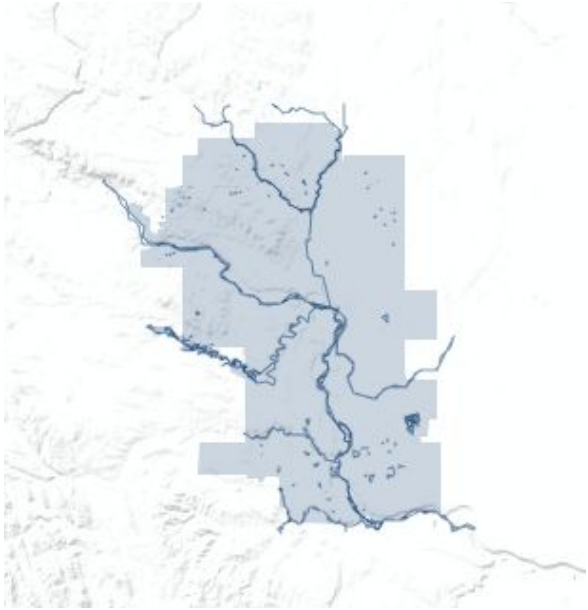
Flood Inundation Prediction

in Calgary, Canada

CPLN 675
Haobing Liu & Ling Chen
2023-03-28

Introduction



Floods are one of the most devastating natural disasters, causing widespread damage to communities and infrastructure. As the frequency and intensity of extreme weather events increase, there is a growing need for effective flood risk management strategies. This is where flood inundation analysis comes in.

The purpose of this analysis is to create a predictive model that can estimate the likelihood of flooding in Calgary, Alberta, Canada, based on a range of factors that we have identified as being important. Then we use this model to predict the probability of flood inundation in a comparable city Pittsburgh, Pennsylvania, US, which helps us understand how our model might perform in different contexts.

Ultimately, our goal is that this analysis will provide valuable insights into flood risk management, and help planners make more informed decisions about how to protect their communities from the devastating effects of flooding.

Use Case

From the city of Calgary we can see that Calgary is at its greatest risk of flooding during spring and summer. Additionally, heavy rainfall on the melting snowpack in the Rocky Mountains combined with steep, rocky terrain caused rapid and intense flooding in southern-Alberta watersheds. Flooding disrupted businesses, damaged critical infrastructure and also led to power outages across Calgary.

As a river city, it is important to prepare, respond and adapt to floods. Every spring, the city of Calgary actively monitor the rivers for flooding. They continuously improve the flood forecasting to provide citizens with the earliest possible warning.

Therefore, the information from this analysis can be used by city planners in Calgary to **make informed decisions about land use, infrastructure development, and emergency preparedness.**

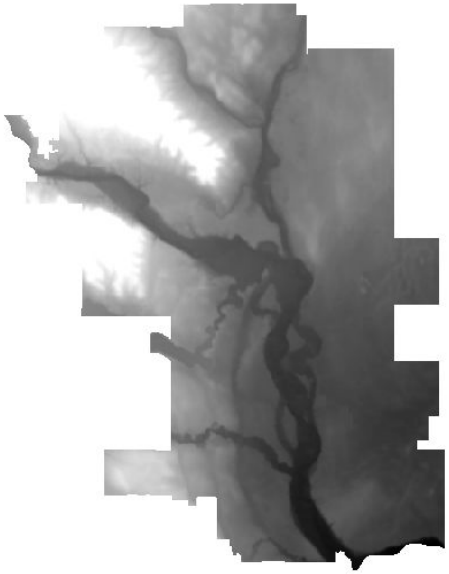
Data

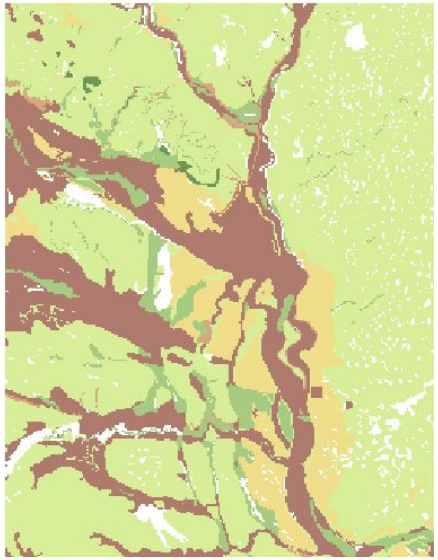
-Hydrology - water bodies and rivers

-DEM -slope

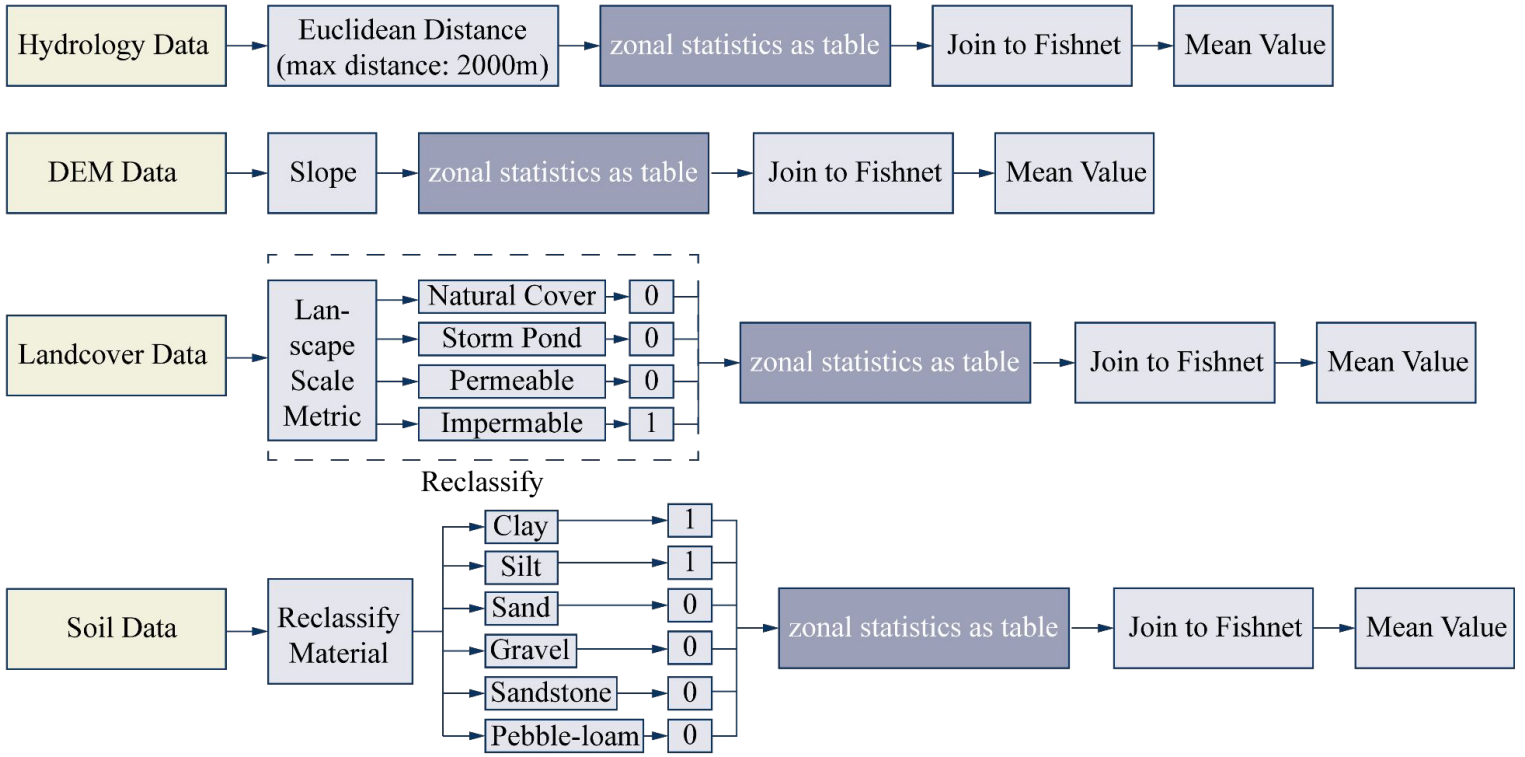
-Citywide Land Cover

-Soil Data - soil materials





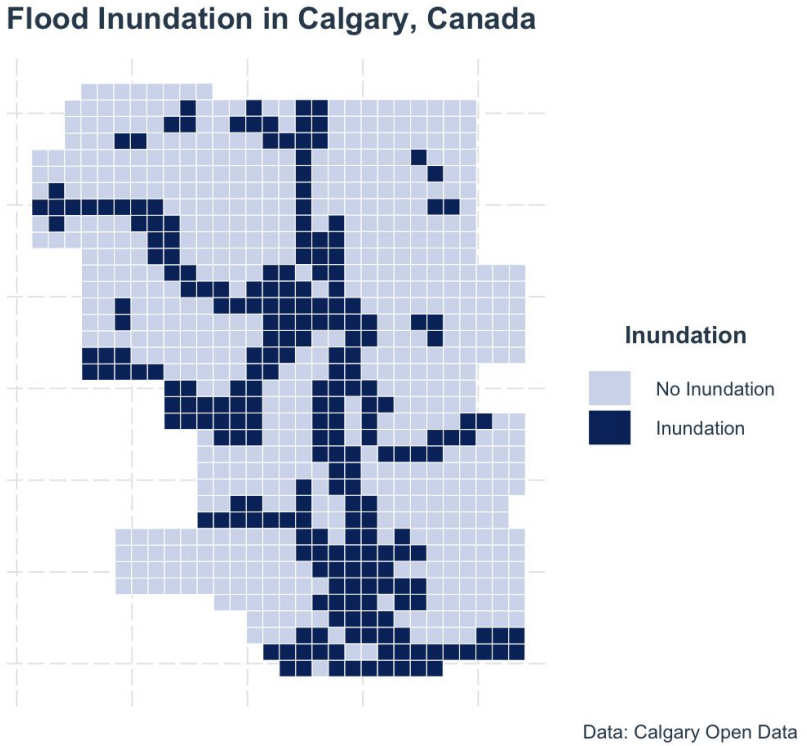
Feature Engineering



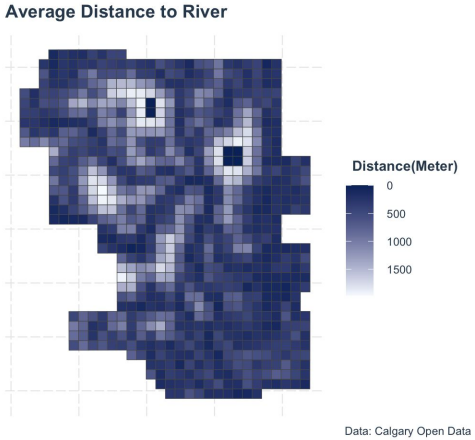
Fishnet

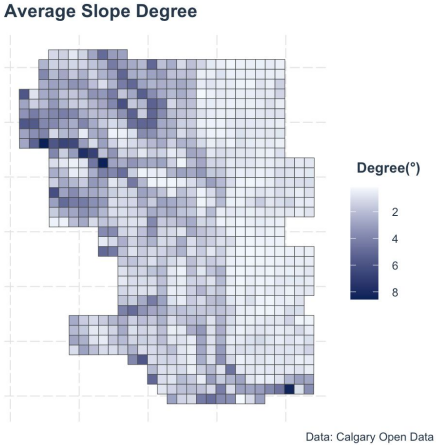
Based on the satellite image with flood inundation, Calgary is receiving a certain amount of flood issues especially along the major hydrology, which are the Bow and Elbow Rivers. Specifically, 237 fishnet cells with darker color are areas with inundation.

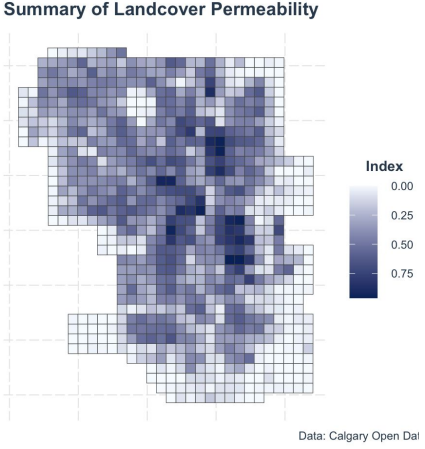
```
## FALSE  TRUE
##   575   237
```

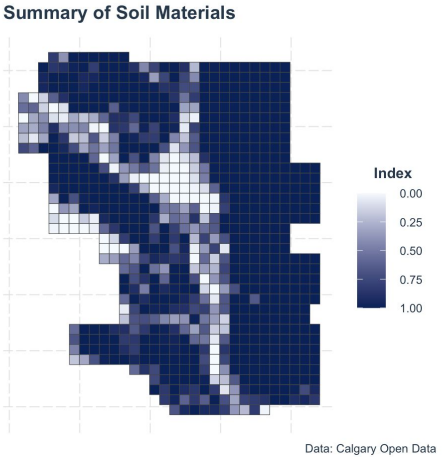


Independent Variables



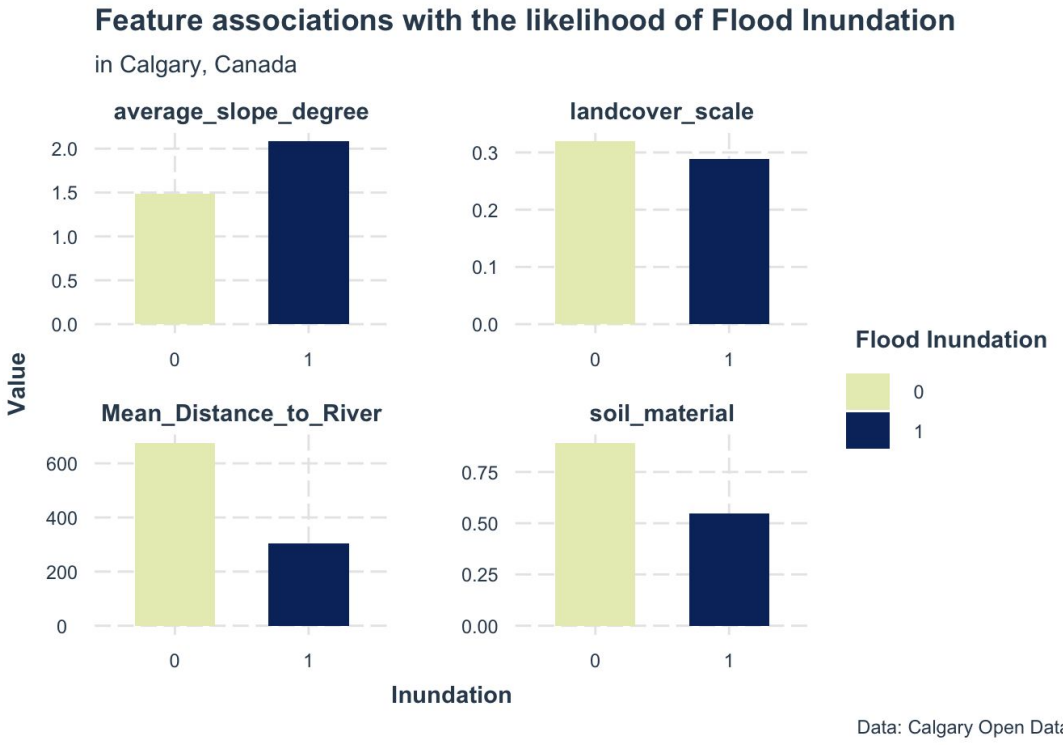






Feature Associations

The chart shows that inundation occurs more in areas that are close to rivers, have high slopes degree, have a single soil type with low permeability.



Regression Model

70% -> traning set (569 rows)

30% -> testing set (243 rows)

Those four features are good predictor of flooding!

The constant term in the model is 2.768, which represents the predicted log-odds of flood inundation when all of the independent variables are equal to zero. The corrected AIC value for the model is 400.5, and the log likelihood is -195.206.

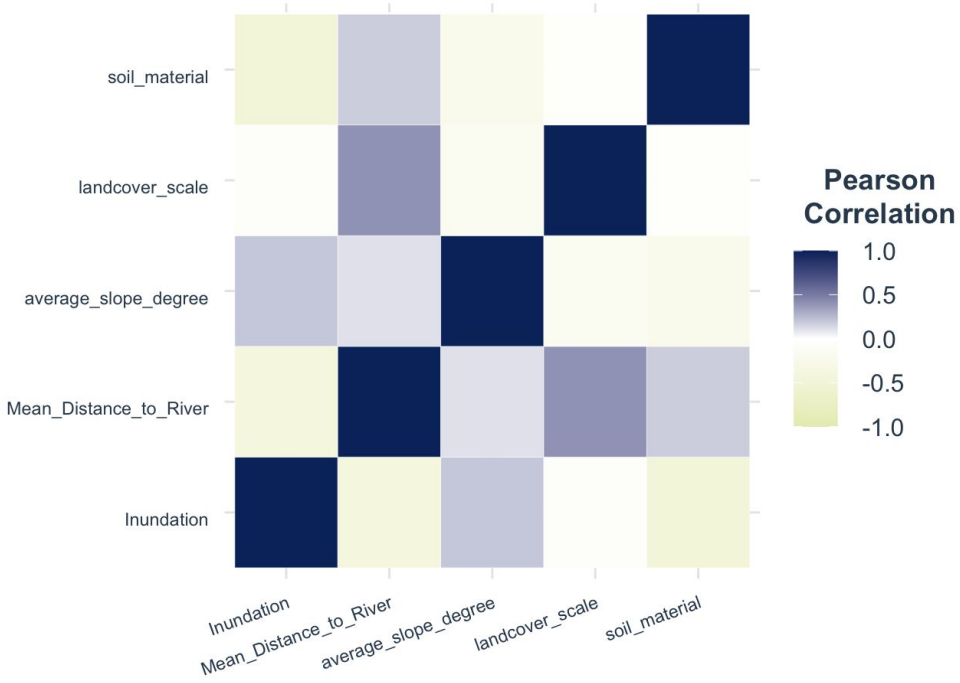
Regression Models Predicting Flood Inundation Probability in Calgary, Alberta

	Dependent variable:
	Inundation Model
Mean_Distance_to_River	-0.008***
	(0.001)
average_slope_degree	0.520***
	(0.101)
landcover_scale	2.938***
	(0.680)
soil_material	-3.181***
	(0.417)
Constant	2.768***
	(0.520)
Corrected AIC	400.5
Observations	569
Log Likelihood	-195.206
Akaike Inf. Crit.	400.412
Note:	*p<0.1; **p<0.05; ***p<0.01

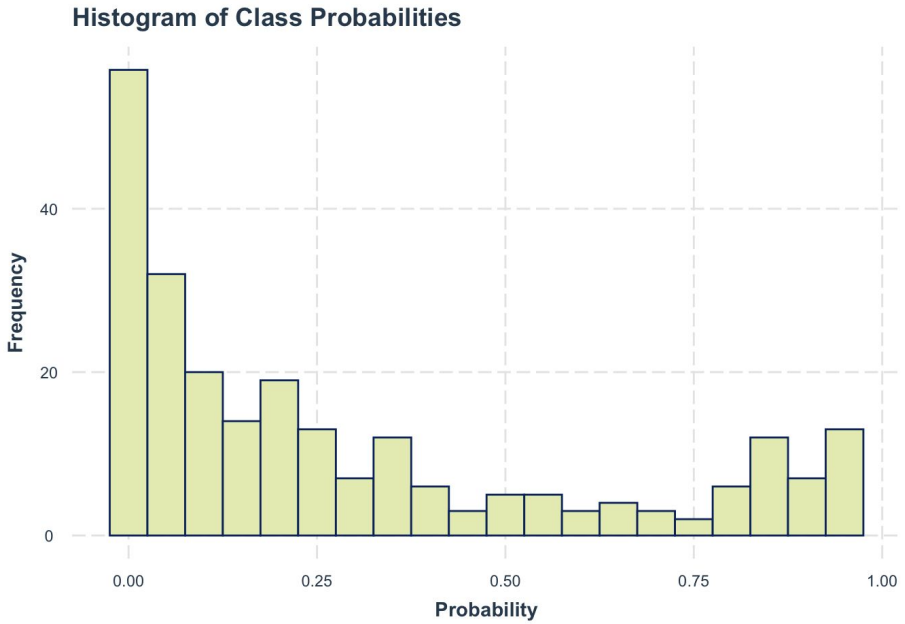
Correlation Matrix

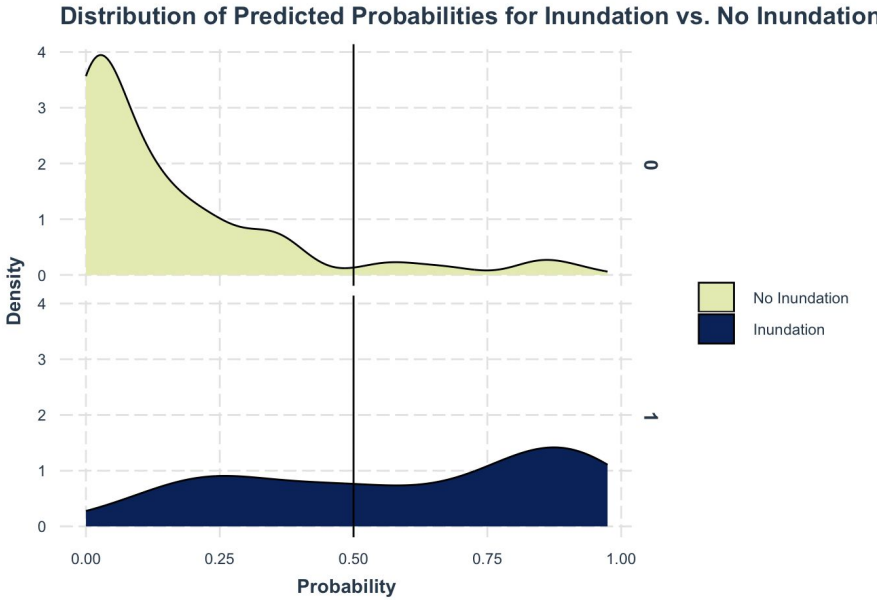
The Pearson correlation coefficients between the variables:

Soil material has significant negative relationship with inundation, the distance to rivers as well. And the average of slope degree has significant positive relationship with inundation.



Classification Probabilities





Confusion Matrix

1. Predicted = 0, Observed = 0 —> **True Negative**
   The model correctly predicted instances belong to the negative class.
2. Predicted = 1, Observed = 1 —> **True Positive**
   The model correctly predicted instances belong to the positive class.
3. Predicted = 1, Observed = 0 —> **False Positive**
   The model incorrectly predicted instances belong to the positive class.
4. Predicted = 0, Observed = 1 —> **False Negative**
   The model incorrectly predicted that 15 instances belong to the negative class.

From the confusion matrix, we know that there are 157 cells with true negative, 42 cells with true positive, 15 cells with false positive and 29 cells with false negative. Meanwhile, the value of model sensitivity is 0.5915, meaning it correctly identified 59.15% of the positive cases (actual 1s), and model specificity is 0.9128, meaning it correctly identified 91.28% of the negative cases (actual 0s).

	Reference	
Prediction	0	1
0	157	29
1	15	42

Accuracy : 0.8189

95% CI : (0.7646, 0.8652)

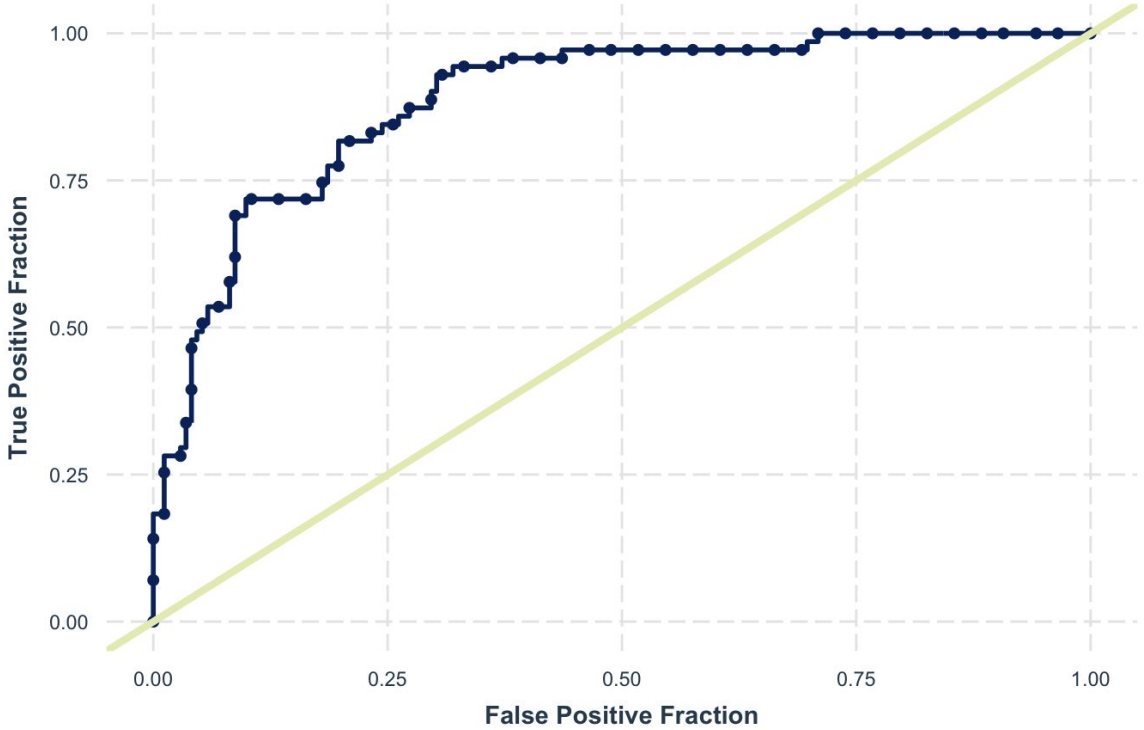
P-Value [Acc > NIR] : 0.00004793

Kappa : 0.5353

Sensitivity : 0.5915

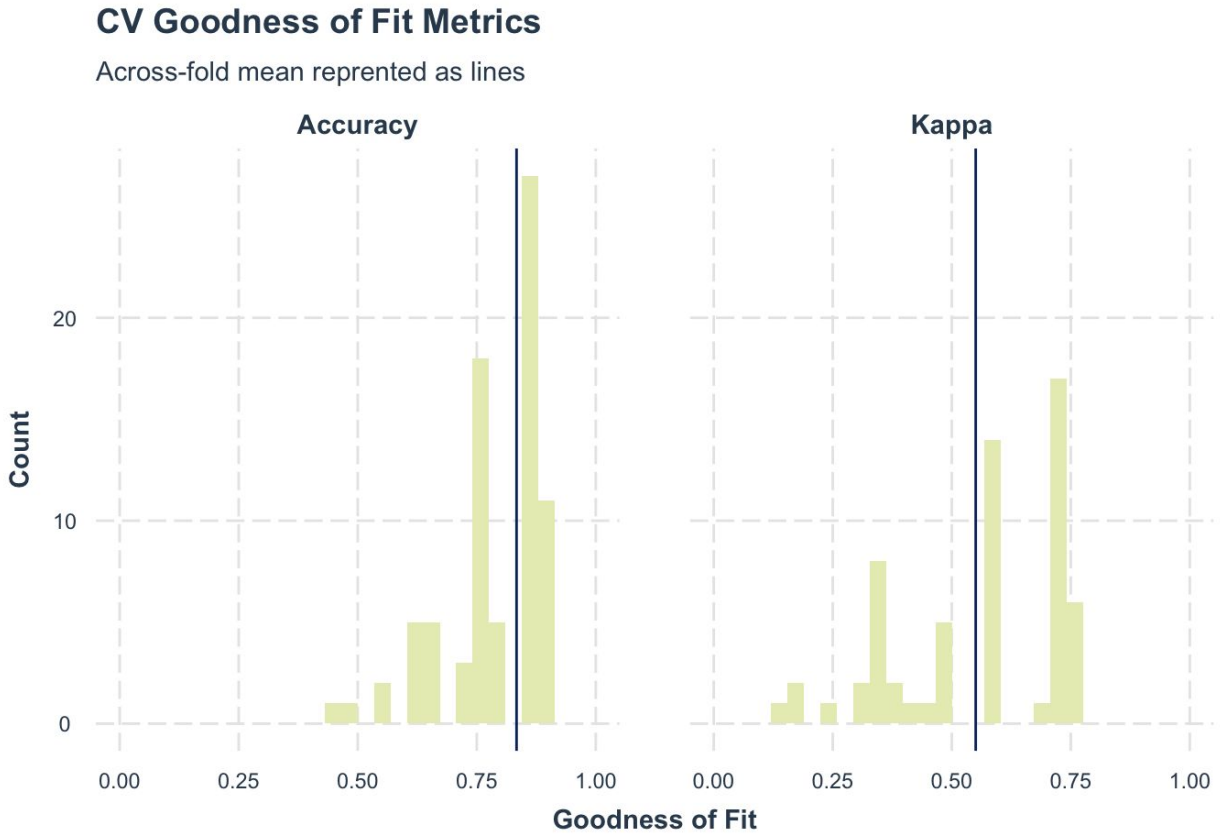
Specificity : 0.9128

ROC Curve



In this case, an AUC (he area under ROC curve) of **0.8867** indicates that the model has a relatively high ability to distinguish between the positive and negative classes, and can classify new observations with a high degree of accuracy.

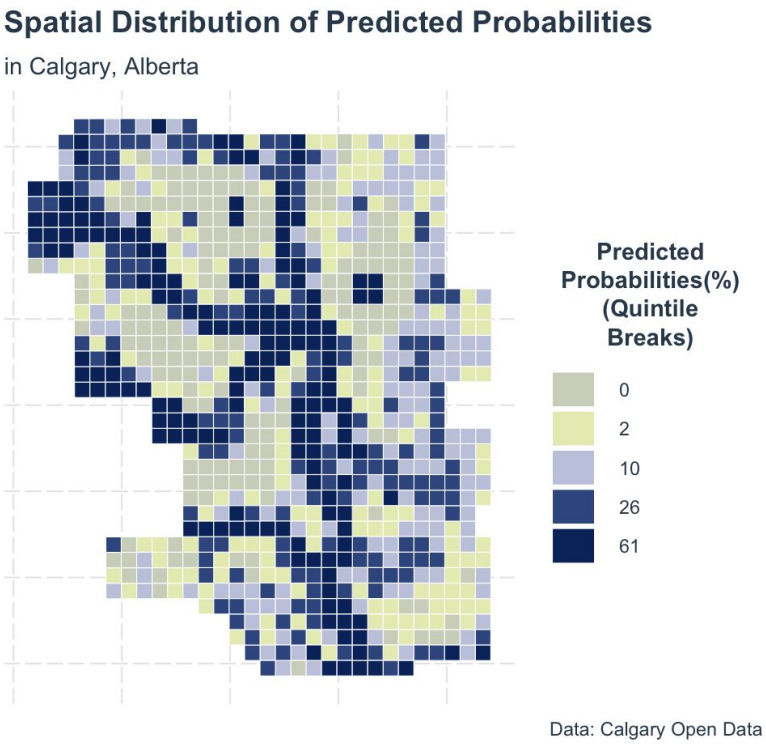
Cross Validation

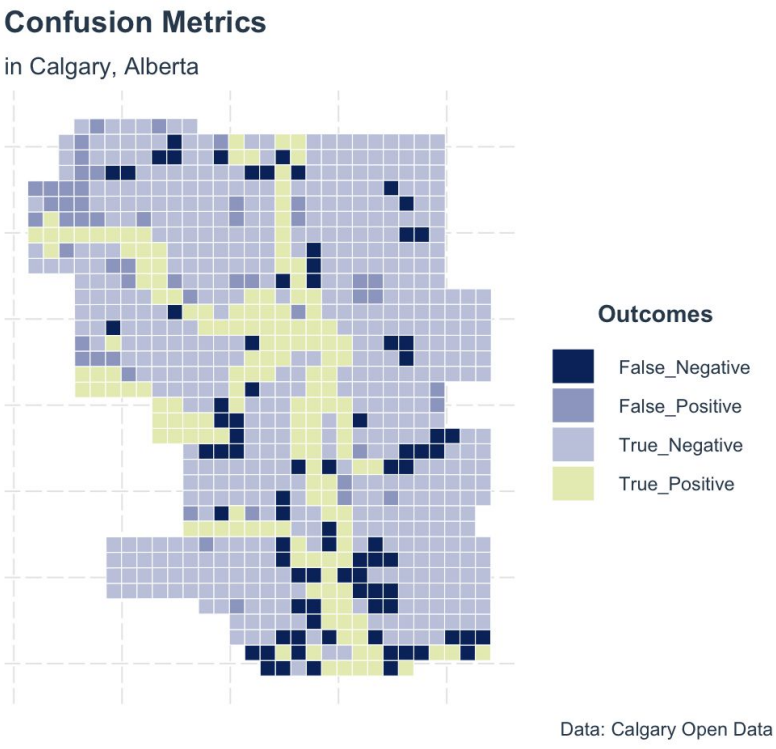


Accuracy: 0.8335913

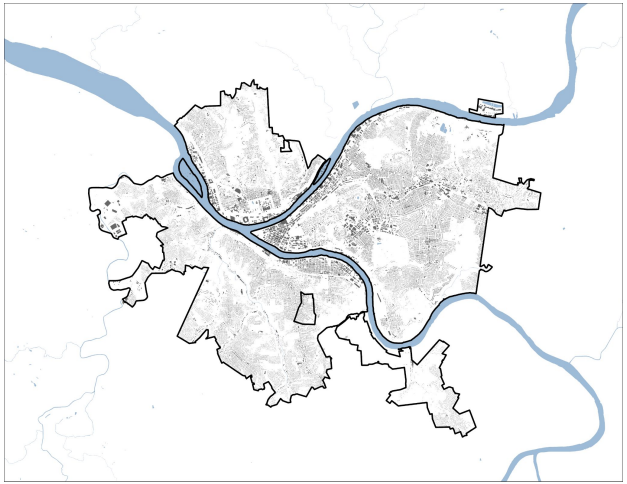
Kappa: 0.550286

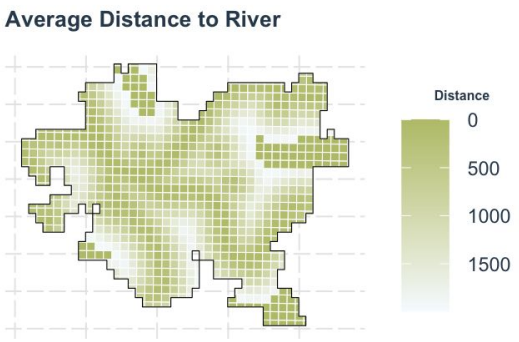
Prediction Map

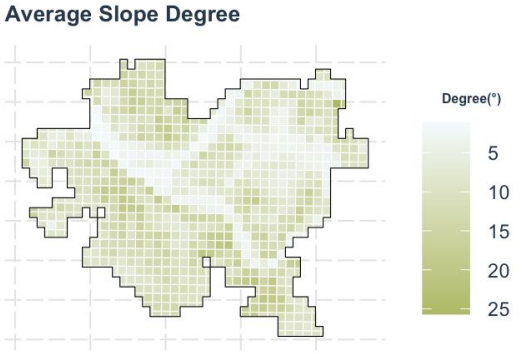


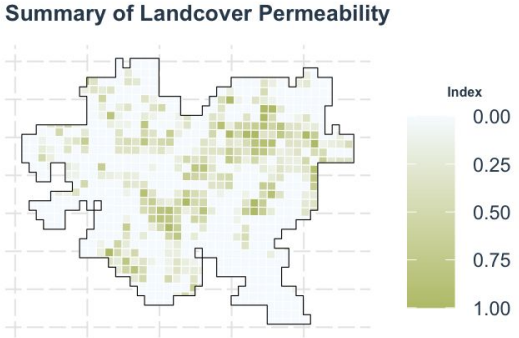


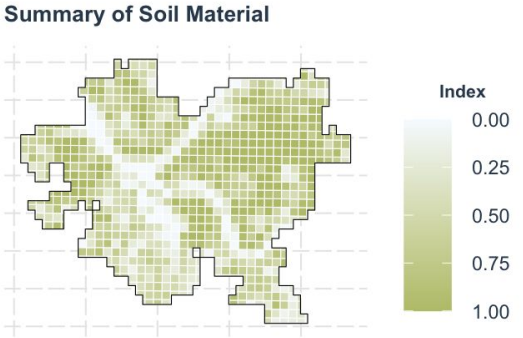
Comparable City - Pittsburgh



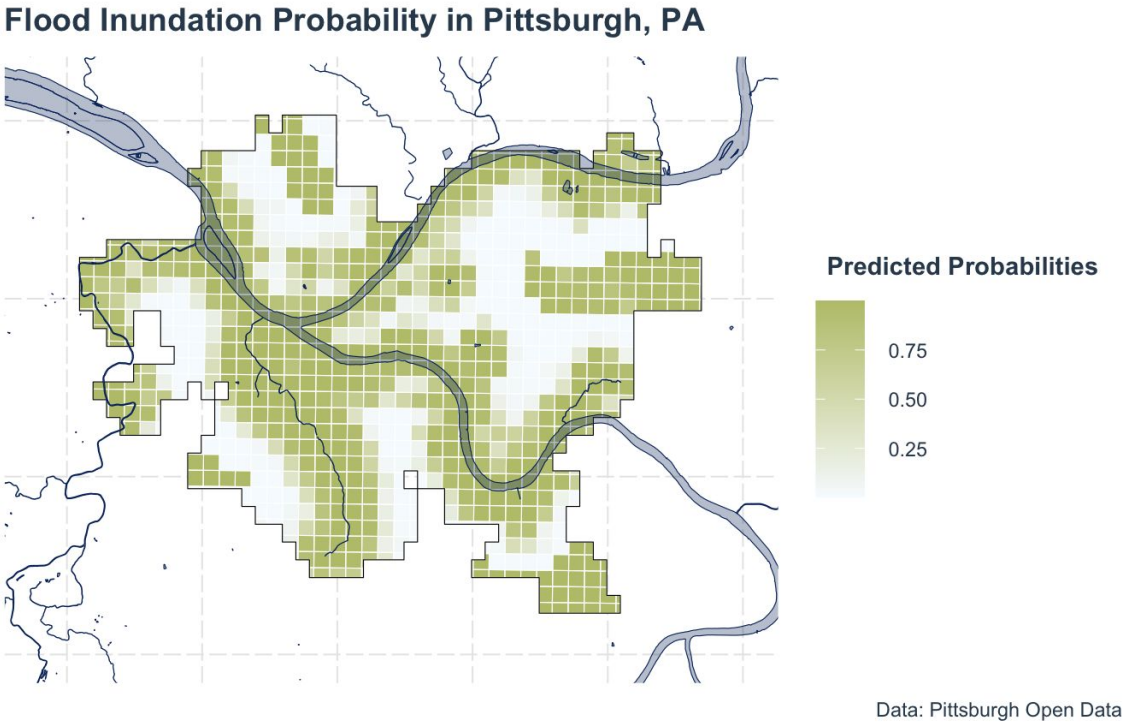








Pittsburgh - Prediction Map



The resulting forecast map shows that the high probability of flooding is concentrated along the Monongahela and Allegheny rivers.

The average of probability is 0.6190202. 25% of the predictions fall below 0.0708185, meanwhile 75% of the predictions fall below 0.9905759.

Thank You!

Flood Inundation Prediction
in Calgary, Canada

Haobing Liu & Ling Chen

2023-03-28