



# Explainable Deep Semantic Segmentation for Flood Inundation Mapping with Class Activation Mapping Techniques

Jacob Sanderson<sup>1</sup><sup>a</sup>, Hua Mao<sup>1</sup>, Naruephorn Tengtrairat<sup>2</sup>, Raid Rafi Al-Nima<sup>3</sup> and Wai Lok Woo<sup>1</sup><sup>b</sup>

<sup>1</sup>Department of Computer and Information Sciences, Northumbria University, Newcastle Upon Tyne, U.K.

<sup>2</sup>School of Software Engineering, Payap University, Chiang Mai, Thailand

<sup>3</sup>Technical Engineering College, Northern Technical University, Mosul, Iraq

**Keywords:** Deep Learning, Semantic Segmentation, Explainable Artificial Intelligence, Flood Inundation Mapping, Satellite Imagery.

**Abstract:** Climate change is causing escalating extreme weather events, resulting in frequent, intense flooding. Flood inundation mapping is a key tool in combating these flood events, by providing insight into flood-prone areas, allowing for effective resource allocation and preparation. In this study, a novel deep learning architecture for the generation of flood inundation maps is presented and compared with several state-of-the-art models across both Sentinel-1 and Sentinel-2 imagery, where it demonstrates consistently superior performance, with an Intersection Over Union (IOU) of 0.5902 with Sentinel-1, and 0.6984 with Sentinel-2 images. The importance of this versatility is underscored by visual analysis of images from each satellite under different weather conditions, demonstrating the differing strengths and limitations of each. Explainable Artificial Intelligence (XAI) is leveraged to interpret the decision-making of the model, which reveals that the proposed model not only provides the greatest accuracy but exhibits an improved ability to confidently identify the most relevant areas of an image for flood detection.


## 1 INTRODUCTION


In recent years, climate change is causing rising sea-levels (Rosier et al., 2023) and extreme weather events, which is resulting in a rapid increase in the occurrence and intensity of flooding (Schreider et al., 2000), with devastating consequences for communities worldwide. Flood inundation mapping has a pivotal role in protecting affected communities by providing insights into potential flood extents, identifying at-risk areas, and enabling effective resource allocation (Sahana and Patel, 2019).

Given the rise in the frequency of floods, the timeliness of flood inundation map generation is of growing importance. Traditional methods often rely on manual preprocessing of images making them labour-intensive and time-consuming to implement (Landuyt et al., 2018). As computational power has increased, deep learning has emerged as an effective technique, with Convolutional Neural Networks (CNNs), providing most effective (Tavus et al., 2022). CNNs, of-

ten in a fully convolutional network (FCN) architecture, segment images into water and non-water pixels, outperforming classical machine learning methods (Gebrehiwot and Hashemi-Beni, 2020). Increasingly, models are employing encoder-decoder architectures, with U-Net and DeepLab variants being popular choices (Katiyar et al., 2021; Ghosh et al., 2022; Helleis et al., 2022; Muszynski et al., 2022; Li and Demir, 2023; Paul and Ganju, 2021; Yuan et al., 2021; Sanderson et al., 2023a; Sanderson et al., 2023b).

The primary drawback of employing deep learning lies in its lack of transparency, with deep neural networks often being referred to as 'black box' models. While these models can achieve impressive accuracy in generating flood inundation maps, understanding the rationale behind their decisions can be challenging. This raises concerns about the trustworthiness of these models, as it becomes difficult to verify that their decisions are not influenced by bias or error. To address this issue, explainable artificial intelligence (XAI) has emerged as a solution. XAI aims to provide human-understandable explanations for the decisions made by deep neural networks and other black box models (Mirzaei et al., 2024). In the

<sup>a</sup> <https://orcid.org/0009-0002-5724-6637>

<sup>b</sup> <https://orcid.org/0000-0002-8698-7605>

context of flood inundation mapping, XAI could offer valuable insights into the factors influencing flood extent prediction, enhancing the trust of stakeholders. Despite this, as far as the authors are aware, research in XAI for flood modelling is limited. Existing studies mainly focus on explaining classical machine learning algorithms using model-agnostic XAI techniques like LIME and SHAP (Prasanth Kadiyala and Woo, 2021; Pradhan et al., 2023). The use of classical machine learning limits their ability to capture complex relationships and results in lower accuracy. The model-agnostic algorithms used in these studies are unable to access the inner workings of the predictive models, limiting the insights they can provide. To address this, model-specific XAI algorithms exist, which can provide more fine-grained explanations, tailored to a specific model. To explain CNN decision-making, techniques like class activation mapping (CAM) (Zhou et al., 2016) and a growing number of modified CAM techniques are employed (Chattopadhyay et al., 2018; Selvaraju et al., 2017; Fu et al., 2020; Wang et al., 2020; Muhammad and Yeasin, 2020; Draeos and Carin, 2020). They visualize important features without significantly affecting model performance. The majority of these techniques are well-demonstrated in image classification, however in semantic segmentation gradient weighted class activation mapping (Grad-CAM) (Selvaraju et al., 2017) is by far the most well-used. In image classification, it has been noted that Grad-CAM has a substantial shortcoming, in the gradient averaging step often regions are highlighted that the model did not use, leading to an unfaithful interpretation. To overcome this, methods such as high-resolution class activation mapping (HiResCAM) (Draeos and Carin, 2020) have been proposed to provide a more faithful explanation, however, the development of this method for semantic segmentation has yet to be investigated, as far as the authors are aware.

In addition to model selection, the choice of input data type is a crucial consideration in developing flood inundation maps. Two of the primary sensor types used on board satellites are synthetic aperture radar (SAR), and optical. The European Space Agency provides free access to images taken by their Sentinel missions, making them a popular and accessible choice for researchers and practitioners in the field. Within the Sentinel missions, Sentinel-1, which makes use of a SAR instrument, and Sentinel-2 equipped with a multi-spectral optical instrument (MSI) are the most commonly used in mapping flood extent. Each of these instruments has its own distinct advantages and limitations. SAR can penetrate through cloud cover and provides its own light source,

meaning that images can be taken in any weather or light conditions. However, it is susceptible to speckle noise, which can make detection of the true signal more challenging. In contrast, MSI delivers imagery with high spatial and spectral resolution, providing a greater wealth of information for land cover and water classification, but is not able to penetrate clouds as effectively (Konapala et al., 2021).

This work aims to develop a novel architecture for flood inundation mapping and demonstrate its effectiveness through comparative analysis with several state-of-the-art models. The influence of input data is also assessed by applying the proposed model to Sentinel-1 and Sentinel-2, analysing both the overall quantitative performance, as well as visually assessing the quality of the generated maps in both clear and cloud-covered conditions. Finally, XAI is employed to interpret the decision-making of the models, as well as to provide deeper insight into how the different input data types influence the behaviour of the model. This study will explore the suitability of HiResCAM as an explainability method for semantic segmentation, in comparison with results from Grad-CAM.

## 2 METHODOLOGY

### 2.1 Dataset

The Sen1Floods11 dataset (Bonafilia et al., 2020) is used in this study, which consists of 446 hand-labelled images from each of the Sentinel-1 and Sentinel-2 satellites, with a total of 892 images from 11 global flood events in Bolivia, Ghana, India, Cambodia, Nigeria, Pakistan, Paraguay, Somalia, Spain, Sri Lanka, and the USA. These flood events were sampled as both Sentinel-1 and coincident Sentinel-2 images were available within a maximum of 2 days of each other.

The Sentinel-1 images are taken with a C-band SAR instrument in the interferometric wide swath (IW) mode, which allows a wide area to be covered, making it suitable for large-scale flood mapping. The ground resolution of these images is 10m. The Sentinel-1 satellite offers different polarization modes, where the images in this dataset are taken in vertical transmit, vertical and horizontal receive (VV + VH), offering enhanced information for improved water detection, with relatively low computational cost. The Sentinel-2 images are taken with an MSI with a 290km field of view, capturing 13 spectral bands in the visible light, near-infrared and shortwave infrared spectrum. The ground resolution of these bands varies, so the images have been re-sampled to

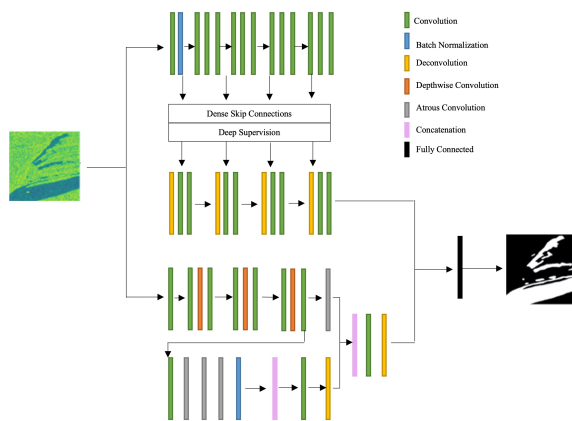


Figure 1: Architecture diagram of the proposed model.

10m for each to ensure consistency in the analysis.

## 2.2 Proposed Model

In flood inundation mapping, conventional fully convolutional neural networks (CNNs) have been widely used but face challenges such as information loss and low-resolution predictions. To address these issues, encoder-decoder architectures are used by most recent works in segmentation, which are able to capture high-level semantic information while preserving spatial detail much more effectively.

The proposed model takes inspiration from this and adopts a dual encoder-decoder architecture, as shown in Figure 1. One encoder is a powerful CNN pre-trained on ImageNet, while the other is a lightweight CNN, ensuring a balance between accuracy and computational efficiency. By employing two base models, the robustness of the final model is improved by providing more diverse input representations to the final classifier and enabling the compensation for limitations in one model with the strengths of the other and vice versa.

The powerful encoder employs convolutional and identity blocks, mitigating issues like vanishing gradients. The lightweight encoder features bottleneck blocks for computational efficiency. In the decoder, transposed convolutions perform upsampling to recover spatial information and performance-enhancing features, including dense skip pathways and deep supervision are integrated to improve the model's accuracy and convergence speed.

Atrous convolution is introduced to the lightweight encoder to address spatial resolution reduction. It enhances the field of view without increasing computational costs. Additionally, spatial pyramid pooling and depthwise separable convolution are employed in its decoder module to capture multi-scale information while managing

computational complexity.

## 2.3 Class Activation Mapping Techniques

CAM is a widely used XAI technique in CNN interpretation. In CAM, the architecture of a CNN is modified by replacing fully connected layers with global average pooling layers, to provide class-specific feature maps, showing the localization of a CNN (Zhou et al., 2016). This modification requires the retraining of the model and also limits the variety of CNN architectures that it will perform well with.

To overcome this, several extended versions of CAM have been introduced, most notably, Grad-CAM (Selvaraju et al., 2017). In Grad-CAM, the gradients of a target concept as they flow into a target layer are used to produce the class activation map, by first finding the gradient for the target class with respect to the feature map activations, then global average pooling to give the importance weight for each neuron. Following this, a weighted combination of the forward activation maps is computed and ReLU is applied to ensure only the pixels with a positive influence on the target class are highlighted.

Due to the gradient averaging step, it has been identified that Grad-CAM does not always reflect the locations of the image that the model used for training, and so can produce misleading explanations, often resulting in smoother heatmaps that suggest a larger area of the image was considered in the model's decision making. To overcome this limitation, HiResCAM (Draeos and Carin, 2020) generates its explanations through element-wise multiplication of the gradients with the feature maps, then summing over the feature dimensions.

## 2.4 State-of-the-Art Models

To best demonstrate the superior performance of the proposed model, a comparative analysis is conducted with two state-of-the-art models, DeepLabV3+ and U-Net++, each with both a ResNet50 and MobileNet\_V2 back-bone. These models have been selected as they remain two of the most sophisticated and commonly used methods for semantic segmentation.

DeepLabV3+ incorporates atrous convolution and spatial pyramid pooling to effectively handle segmentation tasks involving multiple objects at differing scales. A streamlined decoder is then employed, offering control of the density of the encoder features, as well as restoring the precise object boundaries (Chen et al., 2018). U-Net++ involves a CNN

as the encoder for feature extraction, followed by up-sampling by the decoder to increase the output resolution. Dense skip connections are employed to reduce the semantic gap between the encoder and decoder, as well as deep supervision to facilitate versatile operation modes (Zhou et al., 2018).

ResNet50 consists of residual blocks, each involving a sequence of 3x3 convolution, batch normalization and ReLU activation. Skip connections are employed after each block enabling certain layers to be bypassed, addressing issues related to model degradation and vanishing gradients (He et al., 2016).

MobileNet\_v2 is a streamlined architecture featuring depthwise separable convolution, which can reduce computational complexity while maintaining high performance. An inverted residual block with linear bottle-neck is incorporated which takes a compressed low-dimensional representation and expands it into a higher dimension, applying lightweight depthwise separable convolution, before projecting the features back to a low-dimensional representation. This minimizes the memory requirements through the reduction of parameters (Sandler et al., 2019).

## 2.5 Experiments

The Pytorch deep learning framework is used to develop the models used in this study, accelerated with the use of NVIDIA A100 graphics processing unit (GPU), accessed through the cloud. The images and labels in the dataset are split into training, validation and testing sets, with 251 training, 89 validation and 90 testing samples for each satellite. The training images are augmented through random flipping and cropping of the training images create more variation in the data, improving the robustness of the model. The validation and testing images are cropped at a fixed point to ensure consistency in visualizations. All images are normalized by the mean and standard deviation, ensuring all values are within the same scale for effective optimization. A batch size of 16 is used for training the models, with the Cross Entropy Loss function, which measures the similarity between the true and predicted values. The Sen1Floods11 dataset is imbalanced, with a higher number of non-water pixels. To address this, a weighting of 8:1 is applied within the loss function, placing 8 times the importance on the water class to ensure the model places more importance on correctly identifying these pixels.

The optimizer AdamW is used with a learning rate of  $5e-4$ . AdamW is a stochastic gradient descent-based optimization method, where the first-order and second-order moments are adaptively estimated, with

an improved weight decay method through decoupling the weight decay from the optimization steps (Loshchilov and Hutter, 2019). Cosine Annealing Warm Restarts is used to schedule the learning rate, where the learning rate is decreased from a high to a low value, and then restarts at a previous good weight (Loshchilov and Hutter, 2017).

The model was trained over 250 epochs, being evaluated against the accuracy, loss, F1 score, and Intersection over Union (IOU) on both the training, then the validation set to ensure that the model can generalise to unseen data. The model with the best validation IOU is saved and evaluated on the testing data, against the same metrics.

## 3 EXPERIMENTAL RESULTS

### 3.1 Quantitative Results

The performance of each model against the metrics described in section 3.5 is given in Table 1. These results highlight the proposed model's superiority for both image types. In the case of Sentinel-1 images, U-Net++ and DeepLabV3+ yield similar IOU values, with the U-Net++ model employing the ResNet50 backbone and the DeepLabV3+ model with the MobileNet\_V2 backbone slightly surpassing the other two. Furthermore, the DeepLabV3+ model with the MobileNet\_V2 backbone excels in the F1 score. When analysing Sentinel-2 images, performance improves compared to Sentinel-1. Both U-Net++ models outperform the DeepLabV3+ models. It is important to note that performance varies between the two backbones, with ResNet50 excelling in conjunction with U-Net++ with the Sentinel-1 images, while MobileNet\_V2 performs slightly better in all other cases.

### 3.2 Visualizations

Figures 2 to 6 show visualizations of the flood inundation maps generated by each model, and the explanations provided by both Grad-CAM and HiResCAM, with each depicting an example image in clear conditions and a cloud-covered image for both Sentinel-1 and Sentinel-2. These images illustrate that the models achieving the best performance across evaluation metrics, especially IOU, generate highly detailed segmentation masks, resulting in more precise maps. Conversely, models with poorer performance produce coarser masks, featuring an increased number of false positive pixel classifications. Notably, the proposed model generates inundation maps that are



Table 1: Accuracy, loss, F1 score and IOU for each model.

Model	Accuracy	Loss	F1 Score	IOU
Sentinel-1				
U-Net++ ResNet50	0.9413	0.3054	0.6127	0.5642
U-Net++ MobileNet_V2	0.9400	0.2907	0.6070	0.5440
DeepLabV3+ ResNet50	0.9415	0.2740	0.6444	0.5464
DeepLabV3+ MobileNet_V2	0.9371	0.3099	0.6372	0.5660
Proposed Model	0.9730	0.2390	0.7327	0.5902
Sentinel-2				
U-Net++ ResNet50	0.9382	0.1524	0.7081	0.6514
U-Net++ MobileNet_V2	0.9468	0.1345	0.7053	0.6583
DeepLabV3+ ResNet50	0.9525	0.1730	0.7162	0.6427
DeepLabV3+ MobileNet_V2	0.9412	0.1917	0.7209	0.6456
Proposed Model	0.9691	0.1129	0.7894	0.6984

more closely aligned with the ground truth than any other model.

Sentinel-1 demonstrates overall modest performance, particularly struggling in areas with finer detail. In contrast, Sentinel-2, while presenting more distinct flood maps in clear conditions, faces significant challenges in cloud-covered scenarios, where the maps are more poorly defined. Notably, the image taken in clear conditions includes some background areas which appear similar to the water pixels in the Sentinel-1 image, while in the Sentinel-2 image, there is more clear contrast between the two, leading to the Sentinel-1 trained models either falsely classifying background areas as water, or underpredicting the water covered areas.

The class activation maps visualise the important features for each model’s decision-making and demonstrate that the proposed model not only generated the most accurate flood inundation maps, but also made the most appropriate decisions, placing higher confidence on the most relevant pixels, and lower importance on the non-flooded pixels than any other model. It is also evident from these images that Grad-CAM does produce a smoothed class activation map, highlighting a larger area as important, which aligns with the assertion that its gradient averaging causes unfaithful interpretation. HiResCAM provides a class activation map equally understandable to humans as Grad-CAM but presents a more faithful representation of the model’s decision-making process.

### 3.3 Discussion

The model proposed in this work excels in generating flood maps from both Sentinel-1 and Sentinel-2 imagery, with a higher level of accuracy than any of the state-of-the-art models. Notably, there are performance disparities observed between the models

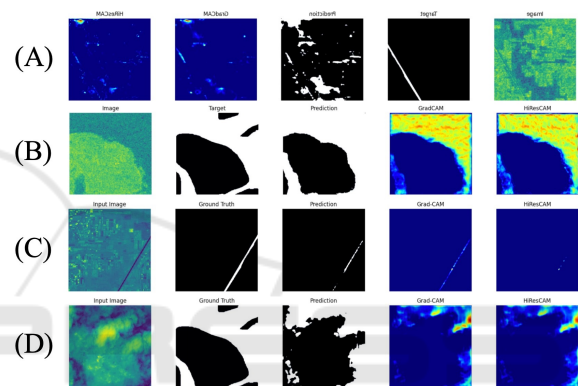


Figure 2: Visualizations generated by U-Net++ with ResNet50 backbone for (A) a Sentinel-1 image in clear conditions, (B) a Sentinel-1 image in cloud-covered conditions, (C) a Sentinel-2 image in clear conditions and (D) a Sentinel-2 image in cloud-covered conditions.

when trained with Sentinel-1 and Sentinel-2, where DeepLabV3+ with MobileNet\_V2 exhibited the highest IOU in Sentinel-1 imagery, while U-Net++ with MobileNet\_V2 was superior with Sentinel-2. The proposed model, outperforming all state-of-the-art models across both image types, stands out as the most versatile option. The visual analysis of the flood inundation maps demonstrates how crucial this versatility is, as each satellite image type demonstrates superior performance in certain conditions, so the ability to easily change the data type without needing to change the model is beneficial. With Sentinel-1 images, the models struggle to detect finer details and frequently overpredict water-covered areas, resulting in a coarser inundation map. Conversely, with Sentinel-2 images the fine detail is much more easily detected, however, cloud coverage significantly impacts performance, resulting in a map with insufficient detail.

The class activation maps provide further insights into the superior performance of the proposed model.

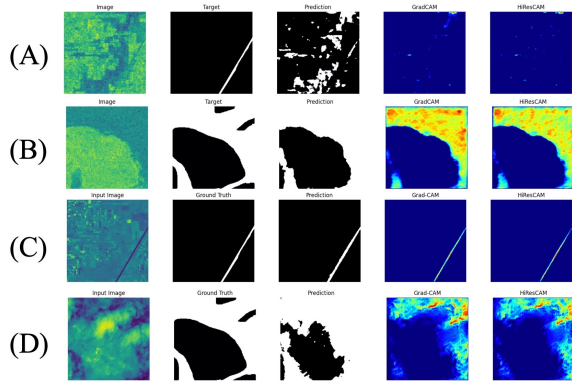


Figure 3: Visualizations generated by U-Net++ with MobileNet\_V2 backbone for (A) a Sentinel-1 image in clear conditions, (B) a Sentinel-1 image in cloud-covered conditions, (C) a Sentinel-2 image in clear conditions and (D) a Sentinel-2 image in cloud-covered conditions.

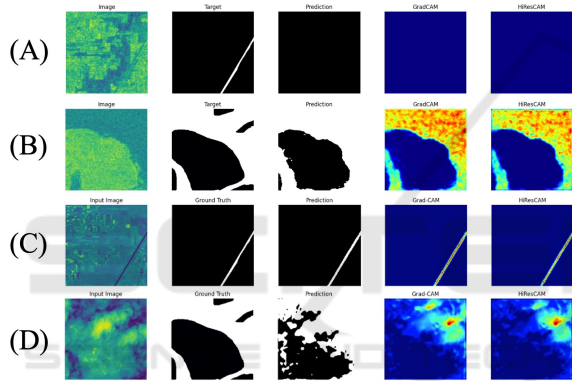


Figure 4: Visualizations generated by DeepLabV3+ with ResNet50 backbone for (A) a Sentinel-1 image in clear conditions, (B) a Sentinel-1 image in cloud-covered conditions, (C) a Sentinel-2 image in clear conditions and (D) a Sentinel-2 image in cloud-covered conditions.

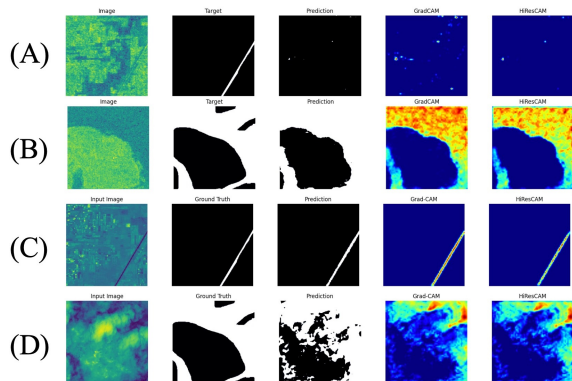


Figure 5: Visualizations generated by DeepLabV3+ with MobileNet\_V2 backbone for (A) a Sentinel-1 image in clear conditions, (B) a Sentinel-1 image in cloud-covered conditions, (C) a Sentinel-2 image in clear conditions and (D) a Sentinel-2 image in cloud-covered conditions.

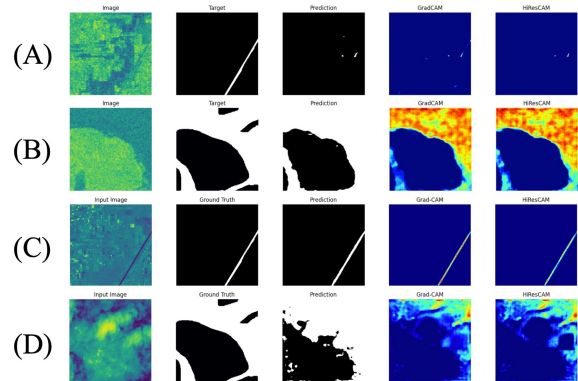


Figure 6: Visualizations generated by the proposed model for (A) a Sentinel-1 image in clear conditions, (B) a Sentinel-1 image in cloud-covered conditions, (C) a Sentinel-2 image in clear conditions and (D) a Sentinel-2 image in cloud-covered conditions.

While many of the flood inundation map visualizations, particularly the cloud-covered Sentinel-1 image, look similar across each model, the class activation maps reveal discrepancies in how the models made the decisions. For instance, the U-Net++ models both have a significantly lower magnitude of importance placed on the relevant pixels for this image. However, they place a higher magnitude of importance on the irrelevant pixels for the Sentinel-1 image in clear conditions. The proposed model consistently assigns a higher magnitude of importance to the flooded areas, and a lower magnitude to the non-flooded areas, demonstrating that it has more effectively learned to delineate floods with greater confidence than any of the state-of-the-art models. This is a crucial consideration for the employment of a model in a flood inundation mapping system, as it shows that the model has more effectively learned how to identify flooded areas, so can be better trusted to correctly identify floods in practice.

It was observed that the class activation maps generated with the use of Grad-CAM are smoothed due to the gradient averaging step, bringing attention to a larger area of the image than the model truly focuses on. This brings into question the faithfulness of the interpretation of the models. The explanations provided by Grad-CAM generally present a more favourable view of the models' performance than HiResCAM, potentially leading to misplaced trust in the models' abilities. This could have potentially catastrophic implications for flood mapping, as this could lead to inaccurate flood maps being generated, leading to inappropriate resource allocation and risk assessment.

## 4 CONCLUSIONS

This study addresses the urgent need for timely and accurate flood inundation mapping in the face of increasing climate-induced challenges. We introduce a novel dual encoder-decoder architecture that consistently demonstrates superiority over state-of-the-art models across both Sentinel-1 and Sentinel-2 images, as evidenced by comprehensive quantitative and visual analyses. The versatility of this model is crucial, highlighted through comparative analyses of each satellite image under different conditions, revealing strengths and limitations in various scenarios.

XAI is leveraged to better understand the decision-making process of these models. It is shown that not only is the proposed model the most accurate, but it has also learned to detect flooded areas more effectively with greater confidence, showcasing its improved trustworthiness for practical applications.

Despite the success of the proposed model, further refinement techniques should be incorporated in the future to enhance segmentation results, such as conditional or Markov random fields. Attention mechanisms have also demonstrated superior results in a range of computer vision tasks, particularly channel and spatial attention. The incorporation of these techniques can enhance the interpretability of the model by highlighting the regions that the model paid attention to, without the need for additional post-hoc XAI algorithms.

The use of additional data is likely to improve the results of the work. Deep learning models continue to improve as more data is added, allowing them to learn more complex feature representations more effectively. The inclusion of different data types, such as Digital Elevation Models (DEM) and Light Detection and Ranging (LiDAR), can provide more detailed information about the topography of an area. This enables the models to generate more precise flood maps, as well as more detailed explanations through XAI regarding how environmental factors impact flood inundation.

## REFERENCES

- Bonafilia, D., Tellman, B., Anderson, T., and Issenberg, E. (2020). 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*, pages 210–211.
- Chattopadhyay, A., Sarkar, A., Howlader, P., and Balasubramanian, V. N. (2018). Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE winter conference on applications of computer vision (WACV)*, pages 839–847. IEEE.
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818.
- Draeos, R. L. and Carin, L. (2020). Use hirescam instead of grad-cam for faithful explanations of convolutional neural networks. *arXiv preprint arXiv:2011.08891*.
- Fu, R., Hu, Q., Dong, X., Guo, Y., Gao, Y., and Li, B. (2020). Axiom-based grad-cam: Towards accurate visualization and explanation of cnns. *arXiv preprint arXiv:2008.02312*.
- Gebrehiwot, A. and Hashemi-Beni, L. (2020). Automated inundation mapping: comparison of methods. In *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, pages 3265–3268. IEEE.
- Ghosh, B., Garg, S., and Motagh, M. (2022). Automatic flood detection from sentinel-1 data using deep learning architectures. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3:201–208.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- Helleis, M., Wieland, M., Krullikowski, C., Martinis, S., and Plank, S. (2022). Sentinel-1-based water and flood mapping: Benchmarking convolutional neural networks against an operational rule-based processing chain. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:2023–2036.
- Katiyar, V., Tamkuan, N., and Nagai, M. (2021). Near-real-time flood mapping using off-the-shelf models with sar imagery and deep learning. *Remote Sensing*, 13(12):2334.
- Konapala, G., Kumar, S. V., and Ahmad, S. K. (2021). Exploring sentinel-1 and sentinel-2 diversity for flood inundation mapping using deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 180:163–173.
- Landuyt, L., Van Wesemael, A., Schumann, G. J.-P., Hostache, R., Verhoest, N. E., and Van Coillie, F. M. (2018). Flood mapping based on synthetic aperture radar: An assessment of established approaches. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2):722–739.
- Li, Z. and Demir, I. (2023). U-net-based semantic classification for flood extent extraction using sar imagery and gee platform: A case study for 2019 central us flooding. *Science of The Total Environment*, 869:161757.
- Loshchilov, I. and Hutter, F. (2017). Sgdr: Stochastic gradient descent with warm restarts.
- Loshchilov, I. and Hutter, F. (2019). Decoupled weight decay regularization.

- Mirzaei, S., Mao, H., Al-Nima, R. R. O., and Woo, W. L. (2024). Explainable ai evaluation: A top-down approach for selecting optimal explanations for black box models. *Information*, 15(1).
- Muhammad, M. B. and Yeasin, M. (2020). Eigen-cam: Class activation map using principal components. In *2020 international joint conference on neural networks (IJCNN)*, pages 1–7. IEEE.
- Muszynski, M., Hölzer, T., Weiss, J., Fraccaro, P., Zortea, M., and Brunschwiler, T. (2022). Flood event detection from sentinel 1 and sentinel 2 data: Does land use matter for performance of u-net based flood segmenters? In *2022 IEEE International Conference on Big Data (Big Data)*, pages 4860–4867. IEEE.
- Paul, S. and Ganju, S. (2021). Flood segmentation on sentinel-1 sar imagery with semi-supervised learning. *arXiv preprint arXiv:2107.08369*.
- Pradhan, B., Lee, S., Dikshit, A., and Kim, H. (2023). Spatial flood susceptibility mapping using an explainable artificial intelligence (xai) model. *Geoscience Frontiers*, 14(6):101625.
- Prasanth Kadiyala, S. and Woo, W. L. (2021). Flood prediction and analysis on the relevance of features using explainable artificial intelligence. In *2021 2nd Artificial Intelligence and Complex Systems Conference*, pages 1–6.
- Rosier, S. H. R., Bull, C. Y. S., Woo, W. L., and Gudmundsson, G. H. (2023). Predicting ocean-induced ice-shelf melt rates using deep learning. *The Cryosphere*, 17(2):499–518.
- Sahana, M. and Patel, P. P. (2019). A comparison of frequency ratio and fuzzy logic models for flood susceptibility assessment of the lower kosi river basin in india. *Environmental Earth Sciences*, 78:1–27.
- Sanderson, J., Mao, H., Abdullah, M. A. M., Al-Nima, R. R. O., and Woo, W. L. (2023a). Optimal fusion of multispectral optical and sar images for flood inundation mapping through explainable deep learning. *Information*, 14(12).
- Sanderson, J., Tengtrairat, N., Woo, W. L., Mao, H., and Al-Nima, R. R. (2023b). Xfimnet: an explainable deep learning architecture for versatile flood inundation mapping with synthetic aperture radar and multispectral optical images. *International Journal of Remote Sensing*, 44(24):7755–7789.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2019). Mobilenetv2: Inverted residuals and linear bottlenecks.
- Schreider, S. Y., Smith, D., and Jakeman, A. (2000). Climate change impacts on urban flooding. *Climatic Change*, 47:91–115.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626.
- Tavus, B., Can, R., and Kocaman, S. (2022). A cnn-based flood mapping approach using sentinel-1 data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3:549–556.
- Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., Mardziel, P., and Hu, X. (2020). Score-cam: Score-weighted visual explanations for convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 24–25.
- Yuan, K., Zhuang, X., Schaefer, G., Feng, J., Guan, L., and Fang, H. (2021). Deep-learning-based multispectral satellite image segmentation for water body detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:7422–7434.
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929.
- Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., and Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pages 3–11. Springer.