Fully convolutional neural networks on semantic segmentation of flooded areas

Samuel G. Ribeiro[1]**, Marcelo M. S. de Souza** [1]

[1]Universidade Federal do Ceará (UFC)
Engenharia da Computação–Sobral–CE–Brasil

`samuelgr@alu.ufc.br, marcelo@ufc.br`

Abstract. *Floods cause extensive economic damage and loss of life worldwide. Thus, automatic image detection is valuable for effectively minimizing response time to these impacts. Synthetic Aperture Radar (SAR) imaging has proven to be an important resource in flood management, as this remote sensing technology is highly sensitive to water. This study applies Fully Convolutional Neural Networks (FCNN), particularly U-Net and U-Net++ topologies, to semantic segmentation of flood-affected regions in Sentinel-1 satellite images from Cloud to Street - Microsoft floods dataset. The U-Net++ architecture demonstrates a high capability in identifying flooded areas, achieving an Intersection over Union (IoU) metric of 0.8280, F1 score of 0.9053, and sensitivity of 0.9001.*

Resumo. *Inundações causam grandes danos econômicos e perdas de vidas em todo o mundo. Assim, a detecção automática de imagens é valiosa para minimizar efetivamente o tempo de resposta a esses impactos. A tecnologia de Radar de Abertura Sintética (SAR) é crucial na gestão de inundações, sendo muito sensível à água. Este estudo usa Redes Neurais Totalmente Convolucionais (FCNN), especialmente as arquiteturas U-Net e U-Net++, para segmentar áreas afetadas por inundações em imagens do satélite Sentinel-1 do conjunto de dados Cloud to Street - Microsoft floods. A arquitetura U-Net++ se destaca na identificação de áreas alagadas, com métricas de Interseção sobre União (IoU) de 0,8280, pontuação F1 de 0,9053 e sensibilidade de 0,9001.*

1. Introduction

The World Meteorological Organization documented increased global natural disasters, noting 11,778 adverse climatic events from 1970 to 2021. These events caused about $4.3 trillion in economic losses and 2 million human deaths. Floods account for 45% of the occurrences, with 32% of financial losses and 16% of fatalities [WMO 2023].

Floods are considered one of the most devastating natural disasters due to their frequency, widespread occurrence, and how many people it affects [Sausen and Lacruz 2015]. Climate changes have intensified the natural hydrological cycle, with heavier rainfall and escalating risk of severe floods, particularly in coastal and low-lying areas [Montello et al. 2022]. Despite the advances in flood alerts and prevention, accurate flood detection is crucial for effective disaster management and mortality reduction [WMO 2023].

Modeling hydrological phenomena in large areas is difficult due to the model complexity, computational costs, and lack of high-resolution data [Almeida et al. 2018]. Nevertheless, Synthetic Aperture Radar (SAR) is effective for monitoring floods regardless of

weather and sunlighting conditions [Tanim et al. 2022], providing comprehensive detection and delineation of flood regions.

Convolutional Neural Networks (CNNs), conventionally used for image classification [Rambour et al. 2020], have expanded into segmentation tasks, where pixel-level classification occurs to delineate regions corresponding to categories [Ronneberger et al. 2015]. Contrasting with instance segmentation, which distinguishes between individual occurrences of categories within an image, semantic segmentation assigns the same label to all category instances [Kumar 2023]. Fully Convolutional Neural Networks (FCNNs) are specifically designed for such tasks, replacing fully connected layers with more convolutional layers. These architectures were successfully applied in detecting oil spills [Orfanidis et al. 2018], generating urban planning maps [Guo et al. 2018], and detecting floods [Bahrami and Arbabkhah 2024].

This study investigates the application of deep learning models for flood detection using SAR images from the Sentinel-1 satellite integrated with Digital Elevation Models (DEM). It assesses the efficacy of FCNNs, specifically U-Net [Ronneberger et al. 2015] and U-Net++ [Zhou et al. 2018], in performing semantic segmentation of flood areas. For this purpose, image pre-processing, data augmentation, and metrics evaluation techniques were applied.

This paper is organized as follows: section 2 outlines the dataset. Section 3 explains the methodology of the work to segment the flood images. Section 4 presents and discusses the obtained results. Finally, section 5 concludes the paper.

2. Dataset

The dataset Cloud to Street - Microsoft Floods (C2S-MS Floods) [Street et al. 2022] covers 18 flood events worldwide, consisting of 900 images captured by the Sentinel-1 from 2016 to 2020. This satellite embeds a SAR sensor, which provides data in the VV (Figure 1b) and VH (Figure 1c) bands. False color images (Figure 1a), created by combining both bands, are available for quick dataset exploration, where water and smooth surfaces are blue, while flooded vegetation is orange and red. Additionally, Sentinel-1 data includes manually annotated water masks (Figure 1d) that let flood extent identification.
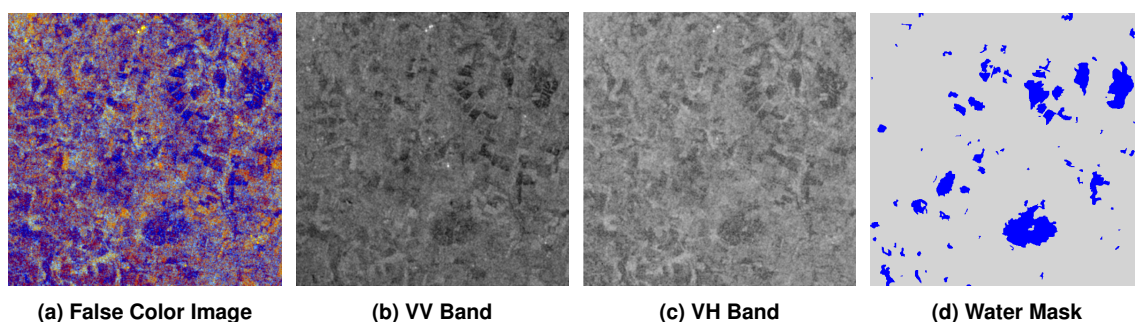

(a) False Color Image   (b) VV Band   (c) VH Band   (d) Water Mask

Figure 1. Sentinel-1 satellite images from dataset

Supplementary elevation data of regions is available in SpatioTemporal Asset Catalog (STAC) [Newman and ESDIS Standards Coordination Office (ESCO) 2023]. This catalog, hosted in Microsoft Planetary Computer [Microsoft Open Source et al. 2022], simplifies query data in STAC standards meeting specific spatial and temporal criteria.

3. Methodology

This section describes the methodology proposed in this work to address the flood detection problem in SAR images.

3.1. Pre-processing

Image pre-processing involves edge and thermal noise removal filters, radiometric calibration, terrain correction, and conversions. This operation is performed with the Sentinel-1 Toolbox (S1TBX) [ESA 2015] and the PyroSAR framework [Truckenbrodt et al. 2019], as illustrated in Figure 2. The former is known to support the extensive datasets of ESA's SAR missions in SAR image operations of calibration, noise filtering, orthorectification, mosaicking, data conversion, polarimetry, and interferometry, and the latter is a Python framework for processing large-scale SAR satellite data.
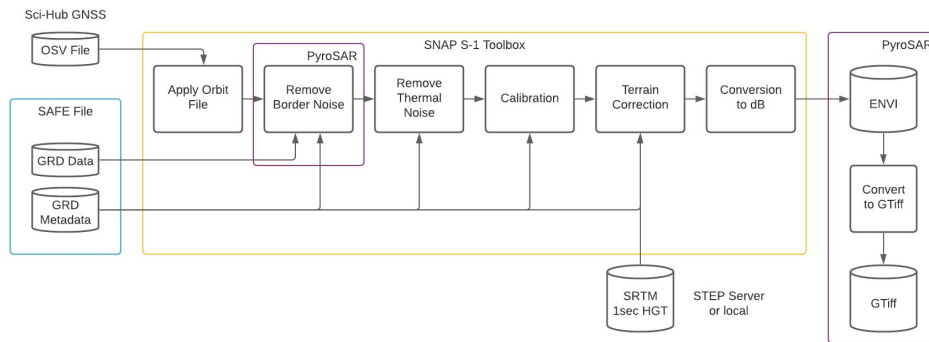

Figure 2. Pre-processing of Sentinel-1 images [Street et al. 2022]

The input raw images correspond to VH and VV bands of SAR and elevation data. These are initially read and converted into arrays, stacked to form three distinct channels, and transformed into tensors. Similarly, each flood mask undergoes the same operation. Finally, image channels are normalized to zero mean and one variation.
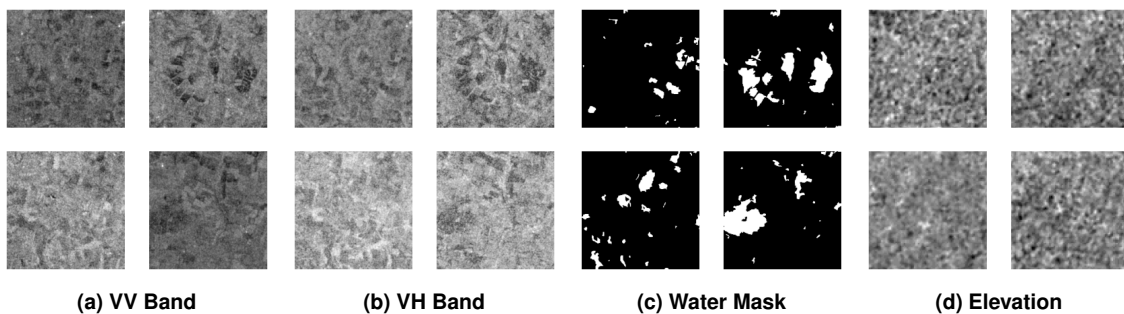

(a) VV Band   (b) VH Band   (c) Water Mask   (d) Elevation

Figure 3. Radar images cropped from dataset

Finally, data augmentation is applied to images to reduce training overfitting and improve model robustness and generalization. Image-splitting is applied to each channel and mask, dividing them into four equal parts, as illustrated in Figure 3. Moreover, images underwent additional transformations with a 50% chance of vertical or horizontal flipping to increase variability. That produced 3,600 training images, 70.42% with floods and 29.58% without floods.

3.2. Models description

The models considered in this work for semantic segmentation of floods are the U-Net (Figure 4a) [Ronneberger et al. 2015] and U-Net++ (Figure 4b) [Zhou et al. 2018]. These two FCNNs follow an encoder-decoder structure with skipped connections. The encoder (left structures in Figures 4a and 4b) comprises a sub-network, where convolution and pooling layers reduce feature map resolution and increase the number of channels. The decoder (right structures in Figures 4a and 4b) incorporates layers that increase the resolution of the feature maps and reduce the number of channels, producing the segmented input image as output. Skipped connections between these sub-networks allow the incorporation of high-level information from the encoder to the decoding process, improving segmentation accuracy [Kumar 2023].

Designed for medical image segmentation, U-Net became popular because it preserves image details and provides accurate segmentation with limited labeled data. U-Net++ is a U-Net redesign that enhances accuracy by addressing the semantic gap between the encoder and decoder.
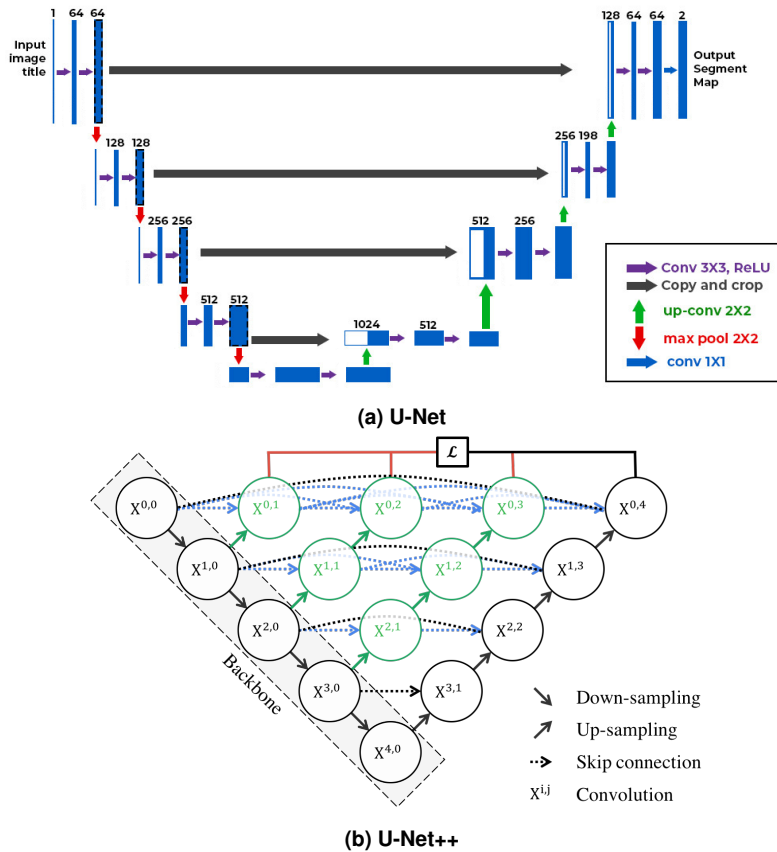

(a) U-Net

(b) U-Net++

Figure 4. U-Net e U-Net++ topologies [Zhou et al. 2018, Ronneberger et al. 2015]

In this work, the backbone of both FCNNs for feature extraction is the ResNet-18 [He et al. 2016]. This choice is because the ResNet-18 mitigates the gradient vanish problem in training.

3.3. Model validation

The model is cross-validated, with images randomly distributed into the train, validation, and test sets in 70%, 15%, and 15% ratios, respectively. Each validation comprises four runs for distinct train, validation, and test random configurations.

The pre-trained weights of ImageNet are considered for starting the training with the Adam optimizer, a learning rate of 0.001, and a maximum of 100 epochs. The choice of optimizer and learning rate was empirical; repeated training experiments were conducted, and the results were analyzed. Early stopping is employed to adjust the learning rate and prevent overfitting during training dynamically. That helps find an optimal point where the model has learned sufficiently without becoming excessively tailored to the training data.

Three loss functions, Binary Cross-Entropy (BCE), Dice Loss Coefficient (DLC), and Tversky Loss Index (TLI), are considered to assess the one that yields the best result. Their hole is to quantify the disparity between predicted and actual masks, guiding the model through the optimization process to learn to segment areas of interest in images accurately. Lower values indicate better alignment and zero loss signifies perfect predictions [Zhang et al. 2023].

Cross-entropy [Shruti 2020] is a widely used information theory measure for image segmentation tasks. It quantifies the discrepancy between the model's classifications and the ground truth labels. For binary segmentation tasks, it is expressed as:

$$BCE(y, \hat{y}) = -(y \cdot log(\hat{y}) + (1 - y) \cdot log(1 - \hat{y})) \tag{1}$$

where y represents the ground truth label, $\hat{y}$ the model's prediction and $log(.)$ the logarithmic function to base 2.

Dice coefficient (DC) [Shruti 2020] evaluates the overlap between predicted ($\hat{y}$) and true masks (y). It is calculated as twice the intersection of the masks divided by their sum. The Dice loss coefficient (DLC) derives from DC as:

$$DLC(y, \hat{y}) = 1 - DC = 1 - \frac{2 \times |y \cap \hat{y}|}{|y| + |\hat{y}|} \tag{2}$$

Tversky [Shruti 2020] is an asymmetric similarity measure known as the Tversky index (TI). It generalizes the DC and the Jaccard index (IoU). Thus, the Tversky Loss index (TLI) is defined as:

$$TLI(y, \hat{y}) = 1 - TI = 1 - \frac{|y \cap \hat{y}|}{|y \cap \hat{y}| + \alpha \cdot |y - \hat{y}| + \beta \cdot |\hat{y} - y|} \tag{3}$$

where α and β regulate the weight of false positives and false negatives in the metric, adjusting its sensitivity. When both are set to 0.5, it reduces to DC. When set to 1 it reduces to the Jaccard loss (JL). High values of β relative to α heavily penalize false negatives when comparing y and $\hat{y}$.

3.4. Evaluation Metrics

The quality of segmentation prediction is evaluated with quantitative metrics IoU, F1 score, and sensitivity.

IoU is valuable for segmentation assessment ranging from 0 to 1, where 0 indicates no overlap between two sets and 1 complete overlap. It is defined as the division of the intersection area by the size of the union of the sets.

$$IoU(y, \hat{y}) = \frac{|y \cap \hat{y}|}{|y \cup \hat{y}|} \qquad (4)$$

Sensitivity, also known as recall, measures the model's ability to identify all positive examples correctly. Precision measures the model's ability to identify only the positive examples correctly. The F1 score combines precision and sensitivity into a single measure. When there is an imbalance between classes, this metric can be more reliable for evaluating model performance [Zhang et al. 2023].

4. Results

Table 1 shows the performance of the U-Net and U-Net++ in the experiments of flood segmentation. Each row corresponds to the mean and standard deviation values of metrics IoU, F1 score, and sensitivity for each loss function considered. That also reveals how loss functions impact the performance of the networks.

The BCE is the most consistent loss function, with minimal variations between runs for both networks. The DLC exhibited more variability between runs, especially regarding sensitivity. This is because of the high sensitivity of this function to the overlap between segmented and reference areas. The mean results with TLI are similar to those with DLC. The sensitivity of TLI to the overlap boundaries between regions accounts for this variability.

Table 1. Validation and test performances

		U-Net		U-Net++	
		Validation	Test	Validation	Test
BCE	**IoU**	$82.50\% \pm 0.71$	$82.40\% \pm 2.00$	$82.35\% \pm 1.20$	$\mathbf{82.86\%}\pm1.80$
	F1 Score	$90.34\% \pm 0.41$	$90.21\% \pm 1.30$	$90.22\% \pm 0.74$	$\mathbf{90.53\%}\pm1.10$
	Sensitivity	$88.16\% \pm 0.32$	$88.70\% \pm 1.70$	$89.21\% \pm 0.87$	$\mathbf{90.01\%}\pm1.50$
DLC	**IoU**	$81.44\% \pm 1.70$	$\mathbf{81.96\%}\pm2.10$	$81.53\% \pm 2.60$	$81.81\% \pm 2.90$
	F1 Score	$89.68\% \pm 1.00$	$\mathbf{89.97\%}\pm1.30$	$89.72\% \pm 1.60$	$89.86\% \pm 1.80$
	Sensitivity	$89.10\% \pm 1.20$	$\mathbf{89.39\%}\pm1.60$	$88.93\% \pm 2.10$	$89.25\% \pm 2.40$
TLI	**IoU**	$81.17\% \pm 1.90$	$81.71\% \pm 2.80$	$81.79\% \pm 1.80$	$\mathbf{81.80\%}\pm2.20$
	F1 Score	$89.48\% \pm 1.20$	$89.80\% \pm 1.70$	$89.89\% \pm 1.00$	$\mathbf{89.85\%}\pm1.40$
	Sensitivity	$88.18\% \pm 1.40$	$88.73\% \pm 2.10$	$88.56\% \pm 1.80$	$\mathbf{88.76\%}\pm2.20$

Table 2 gives a qualitative view of each network and loss function's actual and predicted flood masks. The white regions of images correspond to flooded areas, and the black ones are non-flooded areas. BCE, DLC, and TLI values for U-Net are 60.90%, 94.88%, and 93.73%, respectively. Meanwhile, the U-Net++ yields 95.83%, 95.23%, and 94.90% for the same loss functions. The U-Net++, using the BCE loss for pixel classification, achieved the best performance on the test set, presenting the highest means and the lowest standard deviations for each evaluated metric.

Table 2. Predictions of the U-Net and U-Net++ models for each loss function

Losses for both training and validation sets decreased, indicating progressive model improvement. This trend was consistent across all experiments. However, occasional disparities between training and validation losses suggested potential overfitting, for which early stopping was beneficial.

5. Conclusion

This study demonstrated the efficacy of FCNNs for semantic segmenting flood-affected areas using Sentinel-1 SAR images. U-Net and U-Net++ architectures were employed, with U-Net++ showing superior performance in accurately identifying flooded regions. The choice of loss functions impacted the performance of the models. BCE provided consistent results, while the DLC and TLI introduced variability, particularly in sensitivity. Table 3 compares the performance metrics of the optimal model presented in this study with those of the top models from other related research.

Table 3. Result of the proposed model and other results

	Proposed Algorithm	IoU	F1 Score	Sensitivity
[Tanim et al. 2022]	Change Detection	-	0.8500	0.9000
[Montello et al. 2022]	ResNet-50	0.6300	-	-
	TResNet	0.5900	-	-
Our Proposal	U-Net++	0.8280	0.9053	0.9001

Data augmentation techniques and early stopping helped mitigate overfitting and improved model generalization. Future work could explore integrating additional data sources, such as optical imagery and hydrological data, to enhance the model's robustness. Investigating other advanced FCNN architectures and loss functions may further improve segmentation performance. Overall, this study underscores the potential of FCNNs in flood management and their utility in leveraging SAR data for effective disaster response and management.

References

Almeida, G. A. M. d., Bates, P., and Ozdemir, H. (2018). Modelling urban floods at submetre resolution: challenges or opportunities for flood risk management? *Journal of Flood Risk Management*, 11:S855–S865.

Bahrami, B. and Arbabkhah, H. (2024). Enhanced flood detection through precise water segmentation using advanced deep learning models. *Journal of Civil Engineering Researchers*, 6(1):1–8.

ESA (2015). The sentinel-1 toolbox. `https://sentinel.esa.int/web/sentinel/toolboxes/sentinel-1`. Accessed: 2023-11-02.

Guo, Z., Shengoku, H., Wu, G., Chen, Q., Yuan, W., Shi, X., Shao, X., Xu, Y., and Shibasaki, R. (2018). Semantic segmentation for urban planning maps based on u-net. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 6187–6190.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Kumar, A. (2023). CNN basic architecture for classification & segmentation. `https://vitalflux.com/cnn-basic-architecture-for-classification-segmentation`. Accessed: 2023-09-30.

Microsoft Open Source, McFarland, M., Emanuele, R., Morris, D., and Augspurger, T. (2022). Microsoft Planetary Computer Data Catalog. `https://planetarycomputer.microsoft.com/catalog`. Accessed: 2024-07-02.

Montello, F., Arnaudo, E., and Rossi, C. (2022). Mmflood: A multimodal dataset for flood delineation from satellite imagery. *IEEE Access*, 10:96774–96787.

Newman, D. J. and ESDIS Standards Coordination Office (ESCO) (2023). Spatiotemporal asset catalogs (stac). `https://www.earthdata.nasa.gov/esdis/esco/standards-and-practices/stac`. Accessed: 2025-07-02.

Orfanidis, G., Ioannidis, K., Avgerinakis, K., Vrochidis, S., and Kompatsiaris, I. (2018). A deep neural network for oil spill semantic segmentation in sar images. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 3773–3777.

Rambour, C., Audebert, N., Koeniguer, E., Saux, B. L., Crucianu, M., and Datcu, M. (2020). Flood detection in time series of optical and sar images. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2020:1343–1346.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, volume 9351, page 234–241.

Sausen, T. M. and Lacruz, M. S. P. (2015). *Sensoriamento Remoto para Desastres*. Oficina de Textos, São Paulo.

Shruti, J. (2020). A survey of loss functions for semantic segmentation. In *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pages 1–7.

Street, C. t., Microsoft, and Foundation, R. E. (2022). A global flood events and cloud cover dataset. `https://collections.eurodatacube.com/microsoft-floods-s1`. Accessed: 2023-08-25.

Tanim, A. H., McRae, C. B., Tavakol-Davani, H., and Goharian, E. (2022). Flood detection in urban areas using satellite imagery and machine learning. *Water*, 14(7):1140.

Truckenbrodt, J., Cremer, F., Baris, I., and Eberle, J. (2019). Pyrosar: a framework for large-scale sar satellite data processing. In *P. Soille, S. Loekken, and S. Albani, editors, Big Data from Space*, page 197–200.

WMO (2023). Atlas of mortality and economic losses from weather, climate and water-related hazards (1970-2021). `https://wmo.int/publication-series/atlas-of-mortality-and-economic-losses-from-weather-climate-and-water-related-hazards-1970-2021`. Accessed: 2024-06-22.

Zhang, A., Lipton, Z. C., Li, M., and Smola, A. J. (2023). *Dive into Deep Learning*. Cambridge University Press.

Zhou, Z., Siddiquee, M. M. R., 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 – DLMIA ML-CDS 2018*, volume 11045, page 3–11.