Comparative Analysis of Feature Selection Methods in ML-based Flood Susceptibility Mapping

Heather McGrath[1], Karen E Dunbar[1,2] and Usman Khan[2]
Natural Resources Canada, Ottawa, ON, Canada [1]
E-mail: heather.mcgrath@NRCan-RNCan.gc.ca

Civil Engineering, Lassonde School of Engineering, York University, Toronto, Canada [2]
E-mail: karela@yorku.ca
E-mail: usman.khan@lassonde.yorku.ca

ABSTRACT

Traditional flood susceptibility maps typically rely on static geospatial factors and ignore dynamic meteorological variables, thereby introducing temporal biases. This study addresses these issues by applying different feature selection methods to identify the most influential flood conditioning factors, with a particular focus on including seasonal meteorological data, which is especially important in Canada, where many floods occur during the spring freshet. Three feature selection techniques were used to develop multiple feature sets: Partial Correlation (PC), Partial Mutual Information (PMI), and Combined Neural Pathway Selection (CNPS). In addition to these, two other sets were examined: the full feature set and the base geospatial set (without seasonal meteorological variables). These sets served as inputs for training and evaluating three machine learning models: Extreme Gradient Boosting (XGBoost), an Artificial Neural Network (ANN), and a Convolutional Neural Network (CNN). The PC and PMI feature sets produced results similar to the full feature set, indicating that the feature selection process can reduce model complexity without losing predictive performance. XGBoost performed best, with either the full feature set or the PMI set. Across all models, the highest performance was achieved with the full feature set, highlighting the importance of seasonal meteorological variables for flood susceptibility modelling.

KEYWORDS: Flood susceptibility mapping, feature selection, partial correlation, partial mutual information, combined neural pathway strength

1 INTRODUCTION

Floods are Canada's most expensive hazard, driven by climate change and rapid urbanization (Ajin et al., 2025; Canada, 2025). While Flood Susceptibility Mapping (FSM) is essential for mitigation, traditional methods face challenges regarding data resolution, subjectivity and the exclusion of dynamic variables (El Haou et al., 2025). Most FSM studies rely on static topographic features, overlooking meteorological factors like river gage data, snow accumulation and heavy rain frequency that are crucial for Canadian flood dynamics (Khalid and Khan, 2024; McGrath and Gohl, 2022). Although Machine Learning (ML) methods (e.g., XGBoost, ANN, CNN) excel at modelling non-linear relationships, their performance depends heavily on the quality and quantity of the input data (Bentivoglio et al., 2022; Long et al., 2025). To improve predictive accuracy of ML based FSM, this study focuses on feature selection, which is often overlooked in FSM (Li et al., 2022). By applying Partial Correlation, Partial Mutual Information, and Combined Neural Pathway Strength (Snieder et al., 2020), we aim to integrate seasonal meteorological data in ML models, overcoming the limitations of static susceptibility mapping. To overcome temporal biases in traditional flood modelling, this study focuses on feature selection to integrate dynamic seasonal meteorological variables. By analyzing multi-year flood data using various feature selection methods, this study aims to produce accurate, seasonally aware flood susceptibility maps. The approach will emphasize the critical role of seasonal dynamics in flood susceptibility, enhancing risk mitigation strategies.

2 METHODS

2.1 Study Area and Data Description

The study area is shown in Figure 1 and was chosen due to the availability of historical flood data from the Canadian Flood Archive (Natural Resources Canada, 2018). It comprises of multiple watersheds that cover distinct areas across Canada. Specifically, the study sites range from the mountainous regions of British Columbia and Alberta to the flatlands of the Canadian Prairies to the maritime environment of the Atlantic provinces. The study areas encompass portions of Canada's Major Drainage Areas (Canada, 2024), including the Yukon River (blue), Pacific (orange), Great Slave Lake (red), Mississippi River (purple), Nelson River (cyan), St. Lawrence (pink), and Maritime (brown). The study areas cover 18 and 28 categories of land use and surficial geology, respectively, and include significant topographical variation, with elevations ranging from 0 to 2082 meters and slopes ranging from 0 to 54 degrees.

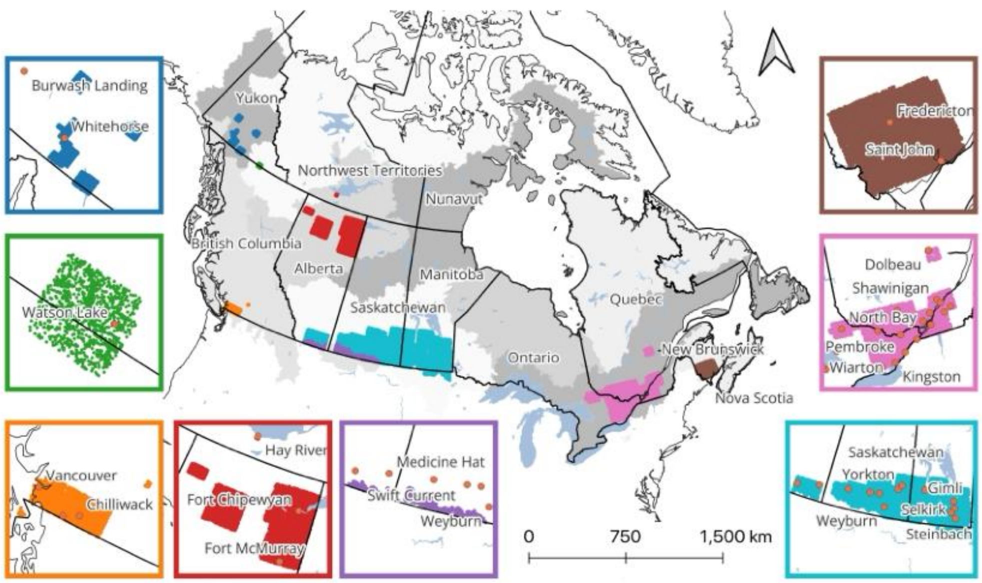

Figure 1: Spatial distribution of the study area, covering multiple zones across Canada. The colours from the inset maps correspond to the specific geographic subsets: Yukon River (blue), Pacific (orange), Great Slave Lake (red), Mississippi River (purple), Nelson River (cyan), St. Lawrence (pink), and Maritime (brown). Shaded areas represent the Major Drainage Areas.

The flood inventory was obtained from the Emergency Geomatics Services (EGS) flood archive, which represents historic flood extents produced using Synthetic Aperture Radar (SAR) and/or optical satellite imagery (Natural Resources Canada, 2018; Natural Resources Canada et al., 2021), with the flood polygon data from 2005 to 2022. There were limited data in Ontario, Northern Quebec and the territories. Overall, 21 data subsets (>200,000 data points) were used in the analysis, distinguished by year, season (Spring, Summer or Fall), and location. One site location can have multiple data points over the study period. An equal number of non-flood points were randomly generated within the study area boundaries to avoid class imbalance, resulting in approximately equal numbers of flood and non-flood points.

FCFs were categorized into geospatial and seasonal climatological groups. Geospatial features included topographical variables (e.g., DTM, slope, aspect, HAND, NDVI, land use), spatial indices (latitude, longitude, Morton/Hilbert indices), and the National Burn Area Composite, reflecting the link between the wildfires and post-fire flood risk (Abogadil and Khan, 2023; Saxe et al., 2018). While often excluded from FSM, meteorological data have recently been shown to improve predictive performance (Khalid and Khan, 2024; McGrath and Gohl, 2022). Consequently, this study integrated seasonal (fall, winter, spring) climatological variables, including total precipitation, temperature extremes, and mean vapour pressure. The full list of FCFs in this study are shown in Figure 2

2.2 Feature Selection Methods

Three feature selection methods were used: Partial Correlation, Partial Mutual Information and Combined Neural Pathway Strength, yielding three ranked feature subsets (in addition to the full set of features and base (only geospatial) to provide insight into the relative strength or influence of each feature on the output. Each feature selection method was use individually for each of the 21 data subsets. The feature subsets were determined using the Natural Breaks optimization (the *jenkspy* algorithm), which calculated intrinsic thresholds for the importance scores.

Partial correlation (PC) measures the direct linear relationship between the target and feature A, while removing the influence of all other features. For instance, the flood label is first predicted using *feature A*, and then using all features other than *A*. The features are ranked by higher linear correlation. This process is repeated for all features until a termination criterion is met. In this case, the process was applied to all features, and the features were selected using Natural Breaks. Partial Mutual Information (PMI) measures the linear or non-linear relationship between the target and feature A, while removing the influence of all other features. The calculation process is similar to PC; however, it uses a non-parametric kernel regression. The residuals from the two models are used to calculate mutual information using the *mutual_info_regression* function from the Sci-Kit (version 1.7.1) package. The features are ranked by higher mutual information. The process is repeated for all features and the final set of features chosen by Natural Breaks. The combined neural pathway strength (CNPS) (Snieder et al., 2020) was used to examine the influence of each feature during model training. In contrast to PC and PMI, CNPS requires a trained ANN model before it can analyze the impact of each feature on the target.

After model training, the weights or strength of pathways for each layer are recorded, and then the input-hidden and hidden-output weight matrices for each fold are multiplied. An alpha variable (Equation 1) measures the consistency of the pathway behaviour. An alpha value of 0.50 means the feature is either excitatory or inhibitory for half of the folds, essentially cancelling its contribution. An alpha value of 1.0 means the feature is either excitatory or inhibitory across all folds and therefore exhibits very consistent pathway behaviour.

$$\alpha_j = \frac{max(\sum(CNPS_j > 0), \sum(CNPS_j < 0))}{n} \tag{1}$$

Multiple features may have an alpha value of 1.0, so a score (Equation 2) is calculated to rank the features. The score used in (Snieder et al., 2020) has been modified so that a smaller range between the minimum and maximum CNPS values yields an optimal score of 1.0, which represents more consistent pathway behaviour.

$$score = 1 - \frac{\max(CNPS_j) - \min(CNPS_j)}{|\max(CNPS_j)| + |\min(CNPS_j)|} \tag{2}$$

To balance the CNPS weights with the score, a combined metric (Equation 3) was used to rank the features, assigning equal weight to the score and the CNPS weights (w_1 and w_2).

$$Combined\ Metric = w_1 \cdot Score + w_2 \cdot |Average\ CNPS| \tag{3}$$

2.3 Model Development and Training

Three model algorithms were developed: Extreme Gradient Boosting (XGBoost), Artificial Neural Network (ANN) and Convolutional Neural Network (CNN). All models produced flood probability between 0 and 1, which were subsequently converted to a binary flood/non-flood classification threshold of 0.5. An ensemble of 10 XGBoost models was randomly trained on 70%, validated on 15% and tested on 15% of the dataset. Each tree had a maximum depth of 20, and a maximum of 200 estimators (trees built sequentially). The *subsample* and *colsample_bytree* parameters were 80%; the models were randomly trained on 80% of the rows using 80% of the features (columns). The L1 (*reg_alpha*) and L2 (*reg_lambda*) regularization parameters were set to 0.5 and 5. The model was evaluated using *logloss*.

An ANN is trained using backpropagation, which repeatedly presents the model with data, measures the prediction error, and then adjusts its parameters to reduce the error. An ANN with two hidden layers and 50 neurons was chosen for its higher performance compared to other hidden-layer neuron configurations. The model was trained for 100 epochs with a batch size of 32 and an early stopping patience of 10 epochs. Adam optimizer with a learning rate of 0.0001 and a dropout rate of 50% was used. The activation function was ReLU, and the loss function was binary cross-entropy with logits loss. The ANN model was cross-validated using leave-two-out cross-validation. Since there were 21 data subsets, each was used as a test set at least once, yielding an ensemble of 21 models.

CNNs are trained using a backpropagation process similar to that of ANNs. Two branches processed two different types of input before combining them as input to the dense layers. The CNN branch processed the static geospatial features, while the LSTM branch processed the dynamic temporal features. In each branch, the 1D input was filtered through 1D convolutional filters, where ReLU was applied as the activation function. One-dimensional feature maps were created during max pooling, then further flattened. The flattened feature maps from each branch were then concatenated and fed into a fully dense 128-neuron layer. A 50% dropout layer was applied to create the output, which was activated by the sigmoid function. The CNN model was randomly trained on 70%, validated on 15% and tested on 15%.

Model performance was evaluated using Accuracy, AUC-ROC, and F1 Score. While Accuracy can be biased in imbalanced datasets (El Haou et al., 2025), this was mitigated by the balancing of the training data (~50% flood/ non-flood). AUC-ROC assesses classification capability across the flood thresholds, ranging from 0.5 (random) to 1.0 (perfect classifier) (El Haou et al., 2025; Khalid and Khan, 2024). Finally, the F1 score is the harmonic mean of precision and recall and was used to quantify the trade-off between detecting floods and minimizing false alarms.

3 RESULTS

3.1 Feature Importance

Figure 2 summarizes the feature importance rankings and final selection results across the PC, PMI and CNPS methods. The heatmap indicates the individual rank assigned by each feature selection method; darker blue indicates higher importance (closer to 1). The features are ordered along the x-axis by their average ranking across all three methods. The ranking shows that the seasonal meteorological (labelled in dark blue), particularly the temperature features, consistently rank higher than most geospatial features (labelled in green), indicating that these are less significant predictors. Several features were strongly predicted as high flood drivers (spring maximum temperature, NDVI, and fall minimum temperature). The red border identifies the selected features for each method. The Natural Breaks thresholds yielded varying numbers of features selected by each method (ranging from n = 18 for PC to n = 21 for CNPS).

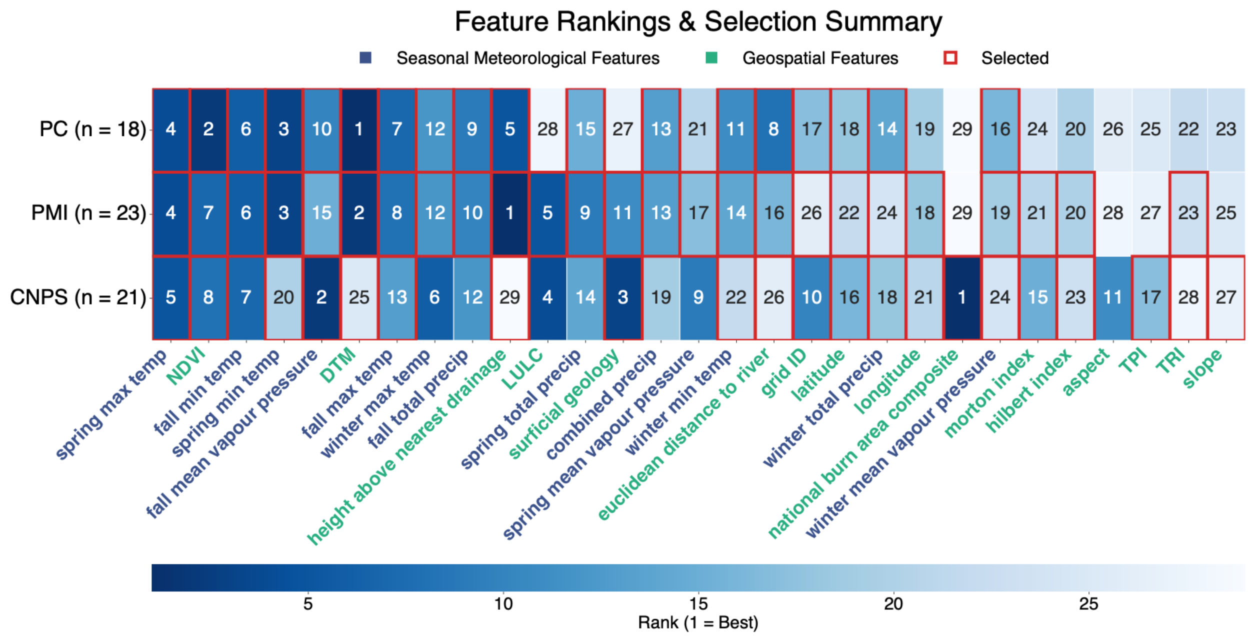

Figure 2: Feature ranking heatmap and selection summary. The rows display the ranks (1-29) assigned to each feature by the PC, PMI, and CNPS methods. The columns are sorted by the average rank across all methods (features on the left had a higher average ranking). The red borders indicate which features are included in each feature subset based on the Natural Jenks thresholds. The count of features per method is represented by n. The text colour distinguishes the seasonal meteorological and geospatial features.

3.2 Model Predictive Performance

Figure 3 displays the AUC-ROC curves for each model across the five feature sets. The top-performing feature set for each model is marked with a star. While the full feature set achieved the highest scores, the performance gap between it and the others is very minimal across all models (< 0.10%); the PMI set had an identical score with the full set with the XGBoost model. Overall, XGBoost was the top classifier (AUC = 0.985), outperforming the top CNN and ANN models by 1.65% and 6.95%, respectively. The top CNN model achieved an AUC of 0.967, while the top ANN model had an AUC of 0.921.

The distribution of performance across the different FS subsets for each model is shown in Figure 4 for the AUC, F1-score and Accuracy metrics. The XGBoost model consistently demonstrated the best performance and stability, achieving the highest mean scores across all three metrics (AUC = 0.981, F1 = 0.932, Accuracy = 0.932), with small variance. The CNN model ranked second-best (mean AUC = 0.962) and demonstrated high stability. In contrast, the ANN model showed the lowest overall performance (mean AUC = 0.847) and high variability in its predictive performance across FS methods.

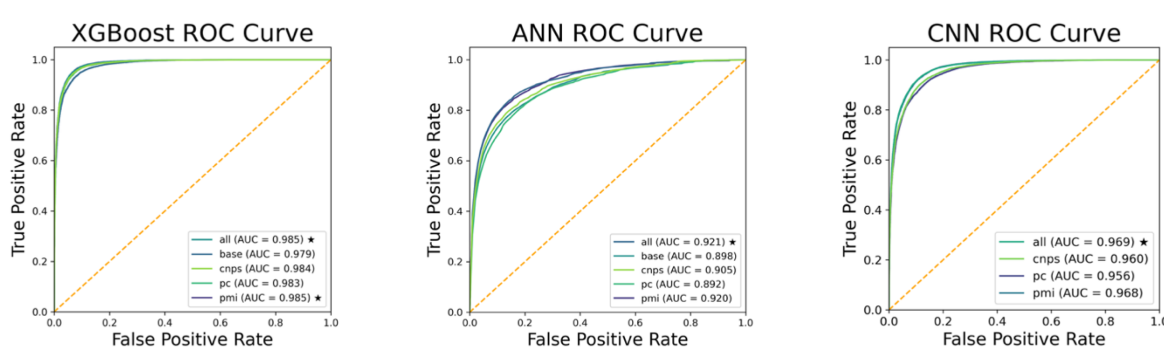

Figure 3: AUC-ROC curves for the three models across the five feature sets

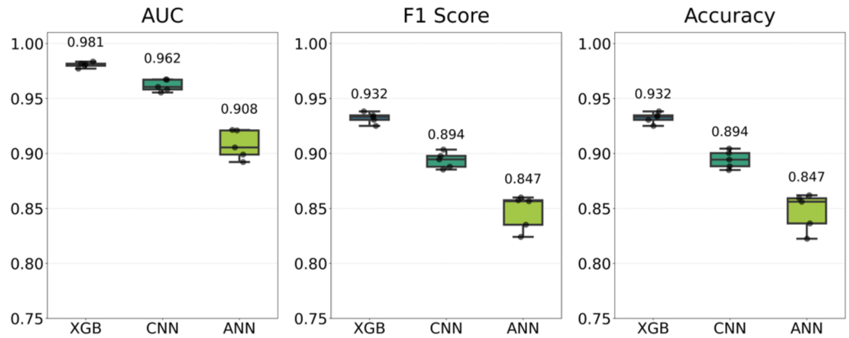

Figure 4: Summary of the model performances for all feature sets across all metrics

3.3 Flood Susceptibility Mapping

The spatial distribution of flood susceptibility for the Spring of 2015 at the New Brunswick study site is shown in Figure 5, generated by the XGBoost model using all features. This area was among the best-predicted of all the study sites. The map shows high susceptibility areas in dark red, strongly correlated with the Saint John River. The near-zero susceptibility areas are transparent.

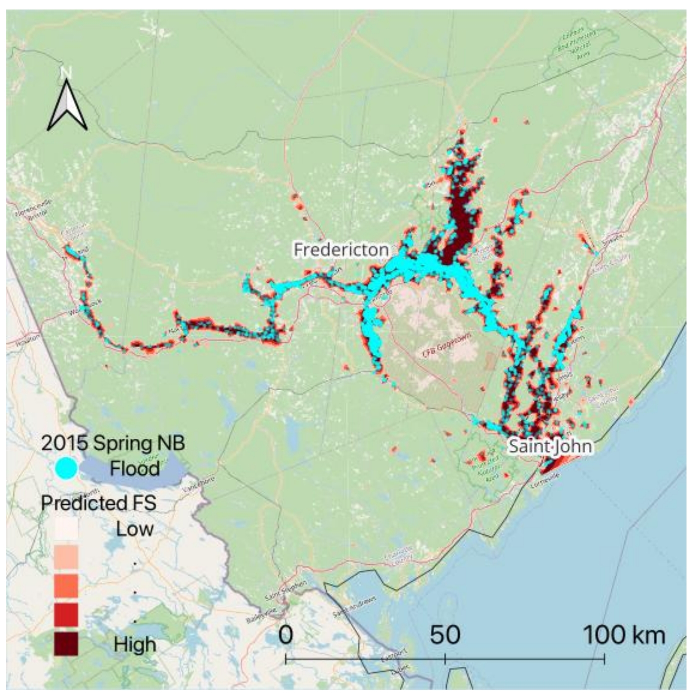

Figure 5: Flood susceptibility output from the XGBoost model using all features.

4 DISCUSSION

Among the feature selection methods, the model trained on the full feature set achieved the highest performance across all model types. However, this came at the cost of greater model complexity, with only a marginal improvement over the models trained on the PC, PMI, and CNPS feature sets. Therefore, these results show that feature selection was highly effective at reducing model complexity without sacrificing predictive power. The PC and PMI feature sets are superior in terms of practicality, requiring less data and achieving faster training and testing times. The feature selection process in this study achieved higher accuracy more efficiently than using the full set of features.

Regarding the specific features selected, it is interesting that not all geospatial features in the baseline set were retained by any of the three feature selection methods; therefore, this baseline set, which is commonly used in FSM studies, is not the optimal feature set. All methods found fewer geospatial indices important (only NDVI, DTM and HAND were included), compared to the higher number of seasonal features. Spatial indices such as latitude, longitude, grid ID, and the Morton and Hilbert Indices were also ranked low, indicating the physical drivers were prioritized over geographic coordinates. This suggests the models prioritized process-based drivers over geographic location which may improve generalizability. All methods excluded aspect. Traditionally important geospatial features such as land use, surficial geology, and derived features such as slope, TPI, and TRI were excluded by either PC or PMI, which could be attributed to existing redundancies with DTM. The importance of seasonal meteorological variables in flood susceptibility is confirmed by its inclusion in all three feature selection methods. There is merit in incorporating seasonal variables that specifically represent local flood characteristics; for instance, this study area may have benefited from temperature and volume variables that capture snowmelt dynamics. PMI and PC selected all (12) except one of the seasonal variables, while CNPS selected seven. The increase in model performance with the inclusion of seasonal variables was the highest for the ANN models, even though the CNPS included the least number of seasonal variables.

When comparing model algorithms, the results of this study show that feature selection methods vary in their rankings, highlighting the different interplay among the features. The predictive power of FCFs varies with the chosen FS method and model, meaning the same variables may contribute differently across models, as observed in other studies (Li et al., 2022; Snieder et al., 2020). On average, the XGBoost models, regardless of the input feature set, performed best (AUC = 0.983, F1 = 0.938, Accuracy = 0.938). However, the best model isn't the one with the highest score, but one that balances performance with simplicity.

Several limitations related to the input data and modelling approach exist for this study. First, the data resolution was 30 meters, which may not be fine enough to capture localized topographical features that contribute to flood patterns; this is highly impactful in areas that have less topographical variability, such as in Manitoba. Further, the absence of urban drainage networks means that the models do not account for infrastructure, which is sufficient for our flood susceptibility focus, but not for hydraulic routing. Regarding the modelling approach, the outputs highlight the flood-susceptible areas; however, they do not provide quantitative measures, such as flood inundation depth, which are important for risk assessment. Since the model outputs were binary, the models were tasked with classification, which makes it easier to achieve a higher performance. Finally, there is inherent difficulty in interpreting the ML models. The feature selection results were complex and, at times, counterintuitive, which could be attributed to the large variability of the study areas in the multiple datasets. This is a common challenge in ML-based FSM studies; employing explainable AI (XAI) methods such as SHAP (SHapley Additive exPlanations) could provide insights into the model's learning process.

Future work can expand on this current study in several ways. Advanced hybrid models, such as CNN combined with Long Short-Term Memory networks (CNN-LSTM), could be tested to capture both the spatial and temporal aspects of flood susceptibility. The generalizability of this model can be examined by training it on a single study area (rather than multiple, as in this study) and evaluating its performance on a second study area. Lastly, incorporating climate change projections would be a valuable next step to help forecast future flood susceptibility, thereby enhancing long-term risk management.

5 CONCLUSIONS

This study developed a data-driven framework for flood susceptibility mapping across diverse Canadian hydroclimatic zones by comparing XGBoost, ANN, and CNN architectures, each combined with three feature selection methods. The results show that, while all models achieved high predictive accuracy, XGBoost consistently outperformed the others. The feature selection analysis demonstrated that the traditional geospatial features and spatial coordinates are often redundant. In contrast, the seasonal meteorological variables are critical for capturing the dynamic nature of flood susceptibility. This research finds that a simplified XGBoost model, optimized by feature selection, results in the highest performance for a national-level flood susceptibility mapping.

REFERENCES

Abogadil, K., Khan, U., 2023. An analysis of 100 years of post-fire streamflow responses of British Columbia watersheds. https://doi.org/10.5194/egusphere-egu23-8043

Ajin, R.S., Costache, R., Bărbulescu, A., Fanti, R., Segoni, S., 2025. Flood Susceptibility Assessment Using Multi-Tier Feature Selection and Ensemble Boosting Machine Learning Models. Water 17, 2041. https://doi.org/10.3390/w17142041

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. Hydrol. Earth Syst. Sci. 26, 4345–4378. https://doi.org/10.5194/hess-26-4345-2022

Canada, 2024. Atlas of Canada - Canadian Drainage Areas [WWW Document]. URL https://atlas.gc.ca/drainage-areas/Atlas_Drainage_Areas_EN.html?_gl=1*1wbttoy*_ga*MzkzNDA3OTIuMTY4MjcwMjc2MQ..*_ga_C2N57Y7DX5*MTc0MTYzMzMxNy4xMjYuMS4xNzQxNjM0MTM5LjAuMC4w (accessed 12.1.25).

Canada, N.R., 2025. Floods in Canada - Archive - Open Government Portal [WWW Document]. URL https://open.canada.ca/data/en/dataset/74144824-206e-4cea-9fb9-72925a128189 (accessed 9.22.25).

El Haou, M., Ourribane, M., Keshavarzi, A., Ismaili, M., Eloudi, H., Hajji, S., El Bouzkraoui, M., Namous, M., Krimissa, S., 2025. Hybrid machine learning models for urban flood susceptibility assessment in semi-arid Beni Mellal, Morocco: a spatial and comparative analysis. Model. Earth Syst. Environ. 11, 403. https://doi.org/10.1007/s40808-025-02584-9

Khalid, R., Khan, U.T., 2024. Flood susceptibility mapping using ANNs: a case study in model generalization and accuracy from Ontario, Canada. Geocarto Int. 39, 2316653. https://doi.org/10.1080/10106049.2024.2316653

Li, J., Zhang, H., Zhao, J., Guo, X., Rihan, W., Deng, G., 2022. Embedded Feature Selection and Machine Learning Methods for Flash Flood Susceptibility-Mapping in the Mainstream Songhua River Basin, China. Remote Sens. 14, 5523. https://doi.org/10.3390/rs14215523

Long, G., Tantanee, S., Nusit, K., Sooraksa, P., 2025. Flood Susceptibility Mapping Using Machine Learning Models with Novel Flood Inventory Sampling Strategies. Sens. Mater. 37, 3829. https://doi.org/10.18494/SAM5586

McGrath, H., Gohl, P.N., 2022. Accessing the Impact of Meteorological Variables on Machine Learning Flood Susceptibility Mapping. Remote Sens. 14, 1656. https://doi.org/10.3390/rs14071656

Natural Resources Canada, 2018. Floods in Canada - Archive - Open Government Portal [WWW Document]. URL https://open.canada.ca/data/en/dataset/74144824-206e-4cea-9fb9-72925a128189 (accessed 8.13.25).

Natural Resources Canada, Strategic Policy and Innovation, Canada Centre for Earth Observation, Emergency Geomatics Services, 2021. Floods in Canada/ International Floods Product Specifications.

Saxe, S., Hogue, T.S., Hay, L., 2018. Characterization and evaluation of controls on post-fire streamflow response across western US watersheds. Hydrol. Earth Syst. Sci. 22, 1221–1237. https://doi.org/10.5194/hess-22-1221-2018

Snieder, E., Shakir, R., Khan, U.T., 2020. A comprehensive comparison of four input variable selection methods for artificial neural network flow forecasting models. J. Hydrol. 583, 124299. https://doi.org/10.1016/j.jhydrol.2019.124299