

Flood Zone Segmentation with U-Net Using Blended Loss Functions

Staten Island Technical High School

Science Engineering Research Program

Dr. John Davis

December 2025

Yan Sharma, Oleg Ponyatyshyn

Vaibhav Dixit, MS

Abstract

Flooding is a major natural hazard that causes widespread damage and requires fast, accurate mapping to support disaster response. Previous research in image segmentation has shown that blending loss functions in U-Net models can improve performance on imbalanced data by combining region-level overlap optimization with pixel-level emphasis on hard examples. While this approach has been effective in biomedical and other remote sensing tasks, its impact on flood segmentation and optimal weighting ratios has been less explored. This study evaluates whether blending Dice and focal loss improves flood segmentation accuracy compared to Dice loss alone. A U-Net model was trained on multispectral satellite imagery to classify pixels as water, land, or cloud, using pure Dice loss as a control and blended Dice–focal loss ratios of 0.5/0.5, 0.75/0.25, and 0.9/0.1. Model performance was assessed using accuracy, Intersection over Union, and qualitative mask analysis. Results show that pure Dice loss achieved high accuracy but consistently overestimated flooded areas, resulting in low water IoU. The 0.5/0.5 blended loss degraded performance and destabilized segmentation. In contrast, the 0.75/0.25 and 0.9/0.1 configurations produced cleaner flood boundaries and higher IoU for water and cloud classes, with the 0.75/0.25 ratio providing the most balanced improvement. These findings suggest that modest focal weighting can meaningfully reduce false positives and improve flood mapping reliability without sacrificing overall stability.

Rationale

Flooding is among the most devastating and frequent natural disasters worldwide, affecting millions of people each year and causing billions of dollars in damage to infrastructure, homes, and agricultural land (IPCC, 2021). Climate change has intensified the frequency and severity of these events, leading to more unpredictable rainfall patterns, sea-level rise, and storm surges. The Intergovernmental Panel on Climate Change reports that extreme precipitation and river flooding events are expected to increase significantly as global temperatures continue to rise. This growing threat gives rise to an urgent need for reliable systems that can detect and map flood-affected areas quickly and accurately.

In many regions, traditional methods of assessing flood damage, such as on-the-ground surveys and manual mapping, are too slow and resource-intensive to meet the demands of modern disaster response (Schumann et al., 2009). They often fail to provide comprehensive spatial coverage, especially in remote or inaccessible areas. Satellite imagery, on the other hand, offers a way to monitor large areas efficiently and in near real-time, making it an invaluable tool for emergency management agencies and policymakers. However, the vast amount of satellite data produced daily requires automated and intelligent systems to analyze it effectively.

Machine learning provides a solution to this challenge by enabling faster and more accurate interpretation of satellite imagery (Schumann et al., 2009). Through automated pattern recognition and classification, machine learning models can identify flood-affected zones far more rapidly than manual methods, supporting early response and recovery efforts. This technological advancement allows for continuous monitoring, better prediction of flood impacts, and improved coordination in relief operations. In particular, the combination of remote sensing

and machine learning enhances the accuracy of flood maps and makes large-scale disaster monitoring more feasible for governments and humanitarian organizations .

Given the escalating risks posed by climate change and the limitations of traditional mapping approaches, developing systems that integrate satellite imagery and intelligent data analysis has become more of a necessity than an option. Such efforts are critical for strengthening disaster preparedness, guiding emergency response, and ultimately reducing the human and economic toll of flooding in vulnerable regions around the world.

Convolutional neural networks (CNNs) are a class of deep learning models designed to process grid-like data such as images (O'Shea & Nash, 2015). Unlike traditional image processing methods that rely on manually defined features, CNNs automatically learn hierarchical patterns directly from raw pixel data through layers of convolution, pooling, and activation operations. This makes them particularly effective for tasks like object detection, classification, and semantic segmentation.

The U-net architecture was first created for biomedical image segmentation in Ronneberger et al. (2015). U-Net is a specific CNN architecture developed for image segmentation tasks that require precise pixel-level predictions. It consists of two main components: an encoder that progressively downsamples the input image to capture contextual information, and a decoder that upsamples the compressed representation back to the original resolution. What distinguishes U-Net from standard CNNs is its use of skip connections, which link corresponding layers in the encoder and decoder. These connections preserve fine spatial details that would otherwise be lost during downsampling, allowing the model to produce accurate segmentation masks that capture both global context and local boundaries.

A loss function in deep learning is a mathematical measure of how far a model's predictions are from the true labels (Azad et al., 2023). During training it tells the optimizer how to adjust network weights to reduce errors. For segmentation it typically evaluates how well predicted masks match ground truth masks. Lower loss means better fit.

Dice loss is based on the Dice similarity coefficient, which measures the overlap between predicted and ground truth regions (Azad et al., 2023). It focuses on the region level, looking at groups of pixels as a whole to improve structural similarity between masks and true objects. It's good at handling class imbalance because it directly optimizes overlap rather than treating each pixel independently.

Focal loss modifies the standard pixel-wise cross-entropy loss (Azad et al., 2023). It down-weights easy, well-classified pixels and boosts the impact of harder, misclassified pixels (especially pixels in minority classes) via a focusing parameter and optional class weights. It operates at the pixel level by adjusting each pixel's contribution based on how confidently it was predicted. This is useful for class imbalance because the many easy background pixels won't overwhelm learning. Xu et al. (2023) for example, compared various loss functions' performance for road segmentation using a D-LinkNet model, a derivation of U-net. The study found that Focal loss performed extremely well, topping various datasets in precision and recall, but possibly overfitted to the training data.

The key difference is that Dice loss measures region-level overlap and balances classes implicitly, while Focal loss emphasizes hard examples explicitly, helping the model learn fine details where segmentation is toughest (Azad et al., 2023). Because Dice does not target hard pixels and Focal does not directly maximize region overlap, combining them can strengthen training: the Dice component encourages good overall overlap, while the Focal component

forces the model to learn from pixels that are currently misclassified. Studies in other domains have shown that a combined Dice+Focal loss can improve segmentation accuracy over using either loss alone, especially with imbalanced classes (Mu et al., 2024; Yeung et al. 2021).

For biomedical image segmentation, there are numerous examples of researchers combining dice and focal loss to great effect. In Yeung et al. (2021) for example, researchers tested models using various loss functions on different class imbalance data sets, finding that their custom combined loss function “consistently outperforms the other loss functions.” Ahmed & Rahmim (2023) also was able to use a “Generalized Loss Function” for a 3D Residual Unet to segment lesions in full body scans.

For flood segmentation, we weren’t able to find many studies specifically looking at combined Dice & Focal loss, especially those that compare different breakdowns of the combined loss. Deb & Verma (2025) combined DINOv2 (vision transformer developed by Meta) and U-net to identify flooded regions. They compared Dice, Focal, and a combined Dice + modified Focal loss. They found that the combined Dice + modified Focal loss achieve the highest mean IoU for the different classifications. Additionally, Hartanti et al. (2025) compared a 50/50 Dice & Cross Entropy loss function with a 70/30 Dice & Cross Entropy loss function. They found that the 50/50 split outperformed the 70/30 according to dice similarity coefficients (DSC). The findings suggested that “balancing the contributions of both dice loss and CE loss is more effective for achieving optimal segmentation.”

Based on this previous research, we believe that there is a need for further analysis of blended Dice & Focal loss functions and their influence on flood segmentation, specifically analyzing different weighing ratios.

This need motivates the central research question of this study: whether blending Dice loss with focal loss can improve flood segmentation accuracy, and which weighting ratio provides the best performance.

This study evaluates multiple blending ratios to identify how different balances between Dice and focal loss influence model behavior, class stability, and segmentation quality. The results aim to clarify not only whether blended loss functions are beneficial, but also how they should be tuned to avoid overcorrection and degraded performance. This work seeks to inform the design of more reliable flood mapping models that can better support time-critical disaster response and resource allocation decisions. By leveraging open-source satellite data and deep learning, the study aims to contribute to the development of accessible, data-driven tools for humanitarian response and climate resilience.

Methods

A U-net CNN was used for this project for its architecture's ability to capture relevant spacial features for flood mapping across different sensors and geographic settings (Pech-May et al., 2024). Since its success in medical imaging, U-Net has been widely adopted in other fields that require semantic segmentation, including flood segmentation from remote sensing images.

This project uses the World Floods Dataset, which contains multispectral satellite imagery from Sentinel-1 and Sentinel-2 satellites designed for land and sea monitoring. Data from these satellites are combined and averaged into "S-2" files, which provide more accurate representations of natural surface conditions. These S-2 images are paired with manually generated ground truth flood masks that serve as reference labels for model training and evaluation.

Images undergo preprocessing to remove low-quality or incomplete data, particularly those with excessive cloud cover or missing spectral bands. All imagery is standardized to a consistent spatial resolution of 10 meters per pixel. Relevant spectral bands are prioritized for model training: near-infrared (NIR), short-wave infrared (SWIR), and visible light channels. These bands are selected for their ability to highlight water features and distinguish flooded from non-flooded terrain, which streamlines the training process. The dataset is automatically split into training (68%), validation (12%), and testing (20%) subsets.

The model will classify each pixel into one of three categories: water, ground, or clouds. The architecture contains approximately 7 to 8 million trainable parameters, with hyperparameters including learning rate, batch size, and number of epochs tuned during training.

The model is first trained using a pure Dice loss function, which serves as the control condition for this study. We will then evaluate blended loss functions that combine Dice loss with focal loss at specific weighting ratios: 0.5 Dice / 0.5 Focal, 0.75 Dice / 0.25 Focal, and 0.9 Dice / 0.1 Focal. These configurations are designed to preserve the region-level strengths of Dice loss while gradually introducing focal loss to emphasize harder and underrepresented examples, such as water and cloud pixels.

Model effectiveness is assessed through various metrics, including Intersection over Union, precision, and accuracy. Intersection over Union (IoU), which measures the overlap between predicted and true flood regions; precision, which indicates the proportion of predicted flood pixels that are actually flooded; and accuracy, which measures the proportion of true flood pixels correctly identified. Using these metrics, we will compare each blended configuration against the pure Dice control, examining how different Dice–focal balances influence segmentation accuracy, false positives, and overall model stability.

Data/Discussion

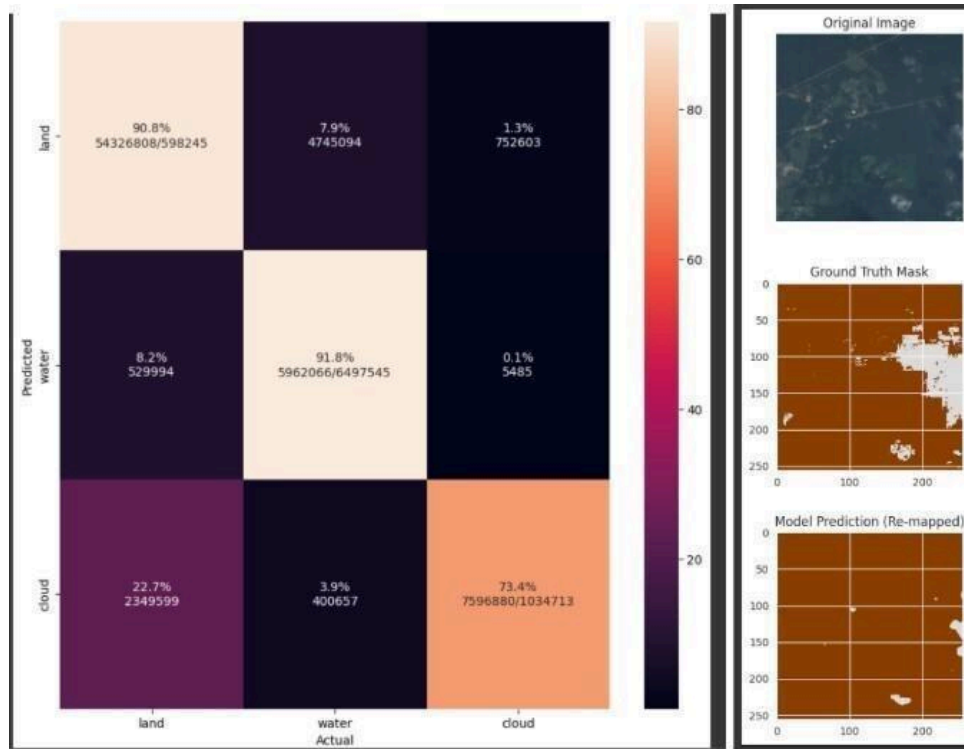


Figure 1 (Image created by author) Confusion Matrix for Pure Dice Loss

When running the U-net model with just Dice loss, we had fairly high accuracy in segmenting the flood zones. The evaluation results indicate that the model performs strongly in distinguishing land and water, but faces more difficulty in accurately classifying clouds. As seen in Figure 1, the confusion matrix shows high accuracy for land (91.4%) and water (91.5%), with cloud detection being lower at 72.9%, partly due to frequent misclassification of land as cloud (23.5%). However, this high accuracy is slightly misleading, as this pattern is inconsistent with the per-class Intersection-over-Union (IoU) values. For the IoU values, land achieves the highest overlap with ground truth (0.87), cloud performs moderately (0.68), and water is the weakest class (0.52). This discrepancy results from the way accuracy and IoU are calculated. Accuracy just comes from calculating what percentage of the true pixels are correctly identified as that

type of mask, not penalizing false positives at all. IoU, however, does penalize false positives, essentially looking at the percentage of things identified as a mask that actually are that category. This means that when just using dice loss, the model does identify most of the water sections correctly, but tends to overestimate the placement of water.

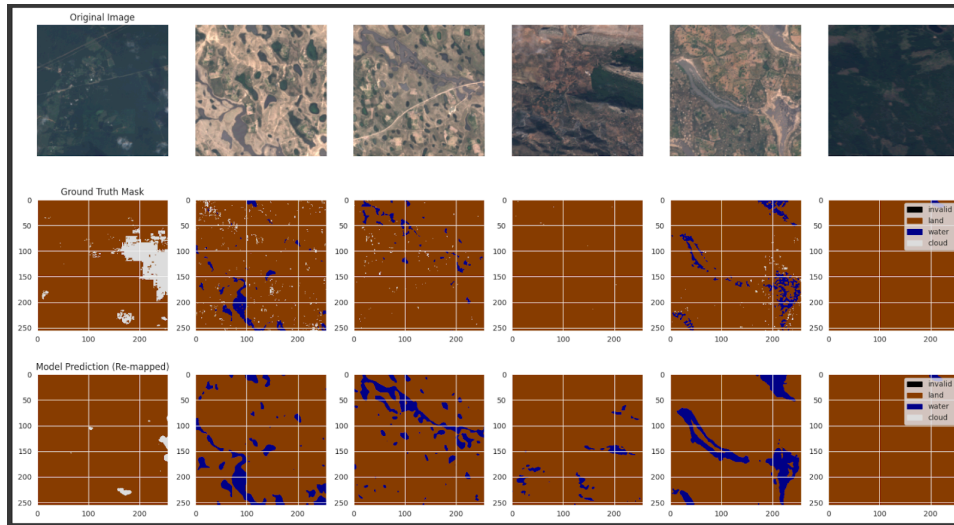


Figure 2 (Image created by author) Predicted Mask Visualization for Pure Dice Loss

This trend can be roughly seen in Figure 2. In the sections where water is prevalent, the model correctly identifies most of the water, but tends to “round it out” in a sense, making many false positives around the area it classifies as water. In the fourth set of images in Figure 2, the model incorrectly predicts an image filled with just land as being dotted with water. The fourth set of images also shows the model struggles with accurately classifying cloud pixels as well. These issues, as well as land having such a comparatively high IoU, likely stem from the relative sample sizes of the different classes. Amongst the database, land is most represented, with water and cloud trailing a distance behind. This means that in the training stages, the model doesn't have nearly as many examples of water and cloud to train off and learn the patterns to classify them.

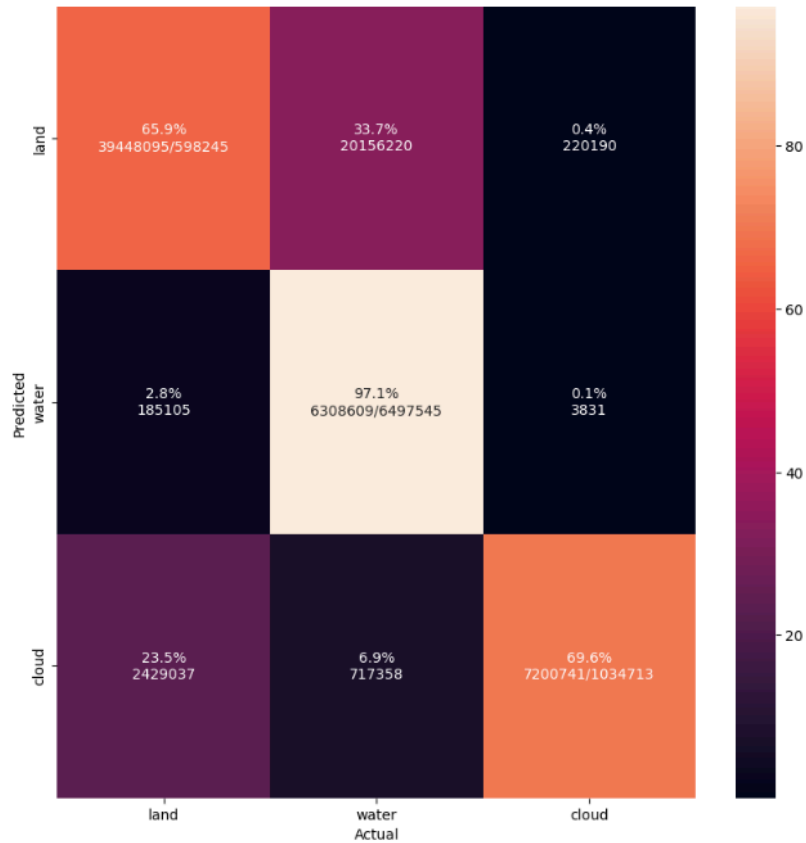


Figure 3 (image created by author) Confusion Matrix for Blended Loss Function (0.5 Dice, 0.5 Focal)

The 0.5, 0.5 blended loss function did not improve results as expected. Although this kind of loss function is typically used to address class imbalance, the confusion matrix in Figure 3 shows that misclassifications remained largely similar to the Dice-only model, and in the case of cloud and land greatly worsened. Land's accuracy dropped to 65.9%, water's accuracy rose to 97.1%, while cloud's accuracy dropped to 69.6%. The IoU values dropped or stayed the same across the board, with land dropping to 0.68, water dropping to 0.23, and cloud staying about the same at 0.68. The model continued to overestimate water in certain regions, and cloud misclassification persisted despite the increased weighting on difficult or underrepresented examples.

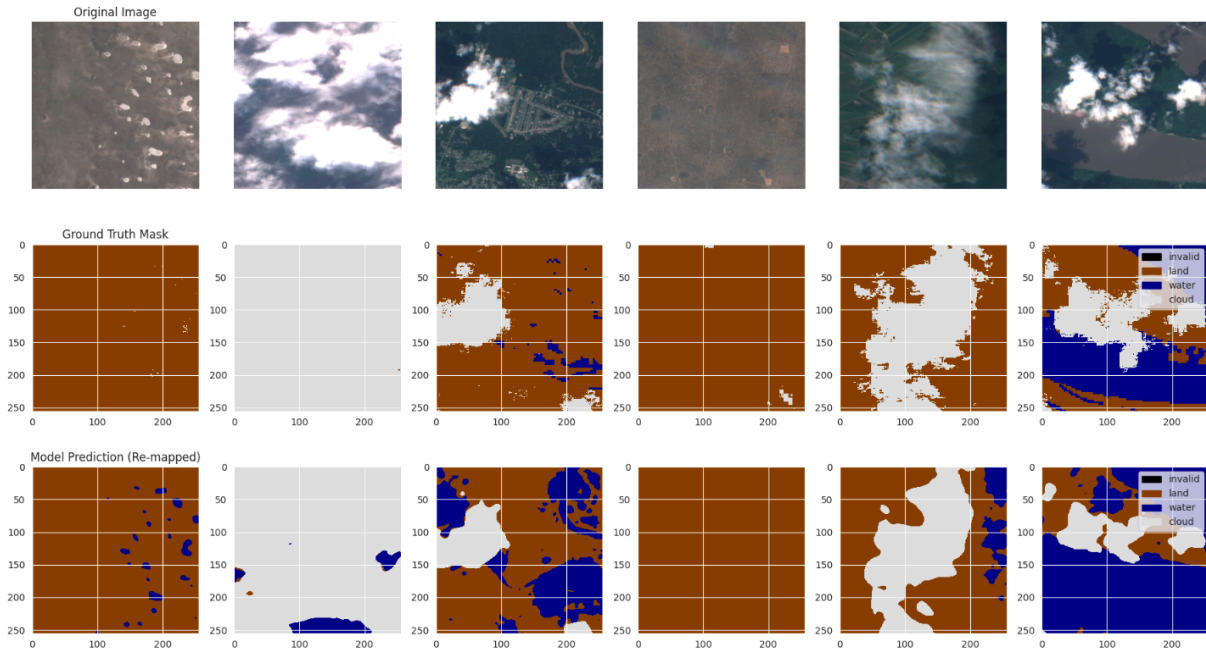


Figure 4 (Image Created by Author) Predicted Mask Visualization for Blended Loss Function (0.5 Dice, 0.5 Focal)

These results suggest that the class imbalance in this dataset is more complex than what the Focal component alone can correct and that weighing it as 0.5 was counterproductive, making the model oversensitive to the underrepresented classes. Because the “hard” examples (water and clouds) vary significantly in appearance across scenes, the model struggled to generalize the patterns emphasized by the focal loss. As seen in Figure 4, the 0.5 0.5 loss model is heavily overpredicting water, “rounding out” existing water and predicting it in place of both cloud and water. Instead of guiding the model toward cleaner segmentation, the blended loss function actually may have over-penalized uncertain predictions, causing the network to over-correct in water while still failing to learn rare examples in cloud and land. This could indicate that the issue relies not only on an imbalance of quality, but also on an imbalance in context and visual variability, which the combined loss function cannot correct for.

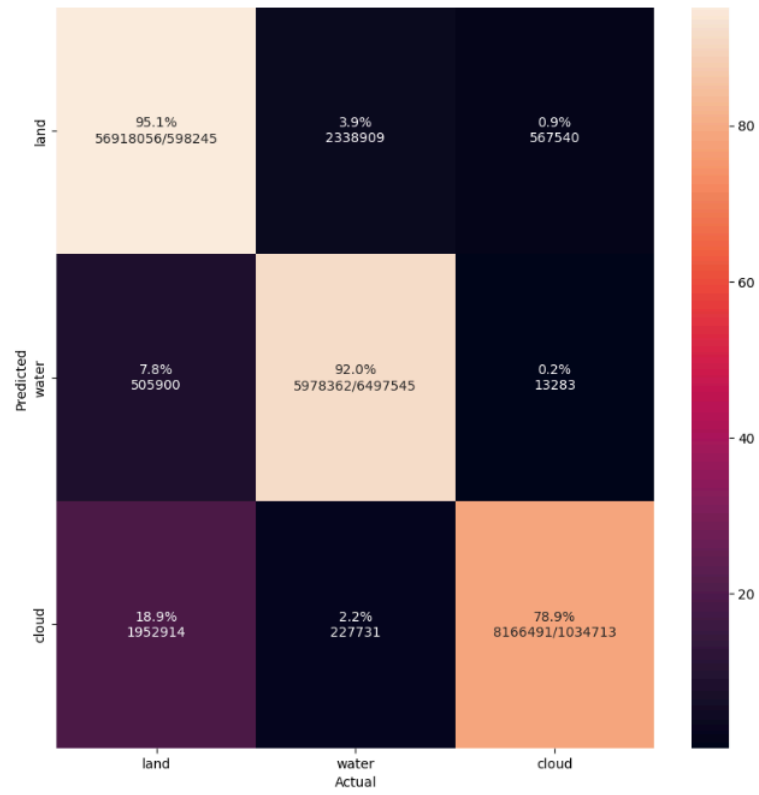


Figure 5 (image created by author) Confusion Matrix for Blended Loss Function (0.75 Dice, 0.25 Focal)

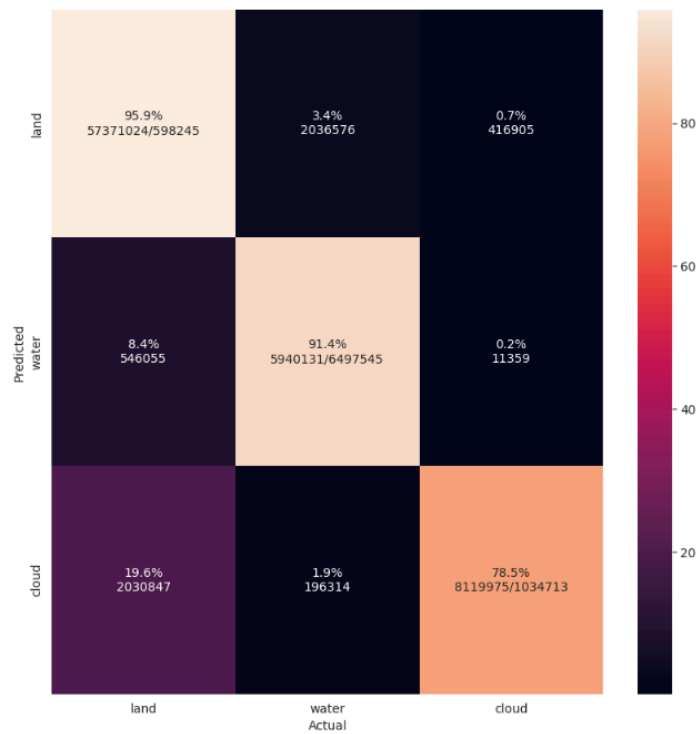


Figure 6 (image created by author) Confusion Matrix for Blended Loss Function (.9 Dice, .1 Focal)

When comparing the 0.75 Dice / 0.25 Focal and 0.9 Dice / 0.1 Focal configurations, both models demonstrate greatly improved stability and segmentation quality relative to the Dice-only and 0.5/0.5 blended models, but with subtle trade-offs. For the 0.75 Dice / 0.25 Focal configuration in Figure 5, the model could accurately classify land with 95.1% accuracy and water to 92.0%, indicating strong stability on the dominant class despite increased focal weighting. Cloud accuracy staggered a little bit, only reaching 78.9%, though it is the highest cloud classification accuracy observed among the evaluated models.

For the 0.9 Dice / 0.1 Focal configuration (Figure 2), the accuracy results are highly similar to those observed with the 0.75 Dice / 0.25 Focal model. Land accuracy remains nearly unchanged at 95.9%, compared to 95.1% in the 0.75/0.25 configuration, while water accuracy is 91.4%, only slightly lower than 92.0%. Cloud accuracy also shows minimal variation, reaching 78.5% versus 78.9% previously.

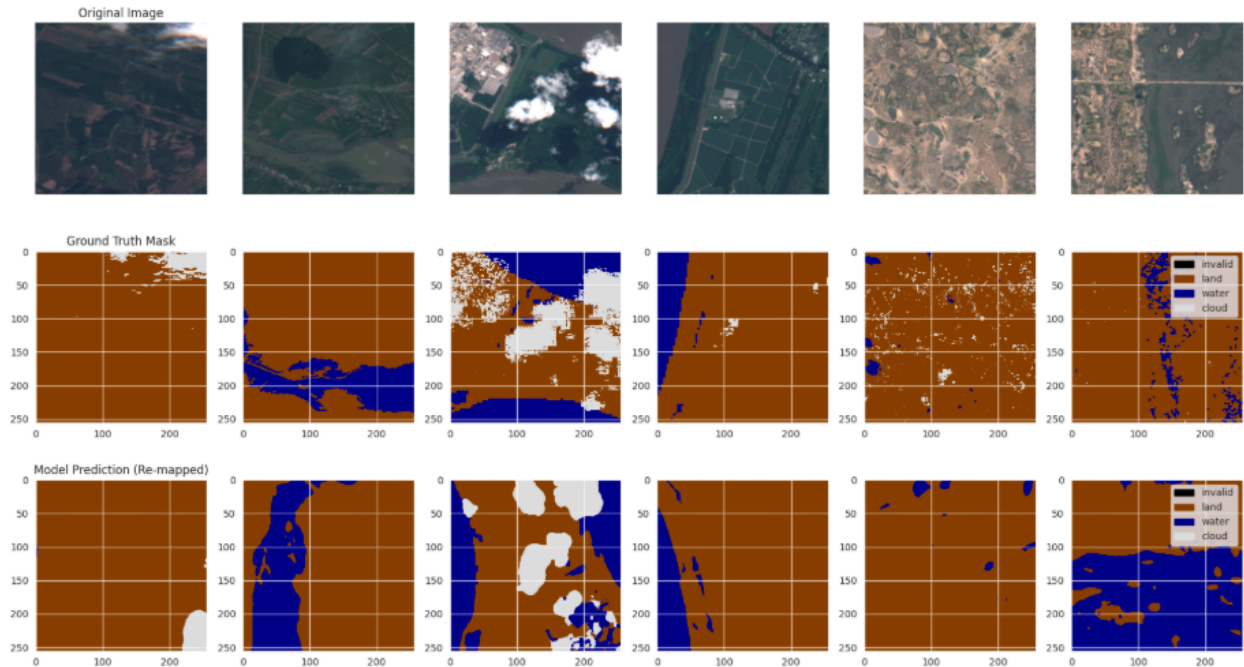


Figure 7 (image created by author) Predicted Mask Visualization for Blended Loss Function (0.75 Dice, 0.25 Focal)

The predicted masks in Figure 7 indicate that water regions are more tightly constrained, with fewer false positives and cleaner boundaries, particularly in areas of mixed land–water transitions. This qualitative improvement is reflected in the per-class IoU values associated with the 0.75 Dice / 0.25 Focal configuration, which achieved IoUs of 0.914 for land, 0.660 for water, and 0.747 for cloud. These values represent a clear improvement over the Dice-only baseline, especially for water and cloud, where false positives had previously inflated accuracy while suppressing overlap quality. Given that the accuracy results for the 0.9 Dice / 0.1 Focal model were shown to be highly similar, its predicted masks exhibit nearly identical spatial behavior. This similarity is reinforced by the IoU scores for the 0.9 / 0.1 model, which reached 0.919 for land, 0.680 for water, and 0.754 for cloud.

Loss Functions	Accuracy (Land, Water, Cloud)	IoU (L, W, C)
Dice	90.8%, 91.8%, 73.4%	0.87, 0.52, 0.68
0.5 Dice 0.5 Focal	65.9%, 97.1%, 69.6%	0.68, 0.23, 0.68
0.75 Dice 0.25 Focal	95.1%, 92%, 78.9%	0.91, 0.66, 0.75
0.9 Dice 0.1 Focal	95.9%, 91.4%, 78.5%	0.92, 0.68, 0.75

Figure 8 (image created by author) Predicted Mask Visualization for Blended Loss Function (0.75 Dice, 0.25 Focal)

When comparing all models, a clear performance trend is seen when comparing loss function configurations. The Dice-only model establishes a strong baseline with high land and water accuracy but suffers from lower water IoU, indicating systematic overestimation of flooded regions. The 0.5 Dice / 0.5 Focal model performed the worst overall, most likely due to the fact that excessive focal weighting destabilizes segmentation, which in turn sharply reduces land accuracy and IoU and causes the model to aggressively water prediction per pixel. By

contrast, the 0.75 Dice / 0.25 Focal configuration achieved the most balanced results, improving both accuracy and IoU for water and cloud while preserving strong land performance and producing cleaner, more constrained flood boundaries. The 0.9 Dice / 0.1 Focal model yielded results nearly identical to the 0.75 / 0.25 configuration, with only marginal metric differences, suggesting that modest focal weighting is sufficient to address class imbalance and that further emphasis on Dice loss provides diminishing returns.

Conclusion

This study demonstrates that U-Net-based segmentation can effectively identify flood-affected areas from satellite imagery, achieving strong performance in land and water classification. The model's ability to process multispectral satellite data and produce pixel-level flood maps shows practical potential for disaster response applications. Automated flood detection systems like this could significantly reduce the time and resources required for damage assessment, enabling faster deployment of emergency services and more efficient allocation of relief efforts in affected regions.

However, the results also highlight implementation challenges that must be addressed before deployment in real-world disaster scenarios. The model's tendency to overestimate water extent, as indicated by the disparity between accuracy and IoU metrics, could lead to false alarms or misallocation of resources if areas are incorrectly flagged as flooded. Similarly, the lower performance on cloud classification creates ambiguity in imagery where cloud cover is present, which is common in the aftermath of storms and heavy rainfall events that cause flooding. For operational use, these limitations mean the system would benefit from human oversight or

post-processing steps to verify predictions, particularly in critical decision-making contexts. Several approaches could address the class imbalance issues observed in this study.

Implementing focal loss or other weighted loss functions would force the model to prioritize learning features of underrepresented classes like water and clouds, potentially improving IoU scores for these categories without sacrificing overall accuracy. Alternatively, targeted data augmentation that oversamples water and cloud examples during training could provide the model with more learning opportunities for these challenging classes.

Expanding the training dataset to include more diverse flood scenarios would also improve generalization. The current model was trained primarily on specific geographic regions and flood types, which may limit its effectiveness in areas with different terrain, water characteristics, or atmospheric conditions. Incorporating imagery from various climates, seasonal conditions, and flood magnitudes would make the system more robust for global deployment. Future studies should also explore post-processing techniques to refine predictions. Morphological operations or conditional random fields could smooth predicted boundaries and reduce false positives in water classification. Additionally, integrating temporal data by analyzing sequences of images before, during, and after flooding events could help the model distinguish between permanent water bodies and newly flooded areas, improving specificity for disaster response.

Finally, testing the model on near-real-time satellite feeds and measuring inference speed would clarify its viability for operational deployment. Disaster response systems require not only accuracy but also rapid processing to be useful during active emergencies. Optimizing the model architecture for faster inference or deploying it on cloud computing infrastructure could make

automated flood mapping a practical tool for agencies coordinating relief efforts in the critical hours following a disaster.

Works Cited

- Ahamed, S., & Rahmim, A. (2023). Generalized Dice Focal Loss trained 3D Residual UNet for Automated Lesion Segmentation in Whole-Body FDG PET/CT Images. ArXiv, abs/2309.13553.
- Azad, R., Heidary, M., Yilmaz, K., Huttemann, M., Karimijafarbigloo, S., Wu, Y., Schmeink, A., & Merhof, D. (2023). Loss Functions in the Era of Semantic Segmentation: A Survey and Outlook. ArXiv, abs/2312.05391.
- Deb, D., & Verma, U. (2025). Leveraging Self-Supervised Features for Efficient Flooded Region Identification in UAV Aerial Images. ArXiv, abs/2507.04915.
- Hartanti, B.S., Krisnadhi, A.A., Rahadiani, L., Susanti, W.D., & Shomim, A.F. (2025). Flood-MATE: A Flood Segmentation Model in Urban Regions through Adaptation of Mean Teacher and Ensemble Approach. IET Image Process., 19.
- IPCC. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 2021.
- Mu, Y., Nguyen, T., Hawickhorst, B., Wriggers, W., Sun, J., & He, J. (2024). The combined focal loss and dice loss function improves the segmentation of beta-sheets in medium-resolution cryo-electron-microscopy density maps. Bioinformatics Advances, 4.
- O'Shea, K., & Nash, R. (2015). An Introduction to Convolutional Neural Networks. ArXiv, abs/1511.08458.

Pech-May, F., Aquino-Santos, R., Álvarez-Cárdenas, O., Arandia, J.L., & Ríos-Toledo, G.

(2024). Segmentation and Visualization of Flooded Areas Through Sentinel-1 Images and U-Net. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 8996-9008.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation.” *Lecture Notes in Computer Science*, vol. 9351, 2015, pp. 234–241.

Schumann, G., Bates, P. D., Horritt, M. S., Matgen, P., & Pappenberger, F. (2009). Progress in integration of remote sensing–derived flood extent and stage data and hydraulic models. *Reviews of Geophysics*, 47(4). <https://doi.org/10.1029/2008rg000274>

Xu, H., He, H., Zhang, Y., Ma, L., & Li, J. (2023). A comparative study of loss functions for road segmentation in remotely sensed road datasets. *Int. J. Appl. Earth Obs. Geoinformation*, 116, 103159.

Yeung, M., Sala, E., Schönlieb, C., & Rundo, L. (2021). Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Computerized Medical Imaging and Graphics*, 95.