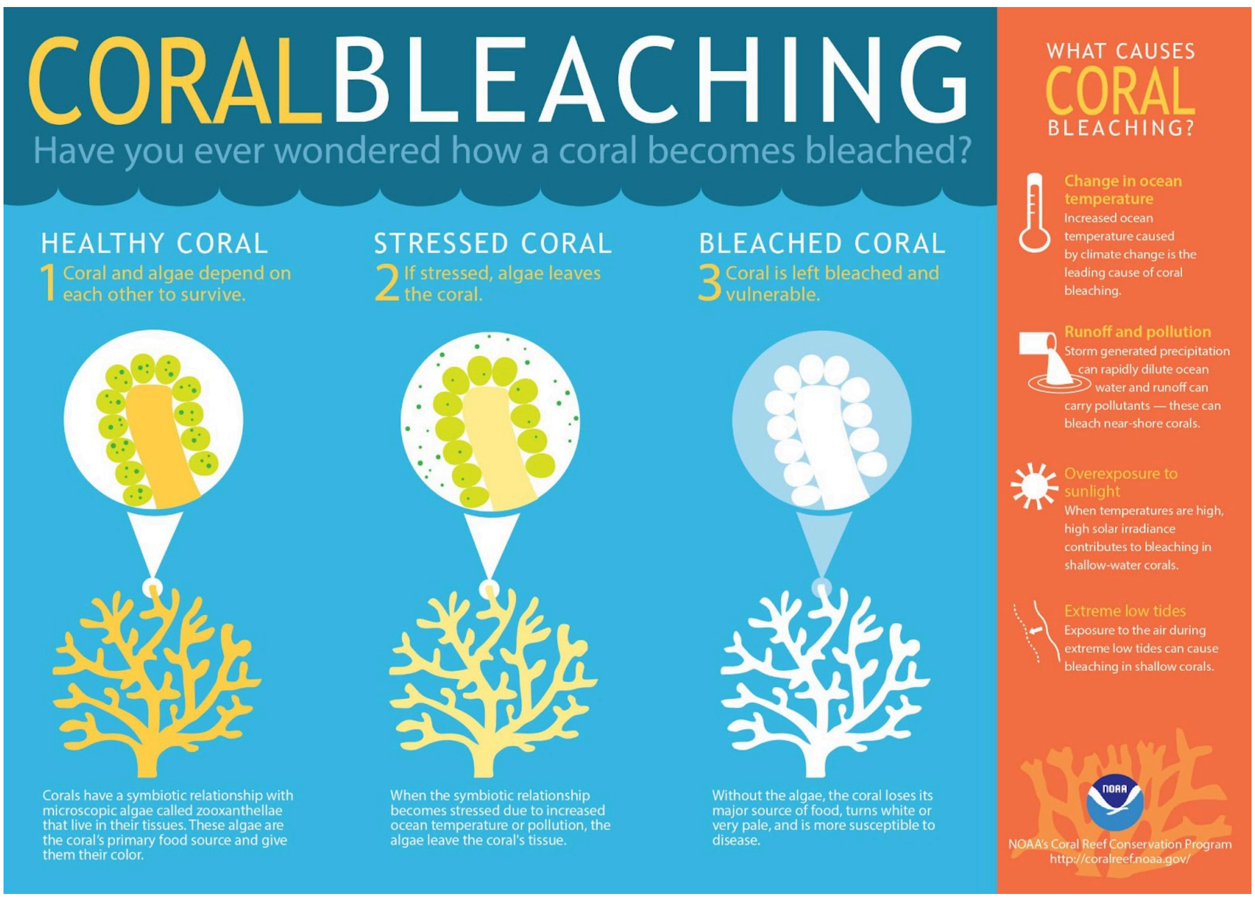
SAKSHI HIRE
Programming Test

Coral Bleaching

Corals and zooxanthellae (algae) have a symbiotic relationship, meaning they both benefit from the association. Corals provide a safe environment for the algae, and in return the algae are responsible for providing corals with color and a major source of food through photosynthesis.

Various factors can stress corals, including rising water temperatures, changes in water quality (Salinity, pollution, and other factors), extreme weather events (Storms and other events). When corals are stressed, they expel the algae, leaving their white skeletons visible. This is called Coral Bleaching.



When a coral bleaches, it is not dead. Corals can survive a bleaching event, but they are under more stress and are subject to mortality.

Reason for Coral bleaching

In 2005, the U.S. lost half of its coral reefs in the Caribbean in one year due to a massive bleaching event. The warm waters centered around the northern Antilles near the Virgin Islands and Puerto Rico expanded southward.
Not all bleaching events are due to warm water.
In January 2010, cold water temperatures in the Florida Keys caused a coral bleaching event that resulted in some coral death. Water temperatures dropped -6.7 degrees Celsius 12.06 degrees Fahrenheit lower than the typical temperatures observed at this time of year. Researchers will evaluate if this cold-stress event will make corals more susceptible to disease in the same way that warmer waters impact corals.
There might also be other factors like pH, marine heatwaves, etc, along with SST(Sea Surface water temperature) affecting this Coral Bleaching process. Also might be that the total number of species in the reservoir might have some influence in coral bleaching.

We are going to find this using Machine learning , where we are going to use various factors to see what influences coral bleaching.

Bleaching severity

Bleaching severity refers to the extent and intensity of coral bleaching. It takes categorical variables: None, Low, Medium, High.

Dataset

I took the Dataset from https://www.kaggle.com/ . My Dataset was named there as *Shifting Seas: Ocean Climate & Marine Life Dataset.*
This Dataset has 9 columns. They are

1. Date: Date of observation
2. Location: Marine location name (e.g., Great Barrier Reef, Maldives)
3. Latitude
4. Longitude
5. SST (°C): Sea Surface Temperature in degrees Celsius
6. pH Level: Acidity level of seawater (lower means more acidic, a sign of acidification)
7. Bleaching Severity: categorical variable: None, Low, Medium, High
8. Species Observed: count of marine species observed during the sampling period
9. Marine Heatwave: boolean flag (True/False) indicating whether SST > 30°C

Programming Task 1

We don’t consider date and location names as they are no way related to Coral bleaching.
Input features include:

Input Feature	Type	Notes
Latitude	Numerical	May correlate with temperature
Longitude	Numerical	May correlate with temperature
SST (°C)	Numerical	Directly related to bleaching
pH Level	Numerical	Indicates acidification, which may relate to stress
Species Observed	Numerical	Biodiversity might be impacted by bleaching
Marine Heatwave	Boolean (0 or 1)	Important indicator of bleaching likelihood

Output Feature : Bleaching Severity where we map None to 0, Low to 1, Medium to 2 and High to 3.

Programming Task 2

Input features include:

Input Feature	Type	Notes
SST (°C)	Numerical	Directly related to bleaching
pH Level	Numerical	Indicates acidification, which may relate to stress

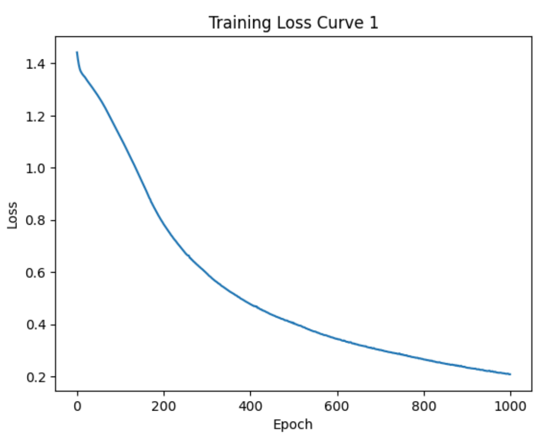
As I consider that Latitude and Longitude is just explaining temperature variation which already present in the input feature as SST. Also Marine Heatwave is true or false for SST > 30°C. So we

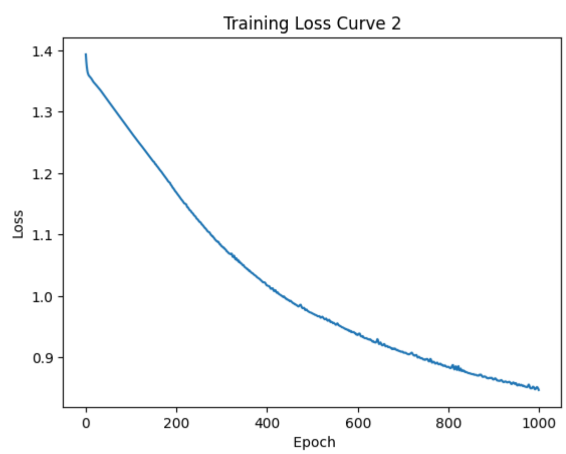
neglect using Marine heatwaves, Latitude and Longitude as it looks like we are reconsidering SST 4 times.
Output Feature : Bleaching Severity where we map None to 0, Low to 1, Medium to 2 and High to 3.

RESULTS A

TRAINING LOSS

Programming Task 1	Programming Task 2
Epoch: 0 and loss :1.4420061111450195 Epoch: 100 and loss :1.1182349920272827 Epoch: 200 and loss :0.7831521034240723 Epoch: 300 and loss :0.593362033367157 Epoch: 400 and loss :0.47636935114860535 Epoch: 500 and loss :0.40481454133987427 Epoch: 600 and loss :0.34382694959640503 Epoch: 700 and loss :0.3008968234062195 Epoch: 800 and loss :0.26452717185020447 Epoch: 900 and loss :0.23309123516082764	Epoch: 0 and loss :1.3933470249176025 Epoch: 100 and loss :1.266249418258667 Epoch: 200 and loss :1.1680711507797241 Epoch: 300 and loss :1.0808213949203491 Epoch: 400 and loss :1.0171141624450684 Epoch: 500 and loss :0.971488893032074 Epoch: 600 and loss :0.9364941120147705 Epoch: 700 and loss :0.9083477854728699 Epoch: 800 and loss :0.8847343921661377 Epoch: 900 and loss :0.8641185164451599





INFERENCE A

Compared to both the cases, we see the model 1 is learning well as loss approaches rapidly to 0 quickly. This means case 1 is favoured over case 2. This means the additional features provide meaningful signals to separate classes of Bleaching Severity (case 1) than just SST and pH(case 2).
Therefore, SST and pH two features are insufficient by themselves for accurate classification. Therefore, here case 1 wins. But,

RESULTS B

Features	Programming Task 1	Programming Task 2
Evaluating Model on Test Data set(Test Loss)	10.4223	3.0440
Accuracy	30%	23%
Confusion Matrix	[[10 12 8 7] [4 10 4 5] [8 7 8 0] [7 8 0 2]]	[[12 6 7 3] [11 6 11 3] [7 4 4 4] [7 4 10 1]]

INFERENCE B

In Task 1, High Test Loss compared to training loss(10.42). This indicates that model 1 has overfit. It fits well to only seen data and not on unseen data in given conditions.
Quite low for both (~30% and 23%), implying models are struggling to classify correctly.
Both confusion matrices show some confusion between classes. There are notable off-diagonal entries, meaning the model often misclassified samples.

IMPROVEMENT

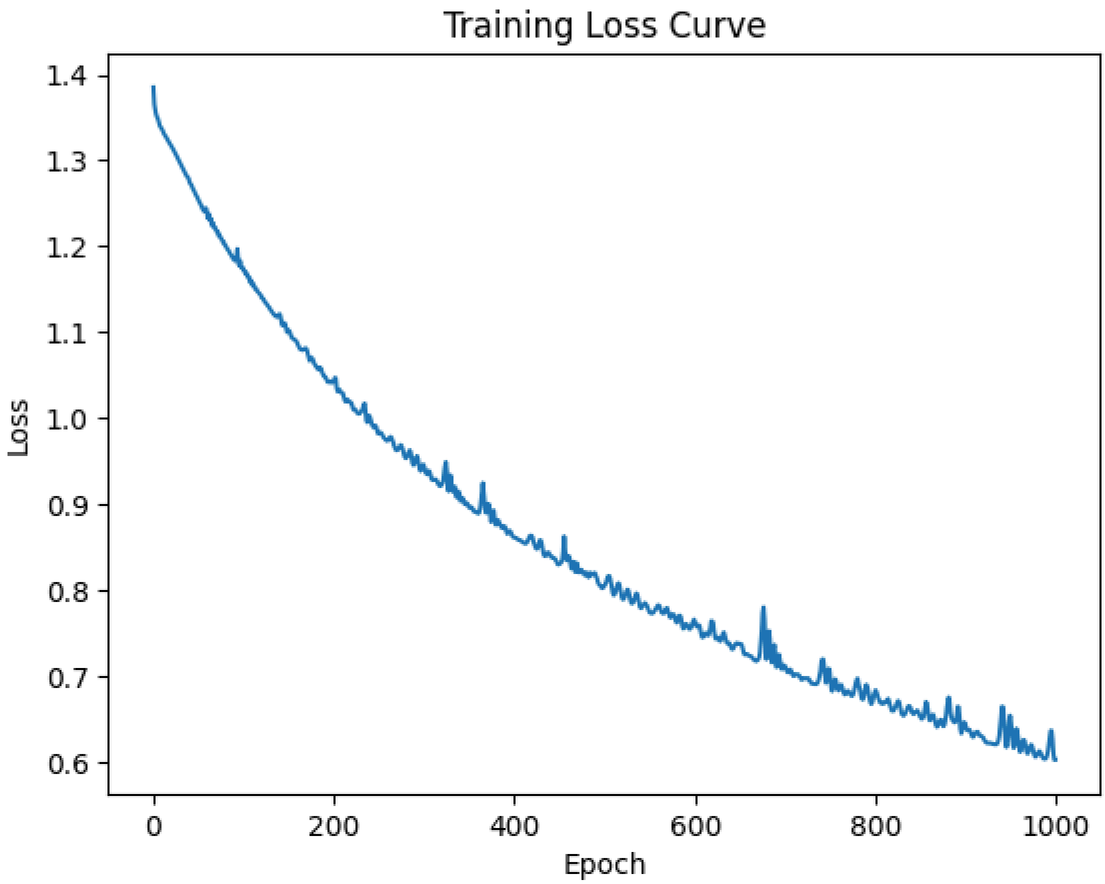
For Task 2, slow loss decrease and low accuracy might mean the model is too simple or the learning rate needs tuning. So, we do Programming Task 3

Programming Task 3

Here we modify the architecture from 2 to 4 hidden layers and increase the number of neurons. I took h1=150, h2=100, h3=880, h4=50.
If we do run for 1000 epoch we get Training Loss as follow:
Epoch: 0 and loss :1.3850839138031006
Epoch: 100 and loss :1.1727712154388428
Epoch: 200 and loss :1.0439242124557495
Epoch: 300 and loss :0.9438678026199341
Epoch: 400 and loss :0.8613293170928955
Epoch: 500 and loss :0.8055977821350098

Epoch: 600 and loss :0.7619010210037231
Epoch: 700 and loss :0.711077094078064
Epoch: 800 and loss :0.6834244728088379
Epoch: 900 and loss :0.6415814757347107

We see loss approaches to 0 as Epoch increases.



Test loss is 3.3422
Accuracy: 33.00%
Confusion Matrix: [[13 7 2 6]
[10 11 5 5]
[6 5 5 3]
[14 1 3 4]]

When we compared all the 3 Cases :

Features	Programming Task 1	Programming Task 2	Improved Programming Task 2
After 1000 epoch Training Loss	0.23	0.86	0.64
Evaluating Model on Test Data set(Loss)	10.42	3.04	3.34
Accuracy	30%	23%	33.00%
Confusion Matrix	[[10 12 8 7] [4 10 4 5] [8 7 8 0] [7 8 0 2]]	[[12 6 7 3] [11 6 11 3] [7 4 4 4] [7 4 10 1]]	[[13 7 2 6] [10 11 5 5] [6 5 5 3] [14 1 3 4]]

INFERENCE

We see by improving architecture i.e., by increasing the number of layers and neurons in each layer, we see progress in our model . We see

1. Data is fitting well on Training data as epoch is approaching 0. After 1000 epochs of Training set it is approaching a moderate value.
2. On Test data, Loss is 3.34 which is lower compared to Task 1 (10.42) , which means our model is not overfitting and getting better.
3. Accuracy has increased to 33%, which is higher among all of them.
4. Also from the confusion matrix we see diagonal elements have increased significantly compared to other cases, suggesting true positives have increased effectively.

CONCLUSION

Therefore, we say that if we increase the neurons in each layer and hidden layers in the architecture we see a significant rise in true positives and accuracy also increases.