Flood Mapping Using Contrastive Learning and Semantic Segmentation

Eliana Martha Bonalde Marcano

A thesis submitted in partial fulfillment of the requirements for the Degree of Master in Applied Mathematics

Director

Luis Núñez, PhD

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Física

Maestría en Matemática Aplicada

Bucaramanga

2025

Contents

Abstract

Title: Flood Mapping Using Contrastive Learning and Semantic Segmentation [*]

Author: Eliana Martha Bonalde Marcano [**]

Keywords: flood, mapping, constrastive learning, semantic segmentation, earth observation.

Description: Flash floods are a common natural disaster worldwide, but on many occasions, they cause many casualties, injuries, and general damage to populations. This is why identifying the areas of susceptibility is vital to prevent greater damage to first aid organizations during flood disasters. On the other hand, remote sensing generates vast amounts of data with significant potential to enhance flood prediction efforts. However, when aiming to use this data in supervised machine learning models, labeled datasets are required, which is often a difficult and costly task. As a result, its application is hindered by the scarcity of labeled data. Also, the existing information is often fragmented, and the atmospheric, geological and topographic conditions vary significantly between regions. These differences make it more challenging to perform analysis on subgroup, particularly when working with aggregated data. This study presents an extensive literature review of current methods, datasets, and results to propose a semi-supervised approach based on transfer learning. The model proposes to use representations learned through contrastive training to improve pixel-wise recognition of flooded and non-flooded areas in the segmentation task. It optimizes flood maps in four regions: Nebraska, Bangladesh, Florence, and North Alabama, using a dataset of synthetic aperture radar images for training. We perform a proof of concept to validate the performance of both the contrastive and segmentation models. Despite computational limitations, experimental results demonstrate its effectiveness, achieving a reasonably

---

[*] Research work proposal to obtain the Masters Degree in Applied Mathematics.

[**] Facultad de Ciencias. Escuela de Física. Director: Luis Núñez

good intersection over union score between predicted and actual flooded areas, achieving an average of 0.52 on the test set. These findings suggest that with adequate resources, the model could reliably identify regions affected by floods.

Resumen

Título: Flood Mapping Using Contrastive Learning and Semantic Segmentation [*]

Autor: Eliana Martha Bonalde Marcano [**]

Palabras clave: flood, mapping, constrastive learning, semantic segmentation, earth observation.

Descripción: Las inundaciones repentinas son un desastre natural común en todo el mundo, pero en muchas ocasiones causan numerosas muertes, lesiones y daños generales a las poblaciones. Por esta razón, identificar las áreas de susceptibilidad es de vital importancia para prevenir mayores daños y facilitar la labor de las organizaciones de primeros auxilios durante desastres por inundación. Por otro lado, la teledetección genera grandes cantidades de datos con un potencial significativo para mejorar los esfuerzos de predicción de inundaciones. Sin embargo, cuando se pretende utilizar estos datos en modelos de aprendizaje automático supervisado, se requieren conjuntos de datos etiquetados, lo que podría ser con frecuencia una tarea difícil y costosa. Además, la información existente está sectorizada, y las condiciones atmosféricas, geológicas y topográficas varían significativamente entre regiones. Estas diferencias hacen que hace más complicado abordar análisis en subgrupos cuando se emplean datos agregados. Este estudio presenta una extensa revisión bibliográfica de los métodos, conjuntos de datos y resultados actuales para proponer un enfoque semi-supervisado basado en transferencia de aprendizaje. El modelo propone utilizar las representaciones aprendidas mediante entrenamiento contrastivo para mejorar la precisión en la tarea de segmentación de píxeles correspondientes a zonas de inundación y no inundación. Se optimizan los mapas de inundaciones en cuatro regiones: Nebraska, Bangladesh, Florencia y Northal, utilizando un conjunto de datos de imágenes de radar de apertura sintética para el entrenamiento.

---

[*] Trabajo de Investigación de Maestría en Matemática Aplicada.

[**] Facultad de Ciencias. Escuela de Física. Director:Luis Núñez, Doctorado en Física.

Realizamos una prueba de concepto para validar el rendimiento de los modelos contrastivo y de segmentación. A pesar de las limitaciones computacionales, los resultados experimentales demuestran su eficacia, logrando una puntuación de intersección sobre unión razonablemente buena entre las áreas inundadas previstas y las reales. Estos hallazgos sugieren que, con los recursos adecuados, el modelo podría identificar de forma fiable las regiones afectadas por inundaciones.

Introduction

Floods are a recurring hydrological phenomena that are potentially destructive and are part of the dynamics of the evolution of a flow. Flooding occurs when an area or land becomes submerged in water due to an overflow of water from rivers, lakes, oceans, heavy rainfall, swift thawing of snow, the rupture of dams or levees, or a combination of multiple causes. Floods can be categorized into several major types according to their causes, characteristics, and sources of water Ruth et al. (2024). Depending on how quickly they occur, can be classified into two types, namely slow floods and flash floods. The latter has a much greater destructive power and is caused by the presence of large amounts of water in a short time; it is cataloged as one of the most common natural disasters that affects people, infrastructures and the environment. Increasing concerns about future flood hazards are linked to human-induced climate change, which is anticipated to raise sea levels and likely intensify the hydrological cycle. Additionally, socio-economic changes are expected to influence vulnerability and exposure. To adapt to these evolving conditions, it is essential to conduct analyses that can accurately predict current hazards with high spatial resolution and provide reliable projections for the future Bates et al. (2020) Ruth et al. (2024).

Taking into account the above and driven by the Sustainable Development Goals, especially 11.5 O'Connor et al. (2020) [1], it is extremely important to early incorporate information on floods.

---

[1] "By 2030, significantly reduce the number of deaths and the number of people affected and substantially decrease the direct economic losses relative to global gross domestic product caused by disasters, including water-related disasters, with a focus on protecting the poor and people in vulnerable situations"

The indiscriminate occupation of riverbanks and streams, along with population growth and associated socioeconomic activities, has exacerbated the negative impacts, including periodic increases in river levels IDEAM (2017).

Nowadays, information collected by Earth observation is used, among other things, to estimate the risk due to natural disasters, including flood monitoring, using diverse artificial intelligence approaches and methods. Among the most widely used are bivariate statistical models, artificial neural networks, support vector machines, random forests, and other machine learning algorithms have been highlighted for their ability to model flood susceptibility Islam et al. (2020), Romulus et al. (2021), Vaibhav et al. (2021), Shahabi et al. (2021), Khosravi et al. (2019), Bahareh et al. (2021). In recent years, advances in artificial intelligence have incorporated more sophisticated approaches such as deep learning semantic segmentation, self-supervised knowledge transfer, and cross modal change detection. These advancements have employed techniques like DeepLabV3+, barlow twins, and the integration of optical and synthetic aperture radar imagery Muhadi et al. (2021), Farshad & Maryam (2023), Chendeb et al. (2023), Pignato & Marino (2024), Wenqing et al. (2024).

Despite these advances, challenges remain. Lack of sufficient data, complexity of study areas, and limitations in the accuracy of flood susceptibility maps hinder the full potential of flood risk assessment. These challenges point to the need for improved tools and techniques to provide reliable information to land planners, decision makers and government agencies tasked with managing areas vulnerable to flood damage.

The success of the methodologies is highly dependent on the quantity and quality of the

available labeled data. However, it is extensive, costly and slow work, often prone to errors due to manual label generation. Additionally, requires specialized knowledge or expensive ground-based sensors Mañas et al. (2021), Jaiswal et al. (2021). In image-based data, limited labeled training samples are available not only due to difficulties with manual annotation through visual inspection, but also with difficulties with complex urban landscapes, dense cloud cover or with viewing angle limitations Hashemi-Beni & Asmamaw (2021).

Deep learning methods allow us to learn and extract, through deep layers, some useful characteristics of the data. For example, spatial patterns that help to delineate water bodies, temporal dependencies that enable the monitoring of flood evolution, and hierarchical features that capture complex relationships between water surfaces and surrounding terrains Binbin et al. (2024). Within deep learning methods, self-supervised learning has recently budded as an effective methodology to learn the underlying representations of large amounts of unlabeled data. One of the approaches of this type of learning is structured learning, which offers a simple method to encode properties in an attained representation. Constrastive learning, briefly said, is about learning through comparisons, where it can be done between positive pairs of similar inputs and negative pairs of different inputs Jaiswal et al. (2021), Le-Khac et al. (2020).

To enhance flood detection and segmentation, we propose a methodological framework where labeled synthetic aperture radar data are leveraged and integrated with these advanced methodologies to optimize pixel-level recognition of flooded and non-flooded areas. Inspired by recent studies Pignato & Marino (2024), Wenqing et al. (2024), Farshad & Maryam (2023), our approach not only aims to exploit the representational power learned through contrastive methods,

but also transfers this knowledge to improve segmentation performance.

Our initial motivation was to apply flood detection techniques to the municipality of Girón, located in the Santander department of Colombia, a region that has experienced recurrent and severe flood events. However, due to the lack of labeled data, a limitation consistent with the challenges discussed above, we opted to use the publicly available dataset provided by the ETCI 2021 competition. This data set contains images with verified flood annotations collected in Bangladesh and in specific regions of the United States, such as Nebraska, North Alabama and Florence, making it a suitable resource for the development, evaluation, and comparison of our proposed model.

This work is organized as follows. Section 1 defines the research problem by outlining its significance, context, and challenges. Section 2 provides a comprehensive review of the related literature, covering methodologies, datasets and results. Section 3 delves into the theoretical foundations behind the methods used in the proposed framework, providing the necessary background. Section 4 details the proposed framework, including data preparation and model architecture. Section 5 is based on the proof of concept, detailing the performance of the model and its effectiveness. Finally, Section 6 concludes the study, summarizing key findings and describing possible future directions.

1. THE PROBLEM AND THE RESEARCH QUESTION

Floods are considered one of the most common and destructive natural disasters that damage communities, infrastructure, and ecosystems. Considering only the study areas, we reaffirm this statement as we shall see below.

The National Center for Environmental Information reports that approximately 7.4% of the total economic losses attributed to natural disasters in the United States are due to flooding NOAA - NCEI (2024). For example, the 2019 flooding in Nebraska, triggered by a combination of heavy rainfall and rapid snowmelt, was one of the costliest inland flooding events recorded. This damage included significant losses to agriculture, roads, bridges, levees, dams, and even military assets such as Offutt Air Force Base. The vulnerability to flooding also extends to coastal regions. Florence, South Carolina, is particularly susceptible to flooding caused by hurricanes and tropical storms. In 2018, Hurricane Florence brought catastrophic flooding to the area, causing widespread devastation. This event followed another major flooding event that occurred in early 2015, which had already caused severe damage and significant disruptions to commerce and infrastructure. Extreme rainfall, over 50 cm, contributed to historic levels of flooding, affecting key transportation corridors such as I-95. NOAA - NCEI (2024)

In contrast, Bangladesh, due to its geographical location and climatic conditions, faces even more severe consequences. The country regularly experiences economic losses, large-scale displacement of people, and destruction of farmland and crops FloodList (2024).

The above highlights the social and economic vulnerabilities in these regions, underscoring

the need for effective tools for flood detection and mapping. In the literature review, we identified several methodologies proposed to address this problem, including traditional remote sensing approaches and techniques based on artificial intelligence. While significant progress has been made, we still face significant limitations. For example, a heavy reliance on labeled data, difficulties in transferring knowledge to regions with distinct geographic conditions, the complexity of models and evaluations often restricted to standard metrics, among others.

In this context, the motivation for this study is how can we develop a methodology that reduces the dependence on labeled data by leveraging contrastive learning to transfer learned representations to a semantic segmentation model? In addition, with this methodology, using pixel-level labels from synthetic aperture radar imagery provided by NASA, we seek to facilitate knowledge transfer to regions with different geographic and environmental conditions.

Building on this motivation, we find the need to address the limitations of existing methodologies and contribute to more timely, cost-effective, and, we hope, more accurate flood detection and mapping solutions. These advances could directly impact disaster risk reduction efforts, supporting the protection of vulnerable communities and the preservation of critical infrastructure.

Considering this question and its relevance, this study aims to propose a methodology that, using synthetic aperture radar imagery, combines the strengths of contrastive learning and semantic segmentation in contexts with scarce labeled data through transfer learning. We believe that if we can validate the effectiveness of this framework using robust evaluation metrics, we can ensure its applicability in real-world flood mapping scenarios, providing information for future advancements in this domain.

2. LITERATURE REVIEW

In this section, we conduct a comprehensive review of the literature on flood detection, starting with the most recent and significant studies. This chronological approach will provide context for the current state of the art and emerging trends in the field. This review covers methods, techniques, data used and the selected regions for study. The basic criteria for the selection were the integration of diverse data sources, consideration of conditioning factors, application of different advanced artificial intelligence techniques and use of standardized evaluation metric. We examine their strengths and limitations to provide a balanced perspective on their impact in addressing flood detection problems.

Recent studies have taken into consideration certain conditioning factors that influence the occurrence of floods. Among them, we have topographic factors such as altitude, slope, aspect, curvature and topographic humidity index; hydrological factors, like stream power index and geological factors: soil type and lithology. Apart from that, the relationship with land use is also studied, as has already been mentioned above. Different approaches and automated methods have been developed for flood detection and mapping.

For example, in 2024, a study consisted of using the barlow twins contrastive self-supervised learning algorithm to derive effective visual representations of flood affected areas from bitemporal remote sensing data Wenqing et al. (2024). This approach, which employs a pretrained CS-DeepLabV3+ network with dual attention mechanisms, demonstrated superior performance in flood extraction compared to traditional methods. Another study leveraged self-supervised learning

to transfer knowledge from optical images to synthetic aperture radar images. A segmentation network trained on optical images is adapted for synthetic aperture radar images through knowledge transfer. Performance is evaluated using metrics such as IoU and F1 Score, showing significant improvements in segmentation accuracy compared to traditional supervised methods Pignato & Marino (2024). Bhattarai et al. (2024) focuses on improving flood susceptibility mapping in the Gandak River Basin, a transboundary region spanning China, Nepal and India. The study employs four machine learning techniques, long-short-term memory, random forest, artificial neural network, and support vector machine, to assess flood susceptibility in this data-scarce region. The study uses open source spatial datasets for modeling, and the results highlight that artificial neural network and support vector machine techniques outperformed other methods in predicting flood susceptibility maps.

Additionally, the integration of multiple data sources has also proven beneficial for enhancing flood detection in complex environments. For instance, in 2023, MSflood leverages remote sensing technologies, using optical imagery, radar imagery and digital elevation models to provide more reliable flood monitoring by processing diverse datasets to create clearer and more accurate flood maps Chendeb et al. (2023). In another study, comparative analyses of real-time semantic segmentation networks were conducted to evaluate their efficiency and accuracy in flood area segmentation. This study compared networks like U-Net, SegNet, and DeepLabV3+, using metrics such as IoU and overall accuracy, highlighting the importance of balancing precision and speed for rapid disaster response Farshad & Maryam (2023).

Several studies have taken advantage of the images provided by NASA, also used in this

work. Among them is the winner of a competition, team Arren from Xidian University, that employed advanced deep convolutional neural networks, combining the U-Net and DeepLabV3+ architectures, achieving an IoU of 76.81% NASA Impact Team and IEEE GRSS Committee (2021). However, more recent studies have developed an approach that, using the same images, propose the use of EfficientNet-B7 as an encoder for feature extraction. This model has been evaluated with metrics such as IoU, F1-score, recall, accuracy and precision, achieving an IoU of 84.77%, which represents a significant improvement Mohaddeseh & Reza (2023).

In the year 2021, in Muhadi et al. (2021) employed two powerful techniques, DeepLabv3+ and SegNet, to estimate water levels from surveillance camera images. However, the DeepLabv3+ network, utilizing ResNet-18 as the backbone, outperformed the SegNet model, achieving over 90% accuracy and IoU in segmenting water areas. This result highlights the effectiveness of using surveillance data for precise flood estimation. Hashemi-Beni & Asmamaw (2021) proposed a convolution neural network method for mapping flooded areas by means of optical images, applying a data augmentation method. When comparing the results obtained with and without it, it is evident that with the increase of data, the precision of the extraction of the flooded areas improves. Additionally, they combine this method with a digital elevation model approach to detect flooded areas under vegetation that are not visible in the images. Shahabi et al. (2021) introduced a new modeling approach based on deep belief network with a backpropagation algorithm optimized by the genetic algorithm. They selected 11 conditioning factors for flooding from topographic, hydrological, geological and land cover factors and used 194 flood and non-flood locations to build the model database.

Bahareh et al. (2021) used artificial neural networks, deep learning neural networks and networks optimized by particle swarm optimization to predict and estimate areas susceptible to future flooding. They prepared two types of data sets. The first one consists of a past flood inventory map of 128 historical Brisbane flood areas dating back to 2001 drawn from high resolution aerial photography after the flood event. The second dataset considers 13 conditioning factors such as the altitude, slope, aspect, curvature, topographic wetness index, stream power index, sediment transport index, soil type, lithology and land use. Romulus et al. (2021) used four ensemble models, based on images provided by remote sensing sensors available on Google Earth and collected data composed of polygons of torrential areas. They evaluated flash flooding in a small catchment from Romania, combining bivariate statistics, deep learning neural networks and alternative decision trees. The factors considered included slope angle, topographic moisture index, topographic position index, contour curvature, convergence index, wind orientation, land use, lithology and hydrological group of soils. Muñoz et al. (2021) employed convolutional neural networks and data fusion, integrating multispectral Landsat imagery, dual-polarization synthetic aperture radar data, and a digital elevation model. The idea of consolidating this information was to leverage the strengths of each data source and achieve improved performance in producing flood maps at moderate spatial resolution (30 m). Radar backscatter data, unlike multispectral imagery, are not limited by adverse atmospheric conditions (for example, shadows or cloud formation), since the radar antenna can emit and receive oscillating signals even during night conditions. Vaibhav et al. (2021) employed two segmentation architectures, SegNet and Unet. This study used the Sen1Floods11 dataset with different combinations of synthetic aperture radar bands for near real-

time flood mapping. The dataset is divided into two parts, one containing data related to floods and the other one containing bodies of surface water. To avoid manual labeling, they propose to use a set of weak labels assigned by means of algorithms.

In 2020, Islam et al. (2020), participed in a contest organized by the technical committee of the IEEE Geo-Science and Remote Sensing Society to perform flood detection. The study employed a dataset containing SPOT-5 images and data from the European Remote-Sensing Satellite-1, employing synthetic aperture radar technology to capture the event both before and after the flooding that occurred in Gloucester, United Kingdom, in 2000. They implemented the semi-supervised domain adaptation method, a combination of convolutional neural network and a semi-supervised Domain Strategy Adaptation. They compare the performance together with other conventional methods used in competition, for example, multi-Layer perceptron and support vector machine. Bui et al. (2020) examined three swarm intelligence optimization algorithms, namely Gray Wolf optimization, grasshopper optimization algorithm and social spider optimization algorithm on a dataset of 11 predictor variables. The results were compared to various methods with common statistical indicators. Dodangeh et al. (2020) worked on hybrid intelligence models based on meta-optimization of support vector regression and group method of data handling using different heuristic meta-algorithms, that is, the genetic algorithm and search harmony. They identified 132 historical flood points, 132 non-flooding points and a geospatial database, but with 9 conditioning factors, namely, degree of slope, aspect, elevation, plane curvature, profile curvature, distance to the river, land use, lithology and rainfall.

In previous years, we also found related work, although in smaller numbers. In 2019, the

article Mahfuzur et al. (2019) presented a hybrid approach that combines machine learning techniques and multi-criteria decision analysis to assess flood susceptibility in Bangladesh. The methods used include artificial neural networks, logistic regression and the analytical hierarchy process, applied to a dataset that encompasses factors such as elevation, slope, land use and proximity to water bodies. The main metrics used to evaluate the performance of the models were prediction accuracy and the receiver operating characteristic curve.

In 2018, Mosavi et al. (2018) published a comprehensive review that provides an overview of flood prediction models based on machine learning. This article highlights the most promising methods for both short-term and long-term predictions, and it evaluates various machine learning algorithms in terms of robustness, accuracy, effectiveness and speed. Key strategies for improving these models, such as hybridization, data decomposition, algorithm ensemble and model optimization, are identified.

Finally, other studies have continued to explore and compare the effectiveness of various ML algorithms in flood susceptibility mapping. These works emphasize that advanced models and hybrid approaches can offer greater accuracy compared to traditional methods, particularly when combined with spatial data and geographic factors. For example, the literature review conducted by Hammed et al. (2024) underscores the importance of integrating optimization strategies and data management techniques into machine learning models to enhance flood prediction accuracy. Similarly, Pham et al. (2021) analyzes the performance of deep learning algorithms, suggesting that these can outperform conventional machine learning models in flood susceptibility modeling. Additionally, Teymoor et al. (2023) provides a detailed comparison of multiple machine learning

algorithms, concluding that models like the cascade forest model and ensemble techniques like stacking offer higher accuracy and stability in predictions. Lastly, Prasad et al. (2022) highlights the potential of hybrid approaches in geospatial modeling for flood susceptibility, showing that the combination of different machine learning techniques with spatial data significantly improves the accuracy of flood risk maps.

3. THEORETICAL FOUNDATIONS

In the previous section, we presented a review of existing studies on flood detection, which allowed us to identify key techniques to address some of the current limitations, particularly the strong reliance on labeled data and difficulties in transferring knowledge to regions with different geographical conditions. In this section, we introduce the theoretical foundation that forms the backbone of our proposed framework. It is structured around three key components, namely contrastive learning, semantic segmentation, and transfer learning between these two methods.

3.1. Constrastive Learning

In the following, we discuss contrastive learning, a method in the realm of self-supervised learning that has proven to be a versatile and effective technique, particularly in scenarios where labeled data is scarce or difficult to obtain.

Self-supervised methods offer a powerful alternative to overcome challenges such as the extensive process of manual labeling of data samples, which tends to be very expensive, slow and error-prone. While a vast amount of data from various sources is available through Earth observations, a major downside is the lack of sufficient labeled data. In studies related to floods, we do not have enough annotations to train and validate the methods. Additionally, while we have labels with broad coverage, they are often very aggregated, and the existing information is sectorized, with varying atmospheric, geological and topographic conditions. These methods, including contrastive learning, learn effective representations with minimal labeled data. Contrastive learning, in particular, focuses on comparing pairs of input samples to learn useful features. The core idea

is to bring similar samples closer together in the embedding space while pushing dissimilar samples farther apart. This is achieved through a discriminative model that measures how close two embeddings are based on their similarity. Le-Khac et al. (2020), Jaiswal et al. (2021).

One of the most widely used approaches is the Simple Contrastive Learning of Representations (SimCLR) Chen et al. (2020), which consists of four main components:

1. The creation of multiple transformations of the same image to build noise-invariant representations. These transformations can include color distortion (e.g., changes in brightness and contrast), geometric transformations, cropping and filling. Such data augmentation provides diverse views of the same underlying data, helping to overcome the limitations of conventional supervised learning.

2. A base encoder that maps inputs to a general representation space, learning to capture a powerful and generalized representation of the data, i.e., extracts representation vectors from augmented data examples.

3. A transformation head that adjusts feature embeddings to align with the chosen similarity metric, allowing it to be efficient and effective in supporting the measurement of distances between samples, i.e., mapping representations to the space where the contrastive loss is applied.

4. A contrastive loss function that minimizes the distance between a query and its positive key, thereby encouraging similar pairs to be close in the embedding space. Simultaneously, it maximizes the distance from negative keys, ensuring that dissimilar pairs are well-separated.

In SimCLR framework, a minibatch of N examples is randomly sampled and two separate data augmentation operators are sampled from the same family of augmentations (denoted as $t \in T$ and $t' \in T$) and applied to each example, resulting in $2N$ augmented data points. These augmentations produce two correlated views for each example. Each view is then processed by an encoder network, producing a vector of unique representations making each element distinct. After encoding, these representations are further processed by a projection head. Instead of explicitly sampling negative examples, the other $2(N-1)$ augmented examples within the minibatch are treated as negative examples. The model is trained using a contrastive loss that encourages the representations of positive pairs (augmented views of the same example) to be closer while pushing apart those of negative pairs (different examples in the minibatch). After training, the projection head $g(\cdot)$ is discarded, and the encoder $f(\cdot)$ along with the representation h are used for downstream tasks, as illustrated in Figure 1 from Chen et al. (2020).

The effectiveness of contrastive learning is primarily assessed by a contrastive loss function, which is a key differentiator between contrastive methods and other representational learning approaches. In general, each contrastive loss function moves positive pairs closer together and negative samples farther apart, reinforcing correct matches and distinguishing different examples. Following this approach, given a set of examples, including a positive pair, the goal of the contrastive prediction task is to correctly identify the matching pair for a given input Chen et al. (2020), Le-Khac et al. (2020).

Contrastive learning shows its true potential when used as a complement to other methods, rather than in isolation. It is generally applied during the pretraining phase of the model, optimizing

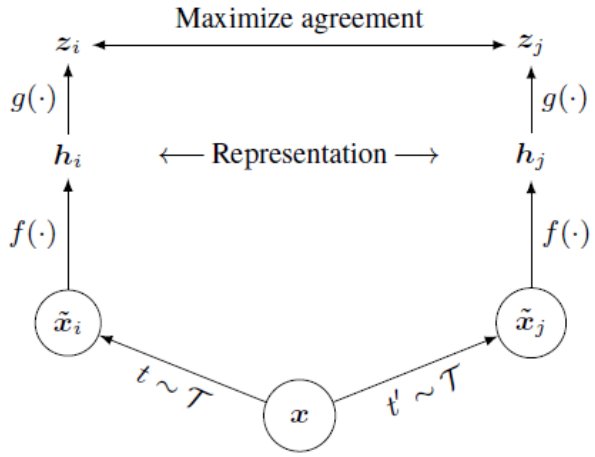


Figure 1. Contrastive Learning Process Implemented in SimCL. The diagram illustrates the main steps, where an input image x is augmented twice to create two views (x_i, x_j). These views are then processed by an encoder $f(\cdot)$ to extract features (h_i, h_j), followed by a projection head $g(\cdot)$ to obtain representations (z_i, z_j). The training objective is to maximize agreement between z_i and z_j in the latent space while minimizing similarity with representations from other samples. Adapted from Chen et al. (2020)

the representation of the data and facilitating its adaptation to new tasks. In our case, this approach is specifically employed in transfer learning towards semantic segmentation, which we describe below.

3.2. Semantic Segmentation

Semantic segmentation is a critical task in computer vision that involves classifying each pixel in an image into a predefined category. This process provides a comprehensive scene description, including details about the object category, location and shape Zhang et al. (2018). It is crucial for understanding of the structure and content of images and has significant applications in environmental monitoring as well as fields as autonomous driving, medical imaging.

The most advanced semantic segmentation approaches are predominantly based on the fully

convolutional network framework, which was a fundamental advancement in this area. The fully convolutional network introduced the concept of performing pixel-to-pixel classification in an end-to-end manner through a fully convolutional architecture. Since then, deep convolutional networks have become dominant in semantic segmentation, with various studies advancing the field through high-resolution representation learning, contextual aggregation and other techniques Zhang et al. (2018), Yuan et al. (2019), Xie et al. (2021).

Traditionally, semantic segmentation relies heavily on supervised learning, which requires large amounts of labeled data. As previously discussed, obtaining such labeled data can be labor-intensive, costly and time-consuming. However, recent advancements in semantic segmentation have focused on incorporating new techniques and methodologies that improve accuracy and efficiency, reducing the dependence on extensive labeled datasets, where the choice of architecture and techniques depends primarily on the ability to work with limited labeled data, but also on other specific requirements of the application, such as the need for high accuracy and efficiency.

The integration of these advanced models with self-supervised learning, such as contrastive learning, and data augmentation techniques can further enhance their performance and applicability. Representations learned through contrastive pretraining can be leveraged and, through transfer learning, improve the ability of the model to generalize to regions with diverse geographic and environmental conditions.

3.3. Transfer Learning between Contrastive Learning and Semantic Segmentation

Inspired by the way humans learn, where we can learn faster, easier and more efficiently using knowledge from previously learned tasks, transfer learning in Deep Learning models applies

this concept. The core idea is to use knowledge from one domain or task, called the source, to improve the performance of a model on another domain or task, called the target Farahani et al. (2021). During the pretraining phase, the model discovers the best representation of the problem by learning the most important general features such as edges, colors and textures. Transfer learning could be used in many scenarios, for example, when there is insufficient labeled data to train your network from scratch, when a pretrained network on a similar task is available or, as in our case, when the source and the target have the same input.

Contrastive learning is utilized to extract meaningful representations from labeled segmented data, capturing essential variations and features in the data through self-supervised tasks. The contrastive model weights, which encapsulate these representations and features, are then used as inputs for a semantic segmentation model, which is trained to assign labels to each pixel in high-resolution images. This process seeks to take advantage of the pixel-level annotations already available, which improves the performance of the segmentation model by providing it with enriched and pretrained representations.

By integrating these methods, we aim to address the specific challenge of creating flood susceptibility maps in data-scarce environments. The idea of this dual approach is not only to improve the quality of extracted features but also to maximize the utility of the limited labeled data available, providing a scalable and efficient solution for real-world applications.

Previous studies have demonstrated the potential of transfer learning in similar contexts. For instance, prior studies have demonstrated improvements in flood extraction and segmentation accuracy by leveraging self-supervised methods and pretrained encoders, such as CS-DeepLabV3+

combined with the barlow twins algorithm, or EfficientNet-B7 for feature extraction Wenqing et al. (2024), Pignato & Marino (2024), Mohaddeseh & Reza (2023). Building on these advances, our work focuses on adapting and extending this methodology by integrating contrastive learning and semantic segmentation models through a transfer learning pipeline tailored to flood mapping in specific regions.

4. METHODOLOGY

This section presents the proposed methodology detailing the approach we have designed with all the key aspects. We will describe the study regions, the dataset, and the architectures employed in the transfer learning strategy, which enables the model to capture features of the data, such as spatial patterns and texture variations in the images.

4.1. Study Areas

Although our initial goal was to apply the model to the municipality of Girón, located in the Santander department of Colombia, the lack of annotated data led us to adopt an alternative dataset. Consequently, we used the ETCI 2021 dataset, which provides detailed and relevant data on flooding NASA Impact Team and IEEE GRSS Committee (2021). The regions included in this database and analyzed in this study were Bangladesh and specific regions of the United States, such as Nebraska, Northern Alabama and Florence.

All of these regions are affected by historical and current flooding problems, which present significant challenges. In Bangladesh, frequent floods caused by its geographical location and climatic conditions severely affect communities, damage infrastructure and significantly impact the economy FloodList (2024). Similarly, regions in the United States have faced substantial losses in economy, infrastructure, agriculture, and transportation due to recurrent flood events, intensified by hurricanes, tropical storms, heavy rainfall, and snowmelt NOAA - NCEI (2024).

4.2. Data

The data used in this research is sourced from NASA Impact Team and IEEE GRSS Committee (2021). This dataset provides extensive remote sensing imagery and associated data for critical region affected by floods. It includes bitemporal remote sensing imagery, both optical and Synthetic Aperture Radar with VV and VH polarizations, as well as topographic and hydrological data. It is important to clarify that we do not use simulated or artificially generated data, the terminology "synthetic aperture radar images" comes from real satellite observations obtained via an active remote sensing technique. An instrument emits a pulse of energy and records the amount of energy reflected after its interaction with Earth. The spatial resolution of radar imagery depends on the wavelength of the sensor and the physical length of its antenna, at a given wavelength, a longer antenna yields finer detail. Because deploying a truly enormous antenna in space is impractical, a large effective aperture is synthesized by coherently combining multiple measurements collected by a smaller antenna along its flight path, thereby achieving high spatial resolution NASA EARTHDATA (2025).

The dataset comprises high-resolution imagery, where each image is tiled into 256×256 pixel segments, with a total of 43,805 tiles distributed across training (33405 tiles) and validation (10400 tiles) sets. The training data is organized into 29 root folders, where each folder contains subfolders for VV, VH, flood_label, and water_body_label, with 2068 files in each. Flood labels indicate the presence (1) or non presence (0) of floods, while water body labels differentiate between water bodies (1) and land (0). As mentioned above, the dataset covers four regions whose respec-

tive areas are: Bangladesh (7,150 sq. km.), Nebraska (1,741 sq. km.), North Alabama (13,789 sq. km.) and Florence (7,197 sq. km.) NASA Impact Team and IEEE GRSS Committee (2021).

4.2.1. Data Processing. NASA provided an initial data processing framework designed to prepare the dataset for training and validation of the flood detection model. In this study, we retained their approach to addressing data quality variability and converting images into a format compatible with our programming language, Python. We then adapted and refined these procedures to align with our research objectives. A detailed outline of both steps suggested by NASA and our modifications is provided below.

- Creating training and validation sets. We organized the data paths for VV images, VH images, flood labels and water body labels. This involves sorting and matching the corresponding VV and VH images, as well as the flood and water body masks. The dataset is then split into training and validation sets. A representative sample of these sets is shown in Figures 2 and 3.

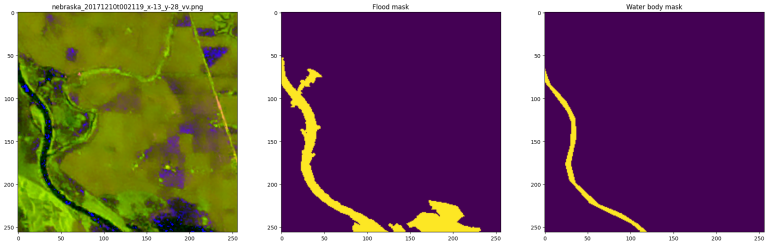


Figure 2. Example of a sample from the training set showing an image of the Nebraska region, the corresponding flood mask indicating flooded areas and the waterbody mask highlighting permanent water bodies.

- Dataset setup for machine learning. Since the validation set provided in the dataset does not

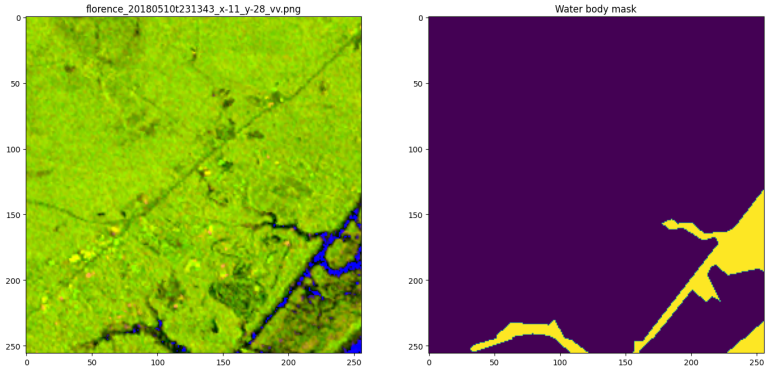


Figure 3. Sample image from the validation set illustrating an image of the Florence region and its corresponding waterbody mask. Unlike the training set, the validation set does not include flooding masks, as these are the results the model is expected to generate.

include flood labels, we followed the suggestion to split the original training dataset, which contains flood masks, into smaller training and development sets. This approach allows for a fair evaluation of the model during training. The splitting method is based on regions, simulating the conditions of the validation set where there are no matches regions between the training and validation data.

- Image Processing for the Model. At this stage, the images are then prepared and processed for use in the machine learning model. This involves the following steps.
  - The VV and VH images are read from their respective paths. These images are normalized by dividing the pixel values by 255.0 to scale them between 0 and 1.
  - The VV and VH images are converted to an RGB format using a specific function, where the ratio of VH to VV is used to generate the third channel.

The following steps are specific to the proposed model.

- To improve the robustness and generalization of the model, we applied data augmentation techniques, using a Python library for image augmentation Buslaev et al. (2020). The techniques we use are as follows:
  - Horizontal and vertical flip. Inverts the image from left to right and from top to bottom, respectively.
  - Random resized crop. Randomly crops a portion of the image.
  - Random brightness contrast. Simultaneously modifies image brightness and contrast by varying lighting conditions.
  - Gaussian blur. Applies a Gaussian filter with a random kernel size and sigma value, smoothing the image by reducing noise and details.
- The processed images and their corresponding labels are structured into a format suitable for model training. For the contrastive learning model, each image x is paired with two augmented versions x_i and x_j, together with the corresponding flood masks.

The final dataset consists of a set of images, original and augmented, along their respective masks, all of which will be considered positive pairs for contrastive learning tasks. Figure4 shows two positives obtained from an image selected from the data set. In the first transformation, we observe flips in both directions and Gaussian blur, while in the second we see only a vertical flip, along with adjustments to brightness and contrast.

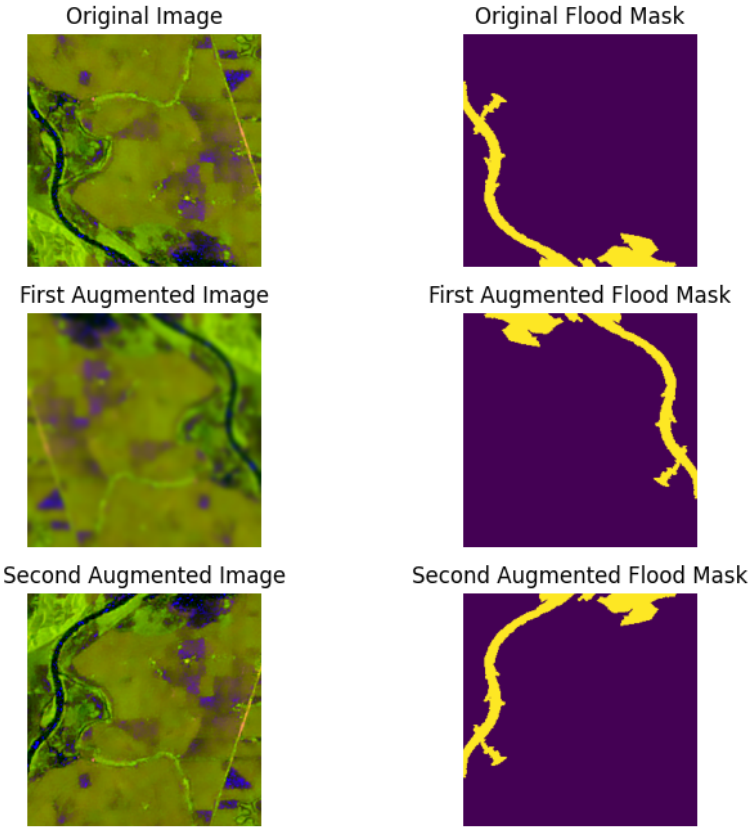


Figure 4. The figure presents a representative sample of the training data set. The first row shows the original composite image together with its corresponding flood mask. Subsequently, after applying two random magnifications to the image and mask, the second row shows a transformation consisting of horizontal and vertical flipping, together with Gaussian blur. While the third row has a different transformation, where we only see a vertical flipping and a modification of brightness and contrast.

4.3. Model Architectures

After exploring different combinations of models for flood mapping, we describe below the final architecture. It mainly composed of the contrastive model and the segmentation model. As mentioned in section 3, both are integrated under the transfer learning framework, where the pre-trained weights of the contrastive model are transferred as inputs to the segmentation model.

4.3.1. Contrastive Model. In subsection 3.1, we examined the SimCLR approach, which serves as the basis for the method followed here. After processing the images, we obtained two versions $\tilde{x_i}$ and $\tilde{x_j}$ of each image x with the set of augmentations described previously. Each version is passed to the encoder, $f(x)$, based on the Residual Networks family of architectures. For the final execution, we selected ResNet50 as the encoder. This step extracts feature maps, resulting in $h_i = f(\tilde{x_i})$ and $h_j = f(\tilde{x_j})$. The encoder is previously trained, but we removed the last layer, which was originally designed for another supervised classification task, leaving the general features that the model has learned from the images.

The feature maps are passed through a projection head to reduce their dimensionality. The projection head consists of two linear layers with ReLU activations that introduce nonlinearity and batch normalization that stabilizes training by normalizing the outputs of each layer. The projection head transforms the encoder output h into a lower-dimensional representation z, which is then L2-normalized. The resulting vectors $z_i = g(h_i)$ and $z_j = g(h_j)$ are compared using cosine similarity, where the objective is to project similar pairs of (positive) images closer together, while different pairs (negative) are separated.

The optimizer determines how the model parameters are updated based on the gradients computed during each iteration. In this study, we employed the Adaptive Moment Estimation (Adam) optimizer due to its efficiency and ability to experiment with different hyperparameters, such as learning rate and weight decay. The learning rate controls the step size of parameter updates during training, while weight decay acts as a regularization technique by penalizing large

parameter values to mitigate overfitting. Through validation, we found that the best results were obtained with a learning rate of 0.0001, along with a weight decay of 0.00001, which together contributed to improved generalization and training stability.

The batch size defines how many image sets will be passed in each iteration. We set the batch size to 64, a value chosen considering both computational limitations and model stability during training. Smaller batches tend to introduce noisy gradient estimates and slower convergence, whereas larger batches can improve gradient estimation but require considerably more memory. In our case, the architecture involves a deep network that combines ResNet50, used as the encoder for contrastive learning, and UNet for segmentation. This results in a large number of parameters being processed, with significant memory consumption. Moreover, an epoch corresponds to passing the entire training dataset through the neural network. We set the number of epochs to 20. We further see, in figure 6, that the loss functions tend to stabilize over epochs.

When sufficient computational resources are available, it is recommended to increase both the batch size and the number of epochs, thus allowing the model to continue its convergence process and achieve higher performance. Previous work supports this behavior, for example, Le-Khac et al. (2020) reports that contrastive learning performance benefits with the incorporation of multiple negative samples and extended training times. On the other hand, Chen et al. (2020) compares training for 100 to 1000 epochs using batches ranging from 256 to 8192 samples, and observes that the performance differences between different batch sizes tend to decrease as training time is prolonged.

4.3.2. Contrastive Loss Function. We use a supervised variant of the Normalized Temperature-scaled Cross Entropy Loss (NT-Xent). Unlike the traditional NT-Xent loss, which is unsupervised, our approach incorporates label information to explicitly define positive and negative pairs. We formulate the contrastive prediction task by sampling a minibatch of $N = 64$ examples. Each example is augmented twice, resulting in $2N = 128$ data points. The loss is calculated as follows:

$$\ell_{i,j} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \mathbf{1}_{[y_k=y_i]} \exp(\text{sim}(z_i, z_k)/\tau)}$$

where $sim(z_i, z_j)$ is the cosine similarity between ℓ_2 normalized vectors z_i and z_j, defined as

$$sim(z_i, z_j) = \frac{z_i^T z_j}{\|z_i\|\|z_j\|}.$$

Here, τ is a temperature parameter that controls the sensitivity of the similarity measure, and we have set it to 0.3. The original NT-Xent loss formulates contrastive learning as a $(K+1)$-way softmax classification problem, considering only augmented views as positive pairs. In our approach, we incorporate label-based supervision, ensuring that pairs with the same label contribute to the numerator of the loss function, while all other pairs contribute to the denominator. The indicator function $\mathbf{1}_{[y_k=y_i]}$ ensure that only samples sharing the same label with z_i are treated as positives. This modification reinforces class separation by maximizing similarity between positive pairs and minimizing it for negative ones, ultimately improving the quality of the learned representations. We shown the detailed configuration of the Contrastive Learning phase in Table 1.

Table 1
Contrastive Learning Phase

Component	**Details**
Encoder	ResNet50, last layer removed.
Projection Head	Two linear layers with ReLU and BatchNorm.
Loss Function	Supervised NT-Xent. Temperature $\tau = 0.3$.
Optimizer	Adam. Learning rate = 0.0001, weight decay = 0.00001.
Batch Size	64
Epochs	20
Augmentations	Two augmented views per image.

4.3.3. Semantic Segmentation Model. In the segmentation model, we use the UNet architecture with a ResNet50 encoder. The encoder is initialized with weights obtained from our contrastive learning model. These pretrained weights encapsulate high-level semantic representations that were learned from the image set, providing a robust initialization for the segmentation task. Unlike the contrastive model, which generates two augmented views per image to facilitate training, the segmentation model applies only a single set of deterministic transformations to the images.

In this architecture, we perform binary segmentation on standard RGB images. To effectively integrate contrastive representations, the encoder extracts feature maps from the input image, while the decoder progressively upsamples these features to produce a segmentation mask with the same spatial dimensions as the input. To improve regularization and enhance training stability, a dropout layer with a dropout rate of 0.2 is applied to the output of UNet, followed by batch normalization to normalize the activations.

We conducted segmentation training with a batch size of 16 using the Adam optimizer with

Table 2
Semantic Segmentation Phase

Component	**Details**
Architecture	UNet. ResNet50 encoder.
Loss Function	Binary Cross-Entropy.
Output Activation	Sigmoid. Threshold = 0.5.
Optimizer	Adam. Learning rate = 0.0001.
Batch Size	16
Epochs	20
Regularization	Dropout, rate = 0.2. Batch Normalization.
Input Type	Standard RGB images.

a learning rate of 0.0001 over 20 epochs. The classification task is binary. For each pixel, the model generates a probability map using the sigmoid function, which outputs a value between 0 and 1. A value close to 1 indicates the "presence of flooding", while a value close to 0 indicates the "no presence of flooding". To perform the final classification, we apply a threshold of 0.5, classifying pixels above this value as flooded. Similarly, we detail the main training parameters and architecture of the Segmentation Segmentation model in Table 2.

4.3.4. Transfer Learning. Our learning transfer strategy implements a full fine-tuning method, where the knowledge acquired by the pretrained contrastive model is leveraged and applied to the flood segmentation task. Specifically, only the ResNet50 encoder from the contrastive model is transferred to the segmentation model, removing the projection head used during the contrastive phase.

During the segmentation training phase, all encoder parameters are updated without constraints, keeping the gradients active in all their layers. This strategic decision not to freeze any

Table 3
Transfer Learning Strategy

Component	**Details**
Strategy	Full fine-tuning (no frozen layers).
Transferred Block	Only the ResNet50 encoder.
Projection Head	Removed before transfer.
Update Policy	All encoder parameters updated during segmentation training.

layer allows the model to progressively refine both low-level features (edges and textures) and high-level features (spatial context and semantic relationships), which were initialized with contrastive pretraining weights.

It should be noted that no additional experiments were conducted with partial or total encoder freezing configurations, however, this remains open for future work to enable systematic comparisons. Finally, we detail the transfer of learning strategy in the Table 3.

In summary, Figure 5 illustrates the complete flow of the transfer learning process described in this section. It encompasses dataset preprocessing, feature extraction using contrastive learning, and integration of these representations into the segmentation model. This approach aims to improve segmentation performance by providing additional context, which is particularly valuable in challenging scenarios such as flood detection in synthetic aperture radar images.

4.4. K-Means Clustering

Since the contrastive model alone does not produce direct predictions, it was necessary to implement a complementary strategy to evaluate whether it was effectively learning meaningful representations before proceeding to the segmentation stage. To this end, we incorporated an

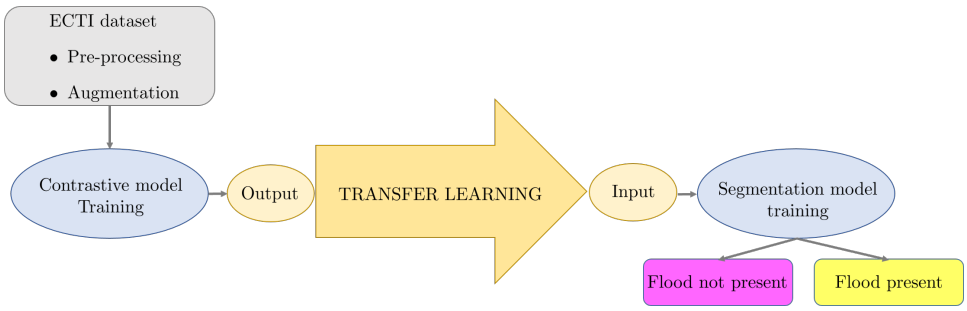


Figure 5. Learning Transfer Process from SimCLR to Semantic Segmentation for flood mapping. The diagram highlights the steps involved, from preprocessing the ETCI dataset and training the contrastive model, to transferring learned features and training the segmentation model to detect flood prone areas.

unsupervised analysis using the K-Means clustering algorithm, applied directly to the projected representations generated by the contrastive model after training.

The K-means clustering algorithm is a machine learning technique used to group data into k distinct groups. It works iteratively to divide a dataset into groups, minimizing the variance within each group. The centroids are at the midpoints around which samples are clustered, and are determined by the algorithm. It involves minimizing the euclidean distance between each point in a cluster and its assigned centroid. The algorithm iteratively adjusts the centroid positions and reassigns points to clusters until convergence is reached. At this point, changes in centroid positions or point assignments stabilize, ensuring optimal clustering.

This method is particularly useful for exploring patterns in the data, such as grouping similar features or detecting underlying structures. In this context, the use of K-means plays a role in tasks like feature extraction and image segmentation, both of which are critical in our study. It provides insight into how the contrastive model separates the data into meaningful groups, potentially corresponding to distinct classes, such as flooded and non-flooded areas.

If the identified clusters reflect a clear separation between classes, this could indicate that the model is capturing relevant patterns from the processed images. Some of these patterns are intensity differences, textural features, edge and boundary information, spatial relationships geometric patterns, contextual features, seasonal or temporal trends, and reflective anomalies. In contrast, a lack of clear separation could suggest the need to tune hyperparameters or the training pipeline.

4.5. Evaluation Metric

The metric we primarily use is the Intersection over Union (IoU), the standard metric for semantic segmentation tasks to analyze the similarity between samples. It basically represents the ratio between intersections and unions

$$IoU = \frac{A \cap B}{A \cap B}$$

where we measure the overlap between the predicted segmentation (A) and the ground truth (B). The intersection is the area where the predicted and true labels match; the union is all the areas predicted or labeled as the positive class.

For binary classification, we first apply the sigmoid function to the output of the model, which gives us probabilities for each pixel, indicating the probability that the pixel belongs to some flood class. We then use a threshold (0.5) to convert these probabilities into binary predictions.

The IoU has scores ranging from 0 to 1, the score will be high if there is a lot of overlap between the predicted segmentation and the ground truth, i.e., there is better performance of the

model. In contrast, if the overlap is low, it will result in a low IoU score and suggests that the model is not properly capturing the relevant areas. Clearly, an IoU score of 1 indicates a perfect match, while a score of 0 means that there is no overlap between the boxes.

In addition to IoU, we evaluate segmentation performance using other popular metrics, namely, Pixel Accuracy, Precision, and Recall. Pixel Accuracy focuses on how many pixels were correctly classified, both flood present and flood not present. Measures the ratio between the number of correctly classified pixels and the total number of pixels in the image. In contrast, Precision and Recall focus specifically on the flooded class. Precision is defined as the number of true positives divided by the total number of pixels predicted as positive, indicating the proportion of predicted flooded pixels that are actually flooded. Recall, on the other hand, is the number of true positives divided by the total number of actual flooded pixels, reflecting the ability of the model to capture all flooded areas.

All these metrics are calculated using a confusion matrix that compares the predicted and true classes for each pixel. In our implementation, IoU is calculated for the whole batch, while Pixel Accuracy, Precision and Recall are calculated for each individual image. The components of the confusion matrix are: true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN).

In the proposed model, these are simply calculated as:

$$IoU = \frac{TP}{TP + FP + FN + s}$$

$$PixelAccuracy = \frac{TP + TN}{TP + TN + FP + FN + s}$$

$$Recall = \frac{TP}{TP + FN + s}$$

$$Precision = \frac{TP}{TP + FP + s}$$

where each variable represents the following:

- TP, pixels correctly classified as flood present.
- TN, pixels correctly predicted as flood not present.
- FP, pixels incorrectly predicted as flood present when they belong to flood not present.
- FN, pixels incorrectly predicted as flood not present when they belong to flood present.
- s, smooth term to avoid division by zero.

5. RESULTS

Next, we present the results of our proof of concept by analyzing the performance of the model. It starts with the loss function in the contrastive model, where we examine how its values change over epochs during training and validation, analyzing trends such as convergence, fluctuations, or potential overfitting. Subsequently, we present the performance of the semantic segmentation model through the IoU metric, which allows us to better understand the quality of the pixel-by-pixel predictions in the context of binary segmentation. Finally, we attach some images that illustrate the predictions made by the model.

5.1. Contrastive Loss

The analysis of training and validation losses provides us with information about the ability of the model to learn robust representations in the latent space, which are generated by the encoder during the training process. In most of the model training cycles, we observe slightly different behaviors, but the general pattern described is consistent in most of them.

Figure 6 shows that there is a reduction in both training and validation losses during the initial epochs. Training loss decreases more consistently in all epochs, starting at 0.5161 and reaching 0.0295, showing a steady learning process. This evolution could indicate that the model is adjusting its parameters to correctly predict these data. Similarly, the validation loss follows a downward trend, decreasing from 0.5891 to 0.0703, indicating an improvement in generalization to unseen data. The rapid drop in losses at the beginning indicates that the model effectively learns contrastive representations during the early stages of training. In addition, the convergence

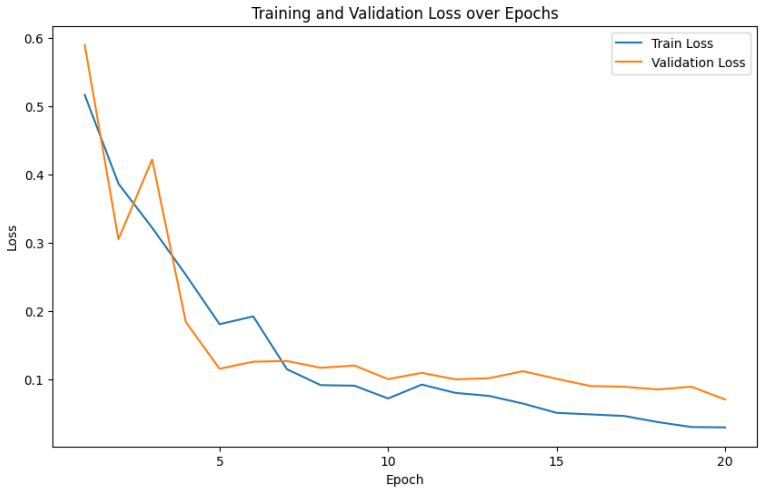


Figure 6. This figure illustrates the contrastive loss values for the training and validation datasets over 20 epochs. The blue line represents the training loss and shows a steady decrease as the model learns to minimize the difference between similar and dissimilar pairs. The red line is the validation loss and follows a similar downward trend, indicating that the model generalizes well to new data.

and proximity observed between the two loss curves suggest that the model is not significantly overfitted.

It is also important to note that given this behavior with such a limited number of epochs, training the model for more epochs could lead to observing the desired behavior beyond a certain point. To gain further insights into the learned representations, we incorporated the K-means clustering algorithm as an additional step to better understand the behavior of the contrastive model.

5.1.1. K-Means Clustering. The plot of k-means from the weights of the contrastive model is shown in 7. The model's output vectors were projected into a two-dimensional space using principal component analysis. The axes represent the first two principal components, which capture the directions of greatest variance in the learned features. Each point in the plot corresponds to a pixel-level representation, and the different colors indicate the cluster assignments

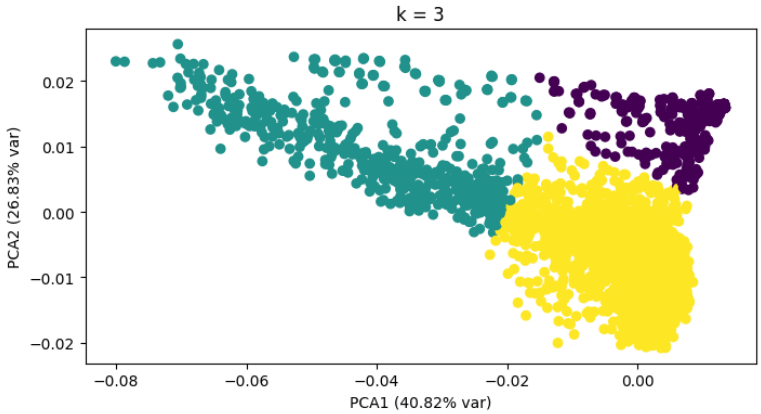


Figure 7. Visualization of the three clusters identified by the K-means algorithm after dimensionality reduction. The colors indicate the cluster assignments, which may correspond to permanent water, flood presence, and flood absence. The axes correspond to the first two principal components of the latent feature space learned by the contrastive model.

identified by the K-means algorithm with $k = 3$.

The presence of three distinct clusters suggests that the model can organize the feature space into separable groups. This visual separation of clusters in the 2D graph reflects only part of the underlying structure of the representations. However, it is useful because it allows us to identify general patterns, assess the quality of the learned representations and understand how the data cluster in the reduced feature space, which could correspond to permanent water and the presence or absence of floods. This kind of structure indicates that the contrastive model is learning representations that encode meaningful differences in the input data, even without access to labels.

Quantitatively, the first two components captured 40.82% and 26.83% of the total variance, respectively. Although the first component contributes the most to the variance, the information is distributed across multiple dimensions, as neither of the two dominant dimensions is sufficient on its own to fully represent the features learned by the contrastive model. The values on the axes

correspond to the projections of the representations onto the principal components, indicating the degree to which each representation aligns with the direction of greatest variance. The observed asymmetric scaling, with PCA1 ranging approximately from -0.08 to 0.02 and PCA2 from -0.02 to 0.02, confirms that the first component captures a greater spread of the data and therefore explains a larger portion of the total variability.

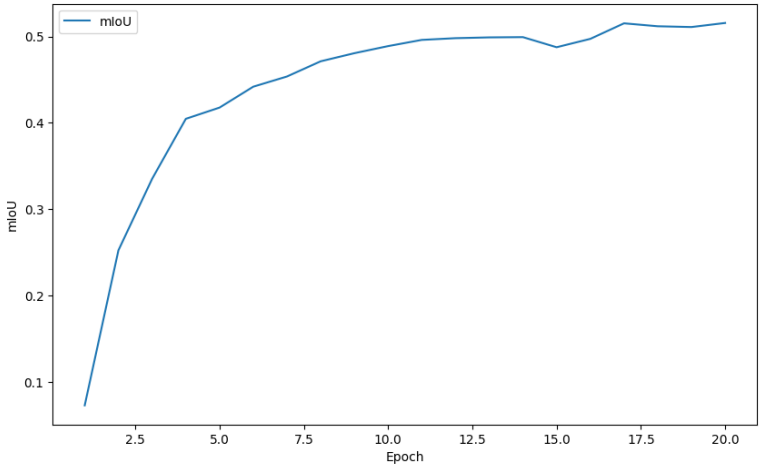
Note that some dispersion is observed in the groupings, which could indicate generalization or overfitting problems at specific training times. This behavior is reflected in the contrastive fluctuations in some epochs. Recall that the model has been trained with a very limited number of epochs. While the model is learning, it has not yet reached an adequate balance between the accuracy of the training set and its ability to generalize to new data. Although this clustering does not constitute part of the final segmentation pipeline, it serves as a useful diagnostic tool. It helps us verify whether the contrastive model is learning features that could later support effective class separation when passed to the segmentation model.

5.2. IoU Metric

To validate the evaluation of segmentation performance, we use the IoU metric that quantifies the overlap between predicted segmentation and ground truth. During model training, the IoU metric is tracked per epoch to assess how well the model is learning to distinguish between different classes.

Figure 8 illustrates the evolution of the IoU metric. During training, the metric follows an upward trend, indicating that the model is successfully generalizing to unseen data. The average IoU stabilizes around 52.20%, demonstrating the ability of the model to identify regions of interest

for both predicted classes.



centering

Figure 8. This figure illustrates the evolution of the IoU metric for the validation dataset over 20 epochs. The line following an upward trend and stabilizing at an average of 0.52, indicating consistent generalization to unseen data.

It should be noted that, although this performance is moderate, it does not reach that obtained by Mohaddeseh & Reza (2023), which reports an mIoU of 84.77% using the same data set and using a UNET++ architecture with EfficientNet-B7 as backbone. However, since this model relies on more robust configurations and greater computational resources, we evaluate the contributions of contrastive pretraining by comparing our model with a baseline trained from scratch without transfer learning.

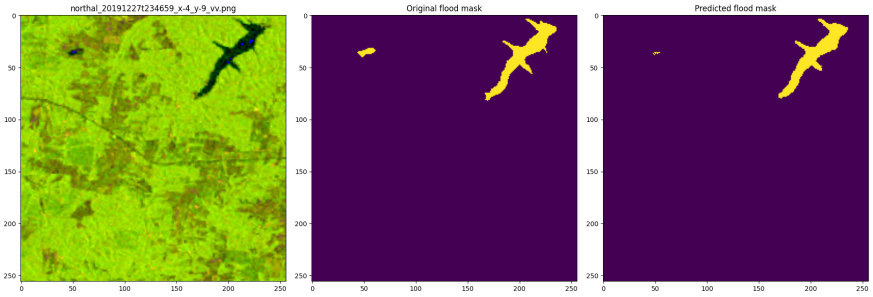
In our evaluation, quantitative results indicate that the model leveraging contrastive learning improves by 16% over the baseline model, which scored an average of 36%. Both models share the same architecture and training setup, the only differences lie in the initialization of the segmentation network and the data augmentation strategy used during pretraining. Furthermore, by limiting our evaluation to models trained under the same computational conditions, we ensure a

fair comparison. This comparison allows us to conclude that the contrastive learning phase demonstrates a trend toward improvement, capturing more robust feature representations and providing a better initialization for the segmentation network.

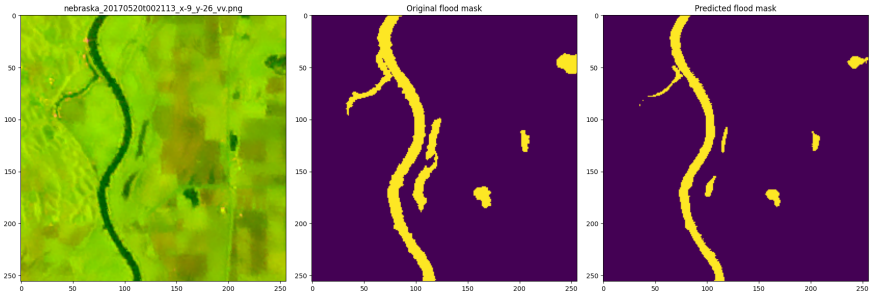
5.3. Segmentation Results Visualization

In the following, we present images that illustrate the results obtained by the model in the segmentation task. Initially, we show results obtained from images in the training set, since flood labels are not available for the validation set, and we cannot make a direct comparison at that stage. Along with each image, we show the Pixel Accuracy, Precision and Recall metrics to provide an idea of the prediction quality. As reminder, that these first results served as a basis for evaluating the ability of the model to correctly segment areas of interest in controlled scenarios.

By comparing the model predictions with the reference masks to visually evaluate it, we observe a partial agreement with the reference mask. In Figures 9a and 9b, the model successfully captures the general shape and extent of the flooded regions in each image, particularly in well-defined areas with clear contrast in the satellite imagery. The table 9c shows that, in terms of Pixel Accuracy, the first one shows higher overall accuracy, indicating that almost all pixels in both flooded and non-flooded areas were correctly classified. We can validate this in the pixel count, as the second one has a higher number of erroneous pixels. With respect to accuracy, Figure 9b shows a higher value, meaning that, of the pixels classified as flooded, a higher proportion are correct, resulting in fewer false positives. However, this comes at the cost of a lower Recall, while in Figure 9a, this metric is notably higher, suggesting that the model detected most of the areas actually flooded.



(a) Prediction of a North Alabama image



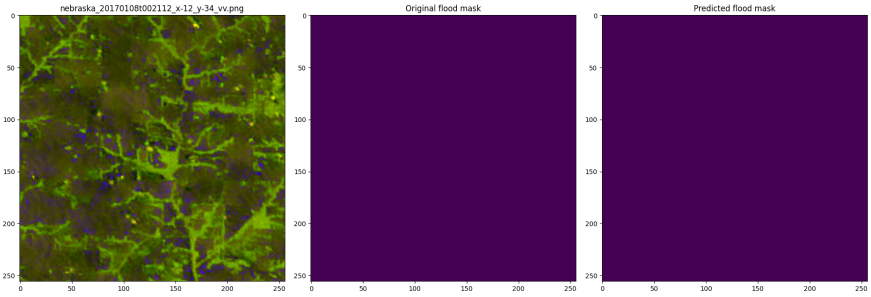
(b) Prediction of a Nebraska image

Metrics	Image 1	Image 2
Pixel Accuracy	0.997	0.983
Precision	0.899	0.936
Recall	0.943	0.832
Correct Pixels	65327	64442
Wrong Pixels	209	1094

(c) Metrics for each image

Figure 9. Segmentation results and corresponding metrics. In these two examples, the model successfully captures the overall shape and extent of the flooded regions. However, when comparing the metrics, (a) has a higher pixel accuracy because it has more correct pixels and detected a larger portion of the flooded areas.

Next, we examine a region where the reference mask does not indicate flooding, as shown in Figure 10a. In this case, the model correctly identifies the absence of flooded areas. This indicates that when there is no actual flooding, it generally avoids producing false positives and successfully classifies non-flooded regions. As can be seen in Table 10b, in this image the model achieves



(a) Prediction of a Nebraska image.

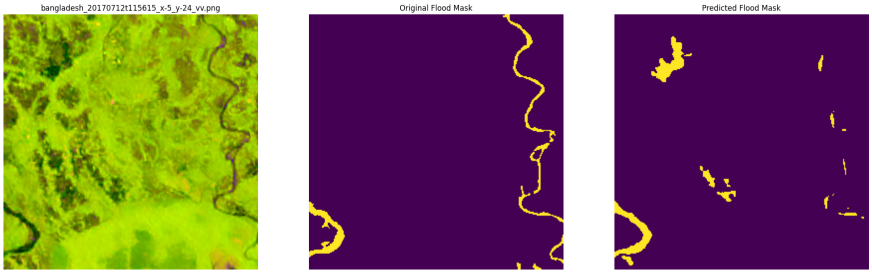
Metrics	Image 1
PA	0.999999
Precision	0
Recall	0
Correct Pixels	65536
Wrong Pixels	0

(b) Image metrics

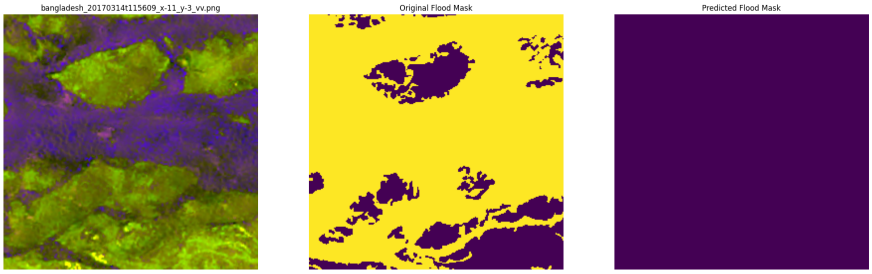
Figure 10. The model correctly identifies the absence of flooded areas, achieving perfect Pixel Accuracy as nearly all pixels are correctly classified. The Precision and Recall values of 0 are expected since there are no flooded areas.

perfect pixel accuracy, since the prediction coincides in its entirety with the reference mask. On the other hand, it is normal that the Precision and Recall values are 0, since, although the model is globally accurate, there are no flooded areas to evaluate its performance in the detection of the positive class.

However, it is clear that there are discrepancies. For example, when comparing the set of images in Figures 11a and 11b, the model does not cover certain flooded areas and, in some cases, confuses non-flooded zones with flooded regions. Figure 11a shows that the model correctly identifies some flooded areas along the river, but fails to detect others, in particular permanent water bodies. In addition, some non-flooded regions are incorrectly classified as flooded, resulting



(a) Prediction of a Bangladesh image



(b) Prediction of a Bangladesh image

Metrics	Image 1	Image 2
Pixel Accuracy	0.969	0.213
Precision	0.402	0
Recall	0.412	0
Correct Pixels	63539	13956
Wrong Pixels	1997	51580

(c) Metrics for each image

Figure 11. Segmentation results and corresponding metrics. In these two examples, the model shows discrepancies by confusing flooded and non-flooded regions. In particular, in the second image, the model fails to identify any flooded pixels, despite the fact that this image contains a higher proportion of flooded areas. As a result, the corresponding metrics are significantly lower compared to the other images.

in false positives. In Figure 11b, it does not recognize any flood pixels, despite the fact that in this image there is a higher percentage of really flooded areas. This behavior could be due in part to a class imbalance in the training data. The model was trained primarily with images in which the vast majority of pixels belong to the non-flooded class, so it may have developed a bias toward

this class. As a result, when confronted with such an image, the model has difficulty detecting them accurately, resulting in significantly lower segmentation metrics. The Table 11c shows the corresponding values, for example, the second image has a very low Pixel Accuracy and both Recall and Precision are at 0, indicating that the model does not classify any pixel as flooded. Since these metrics are presented not only as percentages but also as absolute pixel counts, a classification error involving thousands of pixels implies that a potentially large area of flooding is being missed. The inclusion of absolute values provides a more interpretable measure of impact, especially for operational decision making, as underestimating the extent of flooding can lead to misallocation of resources or delayed emergency response.

Finally, we present some predictions made specifically on the validation set. At this stage, the evaluation is based only on the ability of the model to generalize to unlabeled data, reflecting its performance under more challenging conditions with less information.

In Figure 12 we present predictions for the validation set, for which no reference mask is available. The absence of ground truth makes a quantitative assessment difficult and forces us to perform a qualitative analysis of the results. At first sight, we observe that the segmented areas make sense within the context of the satellite image, since the regions predicted as flooded seem to coincide with typical visual features of water accumulations, such as darker hues or altered vegetation patterns. However, to confirm these observations, it is essential to have the opinion of experts in remote sensing or hydrology. This situation is common in real-world scenarios, as most of the satellite imagery generated continuously for environmental monitoring is unlabeled. In such cases, expert knowledge is essential to validate the consistency of the predictions in relation to

expected flood patterns.

5.4. Computational Constraints

Importantly, the performance of the model is affected by computational limitations, particularly in terms of memory availability and training duration. Due to limited GPU memory, the batch size was reduced to 64 in both models, which is a small value. This reduction may have compromised the stability of the gradient and hindered the optimal convergence. Additionally, training was limited to only 20 epochs, which limited the ability of the model to learn more complex and deeper patterns in the data in the segmentation task, where detailed pixel-level features must be identified. Preliminary tests suggest that extending training to 50 or more epochs could improve the IoU metric by up to 12%, highlighting the potential benefits of additional training time.

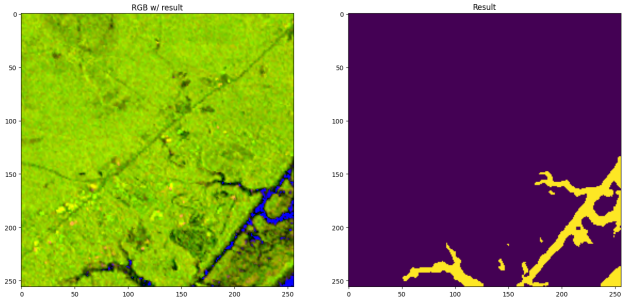
In this context, it would have been desirable to establish a broader comparative baseline that incorporates various approaches reported in the literature, such as a systematic study of architectural components, the exploration of different data augmentation strategies or the incorporation of alternative backbone networks. However, these experiments were not carried out because they fell outside the scope of this research, as technical limitations also influenced the choice of architecture. This study focused on validating the contribution of contrastive learning in a controlled setting, so it was decided to keep both the architecture and training configuration fixed. In all experiments, we used ResNet50 as the backbone network, together with a standard UNet implementation, selected for its low computational cost. The combination of both provided a suitable balance between efficiency and performance.

Nonetheless, there are several opportunities to improve this approach. On the one hand,

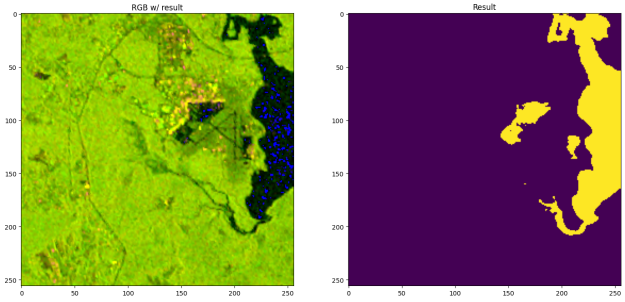
optimization of the contrastive pretraining phase, such as the use of alternative contrastive targets, deeper projection architectures, or different loss functions. This could substantially improve the quality of the learned representations, resulting in better performance in segmentation tasks. On the other hand, the use of more advanced segmentation models could be achieved. However, although they have shown improvements in similar tasks, their implementation involves a considerably greater computational burden. Architectures such as Attention UNet, EfficientNet or UNet++ (used as a reference in our comparison), as well as experimentation with advanced data augmentation techniques, specific loss functions or alternative optimizers, would have significantly increased training time and resource consumption.

While this proof of concept demonstrates the potential of the model, there are clear opportunities for improvement. With increased resources, extended training time and deeper exploration of these components, the performance and robustness of the model could be significantly improved.

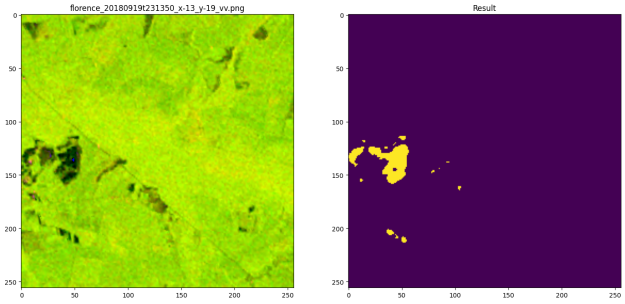
$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$



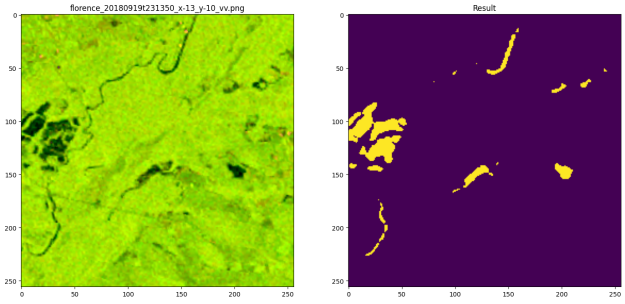
(a) Validation set prediction 1



(b) Validation set prediction 2



(c) Validation set prediction 3



(d) Validation set prediction 4

Figure 12. Results of the model segmentation from the validation set. The left side shows the RGB image of the input satellite image, while the right side shows the predicted flood mask generated by the model. The highlighted regions are the areas identified as flooded, so that a comparison can be made between the visual features in the input image and the segmentation result.

6. CONCLUSIONS

We proposed an approach to flood detection that aims to improve flood mapping through transfer learning from a contrastive model to a semantic segmentation model. This approach was structured as a proof of concept, based on a critical review of the flood detection literature, with a particular emphasis on approaches that utilize machine learning techniques. This review analyzed a wide range of studies, highlighting methodological aspects, results, and potential limitations. Through this process, we identify persistent challenges in the field, such as the scarcity of labeled data and the difficulty of achieving generalization across different geographic regions, which motivated the use of self-supervised and transfer learning techniques.

To evaluate our proposal, we trained the model on images provided by NASA for a flood detection competition, where the data was labeled at the pixel level, indicating the presence or absence of flooding. The results obtained validated the effectiveness of the method through an evaluation framework consisting of two phases. In the first phase, we used a supervised variant of the NT-Xent loss, incorporating label information to explicitly define positive and negative pairs, unlike the traditional unsupervised approach. We achieved a final loss value of 0.0703, with a decreasing trend indicating effective feature learning. As part of our transfer learning strategy, only the ResNet50 encoder from the contrastive model was transferred to the segmentation model, while the projection head was discarded. In the second phase, we evaluated the performance of the segmentation model using the intersection over union score, achieving an average of 0.52 on the test set. All encoder layers were fine-tuned during segmentation training. Compared to a

baseline model trained from scratch, which obtained an IoU of 0.36 under identical computational conditions, our approach demonstrated a 16% improvement in segmentation accuracy. The results demonstrated the coordination between contrastive pretraining and semantic segmentation techniques.

In addition, we assessed the proof of concept by evaluating the performance of the model through predictions made in both the training set and the validation set. Through visualizations, we compared model predictions with reference masks to assess accuracy. Furthermore, by performing predictions on the validation set, we observed that the model managed to capture relevant features of the flooding phenomenon, even without the guidance of direct labels. Training and validation predictions were valuable for identifying successes and potential areas for improvement, such as difficulties in regions with diffuse boundaries or complex textures. These findings will guide future iterations and optimize the performance of the model.

Better results are expected with an increase in computational power, particularly memory and GPU capacity. The current configuration, with a batch size of 16 and only 10 epochs, may have influenced the stability and convergence of the gradient. First, increasing the number of epochs would allow the model to learn more complex and detailed patterns at the pixel level. Second, tuning hyperparameters, such as batch size, could help to improve the generalization of the model and segmentation capabilities. Finally, implementing a deeper and more complex model architecture could further improve performance.

Finally, it would be ideal to perform an external validation with additional data or field measurements that would allow an objective comparison of the results. This step would ensure

a more robust evaluation and help identify possible limitations or areas for improvement in the model.

Bibliography

Bahareh, K., Naonori, U., Vahideh, S., Saeid, J., Fariborz, S., Kourosh, A., & Farzin, S. (2021). Deep neural network utilizing remote sensing datasets for flood hazard susceptibility mapping in brisbane, australia. *Remote Sensing*, 13(13).

Bates, P., Quinn, N., Sampson, C., Smith, A., Wing, O., Sosa, J., Savage, J., Olcese, G., Neal, J., Schumann, G., Giustarini, L., Coxon, G., Porter, J., Amodeo, M., Chu, Z., Lewis-Gruss, S., Freeman, N., Houser, T., Delgado, M., Hamidi, A., Bolliger, I., McCusker, K., Emanuel, K., Ferreira, C., Khalid, A., Haig, I. D., Couasnon, A., Kopp, R., Hsiang, S., & Krajewski, W. (2020). Combined modeling of us fluvial, pluvial, and coastal flood hazard under current and future climates. *Water Resorces Research*, 57(2).

Bhattarai, Y., Duwal, S., Sharma, S., & Talchabhadel, R. (2024). Leveraging machine learning and open-source spatial datasets to enhance flood susceptibility mapping in transboundary river basin. *International Journal of Digital Earth*, 17(1), 2313857.

Binbin, H., Peng, L., Hongyuan, L., Jiamin, Y., Zhenhong, L., & Houjie, W. (2024). Waterdetectionnet: A new deep learning method for flood mapping with sar image convolutional neural network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 14471–14485.

Bui, Q.-T., Nguyen, Q.-H., Nguyen, X. L., Pham, V. D., Nguyen, H. D., & Pham, V.-M. (2020).

Verification of novel integrations of swarm intelligence algorithms into deep learning neural network for flood susceptibility mapping. *Journal of Hydrology*, 581, 124379.

Buslaev, A., Iglovikov, V., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. (2020). Albumentations: Fast and flexible image augmentations. *Information*, 11(2).

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. arXiv preprint arXiv:2002.05709v3.

Chendeb, E. R. M., Muna, D., & Aicha, B. F. (2023). Msflood: A multi-sources segmentation for remote sensing flood images. In *2023 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)* (pp. 298–301). Los Alamitos, CA, USA: IEEE Computer Society.

Dodangeh, E., Panahi, M., Rezaie, F., Lee, S., Bui, D. T., Lee, C.-W., & Pradhan, B. (2020). Novel hybrid intelligence models for flood-susceptibility prediction: Meta optimization of the gmdh and svr models with the genetic algorithm and harmony search. *Journal of Hydrology*, 590, 125423.

Farahani, A., Pourshojae, B., Rasheed, K., & Arabnia, H. (2021). A concise review of transfer learning. arXiv preprint arXiv:2104.02144v1.

Farshad, S. & Maryam, R. (2023). Comparative study of real-time semantic segmentation networks

in aerial images during flooding events. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16, 4–20.

FloodList (2024). *FloodList*. Technical report. `https://floodlist.com/`.

Hammed, A. A., Bashir, A., Qudus, A., Abiodun, S. R., Golden, O., & Sook, C. K. (2024). Integrating machine learning models with comprehensive data strategies and optimization techniques to enhance flood prediction accuracy: A review. *WATER RESOURCES MANAGEMENT*, 38, 4735–4761.

Hashemi-Beni, L. & Asmamaw, G. (2021). Flood extent mapping: An integrated method using deep learning and region growing using uav optical data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 2127–2135.

IDEAM (2017). *Guía Metodológica para la elaboración de mapas de inundación*. Technical report, Instituto de Hidrología, Meteorología y Estudios Ambientales. `http://documentacion.ideam.gov.co/openbiblio/bvirtual/023774/GUIA_METODOLOGICA_MAPAS_INUNDACION_MARZO_2018.pdf`.

Islam, K. A., Uddin, M. S., Kwan, C., & Li, J. (2020). Flood detection using multi-modal and multi-temporal images: A comparative study. *Remote Sensing*, 12(15).

Jaiswal, A., Babu, A. R., Zadeh, M. Z., Banerjee, D., & Makedon, F. (2021). A survey on contrastive self-supervised learning. arXiv preprint arXiv:2011.00362v3.

Khosravi, K., Shahabi, H., Pham, B. T., Adamowski, J., Shirzadi, A., Pradhan, B., Dou, J., Ly, H.-B., Gróf, G., Ho, H. L., Hong, H., Chapi, K., & Prakash, I. (2019). A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. *Journal of Hydrology*, 573, 311–323.

Le-Khac, P., Healy, G., & Smeaton, A. (2020). Contrastive representation learning: A framework and review. *IEEE Access*, 8, 193907–193934.

Mahfuzur, R., Chen, N., Monirul, I. M., Ashraf, D., Javed, I., Ali, W. R. M., & Tian, S. (2019). Flood susceptibility assessment in bangladesh using machine learning and multi-criteria decision analysis. *Earth Systems and Environment*, 3, 585–601.

Mañas, O., Lacoste, A., Giro-i-Nieto, X., Vazquez, D., & Rodriguez, P. (2021). Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data. arXiv preprint arXiv:2103.16607v2.

Mohaddeseh, M. & Reza, S. (2023). Flood detection based on unet++ segmentation method using sentinel-1 satellite imagery. *Earth Observation and Geomatics Engineering*, 7(1), –.

Mosavi, A., Ozturk, P., & wing Chau, K. (2018). Flood prediction using machine learning models: Literature review. *WATER*, 10(11).

Muhadi, N. A., Abdullah, A. F., Bejo, S. K., Mahadi, M. R., & Mijic, A. (2021). Deep learning semantic segmentation for water level estimation using surveillance camera. *Applied Sciences*, 11(20).

Muñoz, D., Muñoz, P., Moftakhari, H., & Moradkhani, H. (2021). From local to regional compound flood mapping with deep learning and data fusion techniques. *Science of The Total Environment*, 782, 146927.

NASA EARTHDATA (2025). Synthetic Aperture Radar (SAR). https://www.earthdata.nasa.gov/learn/earth-observation-data-basics/sar.

NASA Impact Team and IEEE GRSS Committee (2021). ETCI 2021 competition on flood detection. https://nasa-impact.github.io/etci2021/.

NOAA - NCEI (2024). *U.S. Billion-Dollar Weather and Climate Disasters*. Technical report, National Centers for Enviroment Information. 10.25921/stkw-7w73.

O'Connor, B., Moul, K., Pollini, B., de Lamo, X., & Simonson, W. (2020). Compendium of earth observation contributions to the sdg targets and indicators. https://eo4society.esa.int/wp-content/uploads/2021/01/EO_Compendium-for-SDGs.pdf.

Pham, B. T., Luu, C., Phong, T. V., Trinh, P. T., Shirzadi, A., Renoud, S., Asadi, S., Le, H. V., von Meding, J., & Clague, J. J. (2021). Can deep learning algorithms outperform benchmark machine learning algorithms in flood susceptibility modeling? *Jounal oF Hydrology*, 592.

Pignato, S. & Marino, A. (2024). Flood segmentation: Self-supervised knowledge transfer from optical to sar. *Preprints*. https://doi.org/10.20944/preprints202406.1541.v2.

Prasad, M. B., Kumar, G. D., & Prakash, S. D. (2022). Geospatial modeling using hybrid machine learning approach for flood susceptibility. *Earth Science Informatics*, 15(4), 2619–2636.

Romulus, C., Alireza, A., Thomas, B., Bao, P. Q., Thai, P. B., Manish, P., Aman, A., Thuy, L. N. T., & Iulia, C. (2021). Flash-flood potential mapping using deep learning, alternating decision trees and data provided by remote sensing sensors. *Sensors*, 21(1).

Ruth, A., Fei, W., & Jun, X. (2024). History, causes, and trend of floods in the u.s.: a review. *Natural Hazards*, 120, 13715–13755.

Shahabi, H., Shirzadi, A., Ronoud, S., Asadi, S., Pham, B. T., Mansouripour, F., Geertsema, M., Clague, J., & Bui, D. T. (2021). Flash flood susceptibility mapping using a novel deep learning model based on deep belief network, back propagation and genetic algorithm. *Geoscience Frontiers*, 12(3), 101100.

Teymoor, S. S., Yousef, K., Mahdi, H., Roya, S., Jocelyn, C., & Meisam, A. (2023). Comparison of machine learning algorithms for flood susceptibility mapping. *Remote Sensing*, 15(1).

Vaibhav, K., Nopphawan, T., & Masahiko, N. (2021). Near-real-time flood mapping using off-the-shelf models with sar imagery and deep learning. *Remote Sensing*, 13(12).

Wenqing, F., Fangli, G., Chenhao, S., & Wei, X. (2024). Cross-modal change detection flood extraction based on self-supervised contrastive pre-training. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-1-2024, 75–82.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J., & Luo, P. (2021). Segformer: Simple and efficient design for semantic segmentation with transformers. *CoRR*, abs/2105.15203.

Yuan, Y., Chen, X., Chen, X., & Wang, J. (2019). Segmentation transformer: Object-contextual representations for semantic segmentation. *CoRR*, abs/1909.11065.

Zhang, H., Dana, K., Shi, J., Zhang, Z., Wang, X., Tyagi, A., & Agrawal, A. (2018). Context encoding for semantic segmentation. *CoRR*, abs/1803.08904.