Enhancing flood susceptibility modelling in Canada: Integrating seasonal meteorological data, feature selection and machine learning approaches

Natural Resources Canada | Ressources naturelles Canada

Canada

EGU25-3871-A.28

Karen Elaine Dunbar[1,2,a], Heather McGrath[2,b], Usman T. Khan[1,c]

[1]Civil Engineering, York University, Toronto, Canada; [2]Natural Resources Canada, Ottawa, Canada

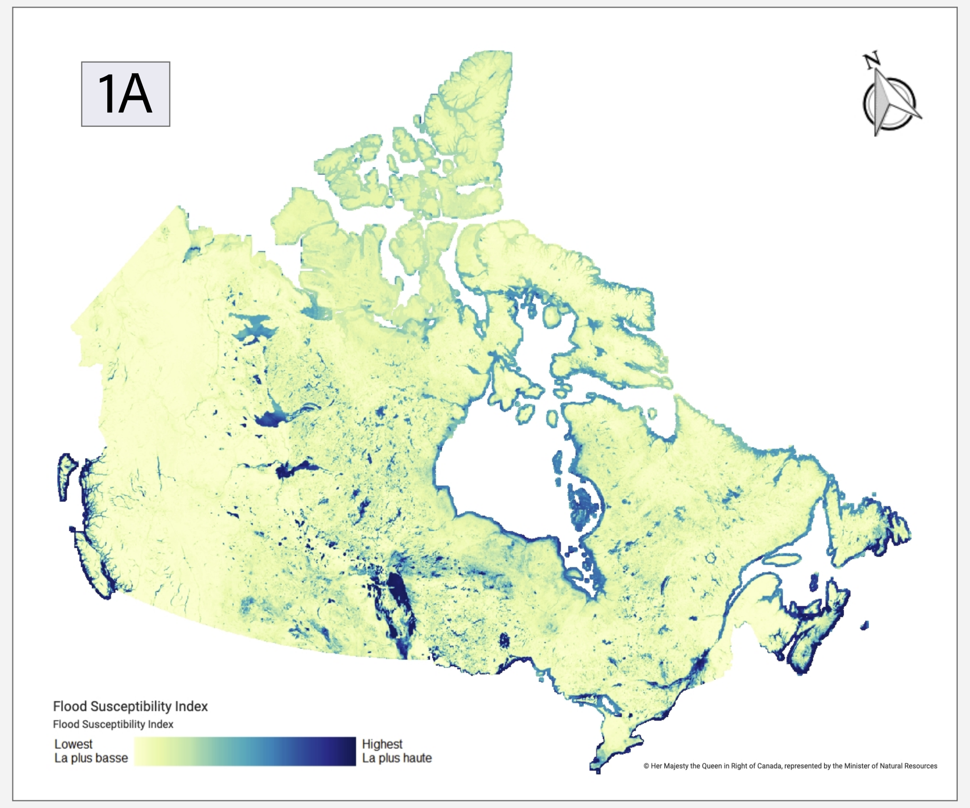
[a]karela@yorku.ca, [b]heather.mcgrath@NRCan-RNCan.gc.ca, [c]utkhan@yorku.ca

LASSONDE SCHOOL OF ENGINEERING | YORK UNIVERSITÉ UNIVERSITY

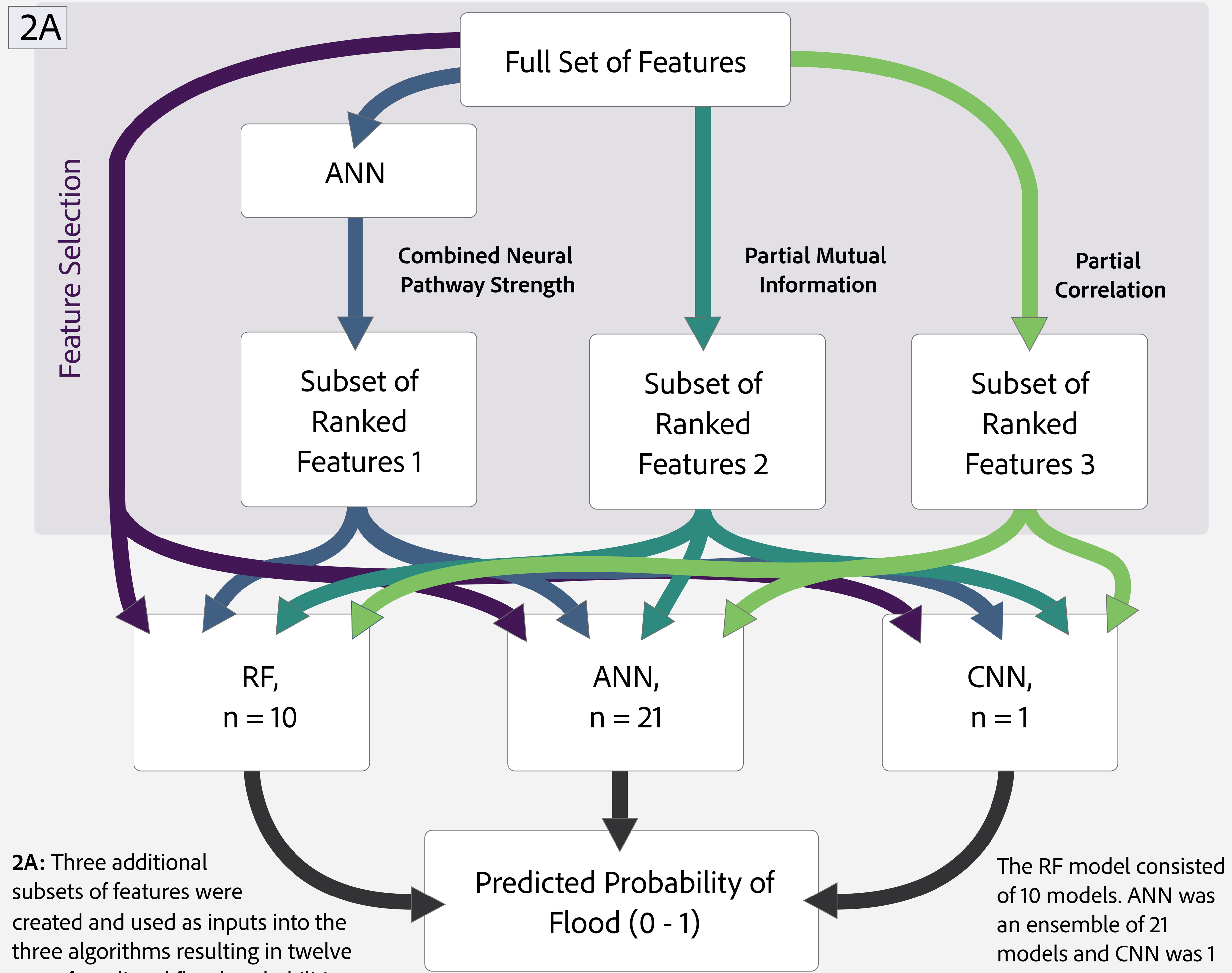
INTRODUCTION

1A: Floods are the costliest natural hazard in Canada in terms of direct infrastructure damage. The Canadian flood susceptibility index (FSI) was developed in 2022 with an overall accuracy of 0.89. This research focused on updating the FSI with the following objectives:

- Inclusion of annual seasonal features and additional geospatial features
- Perform three feature selection methods to determine useful inputs: partial correlation, partial mutual information and combined neural pathway strength
- Compare and optimize the model performance of three models for prediction of flood susceptible areas: random forest, artificial neural network and convoluted neural network



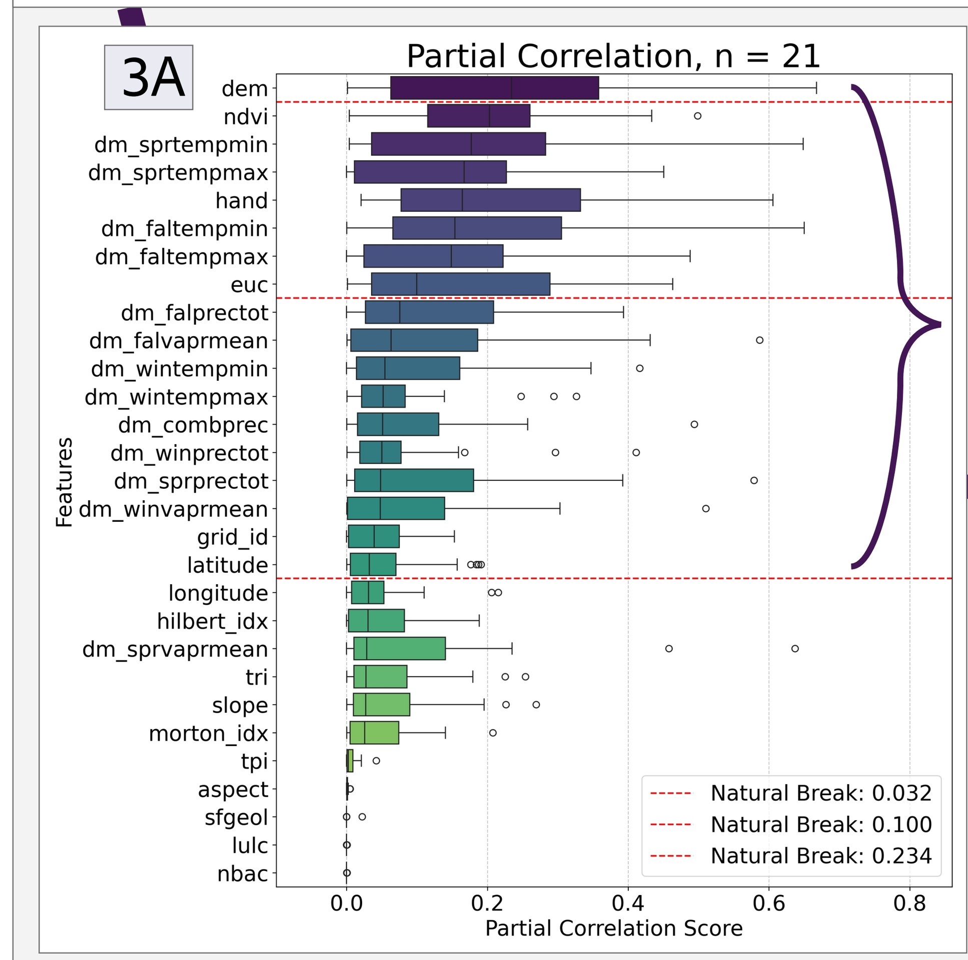
METHODOLOGY

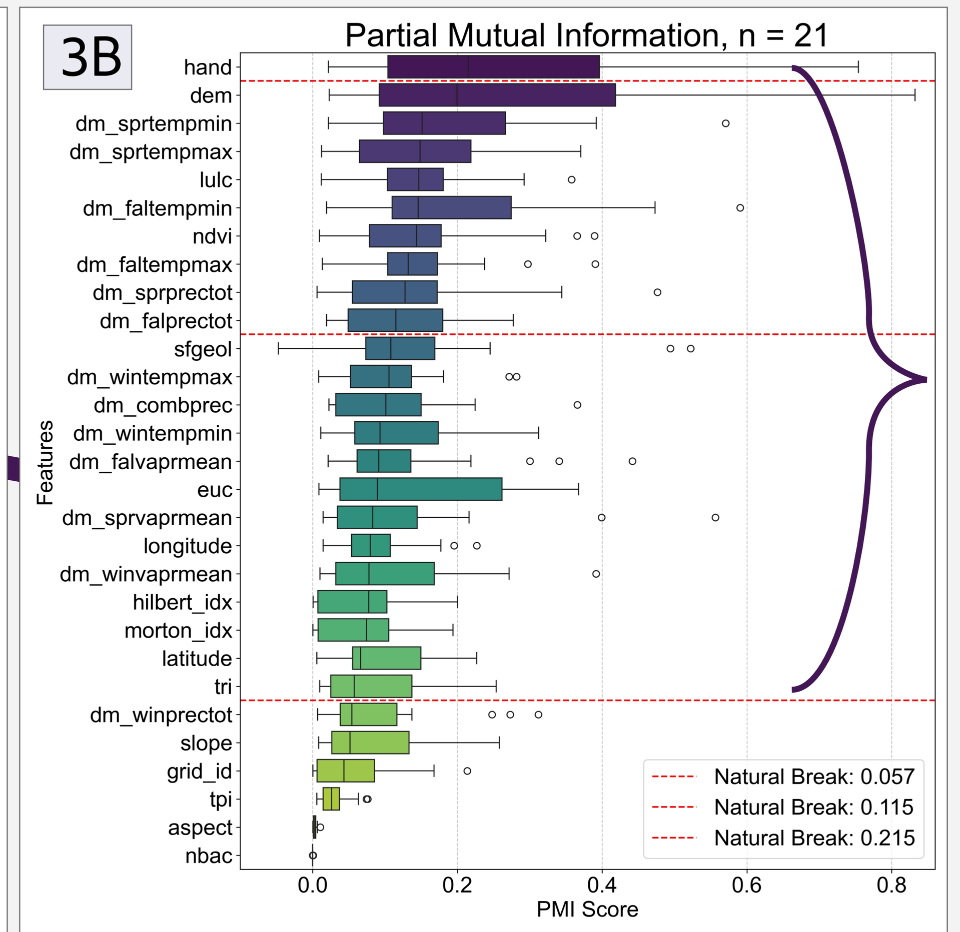


2A: Three additional subsets of features were created and used as inputs into the three algorithms resulting in twelve sets of predicted flood probabilities.

The RF model consisted of 10 models. ANN was an ensemble of 21 models and CNN was 1 model.

RESULTS





3A-D: Feature Selection (FS)

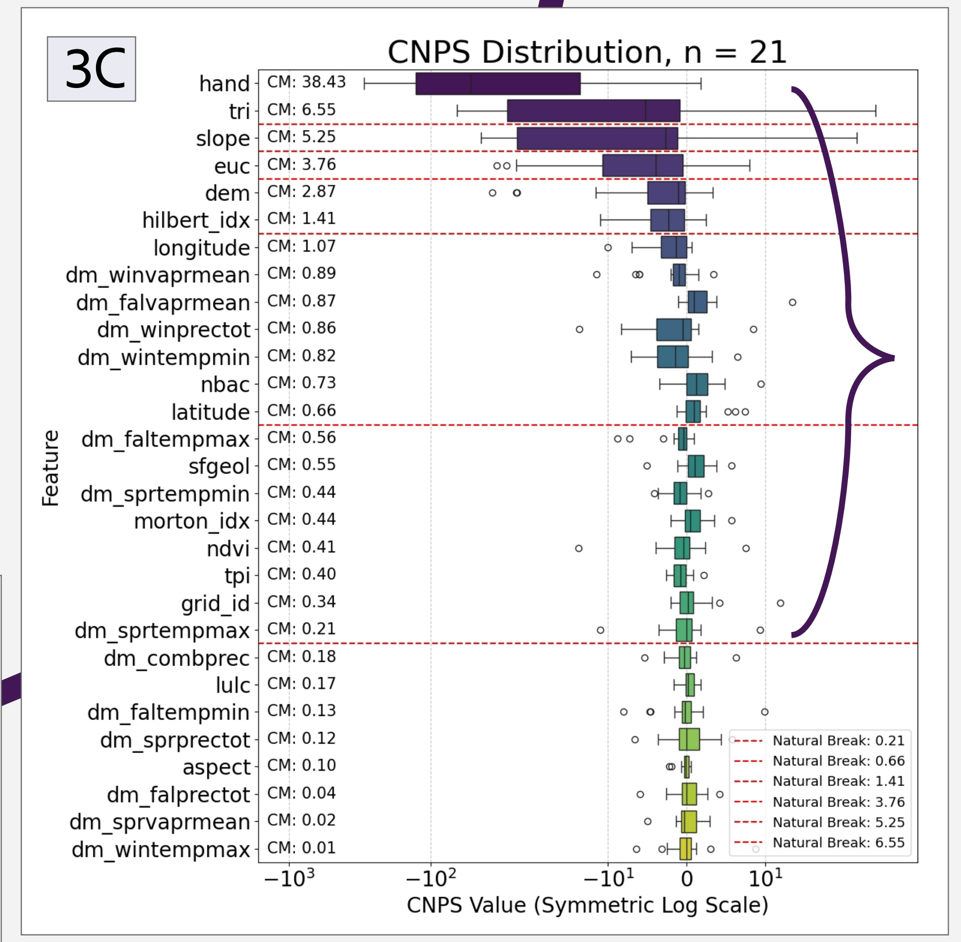
Features are ranked from highest importance to least. Natural breaks *(Fisher-Jenks natural breaks algorithm)* are in dashed red lines, creating groups of features that have similar levels of importance. Features within the curly brackets are features chosen for each group for modelling.

3A: Partial Correlation (PC) ranks features by fitting a linear regression model to predict the target using *feature A* and then again using *all other features*. The PC value is the residual between the 2 models.

3D

Feature (n = 29)	PC (n = 18)	PMI (n = 23)	CNPS (n = 21)
combined precip (dm_combprec)			
fall total precip (dm_falprectot)			
fall max temp (dm_faltempmax)			
fall min temp (dm_faltempmin)			
fall mean vapour pressure (dm_falvaprmean)			
spring total precip (dm_sprprectot)			
spring max temp (dm_sprtempmax)			
spring min temp (dm_sprtempmin)			
spring mean vapour pressure (dm_sprvaprmean)			
winter total precip (dm_winprectot)			
winter max temp (dm_wintempmax)			
winter min temp (dm_wintempmin)			
winter mean vapour pressure (dm_winvaprmean)			
aspect			
digital elevation model (dem)			
euclidean distance to river (euc)			
height above nearest drainage (hand)			
national burn area composite (nbac)			
normalized difference vegetation index (ndvi)			
slope			
topographic position index (tpi)			
terrain roughness index (tri)			
latitude			
longitude			
morton index (morton_idx)			
hilbert index (hilbert_idx)			
grid ID (grid_id)			
land use/ land cover (lulc)			
surficial geology (sfgeol)			

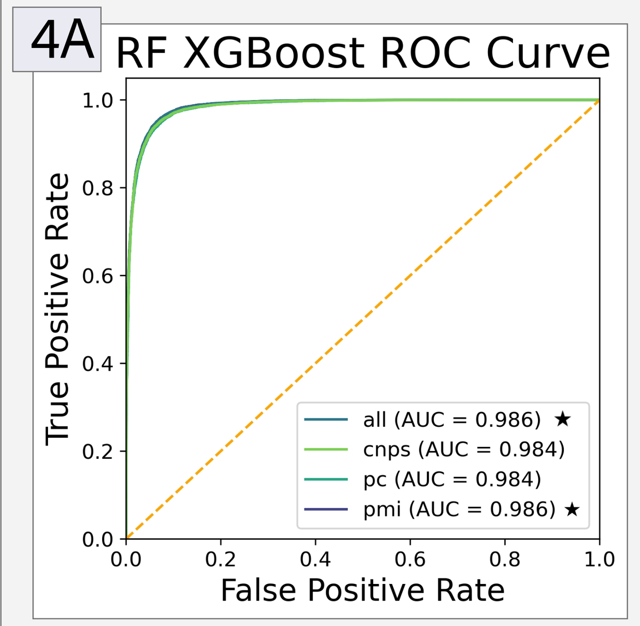
Seasonal Climatological Features — Geospatial Features

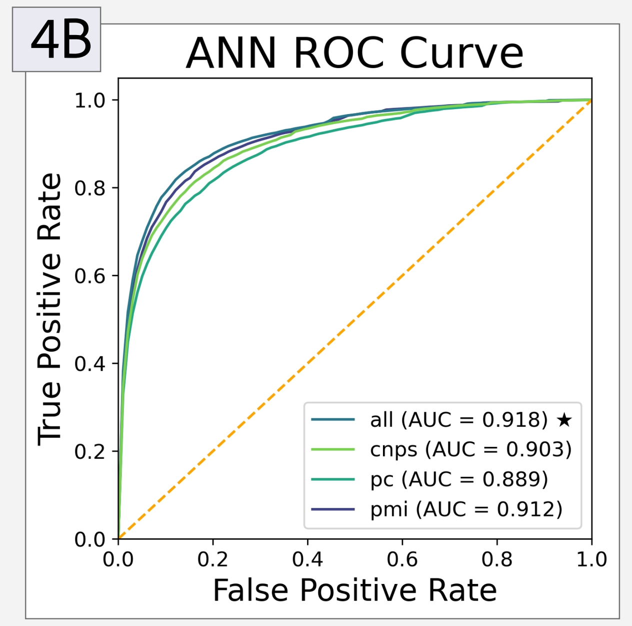


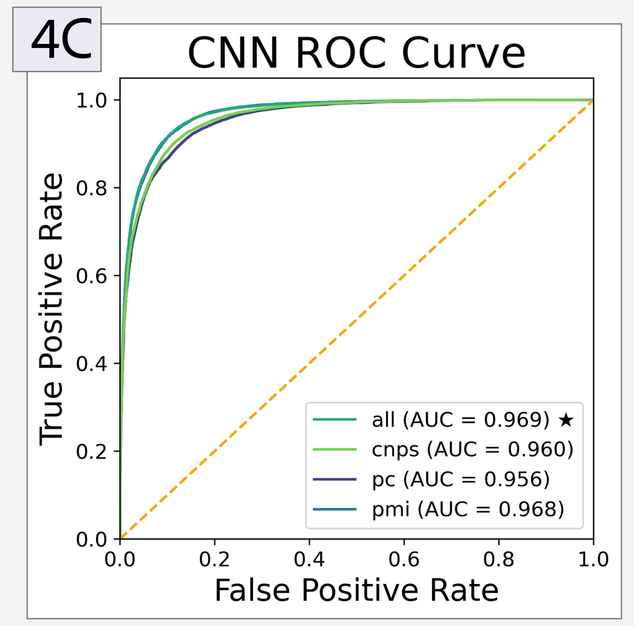
3B: Partial Mutual Information (PMI) ranks features by fitting a non-parametric kernel regression model to predict the target using *feature A* and then again using *all other features*. The PMI value is the residual between the 2 models.

3C: Combined Neural Pathway Strength (CNPS) requires training an ANN model using *all the features* then saving the pathway strength (weights) of the layers. Features are ranked by a combined metric (CM) of pathway consistency and relative range of CNPS values.

3D: Summary of important features across the 3 FS methods.

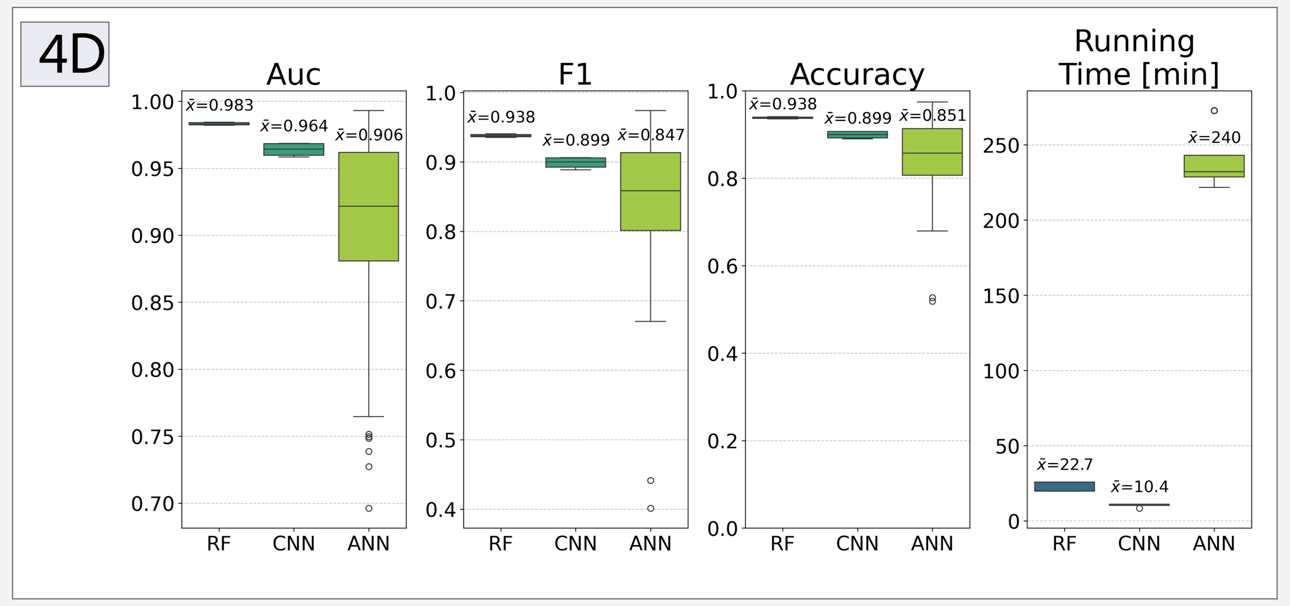






4A-D: Flood Probability Prediction Model Comparison

Performance results of the flood probability models are shown in terms of ROC curves in 4A-4C. F1, Accuracy and Running Time are shown in 4D. The performance for the four subsets of input features are compared for each model algorithm, with the highest performing model emphasized by a star. RF performed the best followed by CNN and ANN. For all algorithms, inclusion of all 29 features resulted in the highest AUC.



4A: All input feature subsets performed well with RF resulting in AUC of over 0.98 for all subsets.

4B: ANN models had the lowest performance with AUC ranging from 0.889 to 0.918.

4C: CNN models resulted in AUC ranging from 0.956 to 0.969

4D: Across all metrics, RF performed the best, while ANN took the longest to develop and performed the worst.

CONCLUSIONS & FUTURE WORK

Inclusion of all features led to the best performance for all 3 model algorithms, which is not surprising since there is a small number of features (29) to begin with, compared to the total sample size. Seasonal climate variables are not often included in FSM, but prove to be useful in model learning. With the preliminary CNN model development and high performance of 1 model, CNN shows great potential. Future work include the continued development of the models and testing on a Canada-wide dataset.

REFERENCES

Bentivoglio, R., Isufi, E., Jonkman, S. N., & Taormina, R. (2022). Deep learning methods for flood mapping: a review of existing applications and future research directions. *Hydrology and Earth System Sciences*, 26(16), 1-50. doi:10.5194/hess-26-4345-2022

Kaya, C. M., & Derin, L. (2023). Parameters and methods used in flood susceptibility mapping: a review. *Journal of Water and Climate Change*, 14(6). doi:10.2166/wcc.2023.035

Khalid, R., & Khan, U. T. (2024). Flood susceptibility mapping using ANNs: a case study in model generalization and accuracy from Ontario, Canada. *Geocarto International*, 39(1). doi:10.1080/10106049.2024.2316653

McGrath, H., & Gohl, P. N. (2022). Accessing the impact of meteorological variables on machine learning flood susceptibility mapping. *Remote Sensing*, 14(1656). doi:10.3390/rs14071656

McGrath, H., & Gohl, P. N. (2023). Prediction and classification of flood susceptibility based on historic record in a large, diverse, and data sparse country. Environmental Sciences Proceedings, 25(1), 18.

Natural Resources Canada. (2023, 07 10). Flood Mapping. Retrieved from Government of Canada: https://natural-resources.canada.ca/science-and-data/science-and-research/natural-hazards/flood-mapping/24223

Snieder, E., Shakir, R., & Khan, U. T. (2020). A comprehensive comparison of four input variable selection methods for artificial neural network flow forecasting models. *Journal of Hydrology*, 583, 124299.