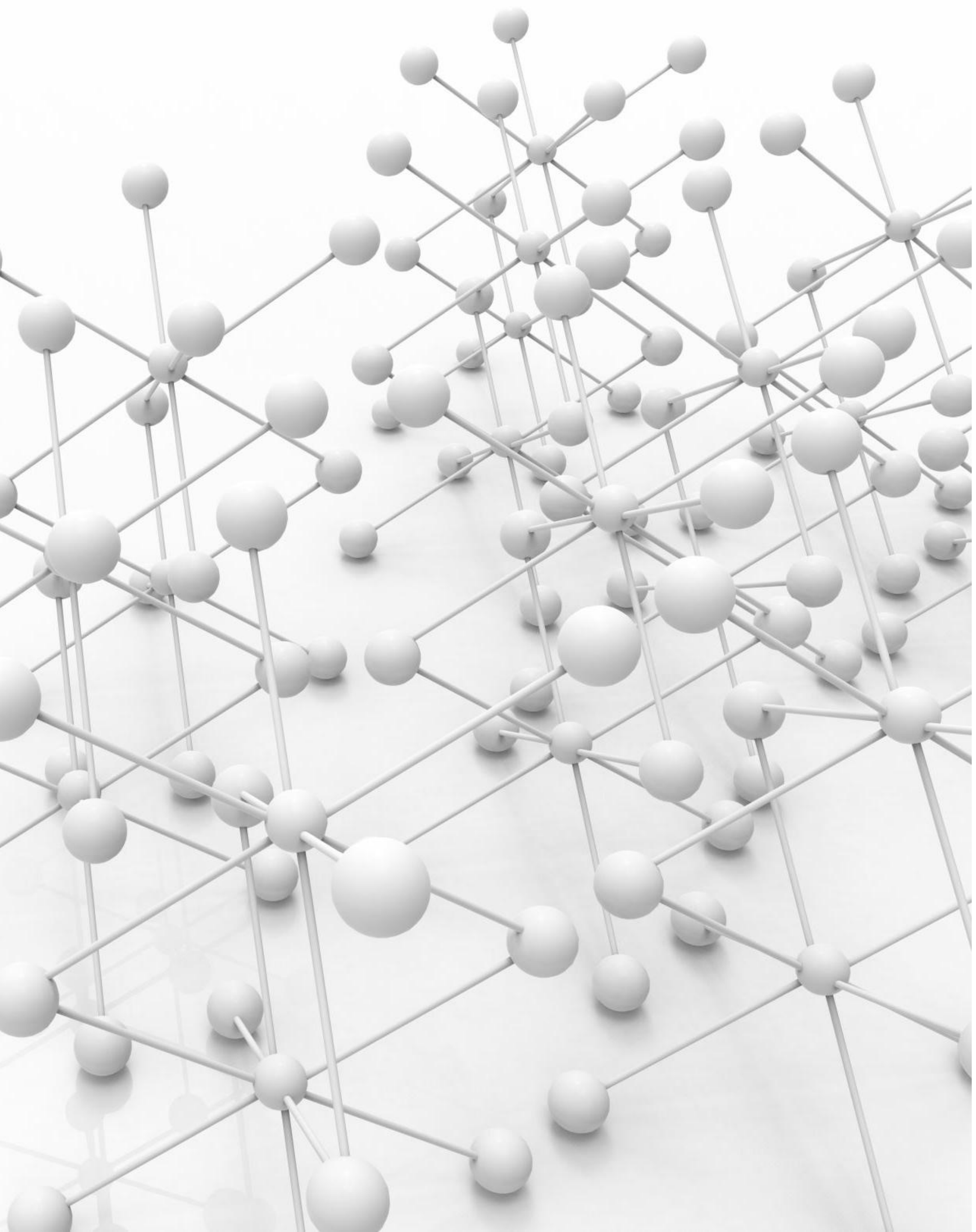
03.06.2025

ClimateWins Objectives

- Identifying weather patterns outside the regional norm in Europe.
- Determining if unusual weather events are increasing in frequency.
- Generate possibilities for weather conditions in the next 25 to 50 years based on current trends.
- Using these predictions, determine the safest places for people to live in Europe in the near future.

Using Machine Learning to Predict Climate Change: 3 Thought Experiments to Consider

- Can a Random Forest model determine whether the weather on a given day is pleasant or not?
- Can a neural network correctly identify and classify weather radar images?
- Using images taken outdoors, can we train a GAN to generate accurate predictions for future weather patterns?

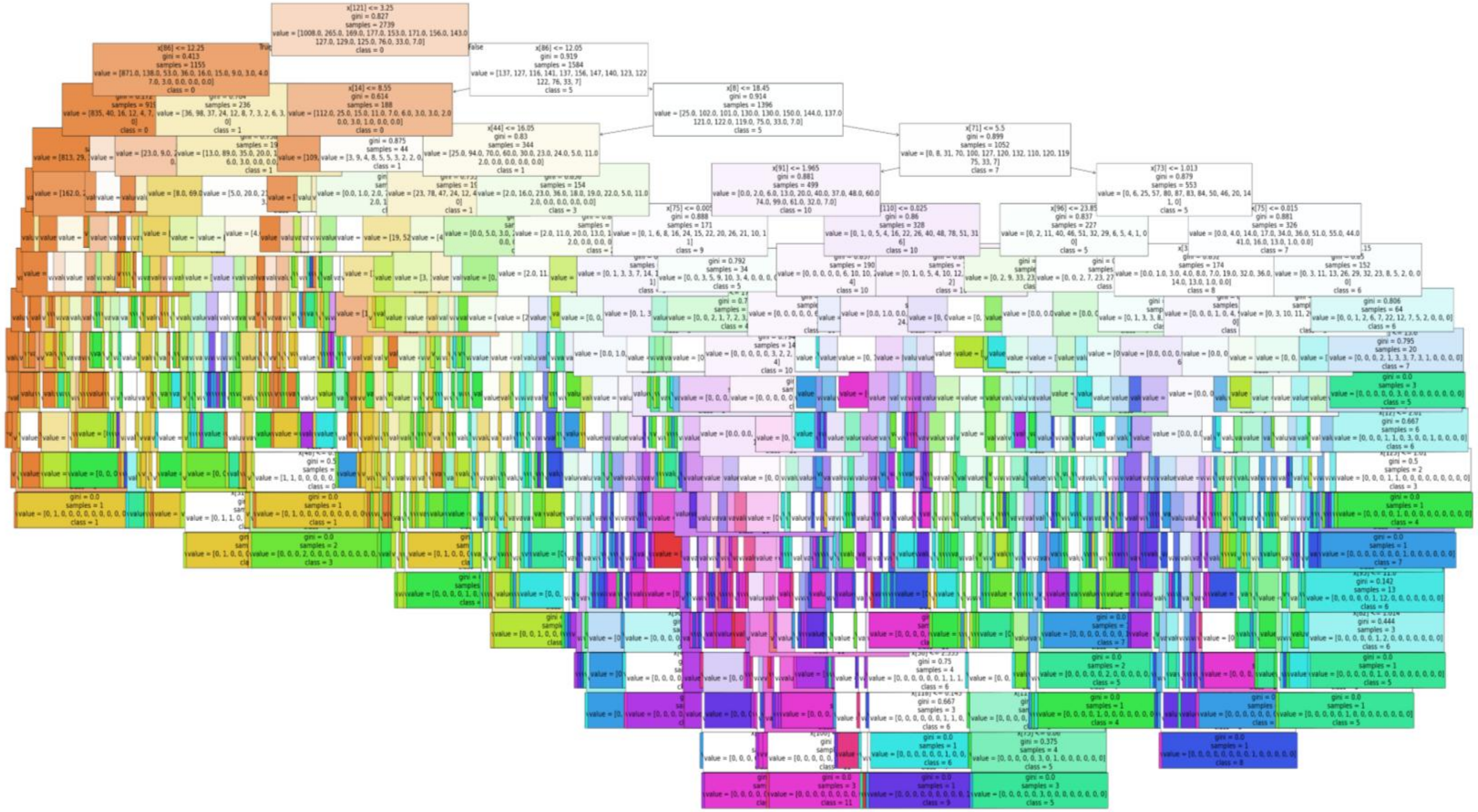


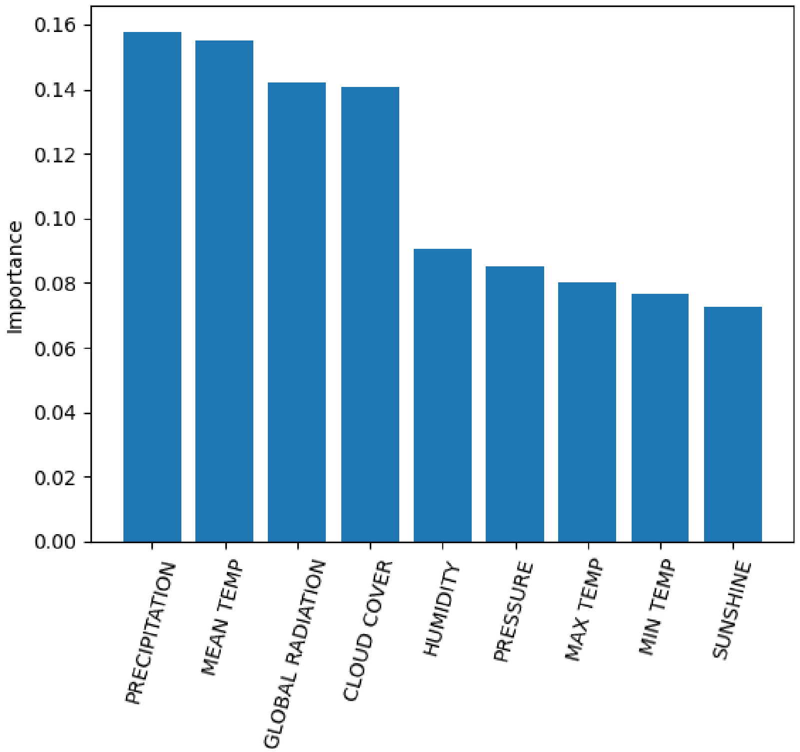
Types of Machine-Learning Algorithms Used

- Random Forest models – collections of decision-trees that solve a single problem using the average of their solutions: an example of ensemble learning.
- Convolutional neural-networks (CNNs) – a type of deep-learning model particularly suited to interpreting visual data.
- Generative adversarial networks (GANs) – using two neural-networks working against each other, a GAN creates artificial data based on real inputs.

Experiment 1: Classifying Pleasant Weather

This experiment uses European weather data and survey data from the same locations to determine whether a random forest model can identify pleasant weather.

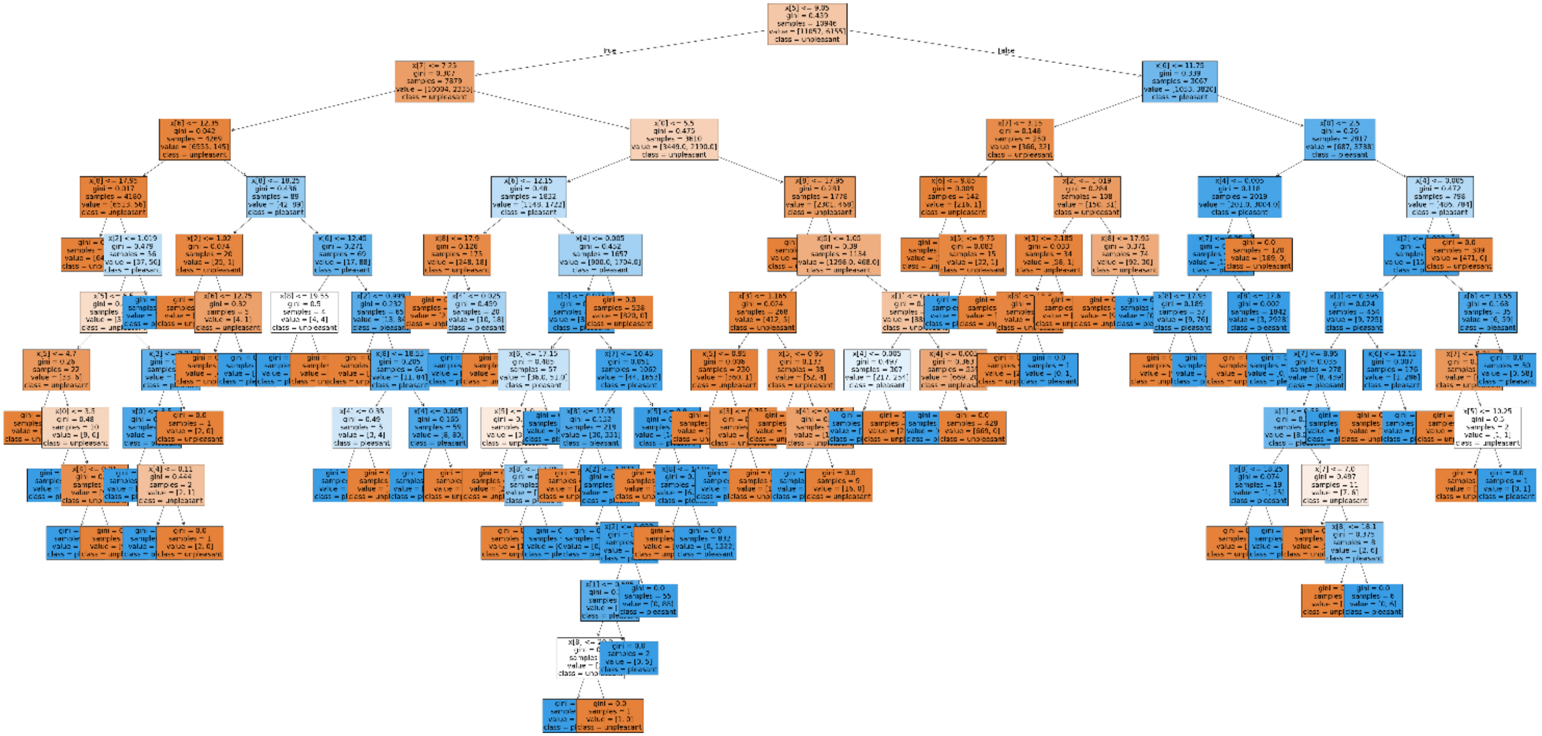


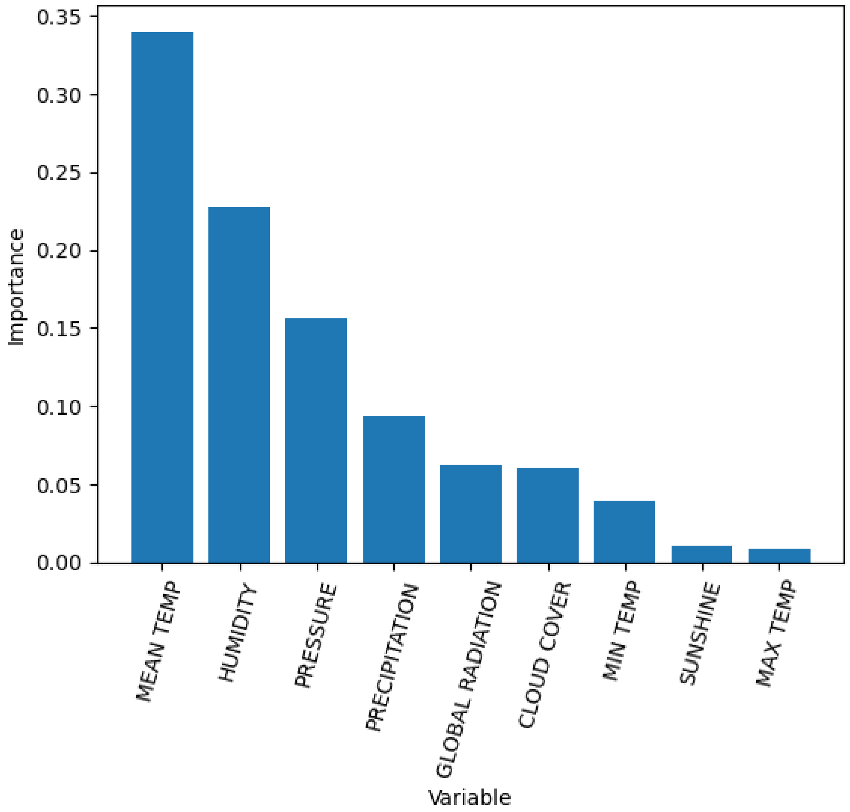


One potential drawback of this type of model is that, particularly with an extensive set of data, the final results can be messy and hard to interpret.

Viewing the feature importance after running this model can also give us an idea about which weather conditions contribute to its pleasantness.
Here we can see that precipitation, daily mean temperature, global radiation levels and cloud cover play the biggest role in determining if the weather is pleasant.

Experiment 1: Classifying Pleasant Weather





Random Forest classifier using data from one station (Belgrade) between 1960 and 2022. The results are somewhat easier to interpret here compared to showing all weather stations, even over a limited period of time.

Running the model for individual locations brings up unique results. Here we can see that the most important weather conditions influencing the Belgrade survey answers are mean temperature, humidity and atmospheric pressure.

Experiment 1: Classifying Pleasant Weather

- Survey answers tend not to be entirely objective and subject to regional bias due to differing cultural attitudes (e.g. what one country may consider pleasant weather would be unpleasant to another).
- Because of this, as well as the varying nature of weather data across a continent, this type of model would be best applied to a smaller region, such as a city or state.
- When applied properly, this model can be used in combination with other relevant data sets to assess other criteria. For example, weather data can be analyzed alongside healthcare data to determine the conditions that pose the biggest health risk for outdoor activities.

Experiment 2: Using a Neural Network to Identify Types of Weather

- A convolutional neural network (CNN) is the ideal type of deep-learning model to identify and interpret complex data.
- I tested a CNN on both visual and numerical data and found that it was able to identify both numerical and visual data with reasonably high accuracy.

Confusion matrices for a CNN trained on weather data from 15 European stations.
With the optimal set of hyperparameter values (right), the model recognized fewer stations, but showed more accurate predictions than with the initial set of hyperparameter values (left).

Pred	BASEL	BELGRADE	BUDAPEST	DEBILT	HEATHROW	KASSEL	LJUBLJANA	\
True								
BASEL	3	1216	309	132	4	3	1431	
BELGRADE	0	783	31	0	0	0	276	
BUDAPEST	0	141	9	0	0	0	64	
DEBILT	0	55	3	0	0	0	23	
DUSSELDORF	0	12	1	0	0	0	16	
HEATHROW	0	26	2	1	0	0	49	
KASSEL	0	8	1	0	0	0	2	
LJUBLJANA	0	28	2	0	0	0	31	
MAASTRICHT	0	2	0	0	0	0	6	
MADRID	0	126	13	13	0	0	253	
MUNCHENB	0	5	1	0	0	0	2	
OSLO	0	3	0	0	0	0	1	
STOCKHOLM	0	3	0	0	0	0	1	
VALENTIA	0	0	0	0	0	0	0	

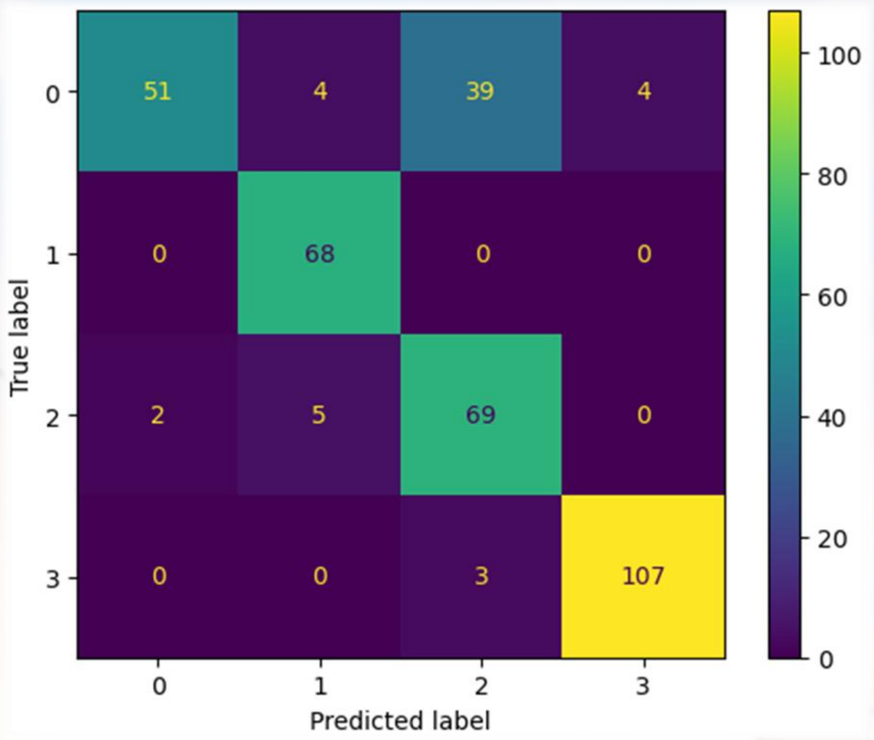
Pred	MAASTRICHT	MADRID	MUNCHENB	OSLO	SONNBLICK	STOCKHOLM	VALENTIA
True							
BASEL	59	294	4	205	10	1	11
BELGRADE	0	2	0	0	0	0	0
BUDAPEST	0	0	0	0	0	0	0
DEBILT	0	1	0	0	0	0	0
DUSSELDORF	0	0	0	0	0	0	0
HEATHROW	1	3	0	0	0	0	0
KASSEL	0	0	0	0	0	0	0
LJUBLJANA	0	0	0	0	0	0	0
MAASTRICHT	0	1	0	0	0	0	0
MADRID	5	48	0	0	0	0	0
MUNCHENB	0	0	0	0	0	0	0
OSLO	0	1	0	0	0	0	0
STOCKHOLM	0	0	0	0	0	0	0
VALENTIA	1	0	0	0	0	0	0

Pred	BASEL	BELGRADE	BUDAPEST	DEBILT	DUSSELDORF	HEATHROW	KASSEL	\
True								
BASEL	3561	66	6	10	7	7	0	
BELGRADE	120	944	7	4	1	2	0	
BUDAPEST	19	23	146	12	1	5	0	
DEBILT	9	5	7	56	2	1	0	
DUSSELDORF	4	2	1	2	5	13	0	
HEATHROW	8	2	1	1	5	55	0	
KASSEL	5	2	1	0	3	0	0	
LJUBLJANA	12	2	4	0	0	2	1	
MAASTRICHT	2	0	0	0	0	1	0	
MADRID	38	11	8	0	0	5	2	
MUNCHENB	5	1	0	0	0	0	0	
OSLO	3	0	0	0	0	0	0	
STOCKHOLM	2	0	0	0	0	0	0	
VALENTIA	1	0	0	0	0	0	0	

Pred	LJUBLJANA	MADRID	MUNCHENB	OSLO
True				
BASEL	1	22	1	1
BELGRADE	3	11	0	0
BUDAPEST	1	6	1	0
DEBILT	1	1	0	0
DUSSELDORF	0	2	0	0
HEATHROW	1	8	0	1
KASSEL	0	0	0	0
LJUBLJANA	26	13	0	1
MAASTRICHT	0	6	0	0
MADRID	5	389	0	0
MUNCHENB	0	0	2	0
OSLO	0	2	0	0
STOCKHOLM	0	0	1	1
VALENTIA	0	0	0	0

Experiment 2: Using a Neural Network to Identify Types of Weather

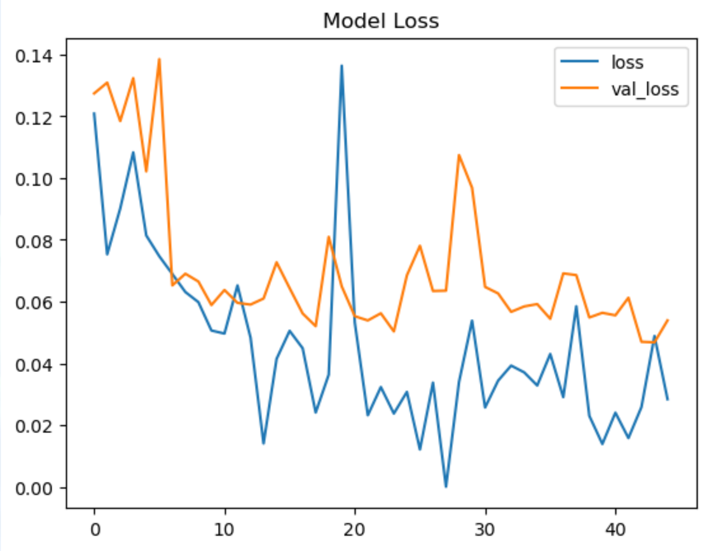
- A similar type of CNN can be trained on historical data taken from weather satellites and radar stations across Europe. CNNs work particularly well with visual data, as they can easily identify edges, shapes and distinct colours .
- This model would be useful for identifying unusual weather conditions and the frequency of their occurrence throughout the region.



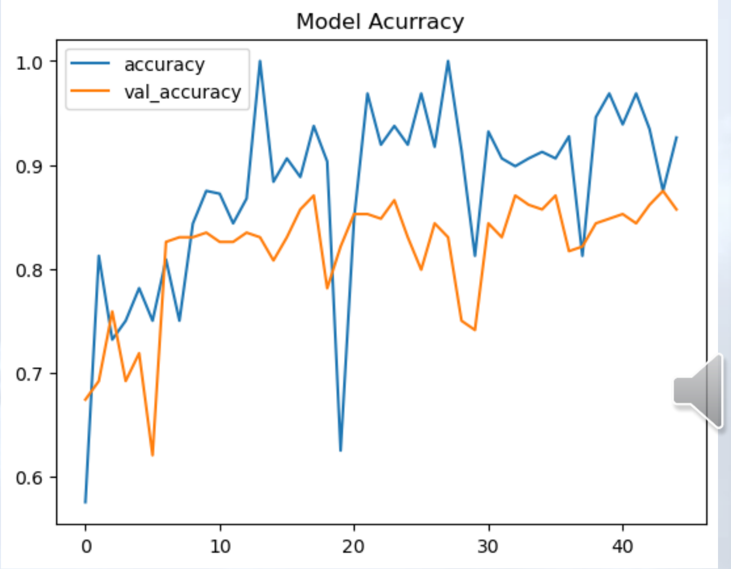
This confusion matrix shows the accuracy of a CNN's guesses after training with image data.
We can see that the model identified almost all images with class 3 correctly (107 out of 110, or 97.3%), and only slightly more than half of those with class 0 correctly (51 out of 98, or 52%).

Experiment 3: Generating Predictions of Future Weather Conditions

- This experiment would apply the CNN mentioned previously and incorporate it as part of a generative adversarial network (GAN) – a model comprised of two neural networks working against each other to produce realistic artificial data based on real inputs.
- The model works as one of the networks – the generator – creates artificial data based on real inputs, while the other network – the discriminator – evaluates the generator's output by comparing it to the real data. A conclusion is reached once the discriminator is no longer able to distinguish between the real data and that produced by the generator.



The loss and accuracy of both components of the model. The levels for both components move closer and further apart as the overall loss decreases and accuracy increases. The ideal outcome of this model is for both lines to converge at a point with maximum accuracy and minimal loss.



Experiment 3: Generating Predictions of Future Weather Conditions

- With the right amount of tuning (e.g. optimized hyperparameter values), a GAN trained on sufficient historical data would be able to produce a highly-accurate visualization of weather conditions within the next 25 to 50 years.
- It is crucial that the model be trained to distinguish types of weather correctly, particularly when identifying outdoor photos, as certain weather conditions are not as easily-identifiable as others.

2 examples of the neural network incorrectly identifying weather conditions in photos retrieved from Kaggle .

Conclusion

- Using survey answers alongside weather data would work well on a smaller, more local level, as this would account for factors such as unique climate conditions or, in the case of the survey answers, cultural attitudes towards the weather.
- Radar and satellite images are relatively easy to identify. Research shows that there are already many examples of using neural networks to analyze and identify this kind of data. This would be useful in determining if unusual weather conditions have increased in frequency in recent times.
- Using images created with a GAN would be helpful in visualizing potential weather conditions in the next 25 to 50 years, as well as determining which locations would be affected the most by extreme weather.

Want to know more?
Check out my GitHub

THANK YOU