**More on Model Performance/Validation**

A blue and white chart

Description automatically generated

**Figure S1.** Confusion Matrix for FNN showing True Positives (TP) (bottom-right cell), True Negatives (TN)(top-left cell), False Positives (FP) (top-right cell) and False Negatives (FN) (bottom-left cell). High True Negatives (TN): The model is very accurate at predicting when there is "No Flood," with 2078 correct predictions.Very Few False Positives: Only 6 instances where the model predicted "Flood" incorrectly, indicating a very low false alarm rate. There are almost no False Negatives i.e. no instances of predicting “No Flood” when there is a “Flood”.

A graph showing a number of blue squares

Description automatically generated with medium confidence

**Figure S2.** Confusion Matrix for CNN showing True Positives (TP) (bottom-right cell), True Negatives (TN) (top-left cell), False Positives (FP) (top-right cell) and False Negatives (FN) (bottom-left cell). High True Negatives (TN): The model is very accurate at predicting when there is "No Flood," with 2081 correct predictions.Very Few False Positives: Only 3 instances where the model predicted "Flood" incorrectly, indicating a very low false alarm rate. However, there are a few False Negatives i.e. Instances showing a prediction of “No Flood” when there is a “Flood”.

A blue and white chart

Description automatically generated

**Figure S3.** Confusion Matrix for LSTM showing True Positives (TP) (bottom-right cell), True Negatives (TN) (top-left cell), False Positives (FP) (top-right cell) and False Negatives (FN) (bottom-left cell). High True Negatives (TN): The model is very accurate at predicting when there is "No Flood," with 2084 correct predictions.No False Positives: 0 instances where the model predicted "Flood" incorrectly, indicating a very low false alarm rate. However, there are almost no False Negatives i.e. Instances predicting “No Flood” when there is a “Flood”.

A graph of a graph

Description automatically generated with medium confidence

**Figure S4.** Precision: The ratio of true positive predictions to the total predicted positives (i.e., precision = TP / (TP + FP)).Recall: The ratio of true positive predictions to the total actual positives (i.e., recall = TP / (TP + FN)).High Precision and Low Recall (Left Part of the Curve): The curve starts with very high precision close to 1 and low recall. This indicates that when the FNN model predicts a flood event, it is very accurate, but it does not predict many of them. Consistency in Precision: For a significant portion of the curve, the precision remains very high as recall increases. This means that the model maintains a high level of accuracy in its positive predictions even as it captures more true positive cases. Sharp Drop in Precision at High Recall: Near the end of the curve, where recall is very high (close to 1), there is a noticeable drop in precision. This indicates that as the model tries to capture almost all flood events (high recall), it starts to include more false positives, thus reducing precision. Model's High Precision: For a large portion of the recall range, the model maintains a precision close to 1. This indicates that when the model predicts a flood, it is almost always correct. Model's Ability to Balance Precision and Recall: The FNN model shows a good balance between precision and recall, maintaining high precision until it reaches the very high recall region. Potential Overfitting or Imbalance: The sharp drop in precision at very high recall suggests that there may be an imbalance in the dataset or the model is becoming too lenient in its predictions to capture all flood events. The precision-recall curve for the FNN model indicates that it performs well in terms of precision across a broad range of recall values but struggles with precision at the highest recall values. This behavior aligns with a model that is generally conservative in predicting flood events, ensuring high accuracy but potentially missing some events (low initial recall).

A graph of a graph

Description automatically generated

**Figure S5.** High Precision and Low Recall (Left Part of the Curve): The curve starts with very high precision (close to 1) and low recall. This indicates that when the CNN model predicts a flood event, it is very accurate, but it does not predict many of them. Fluctuations in Precision: The curve shows significant fluctuations in precision at the low recall end. This suggests instability or variability in the model’s predictions when it starts to detect more flood events. Decreasing Precision with Increasing Recall: As recall increases, precision generally decreases. This means that as the model becomes more lenient in predicting flood events (increasing recall), it also starts to include more false positives, thus reducing precision. Low Precision at High Recall: At the right end of the curve, where recall is high, precision drops significantly, indicating many false positives. The model sacrifices accuracy in its flood predictions to capture more true flood events. Model’s Strengths and Weaknesses: The CNN model is very precise at identifying flood events initially but struggles to maintain this precision as it attempts to recall more flood events. Trade-off Between Precision and Recall: There is a clear trade-off between precision and recall, where improving recall leads to a decrease in precision. Fluctuations Indicate Instability: The fluctuations in precision, especially at low recall, suggest that the model's performance is not stable. This might be due to class imbalance, where the number of flood events is much smaller compared to no flood events.

A graph of a graph

Description automatically generated with medium confidence

**Figure S6.** High Precision and Low Recall (Left Part of the Curve): The curve starts with high precision close to 1 and low recall. This indicates that when the LSTM model predicts a flood event, it is very accurate, but it does not predict many of them. Consistency in Precision: For a significant portion of the curve, the precision remains very high as recall increases. This means that the model maintains a high level of accuracy in its positive predictions even as it captures more true positive cases. Sharp Drop in Precision at High Recall: Near the end of the curve, where recall is very high (close to 1), there is a noticeable drop in precision. This indicates that as the model tries to capture almost all flood events (high recall), it starts to include more false positives, thus reducing precision. Model's High Precision: For a large portion of the recall range, the model maintains a precision close to 1. This indicates that when the model predicts a flood, it is almost always correct. Model's Ability to Balance Precision and Recall: The LSTM model shows a good balance between precision and recall, maintaining high precision until it reaches the very high recall region. Potential Overfitting or Imbalance: The sharp drop in precision at very high recall suggests that there may be an imbalance in the dataset, or the model is becoming too lenient in its predictions to capture all flood events. The precision-recall curve for the LSTM model indicates that it performs well in terms of precision across a broad range of recall values but struggles with precision at the highest recall values. This behavior aligns with a model that is generally conservative in predicting flood events, ensuring high accuracy but potentially missing some events (low initial recall).