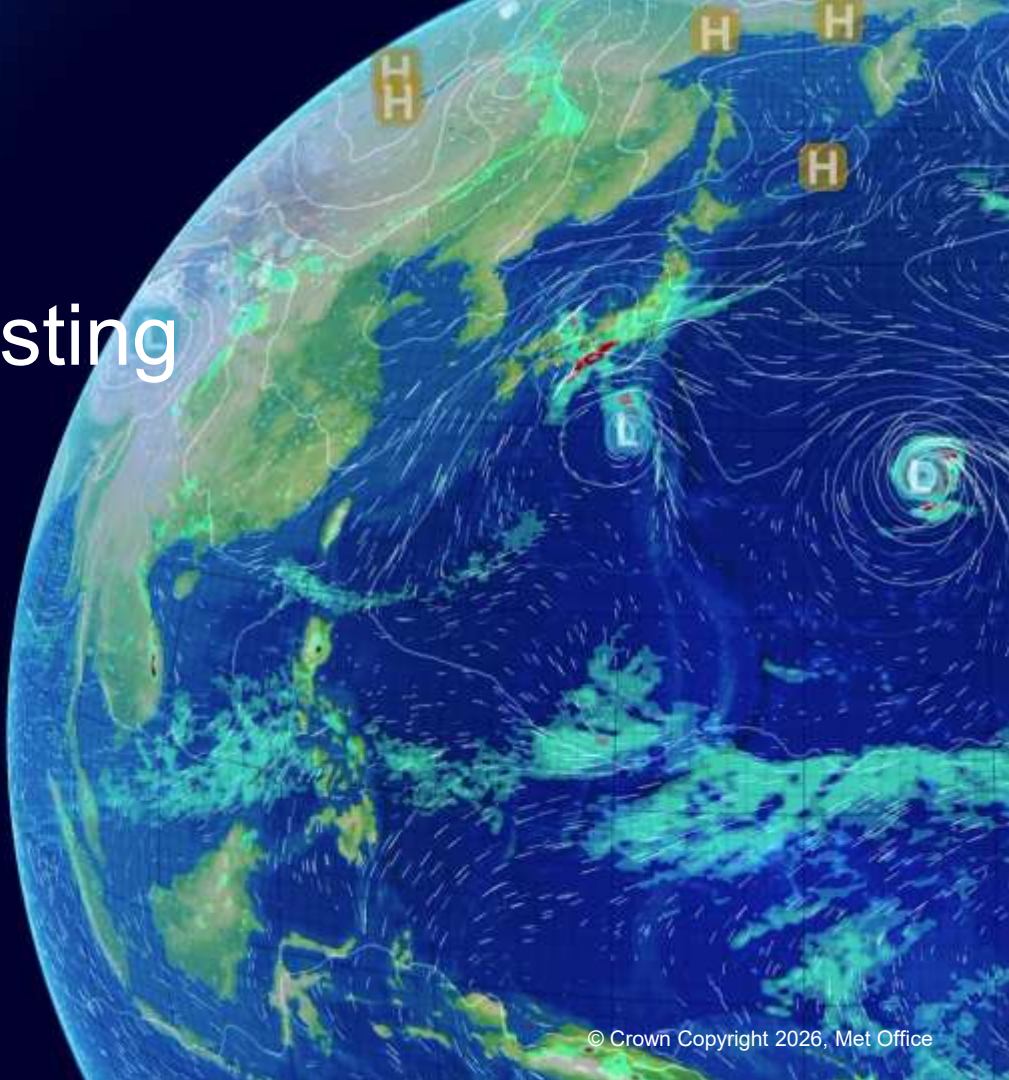


Machine Learning for Visibility Range Forecasting and Classification

Francesco Devoto

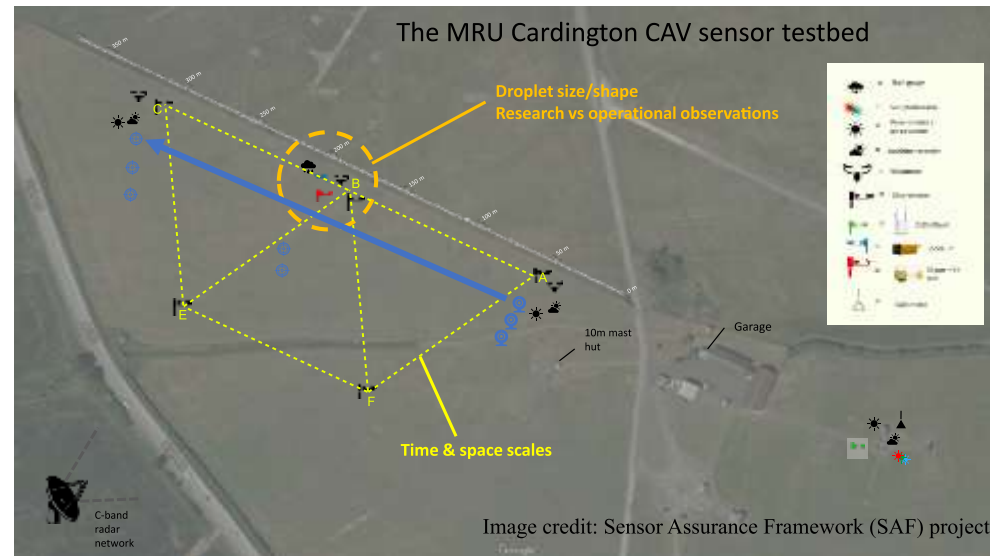
31 July 2026



Cardington Experiment

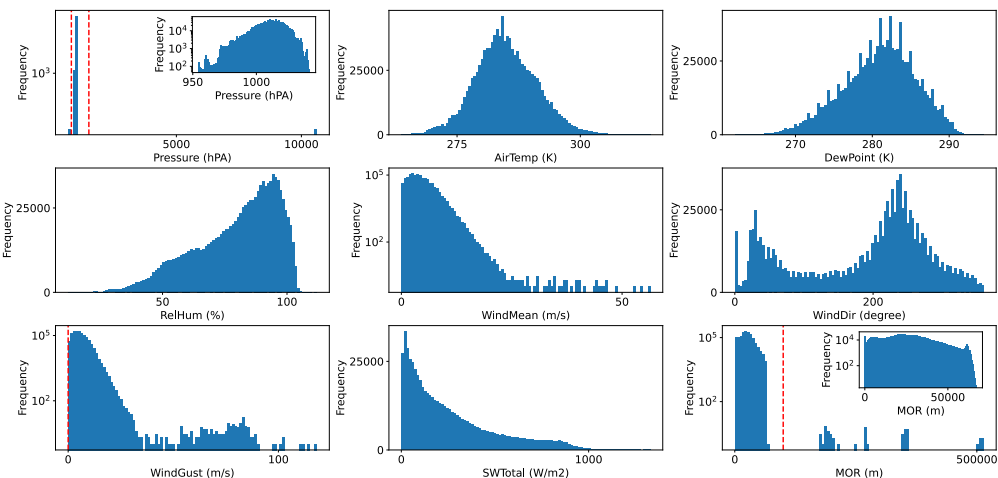
- Met Office and National Physics Laboratory collaboration
- Define a new test paradigm for autonomous vehicles
- Tested sensors in several weather conditions
- Experiment lasted over 2 years: from February 2022 to June 2024

- Sensors in **Area B**
 - Aerosol and fog presence
 - Droplet size distribution
 - Rainfall characteristic
- Sensors across the site
 - Spatial variability of meteorological variables



Exploratory Data Analysis

- Find outliers and missing data for each variable
- Anomalies attributed to sensor/data transmission issues
- Unphysical values replaced with NaNs
- Max continuous missing data interval: 41 days



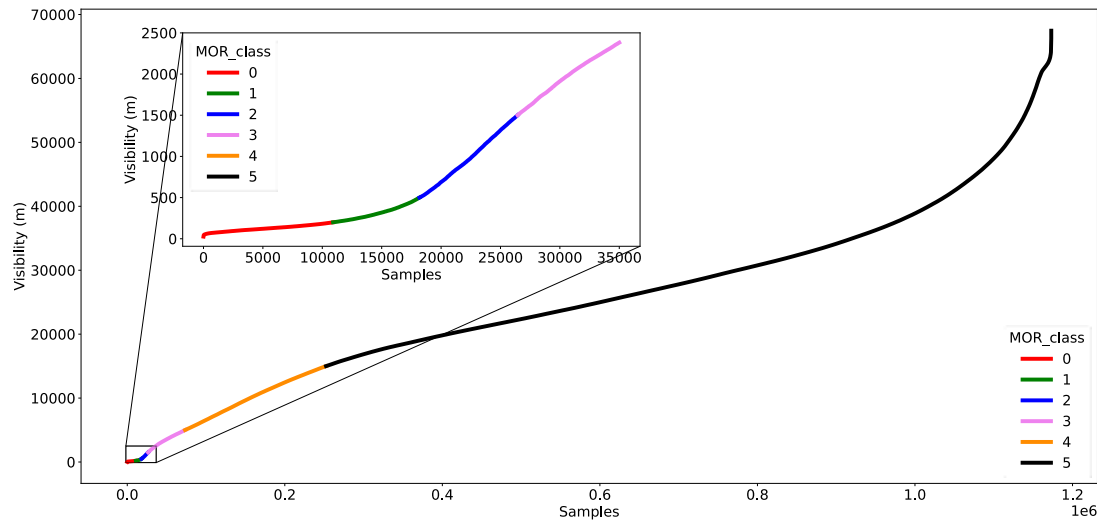
Variables	Min	Max	Miss %
Pressure(hPa)	-8.9e09	10619.0	6.7
AirTemp (K)	263.7	314.2	10.4
DewPoint (K)	261.1	294.5	10.4
RelHum (%)	12.0	112.0	10.4
WindMean (m/s)	0.0	56.4	11.3
WindDir (deg)	0.0	360.0	11.3
WindGust (m/s)	-0.1	118.3	11.3
SWTotal (W/m ²)	0.3	1329.1	52.4
MOR (m)	26.0	514208.0	2.3

Variables	Lower limit	Upper limit
Pressure (hPa)	800.0	1500.0
WindGust (m/s)	0.0	-
MOR (m)	-	100000

Visibility Classes

- Classification model works with discrete variables
- Visibility is a continuous variable
- Discretised visibility into 6 classes

Class	Visibility range (m)	Count
0	$\text{MOR} \leq 200$	10913
1	$200 < \text{MOR} \leq 500$	7269
2	$500 < \text{MOR} \leq 1500$	8260
3	$1500 < \text{MOR} \leq 5000$	46968
4	$5000 < \text{MOR} \leq 15000$	178997
5	$\text{MOR} > 15000$	920590



Machine Learning Models

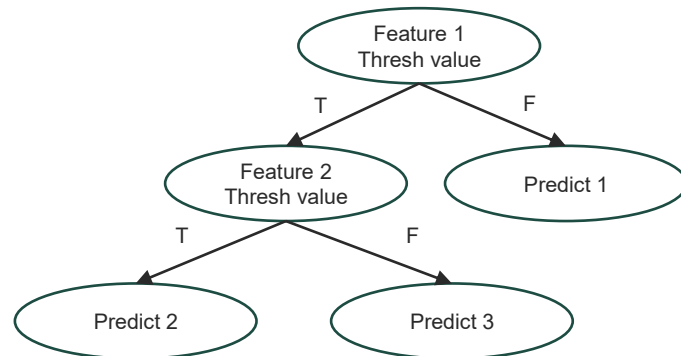
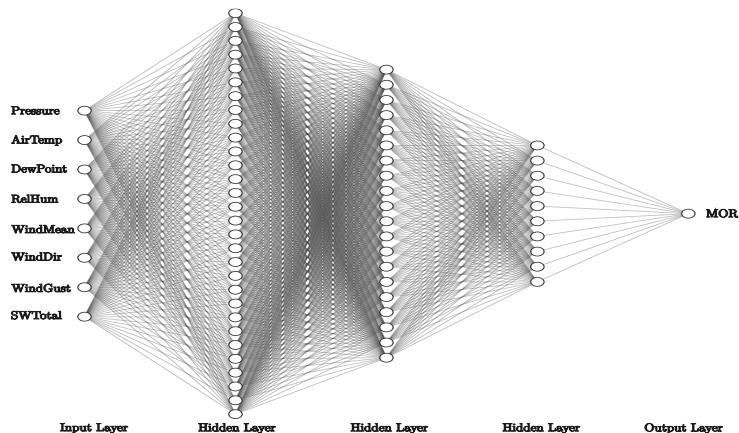
Two Machine Learning models:

- Multi-layers Perceptron regressor: for multivariate time series forecasting
- Histogram-Based Gradient Boosting classifier: for visibility classification

Before ML dataset was standardised using the standard scaler method

- Each input variable contributes equally during the learning process

$$\tilde{x}_j = \frac{x_j - \bar{x}}{\sigma}, \quad \forall j$$



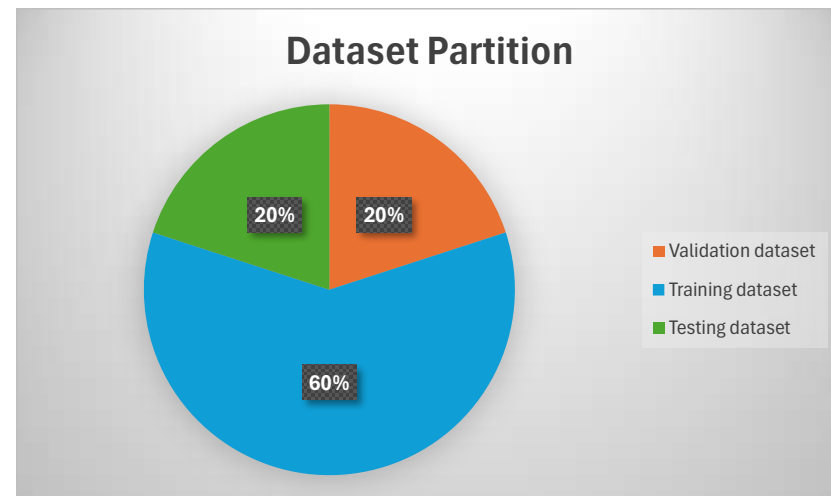
Dataset Partitioning

To avoid over-fitting, the dataset was divided into three sub-datasets:

1. Validation dataset: for ML hyper-parameters tuning, 20%
2. Training dataset: for ML model training, 60%
3. Testing dataset: for ML model evaluation, 20%

Partition strategy:

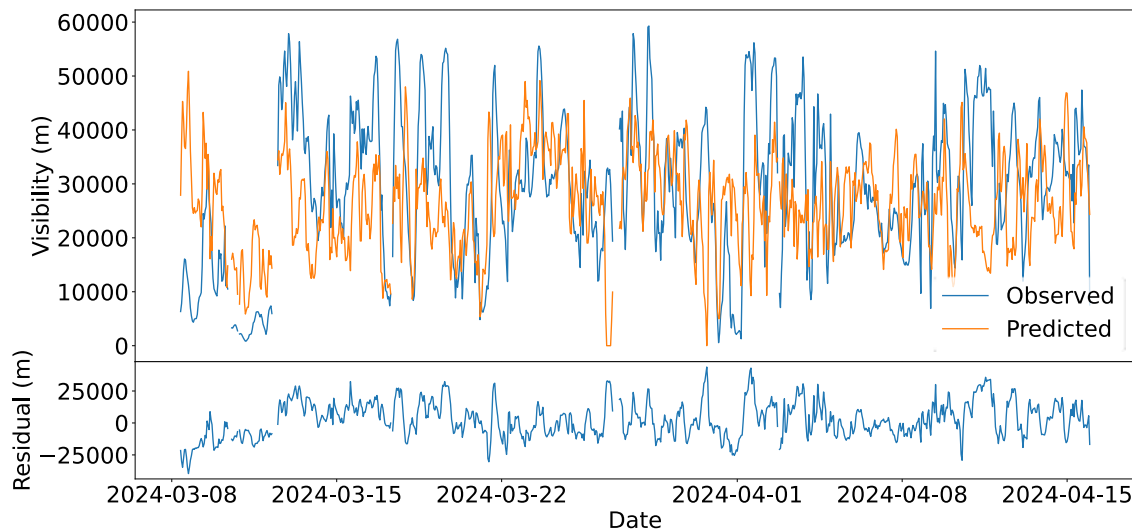
- Regression model
 - Avoid data leakage
 - Preserved temporal order
 - Train on *past* event, evaluate on *future* events
- Classification model
 - Agnostic to temporal order
 - Random selection



MLP Regressor: Results

- Grid search cross-validation approach
- Resampled using a one-hour moving average
- Spearman correlation calculated: **0.39**
- Predicted values broadly follow observed trends
- Evaluate whether the model could reasonably capture the visibility trend

Hyper-parameter	Value
Hidden layers size	(30, 20, 10)
Activation function	ReLu
Alpha	0.01
Solver	adam

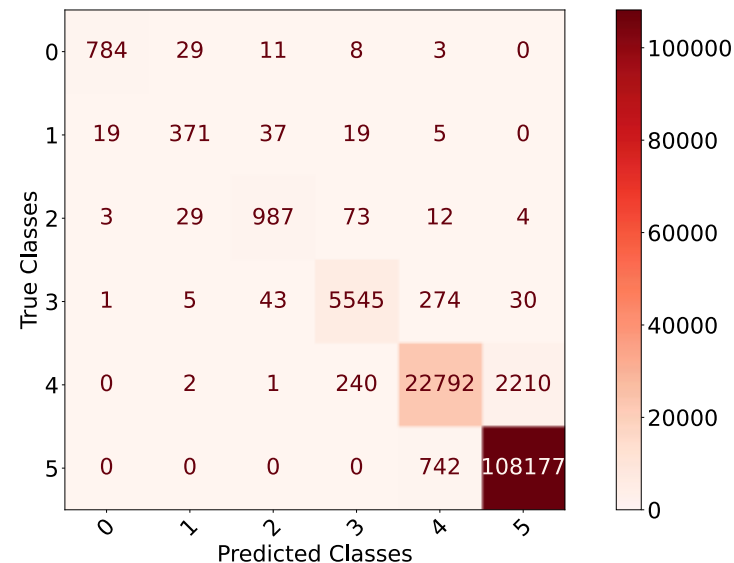


HGB Classifier: Results

- Random search cross-validation approach: 20 sets
- Each class presents a F1-score higher than **0.84**
- Confusion matrix: good agreement between the predicted and the true classes
- Majority of predictions along the main diagonal
- Overall metrics over **0.90**
- Accuracy: **0.97**

Class	Precision	Recall	F1-score	Count
0	0.97	0.94	0.95	835
1	0.85	0.82	0.84	451
2	0.91	0.89	0.90	1108
3	0.94	0.94	0.94	5898
4	0.96	0.90	0.93	25245
5	0.98	0.99	0.99	108919
Macro Avg	0.94	0.91	0.93	142456

Hyper-parameter	Value
Learning rate	0.05
Max leaf nodes	63
Min samples leaf	50
Max depth	7



Conclusions

Overall, the models worked well

- MLP regressor correlation coefficient: **0.39**
- HGB classifier accuracy: **0.97**

Future improvements

- Better correlation study among the meteorological variables
 - Selection of other features
 - Use of SHapley Additive exPlanations (SHAP) methodology
- Better data partition considering the fog season
- Better selection of ML model to use
 - Statistical evaluation metrics: ROC curve, AUC score

Future work

- Starting point for the MSc project:
 - Find correlation between meteorological variables and LiDAR reflectivity
 - Under what meteorological conditions can an autonomous vehicle operate safely?