Imperial College London

SnooCODE

Flood Probability Prediction Based on Machine Learning

Speaker | Chaofan Wu

September 2023

OUTLINE

BACKGROUND

1. Flood Risks in Ghana

1. Annual Seasonal Floods: Ghana experiences almost annual seasonal floods, particularly during the rainy season from May to July [1, 2, 3]. These floods pose severe risks to human lives, property, and infrastructure [4, 5].

2. Underlying Causes: The primary causes of flooding in Ghana are threefold:

- **Climatic Determinants**: Large-scale rainfall, especially during the wet season [1, 2, 3].
- **Geographical Elements**: Features like low-lying terrains and the clayey composition of soil contribute to flooding [1, 2, 3].
- **Anthropogenic Influences**: Human-induced factors such as inadequate drainage systems and improper waste disposal exacerbate flood risks [1, 2, 3].

3. Need for Assessment: There is an urgent need for a comprehensive flood risk assessment. Such assessments would guide urban planners and emergency response teams in formulating strategies to mitigate flood risks and in focused interventions during flood incidents [1, 2, 3].

2. Modelling Approaches

1. 1D and 2D SWE Models:
   - 1D-SWEs: Initially used for flood simulation, easy to calculate but **less effective in complex urban flood environments** [6].
   - 2D-SWEs: An improvement over 1D-SWEs, capable of addressing complex flood situations and can utilize high-precision digital terrain models, but **overly complex and costly to train** [6, 7].

2. Machine Learning (ML) Models:
   - **Quick Results**: ML models are effective for quick predictions and can adapt to changing input parameters [6, 8, 9].
   - **Multi-Dimensional Input**: Capable of incorporating various dimensions like climate, geography, and human activities for more comprehensive modeling [6, 8, 9].

3. Data Requirements and Trends:
   - **Data Intensive**: ML approaches require extensive datasets for training [8, 9].
   - **Mainstream Acceptance**: Methods based on ML have gained traction in recent years, especially those that integrate Geographic Information Systems (GIS) for more nuanced geographical information [8, 9].

Imperial College London

SnooCODE

OBJECTIVES

This project aims to develop a **Machine Learning model** capable of predicting the probability of flood occurrence.

The specific objectives of the project include:

- **Objective 1 (Data Collection):** This encompasses, but is not limited to, the gathering of historical flood data, precipitation data, and geospatial information in Ghana.

- **Objective 2 (Data Processing):** The focus will be on exploratory data analysis (EDA), the design and computation of specific data features such as maximum continuous rainfall, maximum number of continuous rainfall days during the floods period, and data visualisation such as historical flood interactive maps with targeted boundaries and sampling coordinates.

- **Objective 3 (Model Construction):** Experiments with various machine learning models for training and prediction with a consistent training strategy across models. Subsequent evaluations and performance comparisons of the trained models will be conducted.

- **Objective 4 (Model Application):** Utilising the saved best-performing model to execute flood predictions and evaluate the predictive performance. The probability distribution maps of flooding needs to be generated and consulted for the target regions.

Imperial College London

SnooCODE

PROPOSED METHODOLOGY

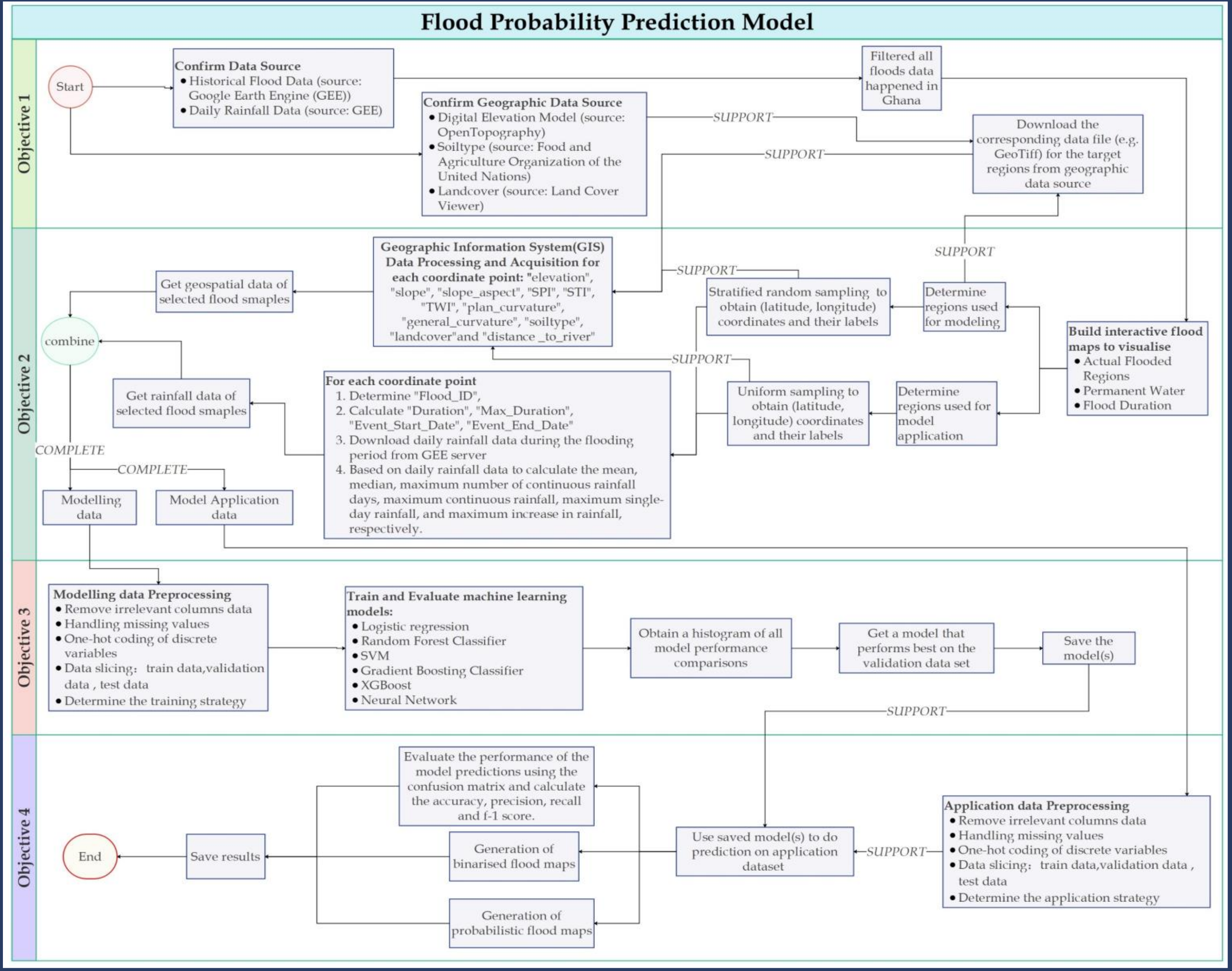

Figure 1: Flow Diagram of the Proposed Methodology.

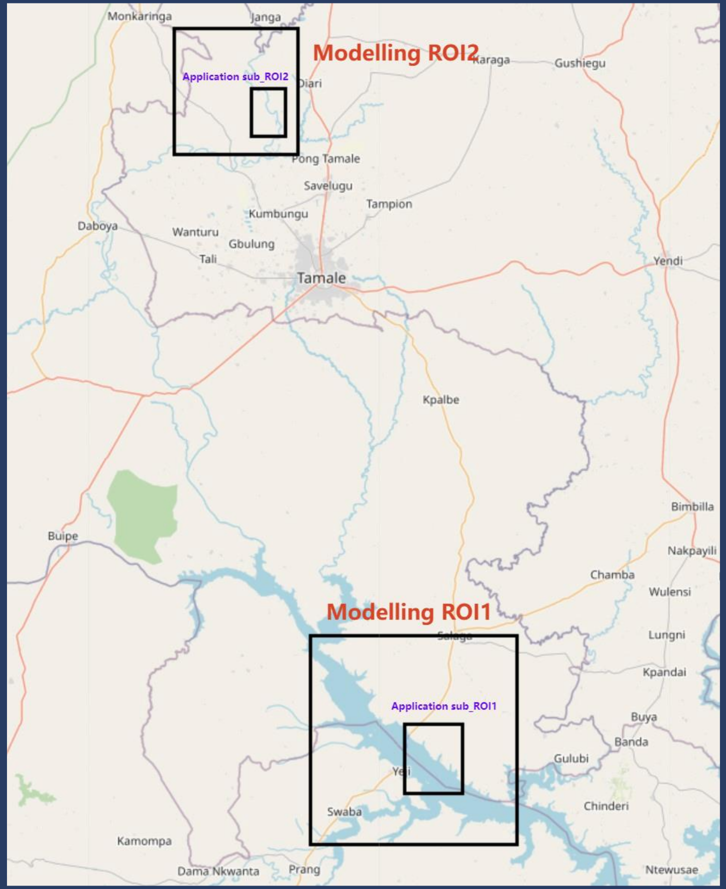

Figure 2: Geographical Representation of Modelling and Application ROI (region of interest)

Areas. Legend: Large Rectangles = Modelling Areas; Small Inner Rectangles = Application Areas.

Imperial College London

SnooCODE

PROPOSED METHODOLOGY

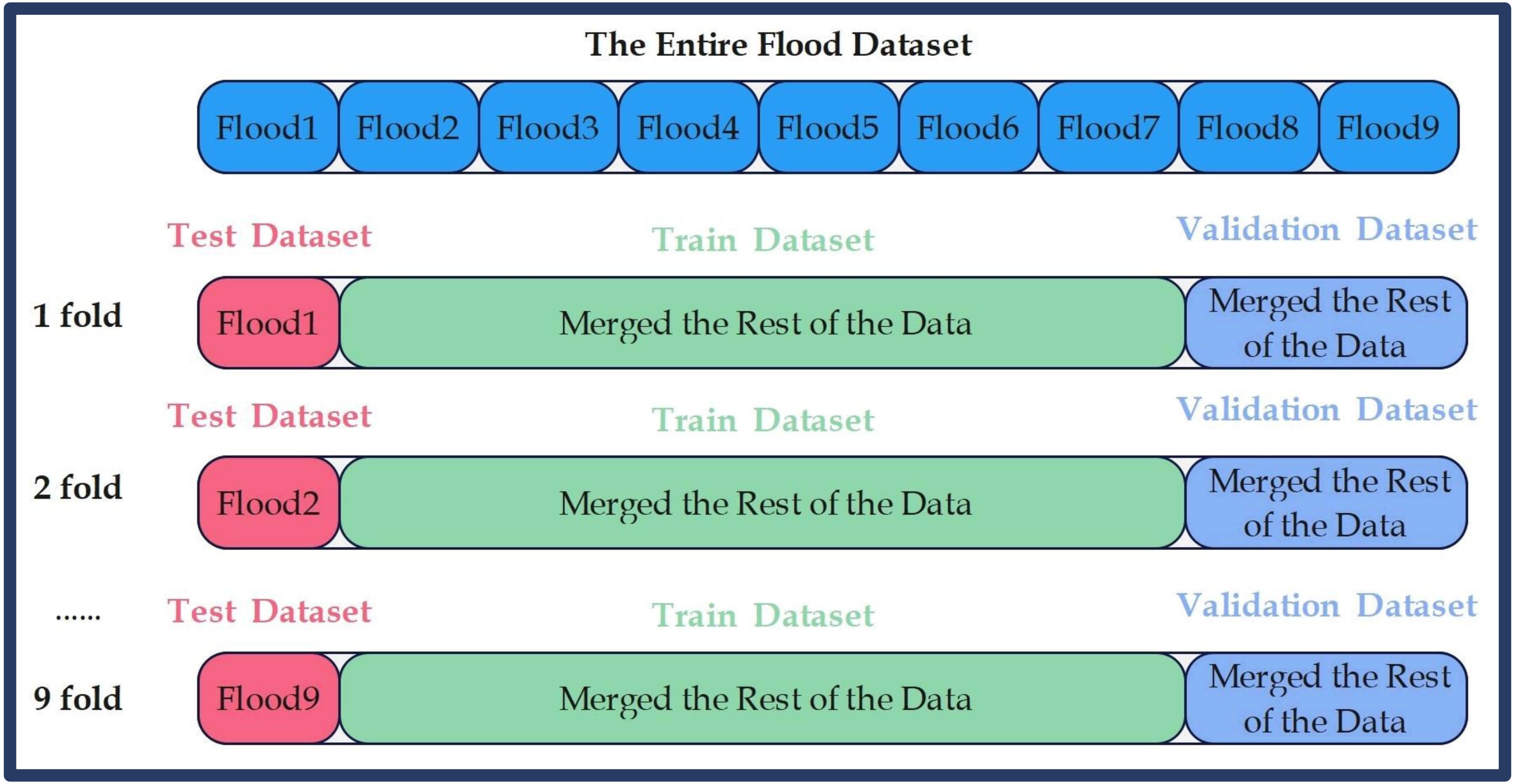

Figure 3: Illustration of Leave-One-Out Cross Validation Strategy. Legend: Test Dataset = Final Model Testing Dataset; Train/Validation Dataset = Model Training and Validation Dataset.

RESULTS

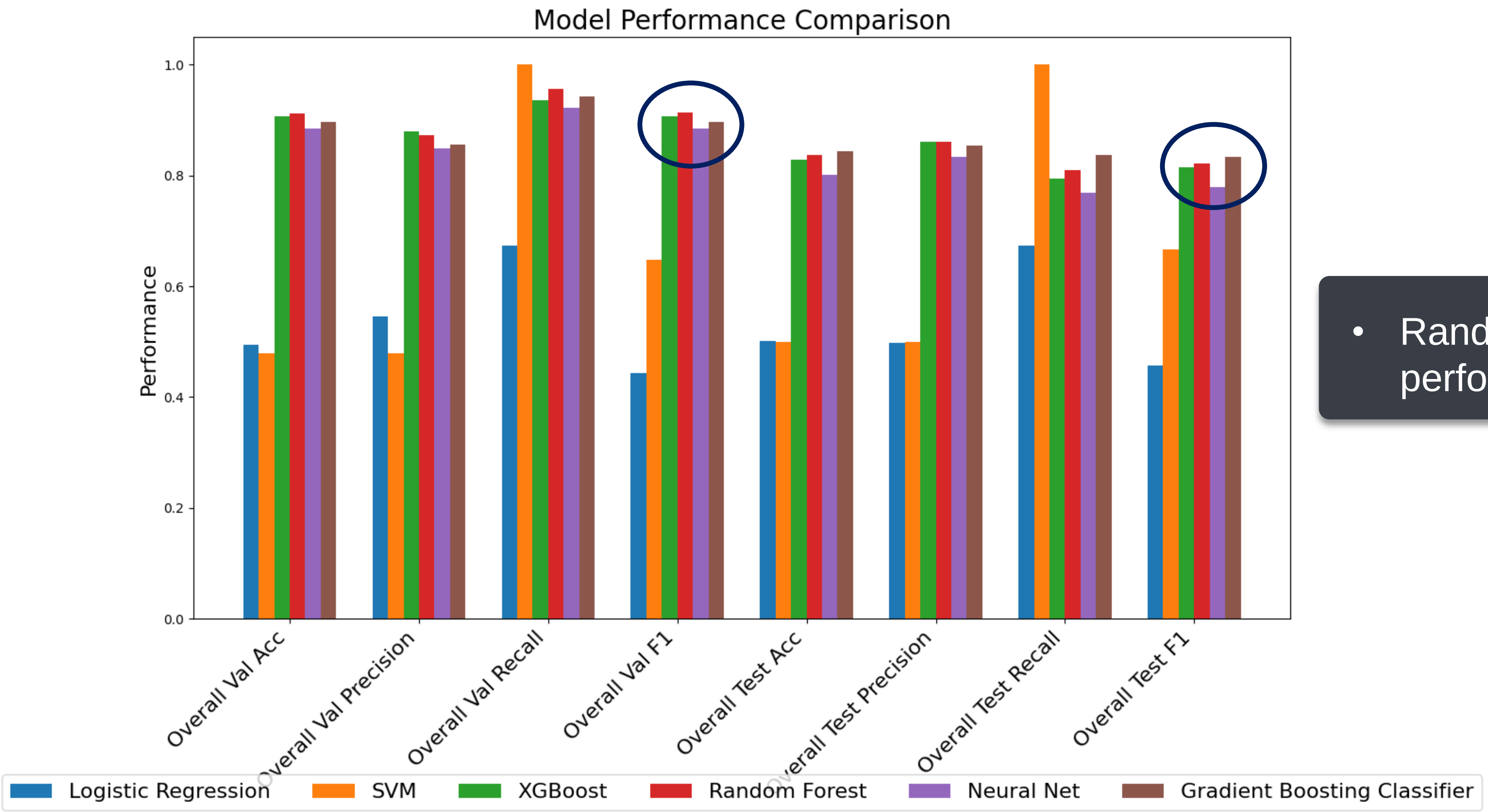

Figure 4: Comparative Analysis of Model Prediction Performances.

- Random Forest had the best overall performance

Imperial College London

SnooCODE

RESULTS

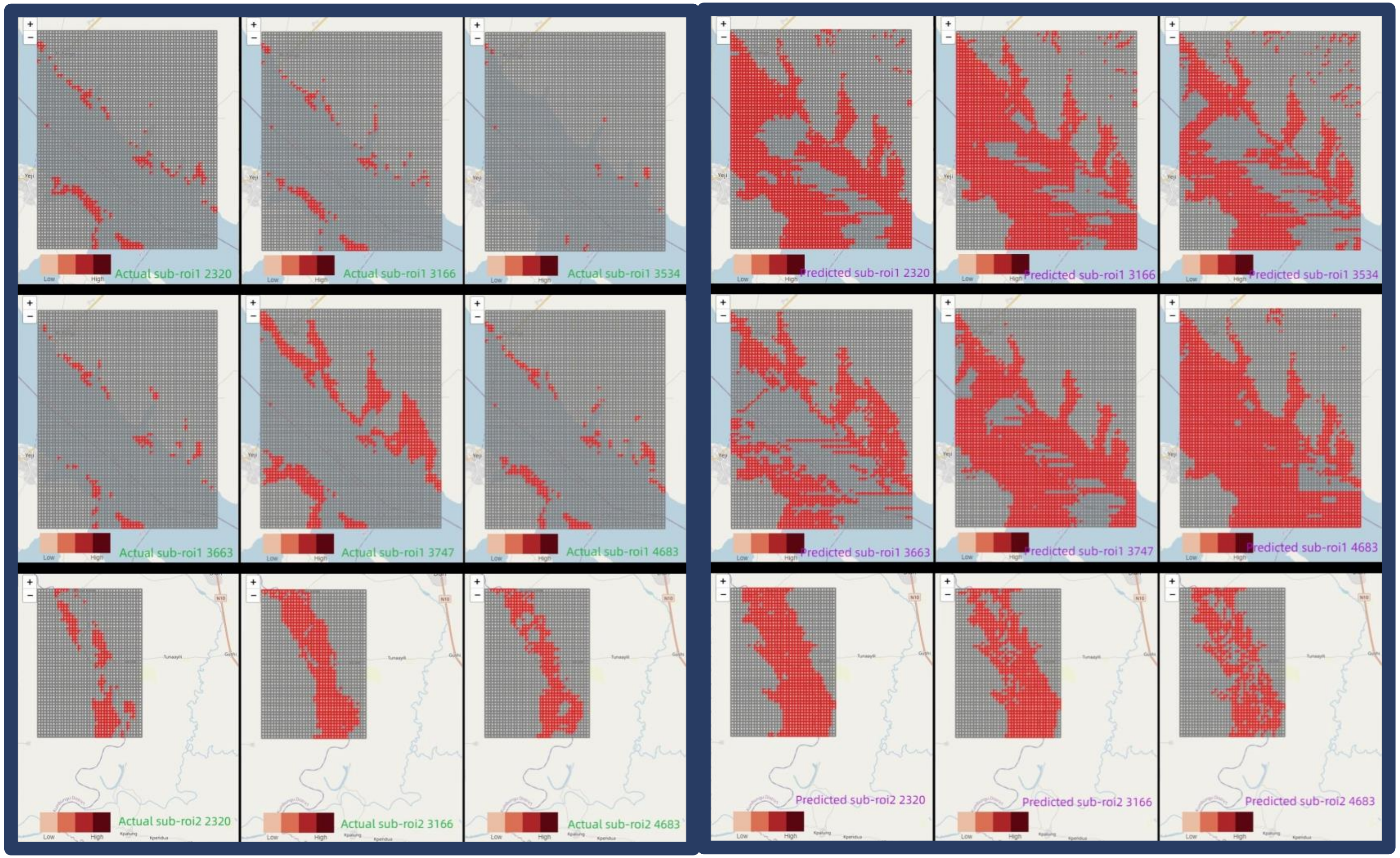

Figure 5: Comparison of Visualised Binary Flood Maps, (left) Actual, (right) Predicted.

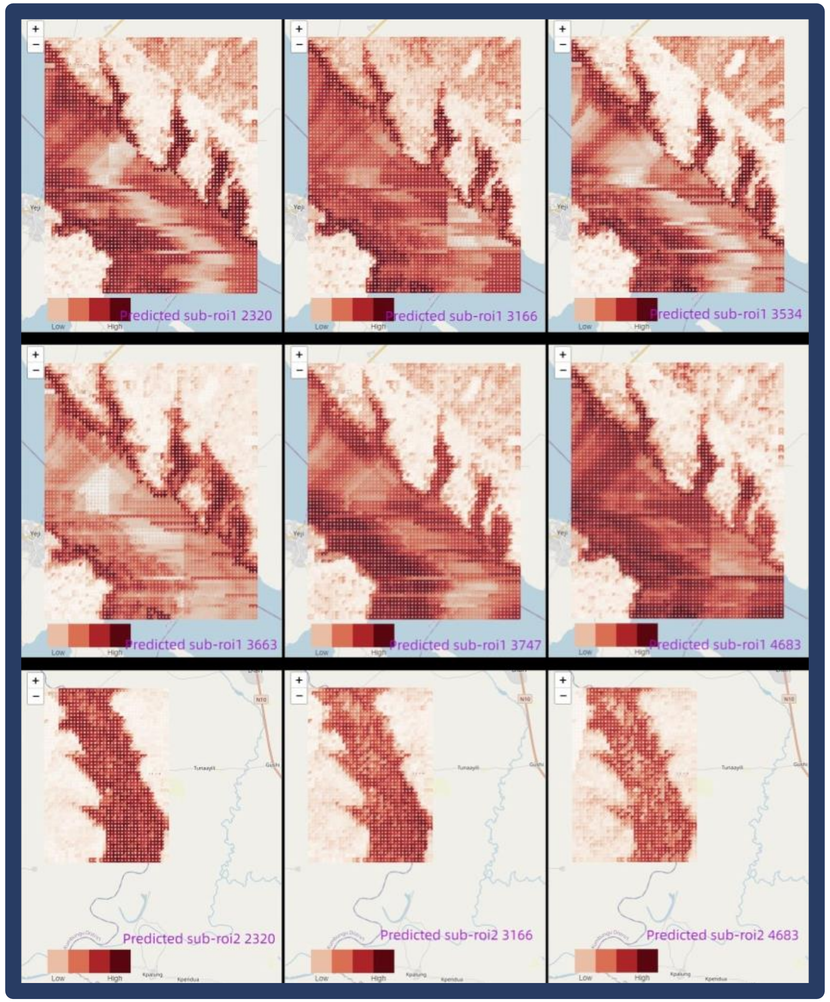

Figure 6: Probabilistic Predicted Flood Maps

Imperial College London

SnooCODE

RESULTS

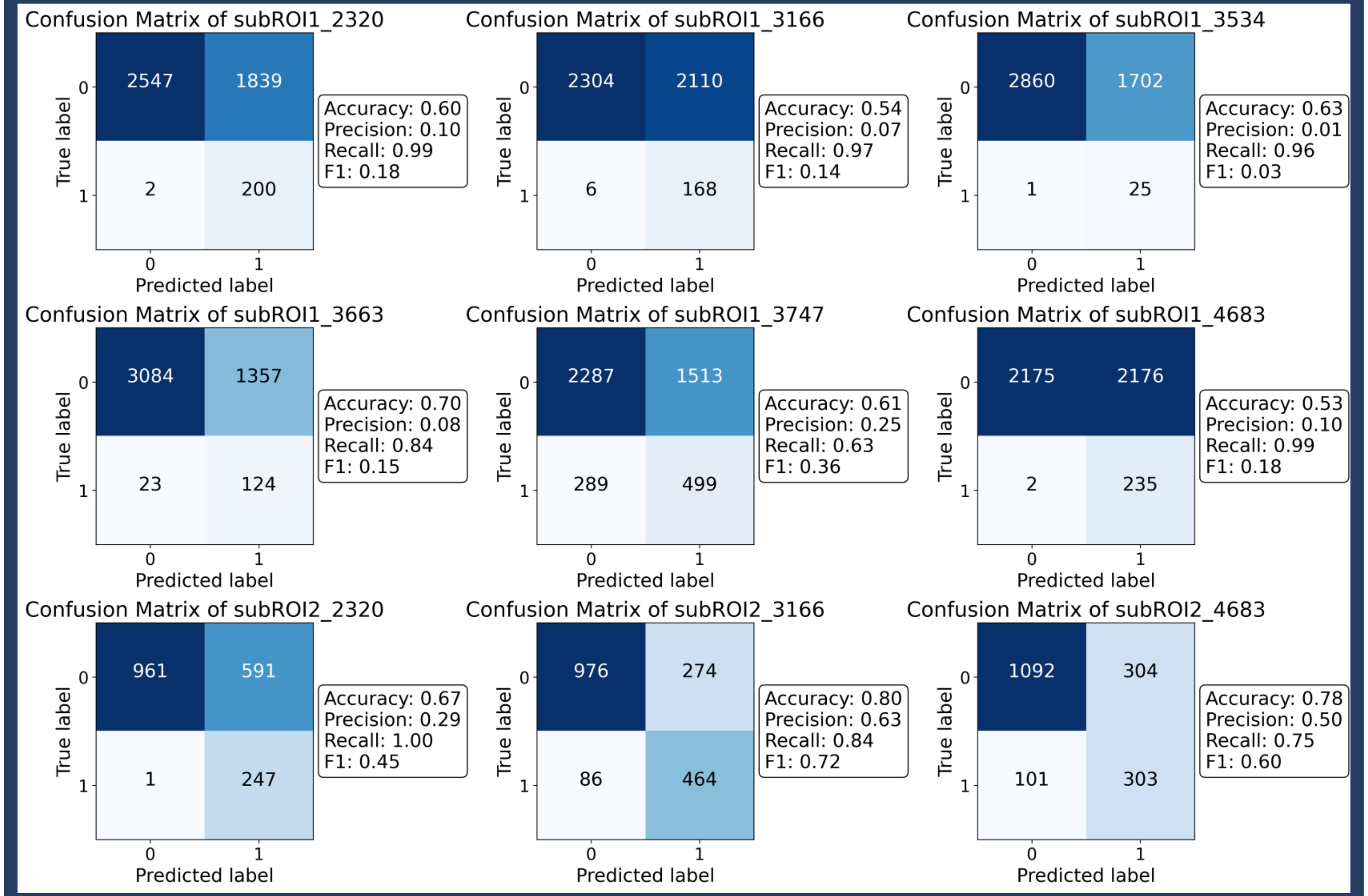

Figure 7: Confusion Matrices for Binary Flood Prediction

It is very clear to see that the model performs exceptionally well on **RECALL** but poorly on **PRECISION**

Imperial College London

SnooCODE

DISCUSSION

1. Model Strengths and Comparison with Previous Work

- High 'recall' metric for effective identification of high-risk flood areas.
- Unique combination of geospatial and rainfall data improves model accuracy.
- Single-data focus in prior research may limit real-world applicability.

2. Limitations in Precision and Resource Allocation

- Model has subpar 'precision', leading to potential resource misallocation.
- Need to improve precision without compromising high recall.
- Inconsistent performance across sub-regions, possibly due to data resampling.

3. Constraints in Real-time Prediction and Potential Solutions

- Model's reliance on historical flood data limits real-time predictive capabilities.
- Lack of precise timing for imminent floods hampers utility in real-world contexts.
- Incorporating a predictive rainfall model could enhance real-time applicability but adds complexity.

4. Challenges in Operational Scalability and Probabilistic Assessment

- Data acquisition system's batch-processing introduces logistical challenges.
- Network latency and server congestion hinder model's scalability.
- Lack of ground-truth probability maps complicates probabilistic assessment validation.

TO SUM UP…

- All 4 previously established objectives were achieved, and the Random Forest model was ultimately used as the final predictive model, based on the performance of all models.
- Model exhibits high recall but lacks precision in flood predictions.
- Study provides guidance for identifying potential flood-affected areas.

Imperial College London

SnooCODE

REFERENCE

1. S. K. M. Agblorti D. Babanawo, P. A. D. Mattah, E. K. Brempong, M. M. Mattah, and D. W. Aheto. Local indicator-based flood vulnerability indices and predictors of relocation in the ketu south municipal area of ghana. Sustainability, 2022.
2. J. Abazaami C. Anab and S. A. Achanso. Moving beyond the ad hoc responses in flood management to a localization approach in ghana. Ghana Journal of Development Studies, 2022.
3. Y. A. Twumasi and R. Asomani-Boateng. Mapping seasonal hazards for flood management in accra, ghana using gis. In IEEE International Geoscience and Remote Sensing Symposium, volume 5, pages 2874–2876, 2002.
4. S. A. Adinku. Disaster preparedness: A sociological study of the flood problem in the odaw catchment in accra. Google Statistics, 1994.
5. Y. A. Twumasi and R. Asomani-Boateng. Planning and managing urban organic solid waste in an african city: linking organic solid waste composting to urban cultivation in accra, ghana. Google Statistics, 1999.
6. L. Cea and P. Costabile. Flood risk in urban areas: Modelling, management and adaptation to climate change. a review. Hydrology, 9, 2022.
7. B. Jamali, E. Haghighat, A. Ignjatovic, J. P. Leit ˜ao, and A. Deletic. Machine learning for accelerating 2d flood models: Potential and challenges. Hydrological Processes, 35, 2021.
8. Z. Wang, X. Chen, Z. Qi, and et al. Flood sensitivity assessment of super cities. Scientific Reports, 13:5582, 2023.
9. W. Fenglin, I. Ahmad, M. Zelenakova, and et al. Exploratory regression modeling for flood susceptibility mapping in the gis environment. Scientific Reports, 13:247, 2023.