•RMSE = Root Mean Squared Error

UNIVERSITY OF CAMBRIDGE

# The rainfall-induced flood prediction on national highways

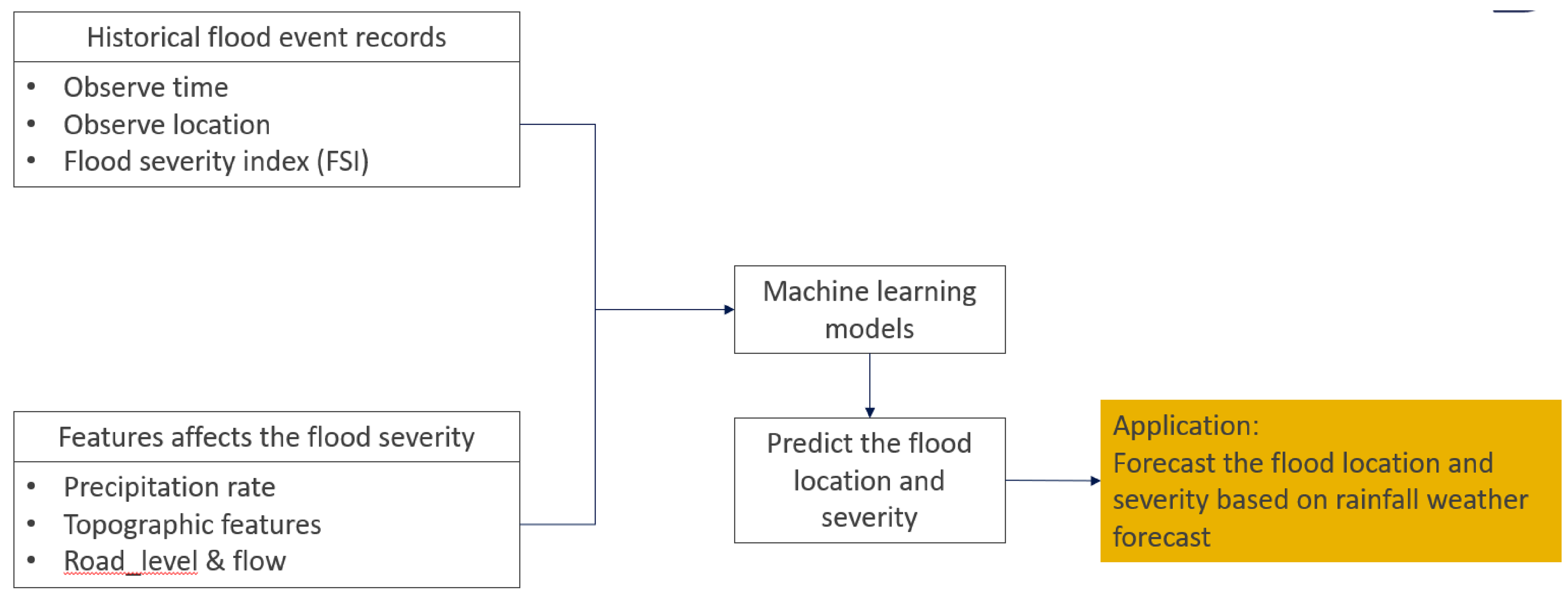

**Fellows: Jie Liu, Zhaojie Sun, Linjun Lu, Zizhen Xu, Xiang Wang**

**Academic Supervisors: Kristen MacAskill & Li Wan**

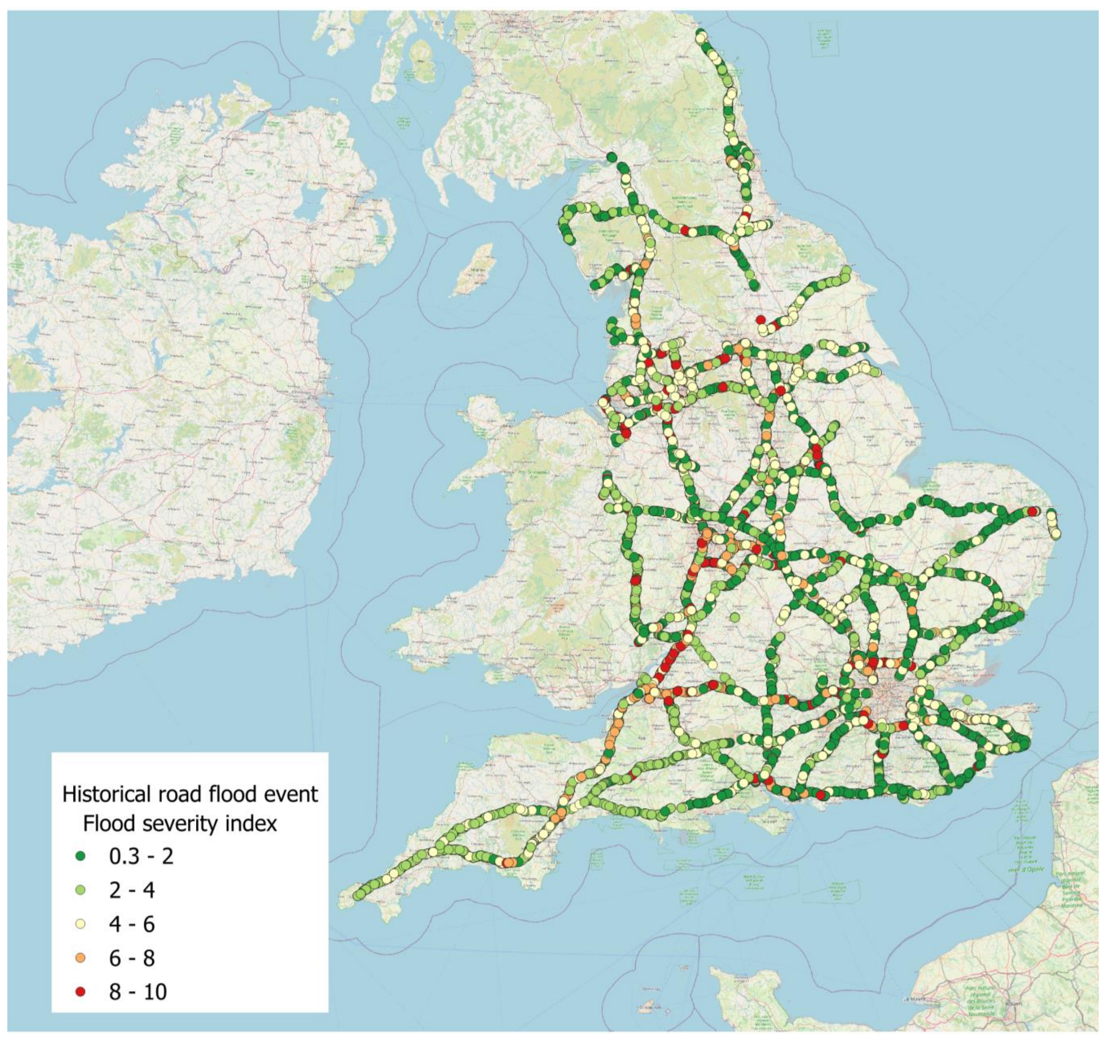
**Industry supervisor: Oliver Thomas, Federico Perrotta**



This research develops a data-driven framework to predict rainfall-induced flood location and its severity using historical flood records, antecedent rainfall prior to historical flood observation time, highway topographic features, and road attributes (e.g., road class, traffic flow, and lane count). We evaluate Random Forest (RF), Support Vector Regression (SVR), and Multilayer Perceptron (MLP) models, incorporating imbalance-aware weighting to better predict flood location and severity. The best-performing RF model achieves strong predictive performance, with a Mean Absolute Error (MAE) of 0.70 (≈7% of the Flood Severity Index (FSI) scale) and a Root Mean Squared Error (RMSE) of 0.97, and demonstrates good calibration across deciles. These results indicate that RF can effectively predict flood location and severity using rainfall forecasts, highway topographic features, and road attributes.


**Research framework**




The historical highway flood incident dataset (35,759 records) is the core input used to model rainfall-induced flooding on highways. Each record represents a reported flood event and includes the Flood Severity Index (FSI) on a 0–10 scale, the observation (reported) time, and the event latitude and longitude.

**Figure historical flood event on highways**

The three categories of acquired features are summarised in Table 1 and used as inputs to the flood prediction models.

**Table. The acquired features for flood prediction**

| |
|---|
| **Rainfall precipitation features** |
| ahead_report_time_3h_total_pricipitation |
| ahead_report_time_3h_max_pricipitation_for_a_hour |
| ahead_report_time_6h_total_pricipitation |
| ahead_report_time_6h_max_pricipitation_for_a_hour |
| ahead_report_time_2h_total_pricipitation |
| ahead_report_time_2h_max_pricipitation_for_a_hour |
| record_month |
| record_hour |
| **Topographic features** |
| DTM_on_road(mean, median, std, iqr, q95, min, max) |
| DSM_off_road (mean, median, std, iqr, q95, min, max) |
| Slope_on_road (mean, median, std, iqr, q95, min, max) |
| Slope_off_road (mean, median, std, iqr, q95, min, max) |
| Roughness_on_road (mean, median, std, iqr, q95, min, max) |
| Roughness_off_road (mean, median, std, iqr, q95, min, max) |
| **Road_level & flow features** |
| roadnumber |
| lanes |
| speedlimit |
| avg_speedm |
| delay_spvp |
| nonrecurre |
| vhd_gdr |

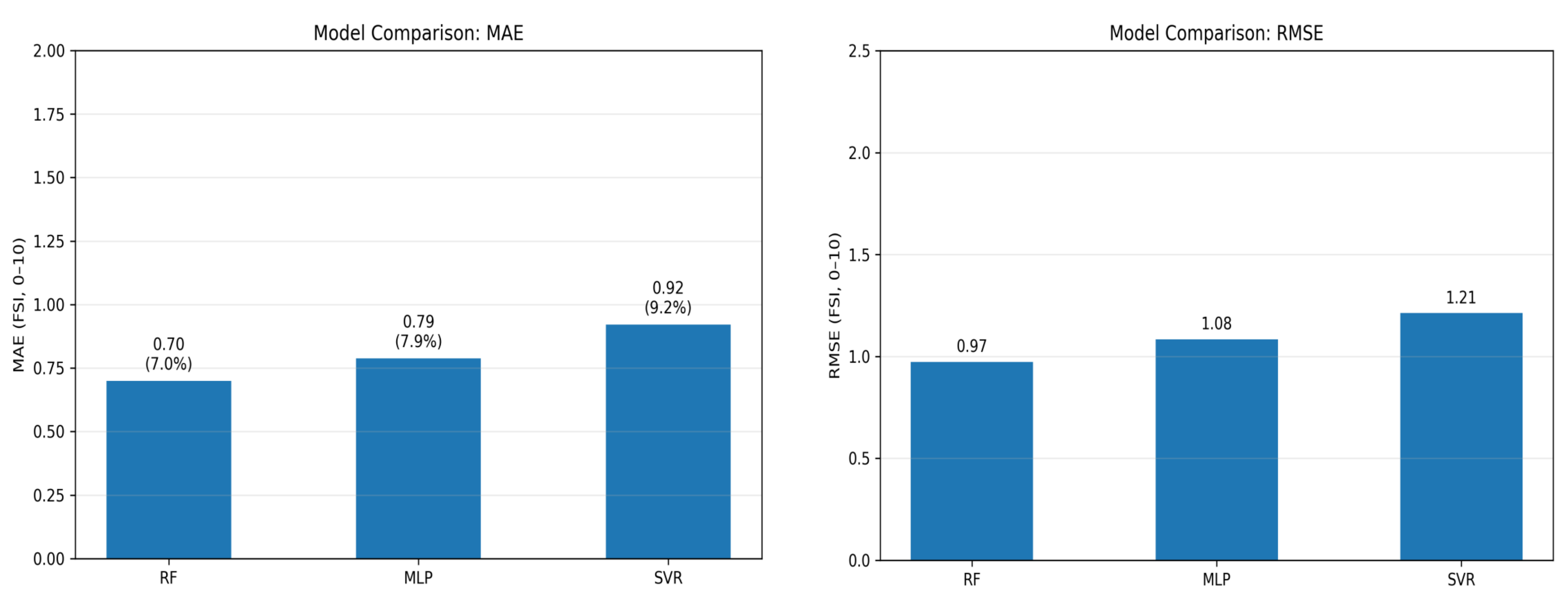



**Figure Prediction comparison**

Overall, the Random Forest (RF) delivers the strongest performance, achieving Mean Absolute Error (MAE) = 0.70 (≈7% of the Flood Severity Index (FSI) scale) and Root Mean Squared Error (RMSE) = 0.97, which indicates the lowest average absolute error and the smallest penalty from occasional large deviations.

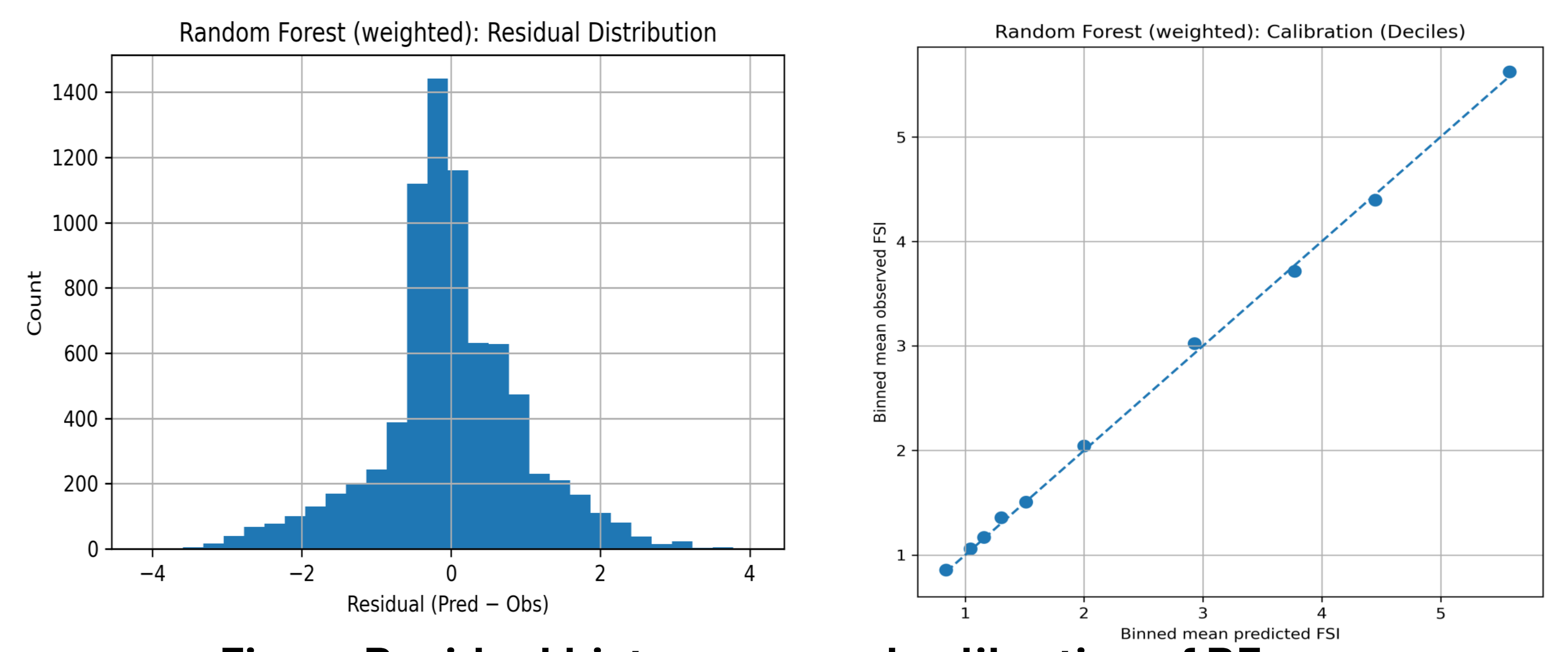



**Figure Residual histograms and calibration of RF**

Random Forest shows the tightest residuals (mostly within −1 to 1), indicating higher precision and fewer extreme errors. It also has the best calibration, with predictions closely tracking observations across the full range, making it more reliable for decision support.

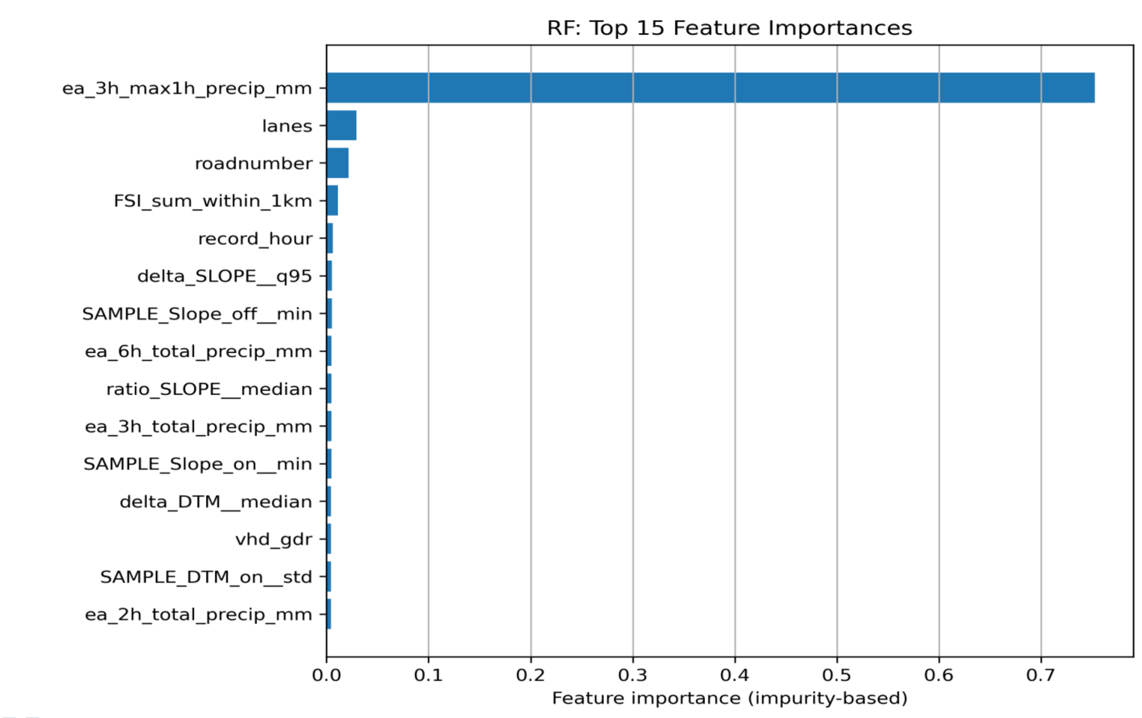



Figure shows the top 15 impurity-based feature importances from the best RF model.

- Short-term rainfall intensity dominates: the maximum 1-hour precipitation in the 3 hours before observation (ea_3h_max1h_precip_mm) is by far the strongest predictor.
- Road attributes (lane count, road class/roadnumber) are the next most important, followed by local context and terrain factors (nearby Flood Severity Index (FSI) within 1 km, time of day, and slope/DTM statistics).
- Overall, Flood Severity Index (FSI) is driven by a combination of recent rainfall forcing, road-network characteristics, and local topographic susceptibility.

COLAS AECOM ferrovial construction TRL