50.039: Theory and Practice of Deep Learning

Team members

Name	Student ID
Chan Wei Jian Ivan	1005924
Tom Manuel Opalla Piccio	1006293
Deshpande Sunny Nitin	1006336

GitHub Link: https://github.com/TomPiccio/Lightning_Risk_Prediction

Deep Learning final report

Topic

Our topic focuses on hyper-local inclement weather forecasting (Lightning Risk Warning)

Application:
This project aims to predict Category 1 Lighting Risk Warning (output) in an area to indicate whether outdoor sports activity can resume using weather station data from strategic locations (input).

Problem Statement:
Many Singaporeans enjoy multiple outdoor activities, especially water sports activities including dragon boating, stand-up paddling, canoeing and other water sports. However, these activities are often disrupted by sudden Category 1 Lightning Risk Warnings, which require all outdoor activities to stop immediately. These warnings can be issued without clear visual indicators, as rain and dark clouds do not always precede a Category 1 risk. As a result, participants have little to no time to adjust their plans.

Hence, this project aims to create a model that predicts the Lightning Risk Warning Category a few hours before so people can anticipate whether their activity can continue and whether they should travel towards the area. It should also be able to predict whether they can expect to get more water time a few minutes ahead.

Target Location: PAssion WaVe @ Bedok Reservoir (16S)
This location was chosen because it is one of the water sports centres that is geographically well-covered by surrounding weather stations.

Input/Output:

Input:

- Rainfall
- Wind speed
- Relative Humidity
- Air Temperature
- Wind Direction
- Date and Time (Current & Target)
- Weather Station Coordinates

Output:

- Lightning Risk Warning Category (0 for Clear or 1 for Lightning present) at 5 different time steps: 0 minutes or current time, 30 minutes, 60 minutes, 90 minutes and 120 minutes from the forecast time

Datasets:

1. Lightning Risk Update Telegram (t.me/Lightningrisk) - provides real-time lightning warnings and identifies affected areas using coded data, updated every few minutes.

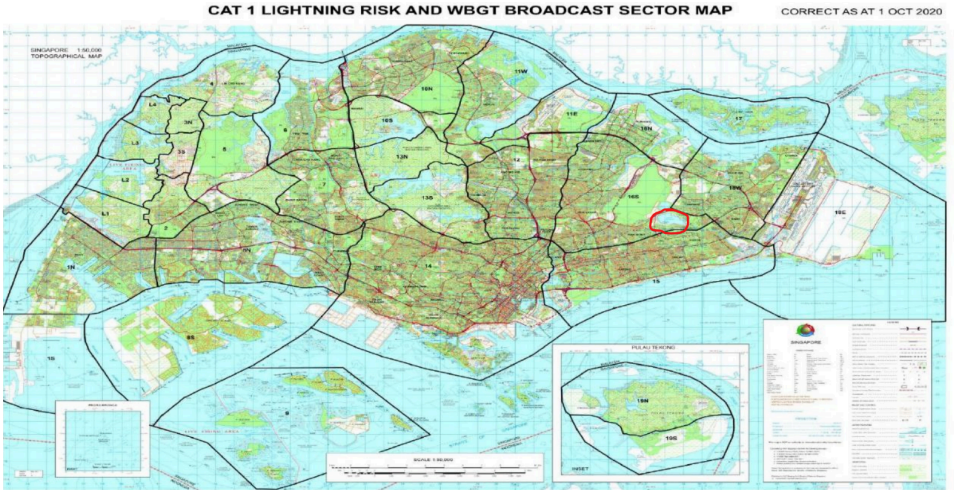


Figure 1: Lightning Risk Warning Zones (target location circled in red)

2. Data Gov SG API (data.gov.sg) - provides access to historical weather data (Rainfall, Wind speed, Relative Humidity, Air Temperature, and Wind Direction) on multiple weather stations in Singapore.

API Weather stations:

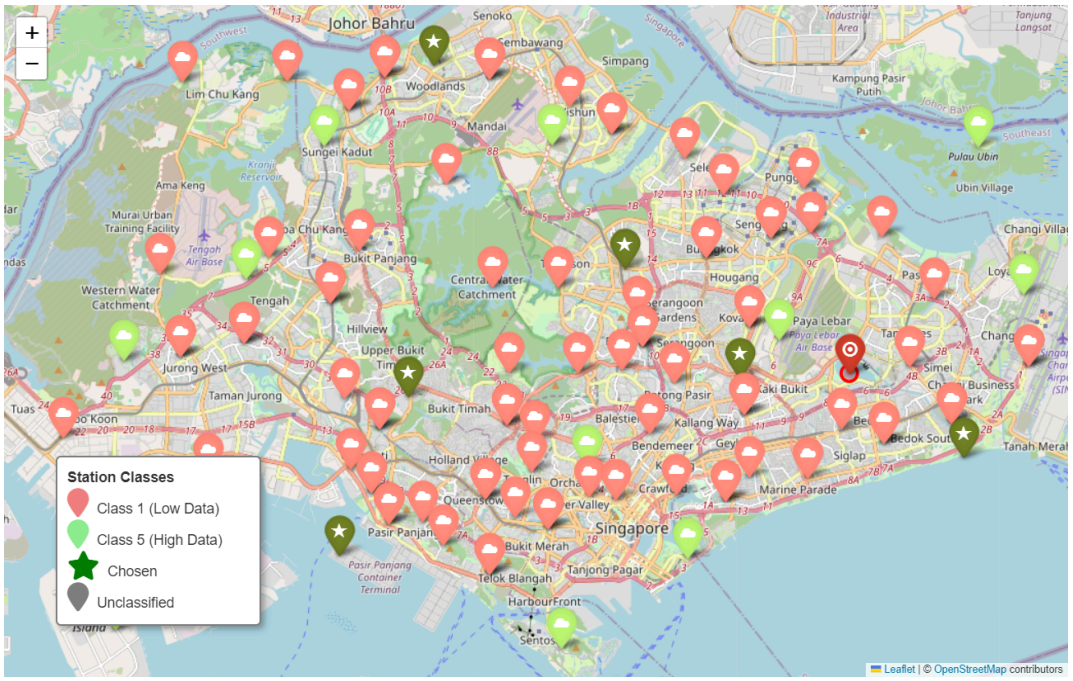


Figure 2: Locations of Weather Stations in Singapore

Data cleaning

We started with 15 stations but we reduced them to 7 different stations due to the large percentage of missing data in the omitted stations

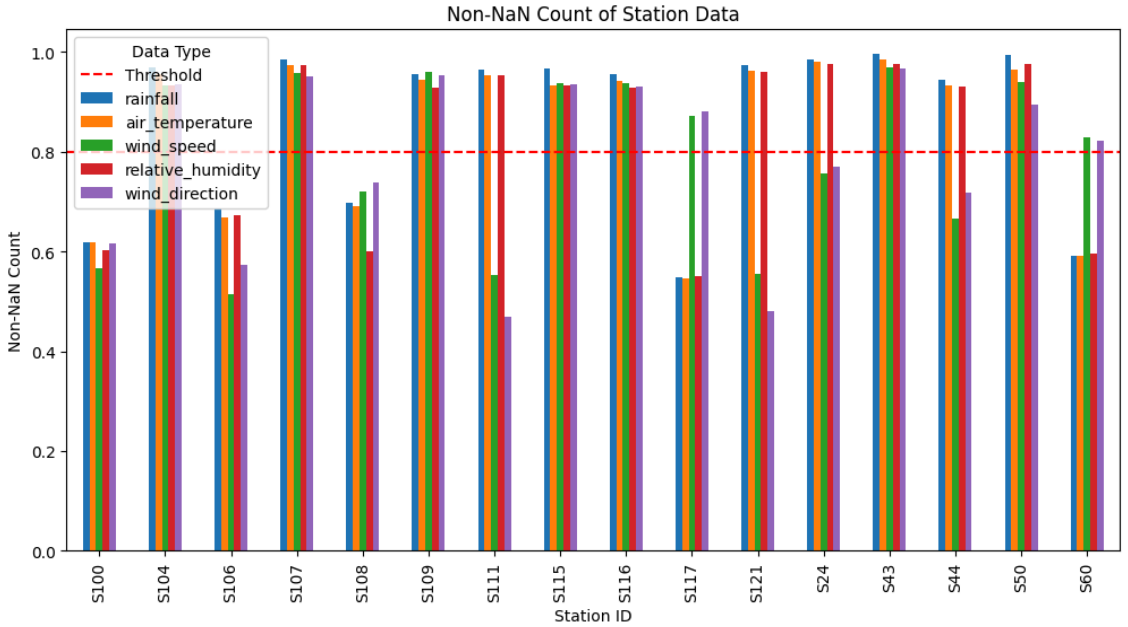
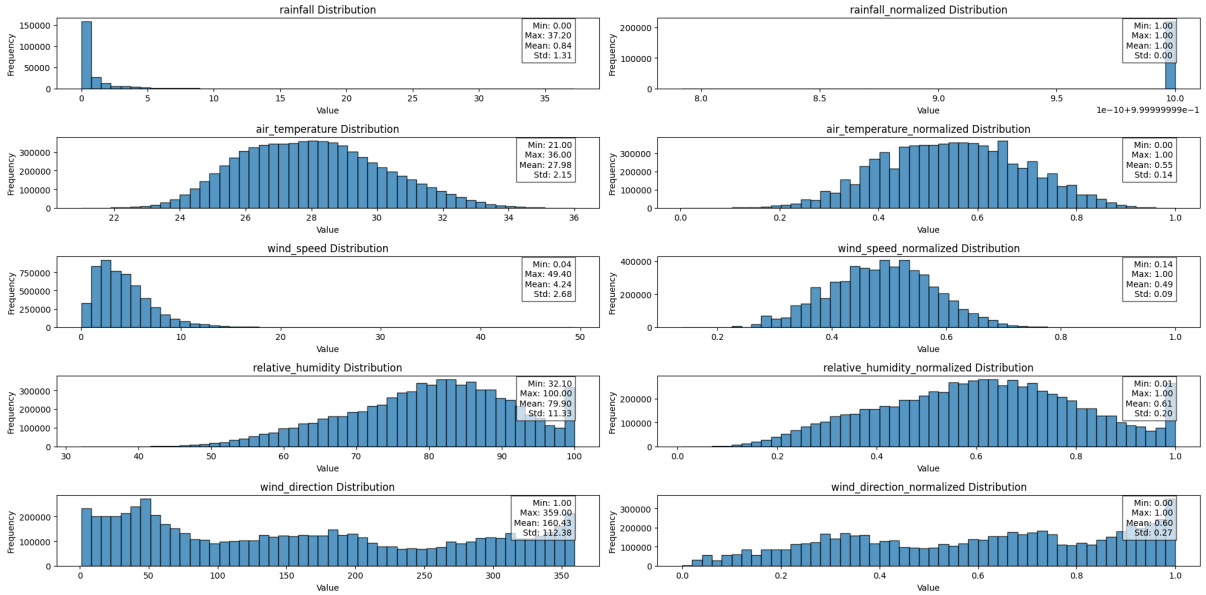


Figure 3: Proportions of Non-NaN data from different datasets per station

We also analysed the data to see that they were not normally distributed. Hence we did the box cox normalisation to normalise them from 0 to 1, hence this would improve the quality of our model training.



Architecture Draft

According to Wang, Hu, Wu, and Zhou (2023), lightning prediction methods commonly employ numerical prediction models, traditional machine learning, convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid neural networks. Traditional methods like support vector machines (SVM), decision trees (DT), and basic artificial neural networks (ANN) use hand-crafted features derived from meteorological parameters. CNNs extract spatial features from satellite images, radar, and acoustic signals, and RNNs, particularly LSTMs, handle temporal data. Hybrid models combine CNNs and LSTMs for better prediction. Common datasets include meteorological parameters, satellite images, electric field data, and WRF simulations. Unlike these rich datasets, the data available for this research is sparse, consisting of only rainfall, wind speed, and relative humidity from various locations, creating an opportunity to develop innovative approaches for lightning prediction.

We have come up with multiple models to address this problem:

- CNN-LSTM
- CNN-RNN
- CNN-GRU
- LSTM multi-label -> we used a 2 hour sequence of data to make a prediction
- LSTM multi-label with attention -> we added in soft attention which is a weighted sum of all the time steps to the prediction.

Results

This is the summary of our results

Model	Test F1	Test Accuracy
LSTM multi-label	0.37	0.85
LSTM multi-label with attention	0.33	0.85
CNN-RNN	0.3	0.93
CNN-GRU	0.3	0.70+

References

Wang, X., Hu, K., Wu, Y., & Zhou, W. (2023). A Survey of Deep Learning-Based Lightning Prediction. Atmosphere, 14(11), 1698. https://doi.org/10.3390/atmos14111698