Flood Prediction Using Multilevel Radar Reflectance: A Random Forest Adaptive Approach

RENJIRO OWEN, KINGSLEY ENG, DINUSHI THATHSARANI JAYASINGHE, and SIDHANT DABAS, University of Waikato, New Zealand

Flood prediction is critical for mitigating the impacts of flooding on communities and infrastructure. This study presents an approach to flood prediction using a comprehensive dataset containing time-series data and various environmental features. The dataset includes timestamped recordings of radar reflectance at three different altitudes (sea level, 2000m, and 4000m), river stage levels, and rain gauge recordings from two locations[Project 2022]. We employ the Random Forest Adaptive algorithm, implemented through the CapyMoa Python library, to model and predict flood occurrences. The radar reflectance data is organised in a 10x14 grid format, providing detailed spatial coverage across different layers, which helps in capturing the potential flooding indicators.

Our approach leverages the strengths of the Random Forest Adaptive algorithm, which combines the robustness of ensemble learning with adaptive techniques to handle the dynamic nature of floods. The model is trained on historical data, learning complex relationships between the environmental features and flood events.

The results demonstrate that our model achieves high accuracy in predicting floods in several hours before. This predictive capability can support early warning systems, enabling timely and effective flood management strategies. The integration of multi-source environmental data and advanced machine learning techniques provides a robust framework for flood prediction, contributing to disaster preparedness and resilience.

CCS Concepts: • **Computing methodologies → Classification and regression trees**.

Additional Key Words and Phrases: Flood prediction, Radar reflectance, Predictive model, Random forest

ACM Reference Format:
Renjiro Owen, Kingsley Eng, Dinushi Thathsarani Jayasinghe, and Sidhant Dabas. 2018. Flood Prediction Using Multilevel Radar Reflectance: A Random Forest Adaptive Approach. *ACM Trans. Graph.* 37, 4, Article 111 (August 2018), 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Due to their destructive effect, floods create a significant impact on human life. In areas prone to flooding, residents face challenges in accurately predicting the risk of floods[Cloke and Pappenberger 2009]. Accurate and timely flood forecasting can enable better preparedness and response, minimising the adverse impacts on affected communities. This project aims to develop a predictive model to forecast the likelihood of flooding within a six-hour timeframe using a Random Forest algorithm.

The data utilised for this project is stored in CSV format and comprises multiple features recorded at various times. Specifically, the dataset includes the following: the Timestamp (UTC time) indicating the exact time when each recording was made; Gridded Radar Reflectance, which encompasses Sea Level (10x14 grid) captured in Columns 1-140, 2000 meters above Sea Level (10x14 grid) in Columns 141-280, and 4000 meters above Sea Level (10x14 grid) in Columns 281-420; River Stage Levels found in Columns 421-422, which capture the height of the river water at two locations; and Rain Gauge Recordings in Columns 423-424, which measure the amount of rainfall at two different locations[Project 2022].

By leveraging the diverse and comprehensive features provided by this dataset, the Random Forest model will analyse patterns and correlations to predict flooding events. The choice of the Random Forest algorithm is due to its robustness, ability to handle large datasets with numerous features, and its effectiveness in classification problems.

The primary goal of this project is to create a reliable flood prediction system that can provide early warnings, thus enhancing the capability to mitigate flood-related risks and manage resources more efficiently.

2 Related Work

Flood prediction systems have seen significant advancements in recent years, particularly with the integration of machine learning and data-driven approaches. Google's operational flood forecasting system, which became operational in 2018, exemplifies this trend by providing accurate real-time flood warnings focused on riverine floods in large, gauged rivers. This system's architecture comprises four key subsystems: data validation, stage forecasting, inundation modeling, and alert distribution. Notably, machine learning models, specifically Long Short-Term Memory (LSTM) networks and linear models, are employed for stage forecasting, while inundation modeling leverages both thresholding and novel manifold models. The LSTM networks have demonstrated superior predictive skills over linear models in stage forecasting, and both thresholding and manifold models have achieved comparable performance metrics in estimating inundation extents[Nevo et al. 2022].

Building on this foundation, our project aims to predict floods using reflectance radar timestamp data, offering a novel approach that could potentially enhance early warning capabilities. While Google's system relies heavily on riverine data and machine learning models, our approach leverages the temporal and spatial patterns captured by radar reflectance data to predict flood events. Reflectance radar data, which captures real-time changes in the surface conditions, provides a rich dataset for machine learning models

Authors' Contact Information: Renjiro Owen, renjiowen71@gmail.com; Kingsley Eng, ke126@waikato.ac.nz; Dinushi Thathsarani Jayasinghe, jayasinghedt@gmail.com; Sidhant Dabas, sd777@waikato.ac.nz, University of Waikato, Hamilton, Waikato, New Zealand.

ACM 1557-7368/2018/8-ART111
https://doi.org/XXXXXXX.XXXXXXX

to identify emerging flood patterns. Previous studies have highlighted the effectiveness of radar data in hydrological applications, and integrating this with advanced machine learning techniques such as LSTM networks could offer a robust framework for flood prediction. This approach not only aligns with the growing trend of utilising diverse data sources for environmental modeling but also promises to extend the predictive accuracy and geographic coverage of flood forecasting systems, particularly in regions where riverine data may be sparse or unavailable.

3 Model

The Adaptive Random Forest (ARF) algorithm is a ensemble learning model designed to handle evolving data streams, combining the strengths of traditional Random Forests with adaptations for streaming data environments. Random Forests, introduced by [Breiman 2001], are an ensemble method that constructs multiple decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees. Each tree in a Random Forest is trained on a bootstrap sample of the data, and at each split in the tree, a random subset of features is considered, which helps in reducing overfitting and improving generalisation.

ARF can be used with streamed data through several key mechanisms, particularly Online Bagging, Feature Subspacing, Drift Detection and Adaption, and Weighted Voting. ARF uses an online bagging approach where each instance is weighted in regard to a Poisson distribution. Specifically, each instance x_t at time t is assigned a weight k drawn from a Poisson(λ) distribution. This weight determines how many times x_t is used to update each tree in the ensemble [Gomes et al. 2017]. In terms of Feature Subspacing, ARF selects a random subset of features at each split in the decision tree, this random selection means that each tree in the forest is diverse, improving the model.

ARF are good at detecting and adapting to concept drifts, which are changes in the underlying distribution over time, which can reduce the performance of static models. ARF can incorporate drift detectors such as ADWIN (Adaptive Windowing) to watch the performance of each tree. When drift is detected a new 'background' tree is trained in parallel as the existing tree continues to make predictions. If the drift is confirmed, this new tree replaces the current tree [Gomes et al. 2017]. In terms of Weighted Voting, in ARF the prediction of each individual tree is combined using a weighted voting mechanism. With each vote being weighted according to its accuracy on recent data, meaning more accurate trees have more influence on the final prediction. This weighting system is shown to help maintain high performance even as the data evolves [Gomes et al. 2017].

These properties make ARF well-suited for real-world applications where data streams are non-stationary and require models that can learn and adapt in real-time. ARF provides a robust and scalable solution for evolving data stream classification, outperforming many state-of-the-art algorithms in various scenarios [Gomes et al. 2017].

4 Proposal

This project aims to develop a predictive model for forecasting floods with a six-hour lead time, leveraging the Random Adaptive Forest algorithm. The model will utilise a comprehensive dataset available from the Time-Evolving Data Science / Artificial Intelligence for Advanced Open Environmental Science (TAIAO), specifically the "River Radar 2015-2018" dataset. This dataset, provided in CSV format, includes a variety of features essential for accurate flood prediction: timestamped radar reflectance data at three different altitudes (sea level, 2000m, and 4000m), river stage levels, and rain gauge recordings from two locations[Project 2022]. By integrating these diverse data points, we can create a robust and dynamic model capable of delivering timely flood warnings.

The Random Adaptive Forest algorithm is particularly suitable for this task due to its ability to handle high-dimensional data[Breiman 2001]. This model will analyse the gridded radar reflectance data (spanning 420 columns), river stage levels, and rainfall measurements to detect patterns indicative of impending floods. The timestamp column will be used to align data points temporally, ensuring that the model considers the chronological sequence of events leading up to flood occurrences. By training the model on historical data and validating its performance, we aim to achieve a high degree of accuracy in predicting floods six hours in advance, thus providing crucial lead time for preventive measures and disaster management.

Implementing this model involves several steps: data preprocessing to handle missing values and normalise the features, training the Random Adaptive Forest on a portion of the data, and then rigorously testing it on unseen data to evaluate its predictive performance. The ultimate goal is to deploy this model in a real-world setting, where it can continuously receive new data, adapt to changing conditions, and provide real-time flood predictions. This approach not only enhances the accuracy of flood forecasting but also offers a scalable solution adaptable to various geographical regions and environmental conditions.

5 Experiments

5.1 Data Processing

The experiment begins by preprocessing raw data stored in CSV format, featuring a dataset with various parameters recorded over time, including radar reflectance, river stage levels, and rainfall measurements. Initial steps involve timestamp extraction from the first column to determine the hour of each recording in UTC time. Following this, radar reflectance values across different altitudes are normalised, with the minimum and maximum value adjusted. The Opotunui column is then removed from the dataset, focusing solely on Tairua for prediction.

In the following steps, adjustments are applied to the target feature by shifting the column by 24 units because each row represents a 15-minute interval. This adjustment, coupled with the subtraction of the current water level, enables the prediction of water movement in Tairua. We believed water movement is favored over water level for flood prediction as it considers the dynamic nature of river systems and accommodates variations in initial water levels across different rivers, providing a more comprehensive understanding of flood dynamics. Additionally, the dataset incorporates the current water level in Tairua as an extra feature, further improving its flood prediction capabilities. These preprocessing procedures establish the groundwork for analysing and predicting water-related events, thereby aiding in informed decision-making and risk management.

5.2 Modelling

Once the data has undergone preprocessing, it is given to the MOA framework. Various machine learning algorithms, including Adaptive Random Forest, SGD, and Perceptron, were employed to evaluate the data. Through experimentation, it was determined that Adaptive Random Forest is the most effective for predicting water movement. Several parameters were tested, revealing that an ensemble size of 25, a window size of 4500, and a grace period of 200 for the fast decision tree yielded optimal results.

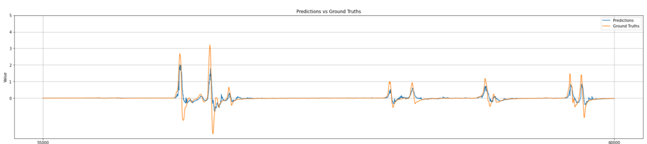

Fig. 1. Prediction Water Movement Compared to Ground Truth

The predictions are first saved in a .txt file and visualised using Python, where they are plotted with the original target values in Figure 1. Moreover, to provide a comprehensive analysis, the original Tairua water level is incorporated and contrasted with the Tairua water level, culminating in the creation of Figure 2.

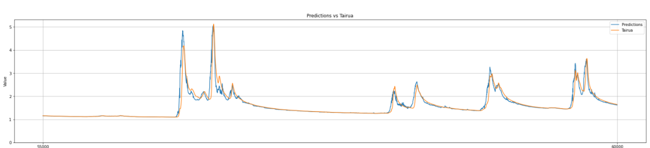

Fig. 2. Prediction Water Level Compared to Tairua Water Level

The performance of the model was comprehensively evaluated using various evaluation metrics. In Figure 3, the RMSE and MAE values are plotted to provide a visual representation of their distribution across the experiments. The RMSE values ranged from 0.00285 to 0.2282, while the MAE values ranged from 0.00269 to 0.07667. Furthermore, the overall RMSE was calculated to be 0.045249, indicating the average magnitude of errors across all experiments. Similarly, the overall MAE was determined to be 0.01746, representing the average absolute difference between the predicted values and the actual values.

We utilised CapyMoa for plotting, which necessitated adjustments to prevent automatic integer conversion. The actual water level data was sourced for analysis, with a specific focus on floods in Tairua, where the threshold is set at 4 meters. Notably, our predictions consistently indicated flood occurrences when the water level surpassed this threshold, demonstrating reliable performance in flood prediction can be seen in Figure 4. Although some of our predicted peaks exceeded the actual levels, this disparity isn't concerning as long as our predictions consistently err on the side of

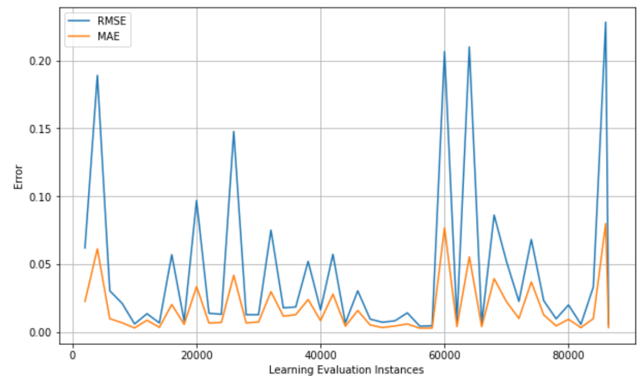

Fig. 3. MAE and RMSE Graph

caution, avoiding underestimation of flood risks. Visualising the predicted water level against the ground truth facilitated a clearer understanding of river dynamics and flood occurrences compared to relying solely on RMSE (Root Mean Square Error), offering a more intuitive insight into the behavior of the river and the timing of flood events.

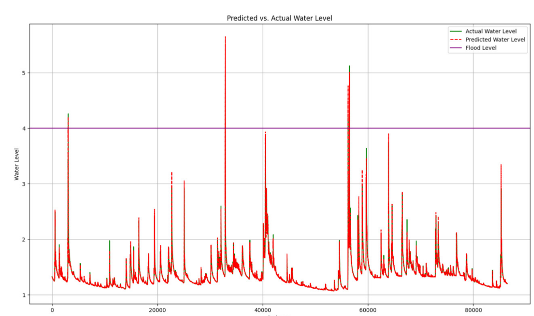

Fig. 4. Water Level vs Predicted Water Level in Tairua in meters

6 Discussion

In this project, our primary aim was to predict water movement, particularly focusing on the potential for flood prediction. Utilising the Adaptive Random Forest algorithm yielded promising results, showcasing its effectiveness in this context. By accurately forecasting water movement, we can enhance our ability to anticipate and mitigate the impact of floods, thereby contributing to improved disaster management strategies. It is essential to acknowledge that despite the overall success of our approach, we encountered instances of mispredictions, particularly when the water level was decreasing. However, our emphasis remained on refining the model's capability to predict flood occurrences, as these events pose significant risks and necessitate proactive measures for mitigation and response. Through continued refinement and possibly incorporating additional features or refining existing ones, we aim to further enhance the accuracy and reliability of our flood prediction model. Furthermore, the low RMSE and MAE values achieved underscore the model's ability to accurately predict water movement, particularly in the context of flood prediction, thereby emphasising its potential utility in real-world applications. Additionally, future work could explore integrating real-time data streams and leveraging advanced modeling techniques to bolster the predictive capabilities of our system, ultimately contributing to more effective flood management practices.

7 Conclusion

In conclusion, our project aims to address the challenge of water movement prediction, with a specific focus on flood forecasting. Through the utilisation of the Adaptive Random Forest algorithm, we have achieved promising results in accurately predicting water movements, thereby enhancing our capacity to anticipate and mitigate flood events. Additionally, the model's performance metrics, including low RMSE and MAE values, highlights its reliability and effectiveness in practical applications.

References

Leo Breiman. 2001. Random Forests. *Machine Learning* 45, 1 (Oct 2001), 5–32. https://doi.org/10.1023/A:1010933404324

Hannah L Cloke and Florian Pappenberger. 2009. Ensemble flood forecasting: A review. *Journal of Hydrology* 375 (September 2009), 613–626.

Heitor M. Gomes, Albert Bifet, Jesse Read, Jean Paul Barddal, Fabrício Enembreck, Bernhard Pfharinger, Geoff Holmes, and Talel Abdessalem. 2017. Adaptive random forests for evolving data stream classification. *Machine Learning* 106, 9 (October 2017), 1469–1495. https://doi.org/10.1007/s10994-017-5642-8

S. Nevo, E. Morin, A. Gerzi Rosenthal, A. Metzger, C. Barshai, D. Weitzner, D. Voloshin, F. Kratzert, G. Elidan, G. Dror, G. Begelman, G. Nearing, G. Shalev, H. Noga, I. Shavitt, L. Yuklea, M. Royz, N. Giladi, N. Peled Levi, O. Reich, O. Gilon, R. Maor, S. Timnat, T. Shechter, V. Anisimov, Y. Gigi, Y. Levin, Z. Moshe, Z. Ben-Haim, A. Hassidim, and Y. Matias. 2022. Flood forecasting with machine learning models in an operational framework. *Hydrology and Earth System Sciences* 26, 15 (2022), 4013–4032. https://doi.org/10.5194/hess-26-4013-2022

TAIAO Project. 2022. River and Radar Sensor Data (2015-2018). https://datasets.cms.waikato.ac.nz/taiao/river_radar_2015_2018/