Development of flood risk map for railway tracks in nationwide floodplain area using digital surface model and machine learning

[1]Watcharapong Wongkaew, [2]Kittipat Phunjanna, [3]Wachira Muanyoksakul
[4]Tanawat Tangjarusritanatorn, [3]Chayut Ngamkhanong

[1]Chulalongkorn University Transportation Institute, Chulalongkorn University, Bangkok, Thailand, Email: watcharapong.w@chula.ac.th
2Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong SAR; Email: 1155131356@link.cuhk.edu.hk
3Advanced Railway Infrastructure, Innovation and Systems Engineering (ARIISE) Research Unit, Department of Civil Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand; Email: 6678002021@student.chula.ac.th, chayut.ng@chula.ac.th
4Department of Water Resources Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand; Email: tanawat.ta@chula.ac.th

ABSTRACT

The monsoon region, where Thailand is situated, experiences frequent heavy rainfall, leading to recurring flooding problems. Furthermore, human activities, such as the construction of railway tracks that obstruct the flow of water and the inadequate natural drainage system, also contribute to the problem. In order to analyze the area-based risk factors that cause railway track flooding, 5 major factors, including average total rainfall, waterway density, land use, slope, and elevation, must be considered. This study employs advanced computer vision techniques, including Digital Surface Models (DSM) and flood simulations, to map the topography of flood-prone areas and railway rights of way and enhance the prediction of flood risk. Specifically, the Extra Trees classifier demonstrated the highest performance, achieving an F1-score of 0.71. This model effectively predicted the Flood Risk Index (FRI), which evaluates risk based on the rules and procedures for train operations during flood conditions. A case study in Thailand further illustrated the practical application of these methods. This research has practical implications for flood risk

METHODOLOGY

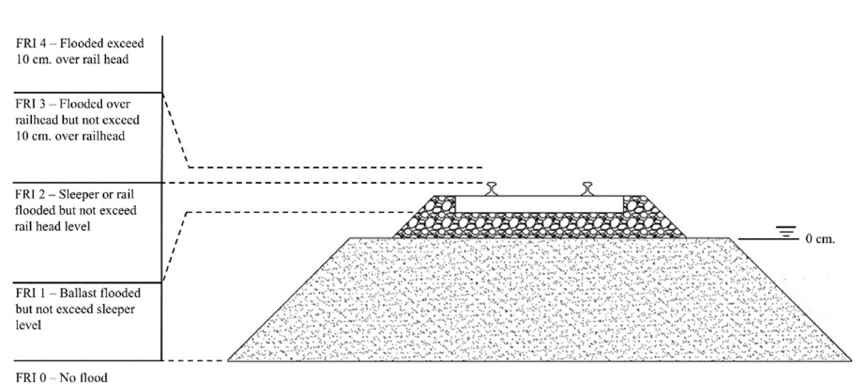
1. Data Collection and Oversampling

The collection of data involved four major risk factors in railway track flooding: historical climate conditions, drainage basin area, slope, and elevation which are provided by various government agencies and university faculty. Due to the inherent class imbalance in the data, where non-flooded areas significantly outweigh flooded ones each year, oversampling techniques can be employed to mitigate bias in the model. To address the class imbalance, two prevalent oversampling techniques, Synthetic Minority Oversampling Technique (SMOTE) and Adaptive Synthetic Minority Oversampling Technique (ADASYN), are chosen Table 1 Proportion of data splitting after oversample

2. Operation-based Flood Risk Index

To estimate the risk of the railway tracks in the floodplain area, we use the operation-based index. Review of the rules for the operation of trains through flood water. Based on current practices of various train operators in the UK, a previous study determined that the critical flood level is at the railhead, with operating speeds being adjusted according to the flood level measured from the bottom of the rail. This implies that trains can safely operate at reduced speeds as long as the flood level remains below the railhead. If floodwater rises above the railhead, many operators recommend halting services. The level of the Flood Risk Index (FRI) and its interpretation is shown in Table 1 and Figure 1

FRI Range	Level of flood	Interpretation
0	No flood	No risk
1	0 - 30 cm.	Ballast level – Sleeper bottom
2	31 – 60 cm.	Sleeper bottom - Railhead
3	61 - 70 cm.	Over railhead (reduced speed)
4	> 70 cm.	No operation



3. Machine Learning Techniques

The input of the model, raster data, Digital Elevation Model (DEM) data, independent variables and precipitation tercile forecasts map, are processed before being integrated into a Dataframe using python. By using canny edge detection, information of an image edges and area edges are collected in an image By mapping the left-most, right-most, and top of an image to raster data in the same area, an approximated pixel to basin can be recognized in raster data. Using this information, the details of precipitation tercile forecasts map can be integrated with raster data followed by DEM data and independent variables using Geopandas Dataframe.

Split Following data pre-processing and organization into a dataframe, the crucial step of data splitting and validation is performed. This process involves separating the 12-year flood history dataset into training and testing sets, 2005 to 2014 for training, 2015 and 2016 for testing. The training and testing sets are chosen based on the data availability.

This study employed a selection of machine learning estimators to investigate the impact of flooding on railway tracks and surrounding areas, with a particular focus on assessing track damage and structural integrity. Flood risk was categorized into five distinct classes, as detailed in Table 1. The models were evaluated using a range of performance metrics .

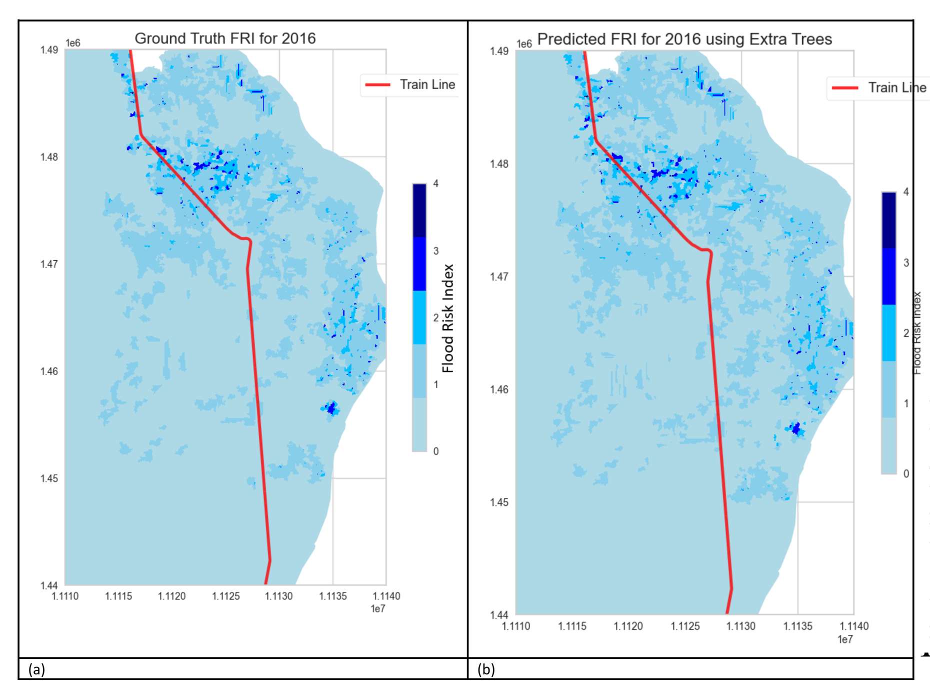
RESULTS

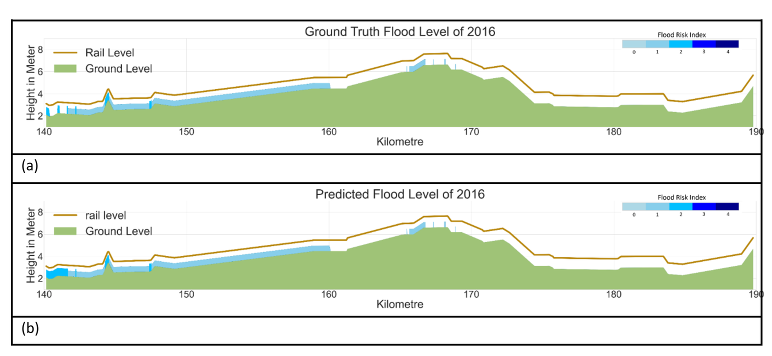
Table 2 displays the performance of each machine learning model using ADASYN as oversampling technique

Model	Average 10-Fold of training set (macro average)				Test set (macro average)				Rank
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score	
extra-tree	0.98	0.96	1.00	0.98	0.74	0.72	0.81	0.71	1
LDA	0.83	0.77	0.96	0.85	0.23	0.11	0.50	0.18	5
LR	0.79	0.74	0.91	0.81	0.72	0.72	0.81	0.70	2
KNN	0.94	0.90	1.00	0.95	0.36	0.53	0.53	0.36	4
SVM	0.55	0.59	0.47	0.40	0.77	0.39	0.50	0.44	3

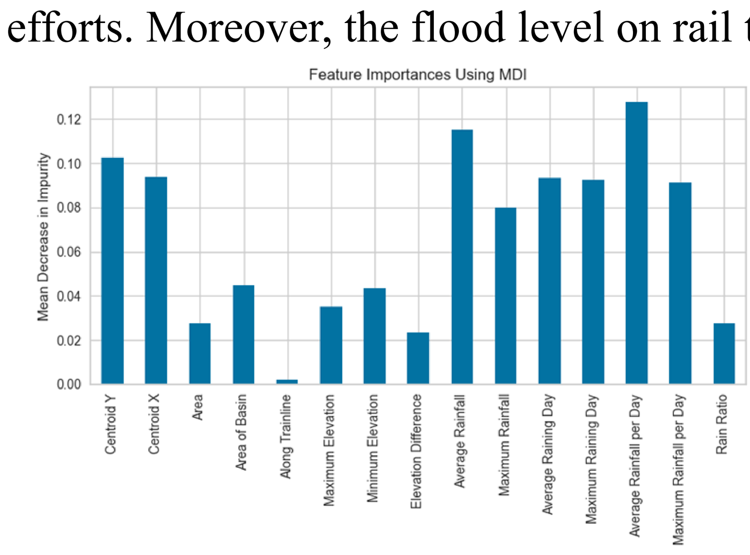
ADASYN Oversampling used

The results show that the Extra Trees, using 100 estimators, the Gini criterion, and bootstrap enabled, achieved the highest precision, recall, and F1 score both without oversampling and with ADASYN (using 5 nearest neighbors and a minority sampling strategy). This indicates that the model is most effective when applied to this category of area with given oversampling technique.





The figure above is the test data, while the predicted data appears to encompass a larger floodplain area compared to the ground truth, it successfully captures nearly all the actual flooded regions. This translates to a high recall value, indicating the model's effectiveness in identifying areas at risk of flooding. This characteristic is crucial for flood preparedness efforts. Moreover, the flood level on rail track is shown from North to South. The ground level is depicted by the green area, while the rail level is indicated by the brown line above it. Various shades of blue represent the Flood Risk Index (FRI), ranging from 0 (no risk) to 4 (highest risk), as shown in the legend. The left figure illustrates the feature importance for various parameters used in the best-performing Extra Trees model for predicting FRI, evaluated using the Mean Decrease in Impurity (MDI) method. The y-axis represents the importance of each feature, with higher values indicating greater contributions to the model's predictive accuracy.



CONCLUSION

The integration of machine learning, particularly the Extra Trees classifier, proved to be the most effective approach for predicting the Flood Risk Index (FRI), achieving the highest F1-score of 0.71. This model enabled the generation of a detailed flood risk map for railway tracks in flood-prone areas, providing valuable insights for infrastructure planning and disaster prevention.

This study provides a better understanding of the relationship between flooding conditions and flood prediction. From mean decrease in impurity (MDI), or Gini Impurity, we can conclude 3 most influential factors that affect flooding in the area namely Average rainfall, Average rainfall per day and Maximum rainfall per day, respectively. Moreover, these aspects can be utilised in developing more effective flood risk management strategies of railway track using forecasting data. Utilizing this process, railway operators will be able to mitigate the risks of flooding and prevent future damages to the railway efficiently. The findings can be applied to other regions with similar flooding problems to develop a more resilient railway infrastructure that can withstand natural disasters.