Project Proposal

Flood Forecasting and River Inundation Modeling

Group Members
AbdeAli (K20-0122)
Syed Sufyan Imran (K20-0386)
Nausherwan Baig (K20-0428)

Department of Computer Science, FAST
Final Year Project

Signature
Project Supervisor
Dr. Zulfiqar Ali Memon

Project Co-Supervisor
Ali Asghar
(Software Engineer at Google)

Contents

Abstract

Addressing the escalating flood risks in Pakistan, this development-centric project focuses on establishing a robust time-series forecasting tool tailored to predict river streamflow's across pivotal Pakistani rivers. The development strategy is mapped out in three stages. In the Data Collection phase, we prioritize procuring daily stream flow measurements and pertinent climatic data, predominantly from the forecasting division of Pakistan. During the Feature Representation and Model Training stage, while core features such as discharge and precipitation will be central, an iterative process will fine-tune additional data features, ensuring optimal model performance. The culmination of this endeavor is the Flood Inundation Map, designed to provide stakeholders with real-time visualizations of regions at risk based on anticipated water levels. This tool, once operational, promises to be an indispensable asset for proactive flood mitigation in Pakistan, with the potential to safeguard both infrastructure and lives.

1 Introduction

Pakistan, graced by the expansive Indus River system, faces the double-edged sword of its waters. While these rivers are vital lifelines, they also bring forth the looming threat of floods, a danger accentuated by the inclined planes in the northern areas. Here, floods are swift, often giving communities and authorities little time to react. Traditional flood forecasting methods, though valuable, now seem inadequate in the face of these rapid challenges.

A crucial gap to note is that, despite its vulnerability, Pakistan currently lacks any AI/ML-driven flood prediction stations. This absence is starkly evident when compared to the global shift towards data-driven flood management, as exemplified by platforms like Google Flood Hub. However, such platforms are not tailored to Pakistan's unique topographical and hydrological nuances.

Enter our project, envisioned to fill this void. Leveraging state-of-the-art machine learning models, we aim to construct a system that factors in the rapid flood movements due to northern inclines and offers predictions custom-built for Pakistan. Our collaboration with the Flood Forecasting Division ensures access to efficient, quantitative, and qualitative data, with a consistent inflow of regular metrics crucial for accurate predictions.

2 Problem Statement

Pakistan frequently faces rapid and potentially devastating floods, especially in its inclined northern regions. Traditional flood forecasting methods, while existent, have shown limitations in their accuracy and timeliness. Despite the availability of consistent flood-related data and global advancements in AI/ML-driven solutions, the country lacks a predictive tool tailored to its specific hydrological and topographical challenges. This gap leaves communities and stakeholders without an efficient mechanism to anticipate and mitigate the impacts of these swift flood events.

3 Scope and Objective

3.1 Data Integration

1. Collaborate with the Flood Forecasting Division of Pakistan to source reliable, consistent, and comprehensive historical data on streamflow and related climatic variables.

2. Ensure the acquired data undergoes rigorous preprocessing to remove inconsistencies and address gaps.

3.2 Model Development & Validation

1. Design a machine learning model tailored to Pakistan's unique topographical and hydrological challenges, emphasizing the rapid flood movements due to northern inclines.

3.3 Application Development

1. Create a user-friendly application interface to showcase real-time flood forecasting results to stakeholders.

2. Integrate the Flood Inundation Map into the application, offering visually representative depictions of flood-prone zones based on forecasted data.

3.4 Continuous Improvement

1. Regularly assess the model's accuracy and fine-tune parameters based on the latest data and feedback.

2. Engage in ongoing research to identify additional features or variables that can enhance prediction accuracy.

4 Limitations

4.1 Data Constraints

- While the Flood Forecasting Division provides consistent data, there might be instances of missing or incomplete records which can affect model accuracy.
- The model's effectiveness is inherently tied to the quality and comprehensiveness of the data acquired.

4.2 Geographical Limitations

- The model's primary focus is on specific river stations and might not account for the entire river system in Pakistan.
- The project primarily concentrates on the rapid flood movements in northern areas, and results may vary for other regions.

4.3 Infrastructure

- Aging infrastructure like dams, canals, and embankments can have undocumented weaknesses or defects that can lead to unexpected flooding.

4.4 External Factors

- Unpredictable climatic events or significant changes in the environment beyond the historical data can influence the model's predictions.
- Sociopolitical or infrastructural challenges might hinder timely data collection or model deployment in some areas.

5 Literature Review

[1] Flood forecasting with machine learning models in an operational framework

- This study presents a comprehensive review of the application of machine learning (ML) models for flood forecasting within an operational framework. The authors discuss the advantages and challenges of using ML models for flood forecasting and highlight the importance of data preprocessing, feature selection, and model evaluation. The study also emphasizes the need for integrating ML models with traditional hydrological models to improve flood forecasting accuracy.

[2] Contemporary Trends in High and Low River Flows in Upper Indus Basin, Pakistan

- The research focuses on analyzing the trends in high and low river flows in the Upper Indus Basin (UIB) of Pakistan. The study uses statistical methods to identify significant trends and attributes the observed changes to climate variability and anthropogenic activities. The findings suggest that there is a noticeable increase in high river flows and a decline in low river flows in the UIB.

[3] Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets

- The paper delves into the challenges of regional rainfall–runoff modeling in hydrological sciences. Traditional models often falter when applied across multiple basins, prompting the authors to explore a data-driven approach using Long Short-Term Memory networks (LSTMs). By training on the extensive CAMELS dataset, which encompasses 531 basins, the authors demonstrate that LSTMs can significantly outperform several benchmark hydrological models. This performance boost is evident even when compared to models calibrated individually for each basin.

  Furthermore, the authors introduce the Entity-Aware-LSTM (EA-LSTM), an innovative adaptation of the standard LSTM. This new architecture is designed to learn catchment similarities, effectively acting as a feature layer in the deep learning model. Impressively, the similarities discerned by the EA-LSTM align well with established hydrological knowledge, underscoring the potential of machine learning in advancing the field of hydrology.

In conclusion, this paper underscores the transformative potential of machine learning, specifically LSTMs, in hydrological modeling. By challenging traditional modeling paradigms and introducing the EA-LSTM, the authors provide a promising direction for future hydrological research.

[4] HydroNets: Leveraging River Structure for Hydrologic Modeling

- Hydrologic models play a pivotal role in applications ranging from water resource management to flood warnings. As the climate undergoes changes, the patterns of precipitation and rainfall-runoff become more unpredictable, making it challenging to obtain accurate training data. The paper introduces HydroNets, a novel family of hydrologic models that harness the structure of river networks. These models are deep neural networks designed to utilize both specific rainfall-runoff signals from basins and the dynamics of upstream networks. By incorporating the structure of river networks, HydroNets can enhance predictions over longer horizons. This approach not only reduces the sample complexity but also offers scalability, making it possible to achieve precise hydrologic modeling even with limited data. Empirical studies conducted over two significant basins in India provide robust evidence supporting the advantages of HydroNets. The research bridges the gap between domain-specific knowledge and data-driven methods, offering a promising direction for future hydrologic modeling endeavors.

6 Tools

6.1 Data Collection & Preprocessing

- **Databases:** PostgreSQL, MySQL (for structured data storage and retrieval)
- **Data Cleaning Tools:** Python libraries like Pandas and NumPy
- **Data Visualization Tools:** Matplotlib, Seaborn (for exploratory data analysis)

6.2 Model Development & Training

- **Machine Learning Libraries:** TensorFlow, Keras, Scikit-learn (for model creation, training, and evaluation)
- **Time-series Analysis Tools:** Prophet, Statsmodels (for specific time-series forecasting models)

6.3 Application Development

- **Backend Development:** Django, Flask (Python-based frameworks for web application development)
- **Frontend Development:** ReactJS, VueJS (for developing interactive UI)
- **Web Servers:** Apache, Nginx (for deployment and hosting)

7 Methodology

7.1 Data Collection and Preprocessing

1. **Data Acquisition:** Collaborate with the Flood Forecasting Division of Pakistan to gather historical data on streamflow, precipitation, and other relevant variables for selected river gauge stations.

2. **Data Cleaning:** Utilize Python libraries to preprocess the data, addressing missing values, outliers, and normalizing features.

3. **Feature Engineering:** Extract or derive new features from the data that can enhance prediction accuracy.

7.2 Model Development and Training

1. **Model Selection:** Begin with evaluating various time-series forecasting models suitable for hydrological data, such as LSTM, ARIMA, and Prophet.

2. **Training and Validation:** Segment the data into training, validation, and test sets. Utilize cross-validation methods to ensure the model generalizes well.

3. **Model Tuning:** Fine-tune hyperparameters of the chosen model(s) to achieve optimal prediction accuracy.

7.3 Model Evaluation

1. **Performance Metrics:** Deploy metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared to evaluate the model's performance on the test set.

2. **Residual Analysis:** Examine residuals to ensure that the model effectively captures the data's patterns.

7.4 Application Development

1. **User Interface Design:** Construct a user-centric interface that displays real-time flood predictions and related information.

2. **Flood Inundation Map Integration:** Use the model's predictions to create a flood inundation map that visualizes potential flood-prone areas.

3. **Deployment:** Host the application on a dependable server to ensure constant access for stakeholders.

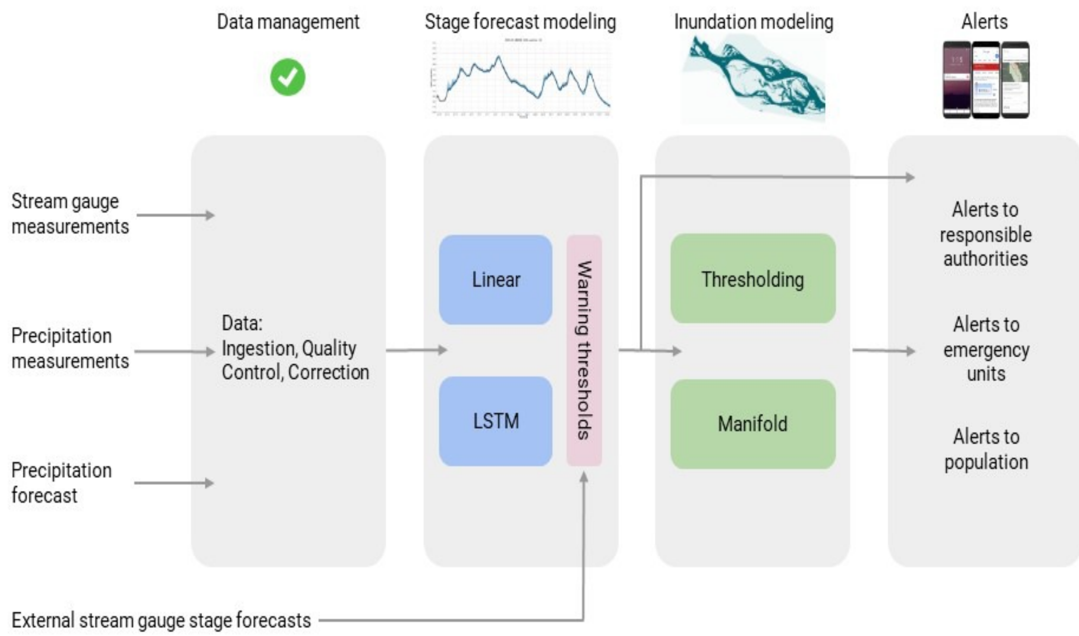

Figure 1: Operational Flood Warning System Scheme

8 Timeline

8.1 FYP-1 (October - December): Documentation, Data Collection, and Model Training

Sprint 1 (October, Week 1-4): Initiation, Data Collection, and Data Preprocessing

- Task 1.1: Begin the SRS (Software Requirements Specification) documentation.
- Task 1.2: Establish a collaboration with the Flood Forecasting Division.
- Task 1.3: Initiate collection of datasets.
- Task 1.4: Preliminary data cleaning.
- Task 1.5: Refine SRS based on initial data insights.

Sprint 2 (November, Week 1-2): SDS (Software Design Specification) Drafting and Feature Engineering

- Task 2.1: Begin the SDS documentation.
- Task 2.2: Identify and extract relevant features for the model.

Sprint 3 (November, Week 3-4): Model Selection, Initial Training, and Optimization

- Task 3.1: Shortlist predictive models suitable for the data.
- Task 3.2: Conduct preliminary training on selected models.
- Task 3.3: Optimize model through hyperparameter tuning.

Sprint 4 (December, Week 1-2): Documentation Refinement and Model Validation

- Task 4.1: Refine both SRS and SDS based on model insights.

- Task 4.2: Validate trained model with test data.

Sprint 5 (December, Week 3-4): Model Iteration and FYP-1 Closure

- Task 5.1: Iterate and refine the model based on November's validation.
- Task 5.2: Prepare for FYP-1 project review and closure.

8.2 FYP-2 (February - May): Flood Inundation Modeling, Deployment, and Finalization

Sprint 6 (February, Week 1-2): Backend Development and Flood Inundation

- Task 6.1: Set up the application backend.
- Task 6.2: Start working on flood inundation modeling.

Sprint 7 (February, Week 3-4): Frontend Development and Inundation Refinement

- Task 7.1: Develop frontend components for the application.
- Task 7.2: Refine flood inundation model.

Sprint 8 (March, Week 1-2): Integration and Testing

- Task 8.1: Integrate backend and frontend components.
- Task 8.2: Conduct initial testing of the system.

Sprint 9 (March, Week 3-4): Stakeholder Engagement & Feedback

- Task 9.1: Demonstrate the application to potential stakeholders.
- Task 9.2: Gather feedback and understand user requirements.

Sprint 10 (April, Week 1-2): Deployment & Iteration

- Task 10.1: Deploy the application on the chosen platform.
- Task 10.2: Iterate based on user feedback and system performance.

Sprint 11 (April, Week 3-4): Final Testing & User Training

- Task 11.1: Conduct final tests of the system.
- Task 11.2: Organize training sessions for stakeholders and users.

Sprint 12 (May, Week 1-2): Final Presentation Preparation

- Task 12.1: Gather all findings, results, and feedback.
- Task 12.2: Prepare and refine final presentation materials.

Sprint 13 (May, Week 3-4): Final Presentation & Project Closure

- Task 13.1: Present the project to examiners and stakeholders.
- Task 13.2: Conclude the project, documenting lessons learned and next steps.

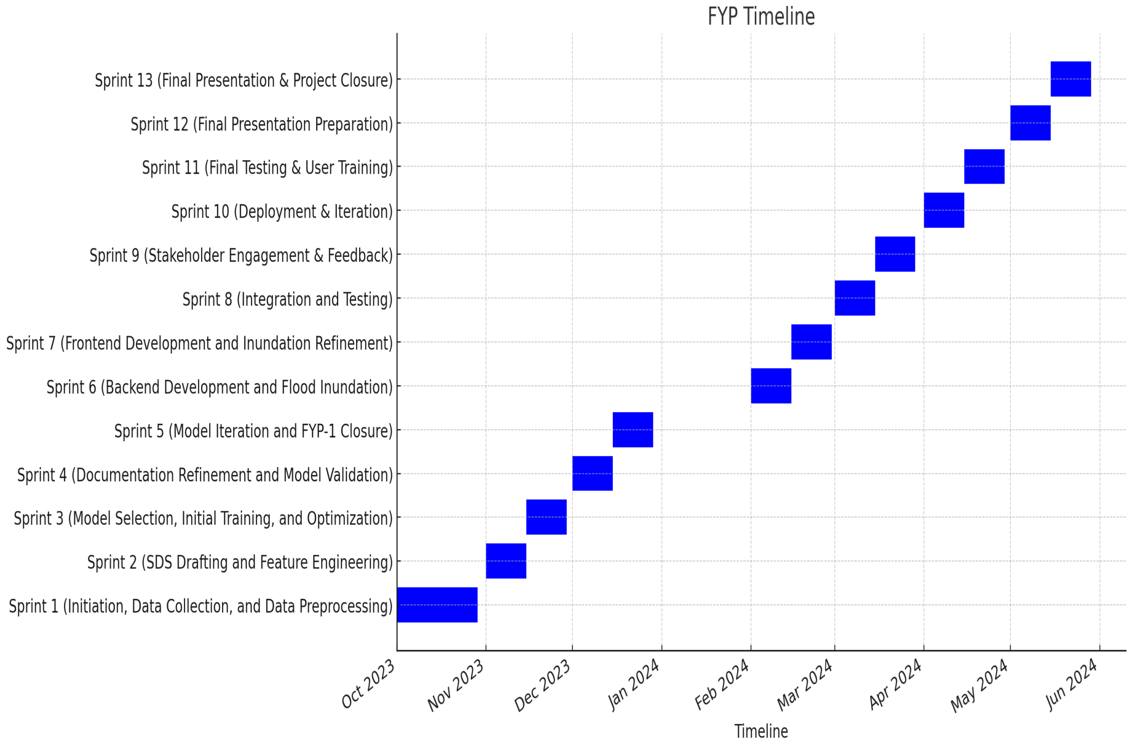

Figure 2: Gantt Chart

9 Expected Outcome

Upon the completion of our Flood Forecasting and River Inundation Modeling project, we anticipate achieving three primary outcomes. First, we'll have a machine learning model tailored to provide accurate streamflow predictions for the selected river gauge stations in Pakistan. This predictive capability will be encapsulated within a user-friendly web application, serving as a real-time tool for stakeholders. The tool will not only offer immediate flood predictions but also visualize potential flood-prone areas through inundation maps. Ultimately, this initiative is poised to equip local communities and authorities with the insights needed

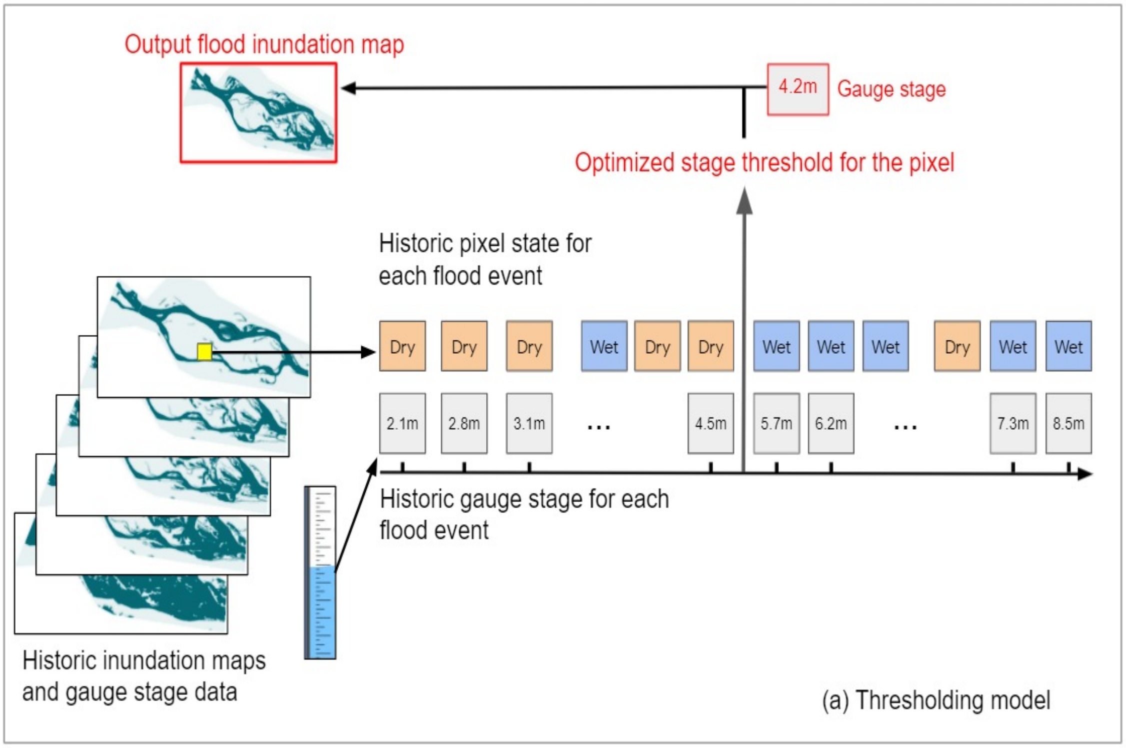

Figure 3: Inundation modeling

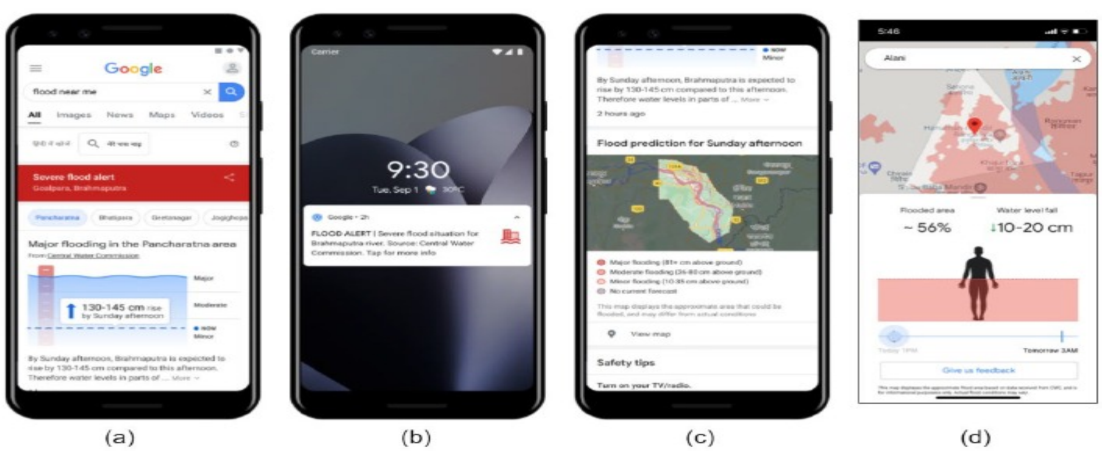

Figure 4: Application for flood alerts

10 Conclusion

This proposal lays the groundwork for a critical initiative aimed at predicting and mitigating flood risks in Pakistan. By harnessing the power of machine learning and leveraging the abundant data collected by the Flood Forecasting Division, our project seeks to transform how we understand, predict, and respond to flood events. The application, once deployed, will offer stakeholders real-time insights and a valuable flood inundation map, bridging the gap between data-driven predictions and actionable responses. As we navigate the challenges ahead, our commitment remains firm: to contribute significantly to the safety and well-being of communities prone to flooding in Pakistan.

References

[1] Iqura Malik, Dipesh Singh Chuphal, Urmin Vegad & Vimal Mishra, **"Was the extreme rainfall that caused the August 2022 flood in Pakistan predictable,"** [Online]. Available: https://www.researchgate.net/publication/369037730_Was_the_extreme_rainfall_that_caused_the_August_2022_flood_in_Pakistan_predictable.March 2023

[2] Muhammad Yaseen, Yasir Latif, Muhammad Waseem, Megersa Kebede Leta, Sohail Abbas, & Haris Akram Bhatti, "**Contemporary Trends in High and Low River Flows in Upper Indus Basin, Pakistan," Water, vol. 14, no. 3.** [Online]. Available: https://www.mdpi.com/2073-4441/14/3/337. 24 January 2022

[3] Ainaa Hanis Zuhairi, Fitri Yakub, Sheikh Ahmad Zaki, & Mohamed Sukri Mat Ali, **"Review of flood prediction hybrid machine learning models using datasets,"** Journal of Environmental Sciences. [Online]. Available: https://iopscience.iop.org/article/10.1088/1755-1315/1091/1/012040/pdf.2022

[4] Fahad Ahmed ,Ho Huu Loc ,Edward Park ,Muhammad Hassan and Panuwat Joyklad **"Comparison of Different Artificial Intelligence Techniques to Predict Floods in Jhelum River, Pakistan,"** [Online]. Available: https://www.mdpi.com/2073-4441/14/21/3533. 3 Nov 2022

[5] Yaseen Muhammad Waseem , Awais Muhammad , Riaz Khuram , Rasheed Muhammad Babar , Waqar Muhammad , Rasheed Sajid **"Artificial Intelligence Based Flood Forecasting for River Hunza at Danyor Station in Pakistan,"** [Online]. Available: https://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-089a05d5-0109-4af3-b218-a4bf06f427a5. 2022

[6] Sella Nevo, Efrat Morin, Adi Gerzi Rosenthal, Asher Metzger, Chen Barshai, Dana Weitzner, Dafi Voloshin, Frederik Kratzert, Gal Elidan, Gideon Dror, Gregory Begelman, Grey Nearing, Guy Shalev, Hila Noga, Ira Shavitt, Liora Yuklea, Moriah Royz, Niv Giladi, Nofar Peled Levi, Ofir Reich, Oren Gilon, Ronnie Maor, Shahar Timnat, Tal Shechter,

Vladimir Anisimov, Yotam Gigi, Yuval Levin, Zach Moshe, Zvika Ben-Haim, Avinatan Hassidim, & Yossi Matias, **"Flood forecasting with machine learning models in an operational framework,"** [Online]. Available: https://arxiv.org/abs/2111.02780. Thu, 4 Nov 2021

[7] Zach Moshe,Asher Metzger,Gal Elidan, Frederik Kratzert, Sella Nevo, and Ran El-Yaniv, **"Hydronets: Leveraging river Structure For Hydrologic Modeling"** Journal of Environmental Sciences. [Online]. Available: https://ai4earthscience.github.io/iclr-2020-workshop/papers/ai4earth04.pdf.2020

[8] Frederik Kratzert, Daniel Klotz, Guy Shalev, Günter Klambauer, Sepp Hochreiter, and Grey Nearing,**"Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets"** [Online]. Available: https://hess.copernicus.org/articles/23/5089/2019/.2019

[9] Yossi Matias **"Tracking our progress on flood forecasting"** [Online]. Available: https://blog.google/technology/ai/tracking-our-progress-on-flood-forecasting/.2019

[10] Sella Nevo **"An Inside Look at Flood Forecasting"** [Online]. Available: https://blog.research.google/2019/09/an-inside-look-at-flood-forecasting.html.2019

[11] Stephen P. Charles, Quan J. Wang, Mobin-ud-Din Ahmad, Danial Hashmi, Andrew Schepen, Geoff Podger, and David E. Robertson **"Seasonal stream flow forecasting in the upper Indus Basin of Pakistan: an assessment of methods"** [Online]. Available: https://hess.copernicus.org/articles/22/3533/2018/. 2018

[12] Yossi Matias **"Keeping people safe with AI-enabled flood forecasting"** [Online]. Available: https://blog.google/products/search/helping-keep-people-safe-ai-enabled-flood-forecasting/. Sep 24, 2018