Spatiotemporal Flood Susceptibility Mapping using a Hybrid CNN-ConvLSTM Architecture

Karen E. Dunbar[1,2], Heather McGrath[2], Usman T. Khan[1]

[1] Lassonde School of Engineering, York University, Toronto, Canada – karela@yorku.ca, usman.khan@lassonde.yorku.ca
[2] Natural Resources Canada, Ottawa, ON, Canada – heather.mcgrath@nrcan-rncan.gc.ca

Keywords: Flood Susceptibility Mapping, Convolutional Neural Network, Long Short-Term Memory, Geospatial Analysis

Abstract

Flood susceptibility mapping (FSM) is a crucial component of flood management strategies; however, traditional statistical and machine learning methods for FSM are limited in their predictive capabilities. FSM approaches typically use static inputs, relying solely on geospatial factors, and typically overlook the spatiotemporal aspects (antecedent conditions) that trigger flood events. This study addresses this gap by developing a hybrid model that combines static geospatial features with dynamic temporal meteorological data, which is often excluded in FSM. The proposed hybrid model consists of two branches: (1) a 2D Convolutional Neural Network (CNN) to extract the features from geospatial inputs (i.e., slope and surficial geology) and (2) a Convolutional Long Short-Term Memory (ConvLSTM2D) network to learn the temporal antecedent conditions from Daymet precipitation, temperature, and snow-water equivalent. This model was trained and tested in the Saint John River basin, New Brunswick, Canada — a region that has experienced significant historical flooding. Three hyperparameters were investigated: temporal sequence length (1–4-month timesteps), resampling ratio (0.1-0.7), and positive class weight (1.5 or 2.0). The optimal model (F1 = 0.89) was achieved with a 3-month timestep, a 0.2 resampling ratio, and a 1.5 positive class weight, capturing the full snowmelt-to-rain spring cycle and outperforming models with 1-, 2-, or 4-month timesteps. The proposed 2D CNN-ConvLSTM2D architecture is effective in simultaneously learning the static geospatial features and temporal meteorological sequences, highlighting the importance of seasonal antecedent conditions in FSM.

1. Introduction

Worldwide, floods are the most destructive (Kundzewicz et al., 2014) and costly natural hazard, particularly in Canada, in terms of direct damage to infrastructure (Canada, 2022). Expanding urban development and rising temperatures due to climate change have made extreme weather events more frequent and intense, increasing flood susceptibility and worsening consequences (Badillo-Rivera et al., 2025; Tellman et al., 2021). This increased cascading hazard necessitates accurate spatial and flood mitigation tools such as flood susceptibility mapping (FSM) for planners, policymakers and emergency services (Mosavi et al., 2018).

FSM is traditionally performed using statistical models (such as the Frequency Ratio and Logistic Regression) and standard Machine Learning Algorithms (such as Random Forest (RF) and SVM) (Khalid and Khan, 2024). These methods are well-established and highly effective (Khosravi et al., 2019) at modelling the relationships between static geospatial inputs (such as slope and land cover) and flood probability (LeCun et al., 2015; Tehrany et al., 2014). However, traditional FSM approaches are static, treating all inputs as fixed in time. Therefore, antecedent conditions, which may be dry or wet (resulting from weeks of heavy rainfall), may be overlooked. The inability to capture temporal dynamics is a significant weakness, as floods are inherently time-dependent events.

Although traditional ML methods, such as RF, Extreme Gradient Boosting (XGBoost), and Artificial Neural Networks (ANNs), have achieved high performance, they overlook spatial relationships among geographic locations (Wang et al., 2025). Preprocessing techniques for these models often require spatial interpolation to account for dependencies; therefore, the resulting flood susceptibility models may exhibit poor generalizability (a flood or non-flood location may appear identical) (Wang et al., 2025). These limitations can be addressed by deep learning models, such as Convolutional Neural Networks (CNN), which can process raw spatial data and thereby capture feature dependencies (LeCun et al., 2015; Wang et al., 2025). CNN can process original raw data directly without requiring feature engineering (Wang et al., 2020; Zhao et al., 2020). CNNs extract features and automatically learn complex spatial relationships from grid-based raster data, making them highly appropriate for FSM (Wang et al., 2020). CNNs in FSM analyzes pixels and predicts the probability of a flood at that pixel. In the context of CNN learning, each pixel contains multiple bands, with each band as a flood conditioning factor (such as slope) (Wang et al., 2020). CNNs have recently been used in flood susceptibility modelling with the use of static inputs (DEM, slope, etc.), showing greater performance over traditional modelling (Wang et al., 2025, 2020; Zhao et al., 2020); however, this superiority is dependent on the CNN's architecture and hyperparameters, and they are still limited in handling temporal features.

Deep learning models that analyze temporal patterns, such as Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) networks, are widely used for modelling time-series data, as they can capture temporal dependencies. LSTM models have been utilized in flow forecasting, with inputs that include time-series data such as rainfall and temperature from gauge stations (Fang et al., 2021; Kratzert et al., 2019; Snieder and Khan, 2025). LSTM models excel at capturing temporal patterns; however, they lack the spatial awareness necessary for FSM.

Current FSM approaches often overlook integrating both the spatial and temporal aspects of flooding. This study tackles this gap by proposing a hybrid spatiotemporal deep learning model. Seasonal meteorological data are typically not included in flood susceptibility mapping (FSM) but have been shown to improve flood probability predictions (Dunbar et al., 2025; McGrath and

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-M-1-2026-1-2026

Gohl, 2022). The inclusion of seasonal meteorological data introduces a temporal aspect to the typically static FSM input dataset, capturing spatiotemporal dynamics of freshet flooding where the timing of snowmelt and spring saturation are critical determinants of flooding onset.

CNN-LSTM modelling has been used widely in spatiotemporal applications, including flood forecasting, traffic prediction, and landslide susceptibility (Fang et al., 2020; Khorram and Jehbez, 2023; Zhao et al., 2021). Since meteorological data are typically not used in FSM, CNN-LSTM have also yet to be applied to FSM. This project aims to utilize a hybrid CNN-ConvLSTM model for integrating meteorological data into flood susceptibility mapping. The 2D CNN branch will extract the spatial features from the static inputs, while the ConvLSTM2D branch will analyze a sequence of 2D rasters (meteorological inputs). This allows the model to learn the spatiotemporal patterns of meteorological data (Shi et al., 2015). When feature vectors from both branches are concatenated, the model can learn how the complex relationship between a specific sequence of weather and the particular landscape results in a flood.

The primary aim of this research is to develop and evaluate the dual-branch spatiotemporal FSM model. Specifically, a hybrid 2D CNN and ConvLSTM2D is implemented, incorporating both geospatial data (in the 2D CNN branch) and seasonal meteorological data (in the ConvLSTM2D branch) to ultimately observe seasonal trends in flood susceptibility. Additionally, the impact of three hyperparameters is investigated: the temporal sequence length, data class imbalance (by resampling), and positive class weighting (or cost-sensitive learning). This approach aims to address two challenges in flood prediction modelling: defining the optimal spatiotemporal window for the model and identifying the most effective strategy to overcome the inherent class imbalance in flood datasets. The model is evaluated using four metrics (AUC, F1 Score, Precision, and Recall), with 95% confidence intervals from bootstrap analysis.

This research focuses on the St. John's River in New Brunswick, Canada. The paper is organized as follows: Section 2 describes the study area, data preparation, model architecture, and experimental design. Section 3 presents the results of the hyperparameter tuning and the performance of the final model. Finally, Section 4 discusses the research limitations and outlines future research directions.

2. Methods

2.1 Study Area

The study area comprises the lower Saint John River basin in New Brunswick, Canada, with the nine Canadian National Hydro Network (NHN) work units depicted in Figure 1 covering approximately 28,446 km^2. This area has experienced significant historical flooding with contributors from heavy rainfall and rapid spring snowmelt.

2.2 Data Collection

The data collected, along with their sources, that are used in this research are shown in Table 1. Meteorological inputs were obtained from the NASA Daymet portal, which consists of daily minimum and maximum temperatures, total precipitation, snow water equivalent, and vapour pressure. The static geospatial inputs were derived from the Canadian Medium Resolution Digital Elevation Model (MRDEM). The spatial reference parameters are shown in Table 2.

The NAD 1983 UTM Zone 19N is the standard for the highest accuracy and lowest projection distortion for the study area (Canada, 2009). Since the original DEM has a 30-metre resolution, the derived features from the DEM were also assigned a 30-m resolution, resulting in dimensions of 9,533 x 8956 pixels, from which 32x32 patches were extracted.

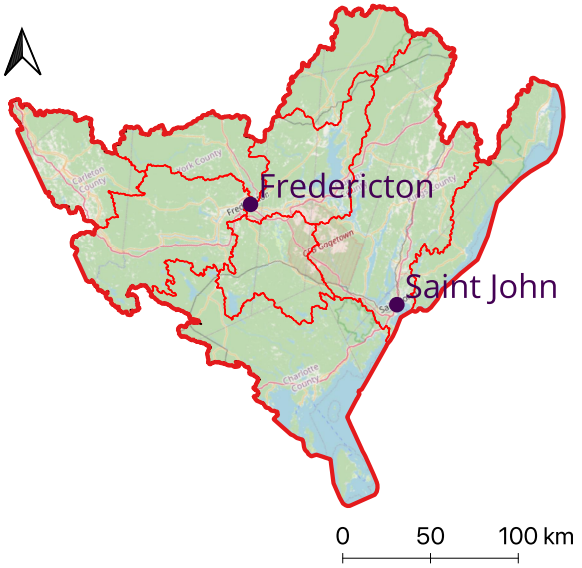

Figure 1: Study area showing nine of the Canadian NHN work units/ basins clipped by the New Brunswick provincial boundary.

Features with lower resolution, such as NDVI and Daymet meteorological data, were resampled to 30 m. Land use/land cover and surficial geology vectors were converted to raster format and then resampled. This step was critical for preparing the data for the CNN modelling since the fully connected dense layer in the model architecture requires a flattened 1D vector and thus a uniform input dimension. The model architecture is explained further in Section 2.4.

Data	Source
Min/ Max Temperature, Total Precipitation, Mean Vapour Pressure, Snow Water Equivalent (SWE)	NASA EarthData portal (Earth Science Data Systems, 2025; Thornton et al., 2021)
DEM, HAND, Euclidean Distance to Rivers (EUC), Slope, Aspect, Topographic Position Index (TPI), Terrain Ruggedness Index (TRI)	MRDEM (Canada, 2025a) Derived from MRDEM using Whitebox Tools (Lindsay, 2016).
Land Use / Land Cover 2020 (LULC)	North American Land Change Monitoring System (Pasos, 2024)
Surficial Geology	Government of Canada Surficial Geology Map (N. R. Canada, 2024)
Normalized Difference Vegetation Index (NDVI)	Government of Canada NDVI (Canada, 2024)
Flood & Non-Flood Inventory Data	Government of Canada Flood Archive (Canada, 2025b)

Table 1: List of data used and their sources

2.3 Input Data

The input used includes both static and dynamic inputs. Static inputs are the properties of the study area that are assumed to remain relatively consistent over time. Dynamic inputs change over time. The combination of static and dynamic inputs necessitates the use of a model that understands the static patterns and the changing inputs.

Spatial Reference Parameter	Value
Datum	NAD 1983 CSRS
Projected Coordinate System	NAD 1983 / UTM Zone 19N (EPSG: 2960)
Cell Size	30 m*

Table 2: Spatial reference parameters used. *NDVI and Daymet meteorological 1 km resolution data were reprojected and resampled using a bilinear technique to create a 30 m grid to match the other raster data

2.3.1 Static Inputs consist of continuous and categorical inputs that are often used in flood susceptibility mapping (Khalid and Khan, 2024). Static continuous inputs are shown in Figure 2.

The categorical inputs, land use/ land cover and surficial geology, were one-hot encoded before model training. For areas that were missing categorical data (surficial geology), these areas were assigned a value of 0.0, indicating that they did not belong to any class.

Aspect, slope, EUC, HAND, TPI, and TRI were derived from the MRDEM using *Whitebox Tools*. Aspect describes the compass direction that the slope is facing, determining the exposure to hydro-geological processes (solar radiation, wind). Slope represents the steepness of the land surface, which influences the velocity and potential for surface runoff. The Euclidean distance to rivers (EUC) is the shortest possible path between a location and the river. Fluvial floods will inevitably inundate areas closer to the river first; therefore, proximity to the river is typically associated with higher flood susceptibility. The height above nearest drainage (HAND) describes the DEM in terms of the drainage network. Finally, the topographic position index (TPI) and terrain ruggedness index (TRI) were also derived to quantify the texture and shape of the landscape. TPI values range from -1 to 1, where negative TPI values indicate areas that are lower on average compared to their surroundings (i.e., stream channels and valleys). Positive TPI values are areas that are higher on average than their surroundings, which is typically associated with lower flood susceptibility. The Terrain Ruggedness Index (TRI) describes the roughness of the terrain, quantifying the elevation change at a given location. Low TRI values represent flat areas, while high TRI values are areas that have many elevation changes. The NDVI data are available for Julian weeks 15 to 41 (growing season) per year.

2.3.2 Dynamic Inputs; Meteorological inputs, shown in Figure 3, are not typically included in flood susceptibility mapping. Few studies have included precipitation in FSM (Khalid and Khan, 2024; McGrath and Gohl, 2022). For areas with missing data, *rasterio* inverse distance weighting was used to interpolate missing values. Meteorological data from the a sequence of four 30-day (monthly) timesteps antecedent to each flood event were used and aggregated into 30-day (monthly) intervals. NDVI data from a four-week window immediately preceding each flood event was used.

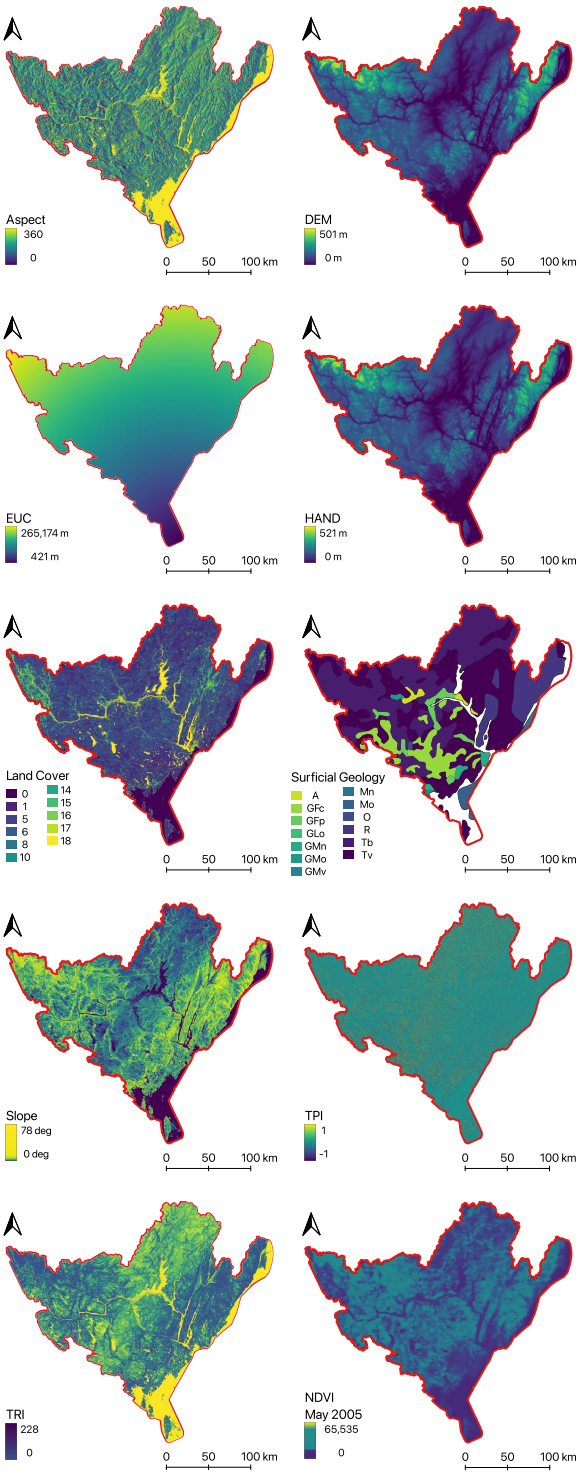

Figure 2: Static geospatial inputs (30-m resolution): aspect, DEM, EUC, HAND, land cover, surficial geology, slope, TPI, TRI, and NDVI (May 2005)

Flood Inventory; The flood inventory was obtained from the Canadian Flood Archive, where flood extent polygons were generated by Natural Resources Canada (NRCan) in response to major flood events, utilizing satellite imagery for emergency response (Canada, 2025c). The flood extent polygons were rasterized, creating binary target maps for model training and validation. Floods from the St John in New Brunswick were included from 2005 to 2019. Flood events that occurred over

multiple days were merged to obtain the maximum flood event, resulting in 15 distinct flood events. All flood events occurred in either April or May. The May 2008 flood event is shown in Figure 4.

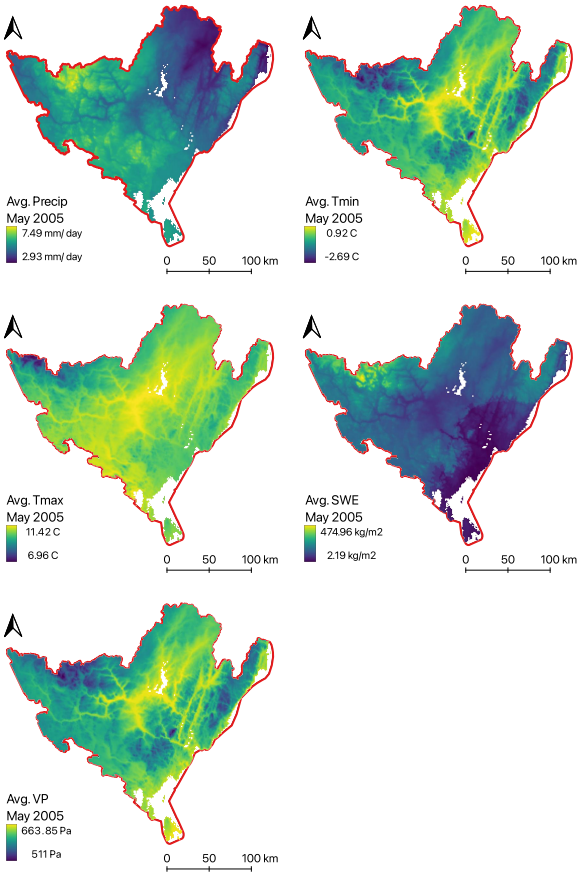

Figure 3: Subset of Daymet precipitation, minimum/maximum temperature, snow water equivalent and vapour pressure. All Daymet data were resampled to 30-m resolution.

2.4 Model Development

The model architecture is shown in Figure 5, which was adapted from Khorram & Jehbez (2023). The CNN-LSTM model was created using the Keras/ TensorFlow package and Python version 3.11.13. The geospatial and meteorological data were handled separately in the model's dual branches.

The CNN branch (left) handled the geospatial inputs, including NDVI. Although processed within the static branch as a single temporal snapshot per sample, the NDVI data were matched to the specific flood year to account for the annual variations in vegetation density. The geospatial raster inputs were processed through the Conv2D layer, resulting in an input tensor of (32, 32, *N*); this corresponds to the 32x32 spatial patch dimensions and the geospatial feature channels that includes the topographic inputs, one-hot encoded surficial geology and land cover categories, and NDVI. A 3x3 kernel was applied to each 32x32 patch, ultimately creating 32 feature maps. The dimensions of the 32 feature maps are reduced by max pooling while retaining the most important information. A 25% dropout was applied, temporarily deactivating 25% of the cells at random within each feature map to prevent the model from relying too heavily on a single feature. This was followed by a second Conv2D layer, which used a 3x3 kernel to detect more complex patterns, producing 64 feature maps. Max pooling and a 25% dropout were used again before flattening the feature maps to a single 1D spatial feature vector. Because the static branch represents one temporal snapshot, the geospatial inputs are processed through the Conv2D layers only once per patch. The static branch is then concatenated with the spatiotemporal vector generated by the ConvLSTM2D branch.

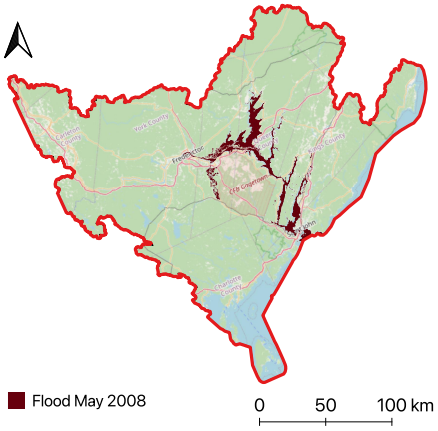

Figure 4: May 2008 flood event

The ConvLSTM branch (right) processes the daily Daymet meteorological data, which consists of five variables: precipitation, maximum temperature, minimum temperature, snow water equivalent, and vapour pressure. To capture seasonal trends, the daily Daymet values were aggregated into antecedent 30-day (monthly) values, where precipitation was calculated as a total sum, while the remaining four variables were averaged. For each iteration, a total of 44,800 patches were extracted across the 15 flood events (with a 70/15/15 training/validation/testing split).

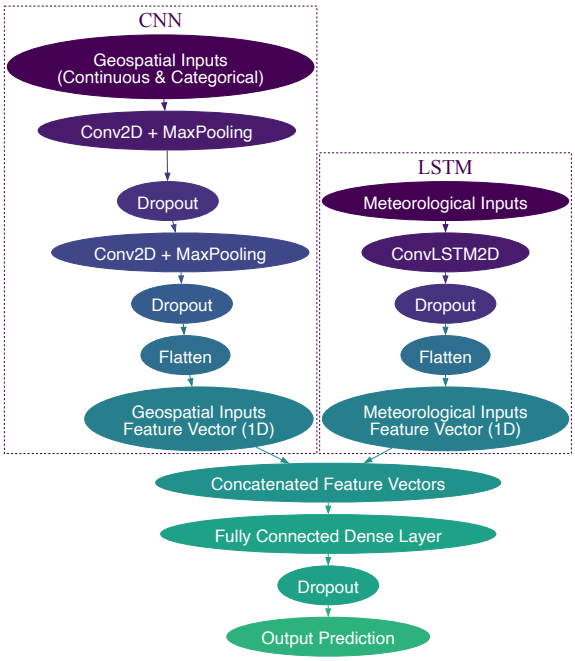

Figure 5: CNN-ConvLSTM2D model architecture

Four timesteps (*t-1*, *t-2, t-3, and t-4*) were tested for the LSTM, resulting in an input tensor of (*t*, 32, 32, 5) per sample; this represents the number of time steps, the 32x32 spatial patch dimensions, and the five meteorological features. The meteorological inputs were processed by the ConvLSTM2D layer, which generated 32 feature maps. A 50% dropout was used before flattening into a 1D feature vector. Following the independent feature extraction in both branches, the geospatial vector from the CNN and the meteorological 1D feature vector from the ConvLSTM were concatenated in a late fusion stage. This integrated feature set was then fed into the fully connected dense layer. A final 50% dropout rate was used before predicting flood susceptibility between 0 and 1 using the sigmoid activation function.

Two parameters (positive class weighting and balance ratio) were investigated to handle the class imbalance inherent in flood data. Given that the flood class is in a significant minority compared to the non-flood class, flood prediction models will be biased towards the non-flood class, resulting in lower performance on rare but critical flood events (Diallo et al., 2025; Khosravi et al., 2019). To overcome this bias, this study investigates and optimizes two strategies for addressing class imbalance: data resampling (using a balance ratio) and cost-sensitive learning (utilizing a positive class weight).

The balance ratio parameter controls the percentage of flooded areas (or data composition) that the model sees during training. A balance ratio of 0.10 represents a training batch that contains 10% flooded patches and 90% non-flooded patches. Balance ratios between 0.10 to 0.60 were investigated. The positive class weighting parameter controls the degree to which the model is penalized for incorrect predictions during training. A class weight of 2.0 means that the loss from a mistake on a flood patch is multiplied by 2.0. Positive class weights of 1.5 and 2.0 were investigated. This parameter indicates that missing flood patches are more critical than missing non-flood patches.

Parameter	Value
Package	Keras/ TensorFlow 2.0
Training/ Validation /Test split	70/15/15 (sequential)
Learning rate	1e-4
Optimizer	Adam
Patch size	32 x 32
Batch size	16 (the number of patches model sees before updating weights)
Epochs	100
Steps per epoch	500
Early stopping	5 (based on validation auc_roc)
Activation function	ReLu (Conv2D and ConvLSTM2D) Sigmoid (output)
Flood threshold	Optimal threshold found from the highest F1 score
Positive class weighting	1.5, 2.0
Balance ratios	0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70
Temporal depths (months)	1, 2, 3, 4

Table 3: Model architecture and hyperparameters used for the hybrid CNN-ConvLSTM.

A third hyperparameter investigated is the temporal window, for the ConvLSTM2D branch, which defines the number of antecedent 30-day averages of the meteorological rasters (precipitation, minimum and maximum temperature, snow-water equivalent, and vapour pressure). For instance, a flood event in April, ts = 3 would create input sequences 30 days before March, February and January, starting from the day before the flood event. The other model parameters are outlined in Table 3.

2.5 Model Evaluation

To evaluate the model's classification capabilities, the following metrics were calculated: F1 score, recall, precision, and AUC-ROC. Recall describes the model's ability to find a flood. A high recall is a model that has few false negatives (missed floods). Precision describes the model's reliability. A model with high precision has a low rate of false positives. There is often a trade-off between recall and precision, which the F1 score can describe. A high F1 score results from having a high recall and precision. AUC-ROC is the area under the receiver operating characteristic curve and describes the model's ability to differentiate between a flood and a dry area. An AUC of 1 is a perfect model, while an AUC of 0.5 is a model that does not perform better than random guessing. To obtain the highest F1 score, the optimal flood threshold was found. The final metrics were then calculated using this optimal threshold. Additionally, bootstrap analysis was performed during testing to obtain 95% confidence intervals for each metric.

3. Results

The performance results across three parameters—timestep (ts), balance ratio (br), and positive class weight (cw)—are displayed in Figure 6.

The F1 score is the harmonic mean of recall and precision, and yields high scores when a high number of true positives is achieved (correct prediction of the floods), making it particularly suitable for imbalanced datasets (Diallo et al., 2025). The highest performing model belongs to the group that had the previous three months of meteorological data (t = 3). Models with timesteps = 1 and 2 perform well, with their peak at 0.88. Four timesteps show a clear decline in model performance. The single best-performing model is the t = 3, bw = 0.2, and cw = 1.5 model, with an optimal F1 score range of 0.87 to 0.91. The lowest performance is achieved with a balance ratio of 0.1, where the model is trained on a dataset containing only 10% flood patches. No clear pattern is observed across the three parameters for the F1 score.

The recall and precision results highlight the trade-off between the two metrics. Like the F1 score, no clear pattern is observable across the three parameters. In the t–4 models, a class weight of 2.0 resulted in higher performance; however, this is not the case for the other timesteps. The performance neither decreases nor increases with undersampling or oversampling. In the t = 3, bw = 0.2, and cw = 1.5 model, the F1 score was the highest due to the significant gain in precision.

The AUC-ROC decreases slightly but consistently as the balance ratio increases (except for t = 2, br = 0.1, which had the highest AUC-ROC). Across all three parameters, all models exhibited high AUC-ROC values (ranging from above 0.95 to as high as 0.99) and demonstrated a strong ability to distinguish between flood and non-flood patches. Unlike the F1 score, where the performance peaked at t = 3, the highest AUC is with t = 1, however, the differences between the timesteps are slight. For AUC-ROC, the performances across the class weighting are

generally in agreement, suggesting the models were not sensitive to the class weights in terms of AUC.

Figure 7 compares the flood susceptibility output (top) with the observed flood extent from April 2019 (bottom). The observed flood extent for April 2019 clearly follows the St. John River along the main river body and its many tributaries. The predicted flood susceptibility map displays the range of susceptibility values between 0 and 1, generated by the best-performing model with parameters ts = 3, cw = 1.5, and br = 0.2 (F1 score = 0.89, recall = 0.86, precision = 0.92, AUC = 0.99). The optimal threshold for this model is 0.36, where a score of 0.36 or higher predicts a flood. A score of 0.36 or lower indicates a low probability of flooding and is depicted in white.

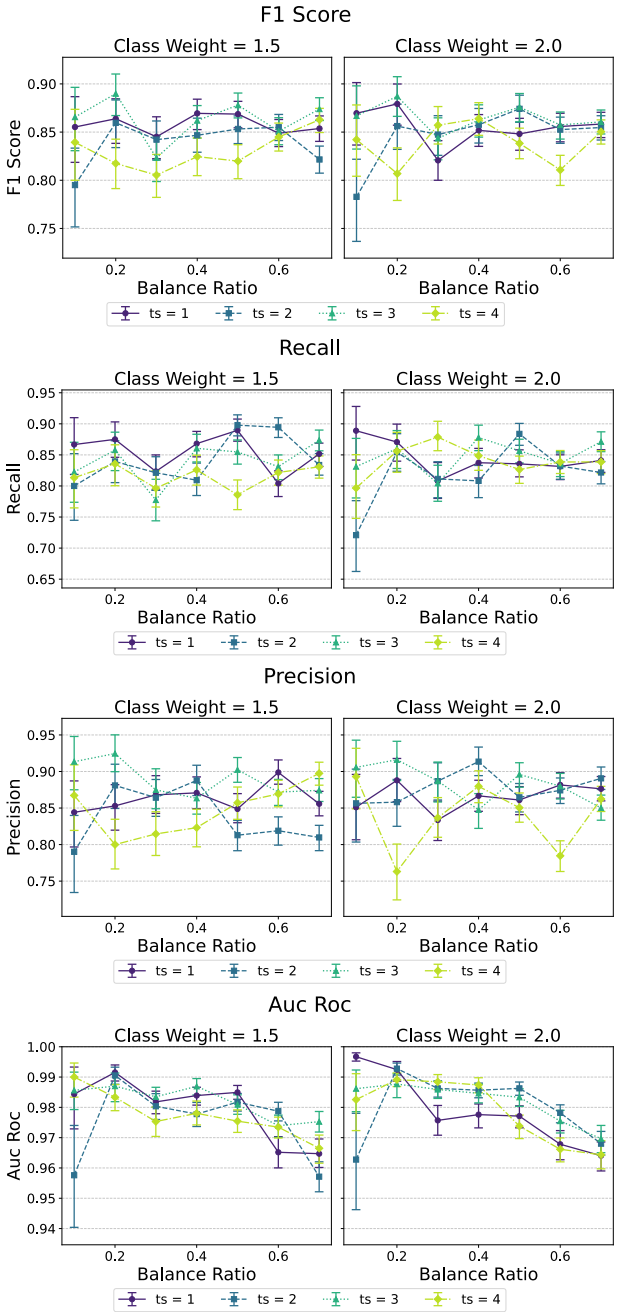

Figure 6: Model performance using the optimal threshold with 95% confidence intervals by metrics: F1, Recall, Precision, and AUC ROC.

The highly susceptible areas align almost exactly with the observed flood extent, showing the model has successfully learned the appropriate relationship between floods and their triggers for this study area. The full extent of the flood was slightly underpredicted (as can be seen visually and inferred from recall = 0.86). The missed regions, specifically in the southwestern part of the study area, include permanent water bodies, such as Oromocto Lake. The observed flood extent, generated from emergency-response satellite imagery, may not differentiate between permanent water bodies and temporary floodwaters. The trained model assigned these under-predicted areas as having low flood susceptibility, demonstrating its ability to distinguish between floodplains and existing water bodies.

4. Discussion

To generate a continuous flood susceptibility map for visual evaluation (Figure 7), a sliding window strategy was used. The best performing model predicts a single susceptibility value for each 32x32 patch, which is assigned to the patch's centre pixel. As the sliding window moves across the study area pixel by pixel, the final flood susceptibility map retains the original 30 m spatial resolution. The highest performance (F1 Score = 0.89) was achieved with a model using a 3-month timestep, a 0.2 balance ratio and a 1.5 positive class weight. This finding indicates that model performance is dependent not only on architecture but also on temporal depth and the method to handle class imbalance (resampling and cost-sensitive learning).

The F1 scores peaked at timesteps = 3, with lower performance at ts = 1, 2, and 4. The previous month or two (ts = 1 and 2) were likely insufficient in predicting floods. For the April Spring flood in New Brunswick, the model performed best when trained on the preceding three months, specifically on the snowpack accumulation (swe) in January and February, before the spring snowmelt (tmax) and rainfall (prcp) in March. The addition of the preceding fourth month may have introduced noise to the model; meteorological data from December may not be as important to an April flood. The higher performance in the 3-month timestep confirms the importance of seasonal consideration for flood susceptibility mapping in this region. This result was expected due to the typical hydrological timeline of a spring freshet, where snow accumulates in winter, followed by melting as temperatures rise and precipitation events occur in spring.

The two options for positive class weighting (penalty for missed floods) resulted in roughly similar precision and recall. The difference in weight may not have been large enough to cause significant differences in model performance. The F1 score results show that a resampling ratio of 0.1 is insufficient. An increase in performance is observed with cw = 0.2, and interestingly, it drops again at cw = 0.3. Higher balance ratios (oversampling) are not necessarily better either. This is particularly the case for AUC, where AUC values drop as resampling ratios increase. There appears to be a sweet spot at a balance ratio of 0.2, where a data composition with fewer flood instances (which represents reality) is more beneficial for model learning, which is advantageous for an imbalanced problem such as flood mapping.

Overall, the high F1 scores and AUC from the group of models show that the hybrid dual-branch model is effective in predicting floods. The 2D CNN branch successfully extracted the static geospatial relationships where water can flow, while the ConvLSTM2D branch successfully extracted the spatiotemporal aspects, learning when the water is likely to flow. Even with a low balance ratio (fewer instances of flood patches), the model

https://doi.org/10.5194/isprs-annals-XI-M-1-2026-1-2026

was able to predict the correct location of the flood as seen in Figure 7. The high performance of the hybrid architecture highlights the necessity of both branches. While a formal ablation study was not conducted, a previous 1D feature selection study has demonstrated that meteorological drivers alone are insufficient for accurate flood susceptibility modelling without the geospatial inputs.

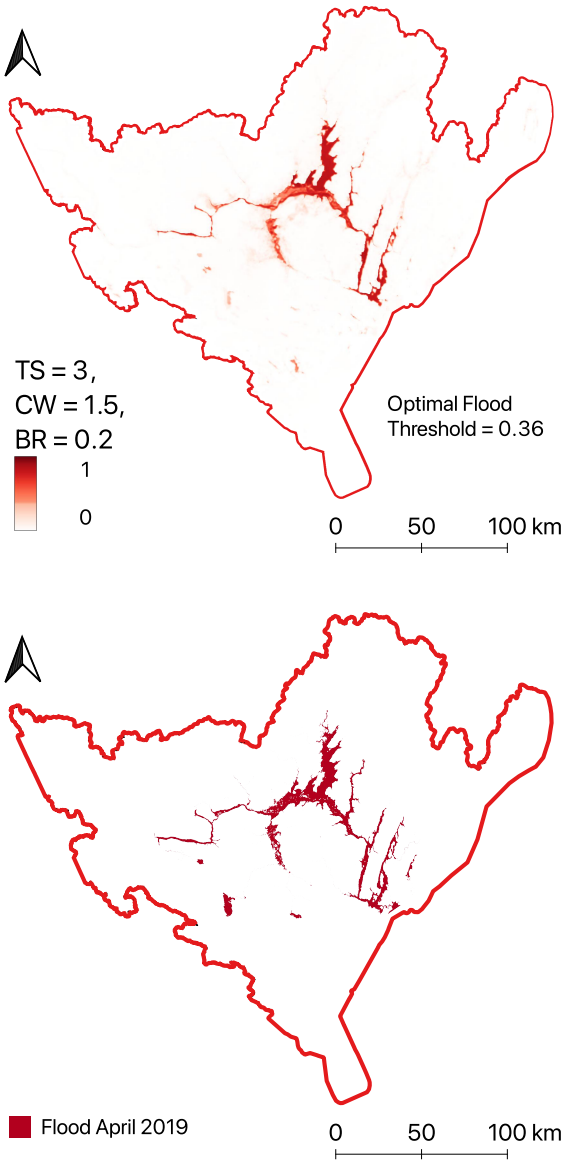

Figure 7: Flood susceptibility map (top) for the April 2019 versus the actual flooded areas (bottom)

5. Conclusion

A hybrid 2D CNN-ConvLSTM2D model was developed for flood susceptibility modelling, showing that the combined use of static geospatial features with dynamic meteorological data is an effective approach for flood susceptibility mapping. The model performance is sensitive to spatiotemporal hyperparameters. The optimal model was obtained using a 3-month temporal window, a 0.2 resampling ratio, and a 1.5 positive class weight, resulting in an F1 score of 0.89. The spatiotemporal approach captures the time-dependent relationship between antecedent conditions and the physical landscape. This approach moves beyond traditional static flood mapping, contributing to more accurate and dynamic flood management.

5.1 Limitations

This research contains multiple limitations. The model combines low-resolution temporal data (1 km) with higher-resolution static data (30 m). The lower-resolution data were resampled to match the highest-resolution data, which is required for CNN modelling. The hybrid CNN-ConvLSTM2D is a complex model; the performance gain should be evaluated to determine whether it is significant enough relative to a simpler model, such as Random Forest, to justify the substantial preprocessing and training time. This study trained and tested the models at a single location, which is subject to a single hydrological narrative characterized by spring snowmelt floods. The models in this study may not perform as well in regions dominated by different hydrological mechanisms.

5.2 Future Work

Future work will include testing the model's spatial generalizability and evaluating it in a different location. The temporal generalizability will also be tested. All the flood events in this study occurred in the spring. The same model may not accurately predict a flood in the autumn due to different hydrological conditions. Previous work trained a standard XGBoost model with both static geospatial and temporal meteorological inputs. The performance of this XGBoost model will be compared with that of the hybrid CNN-ConvLSTM2D model to assess whether the performance gain is sufficient to warrant using a more complex model. Other model architectures, such as a fully connected convolutional network (U-Net) and Transformers, also show high potential for spatiotemporal FSM and should be explored. Additionally, the incorporation of climate change impacts will be analyzed by using climate projections to create future flood susceptibility projections.

6. References

Badillo-Rivera, E., Santiago, R., Poma, I., Chavez, T., Arroyo-Paz, A., Aucahuasi-Almidon, A., Hinostroza, E., Segura, E., Eyzaguirre, L., León, H., Virú-Vásquez, P., 2025. Flood susceptibility mapping in El Niño Phenomenom integrating multitemporal radar analysis, GIS and machine learning techniques, Piura river basin, Peru. Front. Environ. Sci. 13, 1672107. https://doi.org/10.3389/fenvs.2025.1672107

Canada, 2024. Corrected representation of the NDVI using historical AVHRR and VIIRS satellite images (1 km resolution) from 1987 to present - Open Government Portal [WWW Document]. URL https://open.canada.ca/data/en/dataset/44ced2fa-afcc-47bd-b46e-8596a25e446e (accessed 9.22.25).

Canada, N.R., 2025a. Medium Resolution Digital Elevation Model (MRDEM) - CanElevation Series - Open Government Portal [WWW Document]. URL https://open.canada.ca/data/en/dataset/18752265-bda3-498c-a4ba-9dfe68cb98da (accessed 9.22.25).

Canada, N.R., 2025b. 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).

Canada, N.R., 2025c. Floods in Canada – Archive. GEO.CA. URL https://geo.ca/emergency/floods-in-canada-archive/ (accessed 11.14.25).

Canada, N.R., 2024. Surficial geology map of Canada, 1:5 000 000 - Open Government Portal [WWW Document]. URL https://open.canada.ca/data/en/dataset/317ebf06-e053-b10f-6092-bfe0c1af8703 (accessed 9.22.25).

Canada, N.R., 2022. Flood mapping [WWW Document]. URL https://natural-resources.canada.ca/science-data/science-research/flood-mapping (accessed 5.6.25).

Canada, N.R., 2009. About the Canadian Spatial Reference System [WWW Document]. URL https://natural-resources.canada.ca/science-data/science-research/geomatics/geodetic-reference-systems/canadian-spatial-reference-system-csrs (accessed 11.14.25).

Diallo, R., Edalo, C., Awe, O.O., 2025. Machine Learning Evaluation of Imbalanced Health Data: A Comparative Analysis of Balanced Accuracy, MCC, and F1 Score, in: Awe, O.O., A. Vance, E. (Eds.), Practical Statistical Learning and Data Science Methods: Case Studies from LISA 2020 Global Network, USA. Springer Nature Switzerland, Cham, pp. 283–312. https://doi.org/10.1007/978-3-031-72215-8_12

Dunbar, K.E., McGrath, H., Khan, U.T., 2025. Enhancing flood susceptibility modelling in Canada: Integrating seasonal meteorological data, feature selection and machine learning approaches. https://doi.org/10.5194/egusphere-egu25-3871

Earth Science Data Systems, N., 2025. Daymet: Monthly Climate Summaries on a 1-km Grid for North America, Version 4 R1 | NASA Earthdata.

Fang, Z., Wang, Y., Peng, L., Hong, H., 2021. Predicting flood susceptibility using LSTM neural networks. J. Hydrol. 594, 125734. https://doi.org/10.1016/j.jhydrol.2020.125734

Fang, Z., Wang, Y., Peng, L., Hong, H., 2020. Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping. Comput. Geosci. 139, 104470. https://doi.org/10.1016/j.cageo.2020.104470

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

Khorram, S., Jehbez, N., 2023. A Hybrid CNN-LSTM Approach for Monthly Reservoir Inflow Forecasting. Water Resour. Manag. 37, 4097–4121. https://doi.org/10.1007/s11269-023-03541-w

Khosravi, K., Shahabi, H., Pham, B.T., Adamowski, J., Shirzadi, A., Pradhan, B., Dou, J., Ly, H.-B., Gróf, G., Ho, H.L., Hong, H., Chapi, K., Prakash, I., 2019. A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods. J. Hydrol. 573, 311–323. https://doi.org/10.1016/j.jhydrol.2019.03.073

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., Nearing, G., 2019. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. Hydrol. Earth Syst. Sci. 23, 5089–5110. https://doi.org/10.5194/hess-23-5089-2019

Kundzewicz, Z.W., Kanae, S., Seneviratne, S.I., Handmer, J., Nicholls, N., Peduzzi, P., Mechler, R., Bouwer, L.M., Arnell, N., Mach, K., Muir-Wood, R., Brakenridge, G.R., Kron, W., Benito, G., Honda, Y., Takahashi, K., Sherstyukov, B., 2014. Flood risk and climate change: global and regional perspectives. Hydrol. Sci. J. 59, 1–28. https://doi.org/10.1080/02626667.2013.857411

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539

Lindsay, J.B., 2016. Whitebox GAT: A case study in geomorphometric analysis. Comput. Geosci. 95, 75–84. https://doi.org/10.1016/j.cageo.2016.07.003

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

Mosavi, A., Ozturk, P., Chau, K., 2018. Flood Prediction Using Machine Learning Models: Literature Review. Water 10, 1536. https://doi.org/10.3390/w10111536

Pasos, M., 2024. North American Land Change Monitoring System. Comm. Environ. Coop. URL https://www.cec.org/north-american-land-change-monitoring-system/ (accessed 9.22.25).

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W., WOO, W., 2015. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting, in: Advances in Neural Information Processing Systems. Curran Associates, Inc.

Snieder, E., Khan, U.T., 2025. A diversity-centric strategy for the selection of spatio-temporal training data for LSTM-based streamflow forecasting. Hydrol. Earth Syst. Sci. 29, 785–798. https://doi.org/10.5194/hess-29-785-2025

Tehrany, M.S., Pradhan, B., Jebur, M.N., 2014. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. J. Hydrol. 512, 332–343. https://doi.org/10.1016/j.jhydrol.2014.03.008

Tellman, B., Sullivan, J.A., Kuhn, C., Kettner, A.J., Doyle, C.S., Brakenridge, G.R., Erickson, T.A., Slayback, D.A., 2021. Satellite imaging reveals increased proportion of population exposed to floods. Nature 596, 80–86. https://doi.org/10.1038/s41586-021-03695-w

Thornton, P.E., Shrestha, R., Thornton, M., Kao, S.-C., Wei, Y., Wilson, B.E., 2021. Gridded daily weather data for North America with comprehensive uncertainty quantification. Sci. Data 8, 190. https://doi.org/10.1038/s41597-021-00973-0

Wang, J., Sanderson, J., Iqbal, S., Woo, W.L., 2025. Accelerated and Interpretable Flood Susceptibility Mapping Through Explainable Deep Learning with Hydrological Prior Knowledge. Remote Sens. 17, 1540. https://doi.org/10.3390/rs17091540

Wang, Y., Fang, Z., Hong, H., Peng, L., 2020. Flood susceptibility mapping using convolutional neural network frameworks. J. Hydrol. 582, 124482. https://doi.org/10.1016/j.jhydrol.2019.124482

Zhao, G., Pang, B., Xu, Z., Peng, D., Zuo, D., 2020. Urban flood susceptibility assessment based on convolutional neural networks. J. Hydrol. 590, 125235. https://doi.org/10.1016/j.jhydrol.2020.125235

Zhao, Z., Li, Z., Li, F., Liu, Y., 2021. CNN-LSTM Based Traffic Prediction Using Spatial-temporal Features. J. Phys. Conf. Ser. 2037, 012065. https://doi.org/10.1088/1742-6596/2037/1/012065