TOWARDS FLOOD EXTENT FORECASTING: EVALUATING A WEATHER FOUNDATION MODEL AND U-NET FOR FLOOD FORECASTING.

Eric Wanjau & Samuel Maina
Microsoft Research Lab – Africa, Nairobi
{eric.muriithi.22@ucl.ac.uk}, {samuelmaina@microsoft.com}

ABSTRACT

This study explores a data-driven approach that combines flood forcing factors from observation and reanalysis datasets, antecedent flood extent maps, and deep learning to forecast daily flood extents in Rwanda. We extend the architecture used in ClimaX (transformer weather and climate foundation model), investigate its pretrained representations for flood forecasting, and compare performance against a U-Net baseline. Our results demonstrate that a ClimaX variant trained from scratch with a linear projection decoder outperforms the U-Net and other ClimaX variants, highlighting its potential as an effective tool for flood extent forecasting. This work underscores the potential of data-driven deep learning models for flood extent forecasting with implications for improving disaster preparedness and flood risk assessment in vulnerable regions.

1 INTRODUCTION

Floods represent the most widespread and devastating natural disasters, impacting millions annually and causing immense economic losses. In 2022 alone, floods affected over 57 million people globally and resulted in 8,000 fatalities (Fraehr et al., 2023). The frequency and severity of flood-related disasters have more than doubled since 2000 (WMO, 2021), with future impacts expected to worsen due to climate change, population growth, and urbanization (Lee et al., 2023; Nevo et al., 2022). Rwanda is particularly vulnerable due to its mountainous terrain, frequent heavy rainfall, and the impacts of climate change. This, combined with inadequate drainage and deforestation, leads to frequent and severe flooding and landslides, resulting in significant loss of life, displacement, and infrastructure damage. Low and middle-income countries are therefore disproportionately affected due to lack of adequate flood mitigation measures and having densely populated floodplains (Alfieri et al., 2018). As a result, there is urgent need for improved flood forecasting and effective mitigation strategies to minimize the devastating impacts of flooding events.

Flood inundation modeling (FIM), which utilizes hydrodynamic models to estimate critical flood parameters like extent and depth, offers essential insights for scientists and risk managers. Its applications in flood susceptibility mapping and real-time forecasting makes it a valuable tool for flood risk assessment and mitigation. However, large-scale hydrodynamic modeling is constrained by its reliance on many parameters — such as topographic data, initial water levels, resistance factors, and runoff coefficients — which are often unavailable in regions that are poorly monitored, leading to poorly constrained parameters and significant calibration uncertainties. As a result, these physics-based numerical models often fail to capture the non-linear dynamics and complex interactions of flood forcings. Furthermore, the high computational cost of fine-resolution modeling limits their feasibility for large-scale operational applications.

Machine learning methods on the other hand have demonstrated significant potential for flood inundation modeling, providing a computationally efficient alternative to physically-based hydrodynamic models (Kabir et al., 2020; Guo et al., 2021; Liao et al., 2023; Chu et al., 2020; Chang et al., 2018; Sun et al., 2023). However, a key limitation has been the scarcity of accurate flood depth and extent target data, which hinders model development and accuracy (Chu et al., 2020; Guo et al., 2021; Chang et al., 2018). This data limitation has led research to focus primarily on fluvial (river) flooding caused by excessive upstream flow, while pluvial flooding—driven by intense rainfall and poor drainage—has received less attention. As a result, pluvial floods, which are more frequent and can lead to substantial economic losses, remain understudied (Liao et al., 2023). However, the increasing availability of Earth observation data including analysis, forecasts, and climate simulations is opening new avenues for applying machine learning models to flood forecasting (Nearing et al., 2024; Kumar et al., 2023; Kratzert et al., 2019; Brunner et al., 2021) and is mitigating the scarcity of flood depth and extent data (Bonafilia et al., 2020; Misra et al., 2024). These datasets provide extensive spatio-temporal coverage, capturing a broad range of environmental variables that are crucial for understanding and predicting flood events.

In this study, we introduce a transformer-based model to forecast daily flood extent maps in Rwanda using 30 days of past meteorological, land, and geographical forcings. We extend the architecture used in ClimaX, a transformer-based weather and climate foundation model by Nguyen et al. (2023), for flood forecasting and investigate whether ClimaX's learned weather and climate representations can be effectively leveraged for flood prediction. Additionally, we train a U-Net model (Ronneberger et al., 2015) for comparison. Given the increasing need for accurate flood forecasting in a changing climate, our approach aims to enhance disaster preparedness, inform risk management strategies, and support climate resilience efforts.

2 Data

To build effective data-driven models for flood extent forecasting, we need to employ suitable input data that allows the model to learn spatio-temporal patterns, capture complex variable interactions, and generalize to unobserved flood events. Previous studies have identified several key flood-triggering factors which we use in this study. These include:

- **Meteorological factors**. These are weather-related variables that drive atmospheric and hydrological changes, significantly influencing flood occurrence and severity. In this study, we include daily aggregates of total precipitation, 2-m temperature, and surface pressure, all obtained from the ERA5-Land reanalysis dataset (Muñoz-Sabater et al., 2021) and accessed via Google Earth Engine.
- **Geographical factors**. These variables determine how water flows and accumulates in a given area. Elevation and slope, derived from the Shuttle Radar Topography Mission (SRTM) digital elevation dataset (Jarvis et al., 2010), were used in this study and retrieved through Google Earth Engine.
- **Soil and Land use factors**. These variables influence water infiltration and runoff, affecting flood susceptibility. We also included volumetric soil water content from the ERA5-Land reanalysis dataset, also accessed through Google Earth Engine.

To enhance the model's ability to generalize across diverse regions, researchers have incorporated spatial coordinates as additional inputs. This allows the model to establish relationships between input features and their spatial locations, as demonstrated by Sun et al. (2023). The target variable in this study is daily flood extent maps, sourced from Misra et al. (2024). These flood extent maps were generated using a neural network trained on Sentinel-1 SAR imagery, offering nearly a decade of global flood extent data. In this study, all data were standardized to a 1 km resolution ($0.009°$).

3 Methodology

We formulate flood extent prediction as an image-to-image translation task, treating both the input and output as image rasters representing different datasets. This allows for the input data to be mapped into the desired flood extent map using any image architecture such as U-Net (Ronneberger et al., 2015), ResNet(He et al., 2016) or Vision Transformer (Dosovitskiy, 2020). Given the input variables (flood forcing factors + spatial coordinates + antecedent flood extent maps) X_t at a particular time t , the main objective is to predict where a flood will occur at a time t, which can be treated as a binary classification process to label a given pixel with ”flood” or ”non flood”.

To handle the use of multiple flood forcing variables through time, the model needs to be able to combine diverse data types and provide meaningful, physically consistent outputs across space and time. One such available neural network is ClimaX (Nguyen et al., 2023). ClimaX is a weather and climate foundation model built upon Vision Transformers (ViT) and uses *variable tokenization* and *variable aggregation* to handle heterogeneous datasets. Variable Tokenization divides each variable's spatial map ($H \times W$) into $h \times w$ patches, resulting in $V \times h \times w$ patches, where V represents the number of variables. These patches are then embedded into vectors of size D, producing a final output of dimension $V \times h \times w \times D$. Variable aggregation on the other hand uses a learnable query to merge these tokens into a unified representation of size $h \times w \times D$ reducing computational costs and creating unified tokens with universal semantics of all variables involved.

To perform flood extent forecasting, some modifications are made to the original ClimaX architecture. To predict for a specific day, we use a 30-day history of input variables, resulting in an input with dimensions $T \times V \times H \times W$. Each time slice of the input goes through variable tokenization and variable aggregation before additional information such as positional embedding, time history embedding and day of the year embedding are added to the input tokens. The spatio-temporal rich input tokens are then processed by the transformer's attention layers, producing a feature tensor with dimensions $T \times h \times w \times D$, where D represents the embedding size. This tensor is then summarized across its patches using a global average pooling layer, reducing its dimension to $T \times D$. Finally, a cross-attention layer aggregates information across the temporal features and passes this to an MLP decoder which linearly maps the D-dimensional feature vector to a $H \times W$ flood extent map. A sigmoid is then applied to obtain the final flood extent map. These designs are inspired by (Nguyen et al., 2023; Nearing et al., 2024). We refer to this variant as **ClimaX-MLP** (see Figure 2); experiments include training from scratch and fine-tuning a pre-trained ClimaX model (1.40625 °version). In this work, a ViT with 8 attention layers, an embedding size of 1024 and a hidden dimension of 1024×4 was used.

Building on the approach proposed by (Strudel et al., 2021), a second ClimaX variant was developed to generate patch-level class logits. Similar to the first variant, the ClimaX encoder processes input forcing factors resulting in a feature tensor of shape $T \times N \times D$, where $N = h \times w$ represents the number of spatial patches. These features are then aggregated across the temporal dimension using a learnable query, resulting in a spatial feature map $z_L \in \mathbb{R}^{N \times D}$ and passed to a patch-level decoder. The decoder applies a linear layer to the patch encodings to generate class logits for each patch, resulting in $z_{lin} \in \mathbb{R}^{N \times K}$ where K denotes the number of classes. This sequence is subsequently reshaped into a 2D feature map $h \times w$ and bilinealy upsampled to the original image size $H \times W$. A sigmoid is then applied to obtain the final flood extent map. We refer to this version as **ClimaX-Interpolate**.

For a baseline, we considered a U-Net model adapted from (Gupta & Brandstetter, 2022). We trained our models by minimizing a weighted pixel-wise binary cross-entropy loss using the AdamW optimizer at a learning rate of 6e^{-4}, and early stopping to prevent overfitting. Model performance is evaluated using two metrics; **Mean Intersection over Union (mIoU)** which measures the overlap between predicted and actual flood extent maps and **Recall** which assesses the model's ability to correctly identify flooded areas.

4 RESULTS

Table 1 summarizes the results obtained across different model configurations, including a baseline U-Net, ClimaX variants trained from scratch, and fine-tuned versions of ClimaX while Figure 1 provides a qualitative comparison.

Table 1: Performance of various models for flood extent forecasting, evaluated using mIOU and recall metrics.

Model	**mIOU**	**Recall**
Unet	0.5061	0.5144
From scratch ClimaX (ClimaX-MLP)	**0.5292**	**0.5720**
Fine-tune ClimaX (ClimaX-MLP)	0.5002	0.5137
From scratch ClimaX (ClimaX-Interpolate)	0.4986	0.4753
Fine-tune ClimaX (ClimaX-Interpolate)	0.4982	0.4066

5 DISCUSSION AND FUTURE WORK

The results presented in this study highlight the potential of surrogate deep learning models for flood extent forecasting. The ClimaX variant trained from scratch with a linear projection decoder (ClimaX-MLP) achieved superior performance compared to the baseline U-Net, highlighting its ability to effectively learn from diverse input variables and capture complex spatio-temporal dy-

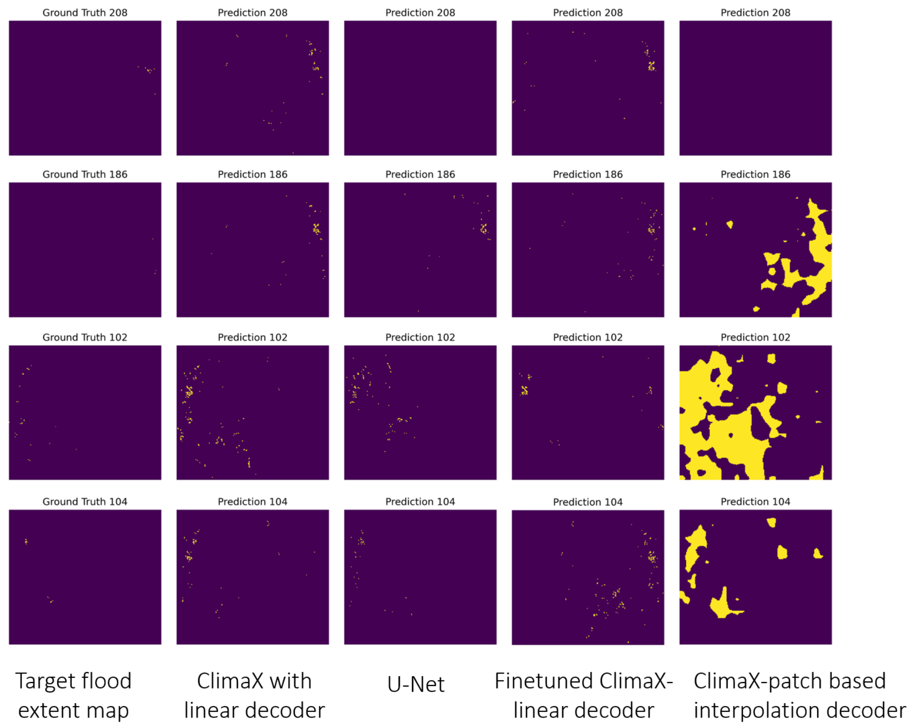

Figure 1: Sample flood extent predictions generated by all evaluated models. The figure qualitatively highlights differences in model performance, particularly in capturing fine-grained flood boundaries and spatial extent.

namics associated with flood events. This indicates that the ClimaX architecture, with its inherent ability to model weather and climate data, can be effectively adapted for flood forecasting tasks. Although the linear projection head decoder performed best, the baseline U-Net remains competitive and could offer a simpler alternative to transformer-based approaches. Future work will involve incorporating more flood forcing factors such as land use factors, hyrological factors such as river discharge, and vegetation and water indexes to improve models' performance. The superior performance of the ClimaX-MLP model compared to the ClimaX-Interpolate model suggests that bilinear interpolation might not have been sufficient to capture spatial interactions or accurately delineate flood boundaries as shown in Figure 1, suggesting further decoder optimization for this task. It is noteworthy that the performance of the fine-tuned ClimaX models was not consistently better than the models trained from scratch which could be attributed to several factors. ClimaX, trained on coarse global spatial resolution of 1.40625°(128×256 grid points), may have had difficulty capturing the fine-grained spatial details crucial for accurate flood prediction at 0.009°in Rwanda. Future work will involve introducing an intermediate pretraining step where the model is exposed to high resolution regional flood datasets to bridge the gap between coarse global patterns and localized flood dynamics.

6 Conclusion

This study demonstrates the potential of data-driven deep learning models for flood extent forecasting, focusing on adapting the ClimaX architecture and a U-Net, to predict flood extent maps in Rwanda. By leveraging the growing availability of observational and reanalysis data, along with advancements in deep learning, this work establishes a foundation for improving flood forecasting models.

References

Lorenzo Alfieri, Sagy Cohen, John Galantowicz, Guy JP Schumann, Mark A Trigg, Ervin Zsoter, Christel Prudhomme, Andrew Kruczkiewicz, Erin Coughlan de Perez, Zachary Flamig, et al. A global network for operational flood risk reduction. *Environmental science & policy*, 84:149–158, 2018.

Derrick Bonafilia, Beth Tellman, Tyler Anderson, and Erica Issenberg. Sen1floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 210–211, 2020.

Manuela I Brunner, Louise Slater, Lena M Tallaksen, and Martyn Clark. Challenges in modeling and predicting floods and droughts: A review. *Wiley Interdisciplinary Reviews: Water*, 8(3):e1520, 2021.

Li-Chiu Chang, Mohd Zaki M Amin, Shun-Nien Yang, and Fi-John Chang. Building ann-based regional multi-step-ahead flood inundation forecast models. *Water*, 10(9):1283, 2018.

Haibo Chu, Wenyan Wu, Quan J Wang, Rory Nathan, and Jiahua Wei. An ann-based emulation modelling framework for flood inundation modelling: Application, challenges and future directions. *Environmental Modelling & Software*, 124:104587, 2020.

Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Niels Fraehr, Quan J Wang, Wenyan Wu, and Rory Nathan. Supercharging hydrodynamic inundation models for instant flood insight. *Nature Water*, 1(10):835–843, 2023.

Zifeng Guo, Joao P Leitao, Nuno E Simões, and Vahid Moosavi. Data-driven flood emulation: Speeding up urban flood predictions by deep convolutional neural networks. *Journal of Flood Risk Management*, 14(1):e12684, 2021.

Jayesh K Gupta and Johannes Brandstetter. Towards multi-spatiotemporal-scale generalized pde modeling. *arXiv preprint arXiv:2209.15616*, 2022.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

A Jarvis, HI Reuter, A Nelson, and E Guevara. Hole-filled srtm for the globe version 4. cgiar-csi srtm 90m database. 2008, 2010.

Syed Kabir, Sandhya Patidar, Xilin Xia, Qiuhua Liang, Jeffrey Neal, and Gareth Pender. A deep convolutional neural network model for rapid prediction of fluvial flood inundation. *Journal of Hydrology*, 590:125481, 2020.

Frederik Kratzert, Daniel Klotz, Mathew Herrnegger, Alden K Sampson, Sepp Hochreiter, and Grey S Nearing. Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, 55(12):11344–11354, 2019.

Vijendra Kumar, Hazi Md Azamathulla, Kul Vaibhav Sharma, Darshan J Mehta, and Kiran Tota Maharaj. The state of the art in deep learning applications, challenges, and future prospects: A comprehensive review of flood forecasting and management. *Sustainability*, 15(13):10543, 2023.

Hoesung Lee, Katherine Calvin, Dipak Dasgupta, Gerhard Krinner, Aditi Mukherji, Peter Thorne, Christopher Trisos, José Romero, Paulina Aldunce, Ko Barret, et al. Ipcc, 2023: Climate change 2023: Synthesis report, summary for policymakers. contribution of working groups i, ii and iii to the sixth assessment report of the intergovernmental panel on climate change [core writing team, h. lee and j. romero (eds.)]. ipcc, geneva, switzerland. 2023.

Yaoxing Liao, Zhaoli Wang, Xiaohong Chen, and Chengguang Lai. Fast simulation and prediction of urban pluvial floods using a deep convolutional neural network model. *Journal of Hydrology*, 624:129945, 2023.

Amit Misra, Kevin White, Simone Fobi Nsutezo, William Straka, and Juan Lavista. Mapping global floods with 10 years of satellite radar data. *arXiv preprint arXiv:2411.01411*, 2024.

Joaquín Muñoz-Sabater, Emanuel Dutra, Anna Agustí-Panareda, Clément Albergel, Gabriele Arduini, Gianpaolo Balsamo, Souhail Boussetta, Margarita Choulga, Shaun Harrigan, Hans Hersbach, et al. Era5-land: A state-of-the-art global reanalysis dataset for land applications. *Earth system science data*, 13(9):4349–4383, 2021.

Grey Nearing, Deborah Cohen, Vusumuzi Dube, Martin Gauch, Oren Gilon, Shaun Harrigan, Avinatan Hassidim, Daniel Klotz, Frederik Kratzert, Asher Metzger, et al. Global prediction of extreme floods in ungauged watersheds. *Nature*, 627(8004):559–563, 2024.

Sella Nevo, Efrat Morin, Adi Gerzi Rosenthal, Asher Metzger, Chen Barshai, Dana Weitzner, Dafi Voloshin, Frederik Kratzert, Gal Elidan, Gideon Dror, et al. Flood forecasting with machine learning models in an operational framework. *Hydrology and Earth System Sciences*, 26(15): 4013–4032, 2022.

Tung Nguyen, Johannes Brandstetter, Ashish Kapoor, Jayesh K Gupta, and Aditya Grover. Climax: A foundation model for weather and climate. *arXiv preprint arXiv:2301.10343*, 2023.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pp. 234–241. Springer, 2015.

Robin Strudel, Ricardo Garcia, Ivan Laptev, and Cordelia Schmid. Segmenter: Transformer for semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 7262–7272, 2021.

Alexander Y Sun, Zhi Li, Wonhyun Lee, Qixing Huang, Bridget R Scanlon, and Clint Dawson. Rapid flood inundation forecast using fourier neural operator. In *Proceedings of the IEEE/CVF international conference on computer Vision*, pp. 3733–3739, 2023.

WMO. 2021 state of climate service (wmo-no. 1278), 2021.

A APPENDIX

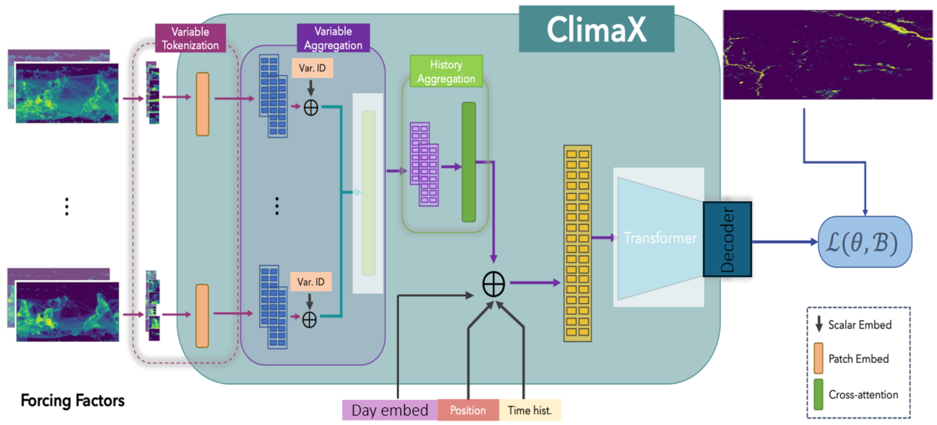

Figure 2: Variant of ClimaX (Nguyen et al., 2023) architecture used in this study for flood extent forecasting.