2021

Disaster Analysis using Satellite Image Data with Knowledge Transfer and Semi-Supervised Learning Techniques

Palavi Jain

2021-05-20

Disaster Analysis using Satellite Image Data with Knowledge Transfer and Semi-Supervised Learning Techniques

Pallavi Jain

Confirmation Report:
Disaster Analysis using Satellite Image Data with Knowledge Transfer and Semi-Supervised Learning Techniques

Pallavi Jain

School of Computer Science

Technological University Dublin

Ireland

pallavi.x.jain@mytudublin.ie

***Supervisors*:**

Dr. Bianca Schoen-Phelan and Dr. Robert Ross

Table of Contents

1 Introduction

With the increase in frequency of disasters and crisis situations like floods, earthquake and hurricanes, the requirement to handle the situation efficiently through disaster response and humanitarian relief has increased. Disasters are mostly unpredictable in nature with respect to their impact on people and property. Moreover, the dynamic and varied nature of disasters makes it difficult to predict their impact accurately for advanced preparation of responses [104]. It is also notable that the economical loss due to natural disasters has increased in recent years, and it, along with the pure humanitarian need, is one of the reasons to research innovative approaches to the mitigation and management of disaster operations efficiently [1].

For many years, satellite images have been used for several earth observation (EO) applications such as crop monitoring, land coverage mapping, oceanography, water bodies mapping, and so forth. Indeed, satellite data processing has a natural application to disaster analysis in cases such as flood monitoring, building damage estimation, volcano monitoring and more. In such cases, satellite data is not only useful in providing initial alerts, but also in the continuous management and review of a situation.

With the rapid increase in the number and availability of satellite platform such as Sentinel, Landsat, Worldview, MODIS, a tremendous increase in EO data has been seen in recent years. This data is available in huge volume and high velocity. But along with the increase in the number of sensors, the variety of data has increased in terms of spectral properties, spatial resolution, temporal and radar information availability and so forth. Spectral information from different satellites typically varies from 4-16 bands, while spatial resolution varies from 31cm-30m, and temporal information from daily to 10 day updates. In addition to optical satellite, data is also available from synthetic aperture radar which is useful in terrain, landforms, flood and volcano mapping. Although many of these varied data sources and types are used in isolation, it is highly advantageous to fuse different data sources to provide high-quality input for disaster analysis.

Given the high volumes and varieties of available data, it is not feasible to analyse disaster situation manually. Given this, the research community has moved to utilise artificial intelligence in almost all EO application for timely and accurate analysis. Labelled data for satellite imaging is however still scarce, especially in disaster analysis. Fortunately in the past few years many competitions were held, which in turn lead to the release of several earth observation datasets that have been investigated using methods from machine learning (ML). Some of these datasets are EuroSat [40], BigEarthNet [93], Sen12MS [87] for scene classification, xView2 [35] for building damage analysis, MediaEval [6] for flood detection, and so on. These datasets focus on particular satellite sensors and specific tasks only, thus they cannot be directly used to create a generalised model for a wide range of satellite data inputs and disaster situations. This fact motivates the main theme of our research work towards generalising methods in the EO domain through recent advances in domain adaptive learning techniques, especially for disaster analysis from satellite images.

One of the most important methods for domain adaptation in the ML community is knowledge transfer within deep learning frameworks. This method has already been shown to be highly advantageous for EO applications [41], [68], but most application utilises existing transfer learning models from ImageNet [21], MS coco [66] etc and do not consider the specifics of the EO domain. The problem with existing pre-trained models is they utilise only three channels, i.e. Red, Green, Blue or RGB, so any application to datasets that do not contain Red, Green and Blue tri-band data is problematic. This challenge suggests the solution of having pre-trained models trained specifically on EO data with multi-spectral, multi-resolution or Radar images. However, to have such pre-trained models traditionally requires a large amount of labelled data across spectral bands, spatial resolutions, and even different sensor types, for example optical or radar, which is difficult to obtain.

Given these challenges, we argue that in order to have useful pre-trained models for the EO domain, techniques which can learn representations from unlabelled or only partially labelled data will be required. For this, we may take advantage of recent advances in semi-supervised techniques which have been shown to allow researchers to generalised models across particular domain [54], [101], [106], [8]. Semi-Supervised learning lies between supervised and unsupervised learning with small labelled datasets being used to bootstrap the learning process. Also, recent advancement in self-supervised learning have also shown greater potential to generalised models across particular domains [13], [33], [16]. Primarily self-supervised learning can be divided into pretext tasks and contrastive learning, which learns the intrinsic invariant representations of the data. Pretext tasks are based on augmenting the images in specific manners such as rotation, cropping, colourisation, jigsaw puzzle etc. [27], [108], [32], as part of a training process. In this type of learning, focus remains on how to learn the representation rather than on the final prediction. These approaches are based on the assumption that augmentation of images does not change their labels. Similarly, recently a major focus has been placed on contrastive learning. Contrastive learning is based on the same broad assumptions but tries to bring similar images (positive pairs) closer whereas dissimilar images (negative pairs) are repulsed [13].

While the potential of these families of methods is clear, it is far from clear how in practice we can apply these methods efficiently and effectively. Indeed, the specifics of the EO domain where multi-spectral and even multi-modal data are a given can offer us both challenges and opportunities. Given this we hypothesis that by merging one or two sources of satellite data across multi-spectral, multi-resolution images through semi-supervised and self-supervised learning techniques, we can provide a generalised learning model for the purpose of knowledge transfer that can in fact be applied to further data sources within the EO domain.

With this basic goal set out, the rest of this report is structured as follows: Section 2, provides a literature review focusing on the previous work done in the area of EO centered disaster analysis across different modalities using artificial intelligence. Building on that, I state my research question in Section 3. Following the research question, the report showcase the work done to date in Section 4, and my plans for future work for the next step of the PhD programme in Section 5. I summarise my research progress and present my achievements in Section 6.

2 Literature Review

With the rise in satellite technology, data is now available in a multitude of different formats. This has led to a tremendous amount of research to date on information retrieval tasks [104]. Disaster or crisis situation is one task wherein the use of satellite imaging has produced significant amounts of data across multi-spectral, multi-resolution, multi-temporal and multi-modal dimensions. This data has been used widely to quickly assess a situation and even for helping rescue teams analyse a situation in more depth, or to perform retrospective analysis.

2.1 Satellite Images and Disaster Analysis

Satellite imaging has huge application to tasks like scene classification [40], crop monitoring [60], [44], oceanography [24], land coverage mapping [39], and disaster analysis [25], [37]. Such data has been used for many years to generate risk mappings, disaster impact assessments, and damage assessments due to its capability of capturing a wide area and being operational in all weather conditions. Recently, satellites have been providing higher resolution images which help in the analysing and mapping of disaster damage at a finer level of detail. Analysing disasters using satellite images has several specific applications such as change detection, flood monitoring, building damage mapping and assessment, and the extraction of road networks, all of which directly benefit rescue planning [82], [75], [85], [65]. Traditionally, disaster analysis through satellites has been performed as a manual process or with hand crafted features [83]. But with the advancement in artificial intelligence the true automation of the process is now much more feasible.

2.2 Satellite Image Data Fusion

With respect to integration across the many different sources and types of satellite data, many data fusion techniques have been proposed, such as the fusion of optical and radar images, the fusion of low spatial resolution (LSR) data with high spatial resolution (HSR) data, to name but two. Data fusion is not new, but there have been considerable amounts of interest lately across the research community where it is hoped that it can improvise information content.

One of the widest applied fusion methods is spatio-spectral fusion, which is used to obtain high spectral and spatial resolution images from combinations of fine and coarse spectral and spatial resolution sources. With respect to spatio-spectral fusion, methods are further categorised into different data fusion techniques such as Panchromatic (PAN)/multispectral (MS) [72], PAN/hyperspectral (HS) [30], and MS/HS [102], [23] fusion. Another fusion method is spatio-temporal fusion, which uses images with fine spatial resolution but less frequent temporal coverage to fuse with coarse spatial resolution and frequent temporal coverage, in order to obtain fine spatio-temporal images [109], [64]. Spatio-temporal fusion has many applications in change detection for land coverage. But it is also highly advantageous in the assessment of disaster damage such as flood detection, building damage and more, due to the changes that occur following disasters [110].

Another type of fusion is cross modality fusion where data from different sensors, namely optical and radar sensors are combined. Where optical images often suffer limitations due to weather effects, radar images suffer instead with low spectral resolution. However, each of their limitation is complimented by the other, which has motivated researchers to obtain high-quality images through the fusion of optical and radar images [59], [82]. The application of this particular fusion method is highly advantageous in monitoring floods, tsunami, hurricane etc. where weather conditions are not favourable and optical images fail to capture fine images [82].

Most of the research work mentioned to this point was based on classical methods, but recently tremendous amounts of interest have been seen in applying deep learning to learn and recognise the intrinsic pattern in the underlying data. Some of the recently proposed networks for spatio-spectral fusion include the multiscale and multidepth convolutional neural network (MSDCNN) [105], and the two-stream fusion network (TFNet) [67]. Similarly for spatio-temporal fusion, Liu et al. [69] proposed the spatial-temporal fusion net (StfNet) with a two stream CNN, and Li et al. [64] proposed a temporal framework based on a deep CNN. A numbers of works on optical/radar fusion has also been investigated with deep learning methods [88], [43], [45].

Although all these methods are studied well, there are still very few works which attempt to combine all into a single model– though some of the work shows the possibility of combining multiple aspects in a model [5], [85], [80]. The importance of the fusion of multiple data sources however is of vital importance as it is expected to lead to more robust modelling as is required for domains such as disaster analysis.

2.3 Satellite Images with Deep Learning

Traditionally, satellite image analysis has been manual or at least benefited from a hand crafted feature extraction process. But with the advancement in machine learning, many researchers explored the possibility of satellite data analysis through machine learning. There are a number of techniques commonly applied to satellite images but the most popular are support vector machine (SVM), random forest (RF) and neural networks. A comparative analysis by Bangira et al.[4] has been performed for different traditional ML algorithms such as decision tree (DT), k-nearest neighbours (KNN), RF, and SVM. These were applied to different water index techniques such as normalised difference water index (NDWI), modified NDWI (MNDWI), which showed that SVM outperformed all other model [4]. SVM is mainly based on establishing a decision boundary with the maximum distance between two classes that minimises misclassifications. Notably, it can handle non-linear and high-dimensional tasks with limited amounts of data. There are many research works that have shown the high performance of SVM in remote sensing tasks [2], [76], [84], but they are highly dependent on hand crafted feature extraction, eg. HOG, Gabor feature and Hough transform, etc. and kernel selection. Also, with the increase in dimensions as seen in hyper-spectral or multi-spectral images, SVMs have increasing issues with noisy data and high computational requirements [76].

More recently, deep learning (DL), which is the form of classical artificial neural network (ANN), has attained great performance in several tasks ranging from computer vision, through speech recognition, natural language processing (NLP), to machine translation. The key reason for these successes is due to the ability of the neural network to self learn complex patterns within data without requiring hand crafted features. Also, they are versatile in nature due to their robustness and scalability, ease in fine-tuning hyperparameters, and knowledge transfer. Although the initial training process for DL models is expensive in terms of time and computational requirements, they can easily be applied to real-world applications with a fast inference time.

Most DL models are based on stacking multiple neural network layers with suitable architectures to gain optimum performance with a minimised error rate. Among the many deep learning architectures, convolutional neural networks (CNNs) have shown excellent performance in computer vision tasks such as image recognition [91], [38], image detection [90], [98] and segmentation tasks [73], [34]. A basic CNN architecture is a multi-tier network, consisting of convolutional layers, pooling layers and one or more fully connected layers (FC) layers. The power of the CNN is based on learning the local features (eg. edges, lines) of images and then combining them to learn high-level features under the assumption that low-level features are spatially invariant, with their position relevant only in terms of relation to other features [62], [61]. The concept of a local receptive field, shared weights and spatial sub-sampling of CNN make them robust and applicable to a number of domains such as medical imaging, remote sensing, face recognition, historical data collection analysis and many more.

Since CNNs are highly scalable, many variant CNN architecture have been proposed over the last ten years, in terms of scaling width and depth, reducing parameters and minimising error rate or loss. Some of the popular architecture in image recognition are AlexNet [58], VGG [91], ResNet [38], DenseNet [42], and EfficientNet [94] etc. With the introduction of image recognition challenge on ImageNet [21] data, which is a library of million images, AlexNet showed the highest initial performance amongst CNNs. Following this, many deep networks (eg. VGG, ResNet) were introduced in order to overcome the drawbacks seen in former models.

The wide use of CNNs image processing has led to many research efforts in their analysis of satellite imagery for disaster analysis [26]. One such application area is change detection in time-series images [57], [99]. Change detection using CNNs significantly helps in automating flood detection and mapping [85], [65], building damage assessment models [28], [55], and extracting road networks and mapping passable roads [75].

Among the popular deep CNNs, VGG and ResNets have been very popular in most computer vision tasks as well as in remote sensing. Although VGG suffers from vanishing gradients, and has a high training time, several applications showed its potential in remote sensing tasks [68], [79], [95]. Building on the early VGG models, the so called ResNets overcame the vanishing gradients issue with skip connections to allow deeper networks with in some cases 100s of layers. ResNets typically reduced the size of parameters to provide faster training time while facilitating in some cases error rates below a human error rate. Considering these factors, many works utilised the skip connections concept in multispectral image processing such as in MRI [47], and reconstruction of hyper-spectral images from RGB images [36]. Also, evaluation of EuroSat data, which consists of a large amount of high resolution images for land cover classification, showed that ResNet50 outperformed other larger models on classification with 98.57% accuracy, in spite of it having many fewer layers than of the competing networks [40].

The relative lack of research in applying CNNs to satellite data for disaster analysis is due to the fact that labelled data is scarce in satellite imaging and especially for disaster analysis. But recently, high-resolution satellite images have been made freely available from satellites like Sentinel, Landsat, MODIS; yet still it is difficult to get annotated data for particular domains of work. To bridge this gap in the availability of annotated datasets for research, many competitions are being held. These competitions include DeepGlobe [20], MediaEval [6], xView2 [35], and many more. Recently, datasets such as EuroSat [40], BigEarthNet [93], Sen12MS [87] have also been released which include diverse land coverage and multi-modality. These competitions and datasets provide labelled datasets for scene classification, building damage mapping, flood detection and mapping, land coverage mapping and classification, cloud segmentation, agricultural land cover and so on. It is notable however that these datasets are still small in terms of the size and variety of labels in comparison to traditional datasets such as ImageNet.

2.3.1 Transfer Learning

Originally deep learning CNN models were designed only for either grey-scale or RGB images but due to the availability of more spectral information their application has expanded to multi-spectral or hyper-spectral imaging [86], [100], [98], [89]. CNNs have shown good performance in multi-spectral image recognition, multi-spectral image segmentation, and in a number of remote tasks in both the EO and medical domain [56], [47], [18]. Many different CNN variants have been proposed for remote sensing classification such as the work by Zhang et al., which showed the potential of combining CNN with a multilayer perceptron (MLP) for processing spectral bands [107]. Another approach proposed by Jiang et al. takes advantage of the fusion of features from RGB and NIR images with double channel CNN models [53] for scene classification. There is however very little research that has utilised the full spectral capacity of multi-spectral images; instead bands are mostly selected manually to enhance performance. While this is not an ideal approach, it has to this point been a necessity due to the low availability of data and the high computational cost of training models from scratch. Within this approach, work by Mahadianpari et al. showed that increasing spectral information improves performance when trained with pre-existing deep CNNs for wetland classification [70]. Another recent work proposed Sen2HSE-Net, which utilises 10 spectral bands of sentinel-2 for mapping human settlement extent with the help of a CNN architecture [81]. Chen et al., meanwhile showed the potential of 3D CNN for feature extraction and scene classification in hyperspectral images [17]. These works clearly highlight the suitability of CNNs for multi-spectral satellite image classification tasks, however, it should be kept in mind that each classification task has different architectural requirements and there is clearly a lack of a more systematic generalised representation, comparable to those derived from the ubiquitous ImageNet [21].

Considering the issue of scarcity of labelled satellite data, transfer learning can have great potential through the application of pre-trained models that have already been trained on large volume datasets from other sources. Some of these large volume datasets as mentioned earlier are ImageNet [21], MS Coco [66], and Pascal VOC [29] which consists of millions of labelled images. These datasets are used to train deep learning models to learn generic features from the data, and then their learned weights can be utilised to bootstrap task-specific models. This phenomena of transfer learning have shown a great advantage in many domains such as medical imaging, self-driving systems, and more. It has also useful for satellite image problems and it has been shown that models trained on ImageNet can give better results than training from scratch [41], [68], in-spite of the fact ImageNet is quite a different dataset from the data obtained in satellite imaging.

The drawback of transfer learning centres on the number of channels, since all the pre-trained models, as mentioned above are trained on three colours only i.e. RGB; it is thus not possible to utilise them directly for multispectral images; instead they can only be applied to a selected three bands. This issue of transfer learning again leads to under-utilisation of multispectral data. The solution to this challenge may be to construct suitable pre-trained networks from multi-spectral and other satellite data. However as indicated previously, having a large amount of labelled satellite data is difficult to amass in the remote sensing domain, which in turn suggests that we explore methods for training with unlabelled data or at least minimal amounts of labelled data.

2.3.2 Semi-Supervised and Self-Supervised Learning

Recently, semi-supervised learning (SSL) has become popular to resolve the traditional dependency on large amounts of labelled data, hence making models more generalisable. Semi-supervised learning lies between supervised and unsupervised learning, which exploits both unlabelled and labelled data. Semi-supervised learning can be broadly divided into inductive and transductive methods, where inductive methods try to optimise over classification model whereas transductive methods optimise over the predicted labels from unlabelled data [96]. Previously wrapper methods and graph based methods were quite popular in SSL, but with the recent advancement of neural networks, perturbation based and pseudo labelling based methods have gained popularity. Perturbation based methods are based on the weak smoothness assumptions, which says that small changes or distortion in data should not change the labels of the data [96]. Whereas pseudo labelling methods are based on assigning the artificial (pseudo) labels to unlabelled data through an initially trained model on small amounts of labelled data. This model is then trained on labelled data, along with pseudo labelled data [92]. Recently hybrid models have been quite popular; in particular, methods that have utilised pseudo labelling and perturbation methods together [8], [92]. These methods have shown state-of-art results against supervised learning by tweaking cost functions with unsupervised loss terms [96]. The main approach of these models remains focused on pseudo labelling, consistency regularisation, entropy minimisation and various augmentation theories [92]. Considering the general concept of perturbation, FixMatch [92] utilises strong augmentation i.e. heavily distorting an image using CutOut [22], CTAugment [7] and RandAugment [19] and weak augmentation by flipping or rotating the images. The performance of FixMatch with different augmentation shows the applicability of smoothness assumption and can be beneficial in the case of satellite data because of variable resolution and spectral data.

Alongside semi-supervised learning, there has been considerable interest in self-supervised learning methods [106]. The goal of self-supervised learning is to learn the invariant representations of data, which then can be used for downstream tasks. The most popular methods for self-supervised learning are pre-text task learning and contrastive learning. These types of learning are again based on the smoothness assumptions that perturbation or augmentation of images do not change their labels. Pre-text tasks are useful in learning semantic invariant representations in visual data by dividing them into learning pre-text tasks such as rotation [32], colourisation [108], solving a jigsaw puzzle [77] and more. Recent works also showed the possibility of combining the self-supervised pre-text tasks along with semi-supervised learning techniques in order to learn better representations [106], [14].

Contrastive learning is based on the concept that an image and its augmented view (positive pairs) should have closer similarity, while two different images (negative pairs) should be much further apart [13]. This concept has led to a surge in various self-supervised architectures recently, such as SimCLR [13], MoCo [15], SwAV [11], BYOL [33], and SimSiam [16]. Where SimCLR and MoCo are based on positive and negative pairs of images, BYOL and SimSiam are based on only positive pairs of images. All these methods show outstanding performances while optimising the learning process by reducing the complexity of an architecture in terms of batch size, memory requirements, and training time.

Although their application in satellite imagery is not well explored, some work have recently shown the potential of self-supervised learning in satellite imagery such as that of Bischke et al. [9] where distance learning was used along with a building footprint segmentation task. Also recently, Vincenzi et al. [97], utilised a colourisation pre-text task to learn representation to recolour images based on the assumption that spectral and semantic connection strongly exists in satellite images [97]. Another work by Ayush et al., [3], utilised temporal information as the basis of a self-supervised task with contrastive learning, which showed the potential of using pre-text tasks with contrastive learning.

While this review has shown a great many advances in the processing of satellite imagery, the automated high-quality processing of satellite data from applications as wide as disaster management to ocean observation is far from a completed task. In particular, since satellite data comes with multi-spectral, multi-resolution, multi-modal, and multi-temporal aspects, fusing these aspects may be essential in providing high-quality semantic information for EO applications. Although previous work highlighted the potential of data fusion to obtain fine spatio-spectral-temporal information from satellite images, there remains a lack of generalisation across different sensors. Also, most of the previous works were limited to the fusion of one or two aspects, and knowledge transfer was limited across spectral and model bands in the EO domain. The major cause of not having generalised models is the limitation of labelled data in the EO domain. So to solve both aspects, that is generalisation and scarcity of labelled data, the idea of knowledge transfer by utilising the concept of semi-supervised and self-supervised learning with a range of data sources is one potential solution.

3 Research Question

We have seen that satellite imagery can be thought of as having multiple aspects, namely, multi-spectral (MS), multi-resolution (MR), multi-temporal (MT), as well as being based upon fundamentally different sensor technologies or models, e.g., radar vs optical imaging. While these are seemingly very different aspects, the resultant data is highly correlated, which can be used to maximise information from satellite imagery for a particular task such as disaster analysis. Many work have shown the potential for the fusion of one or two aspects together, but little work has been done in utilising multiple aspects of satellite data together. There are mainly two challenges in fusing different aspects for a disaster relief task: i) complex data structures, and ii) the scarcity of labelled data. In order to overcome both of these issues, the creation of a generalised model that is trained, specifically on the earth observation (EO) domain can be seen as one solution. Here the concept of building and then applying a generalised model can be interpreted as transfer learning, but usually the construction of the pre-trained model in the EO domain would require a large amount of labelled data, which is scarce. To resolve this scarcity of labelled data, one can leverage self-supervised and semi-supervised learning methods. These techniques recently demonstrated tremendous potential to move machine learning towards generalisation by utilising a large amount of unlabelled data.

Concerning the above issues, in this research work, I aim to investigate and build upon the transfer learning phenomena, with the help of self-supervised learning, in order to solve the scarce labelled data problem in multi-dimensional satellite imagery. As such, I hypothesise that a model trained on two or more aspects of satellite image data can be applied to other aspects as well as for downstream tasks in disaster analysis. This work in return can benefit domain adaptation across different satellite data sources, and more specifically across different disaster analysis domains. With this established, the research question can be phrased more concretely as follows:

How can disaster analysis from satellite data sources leverage transfer learning using semi-supervised and self-supervised learning techniques to provide greater scalability and cross-domain utility while minimising the need for labelled datasets?

In order to approach the research question, it has been broken down into the following sub-questions:

1. **RQ 1: How can traditional satellite data processing methods such as indexing techniques benefit the machine learning approach to disaster analysis?** Among several multi-spectral satellite image processing techniques, indexing techniques such as NDWI, NDVI, Normalised Burn Ratio (NBR) have been very popular and shown good accuracy. Considering the efficiency of these traditional techniques, it would be beneficial to investigate their application further with machine learning for disaster analysis tasks.
2. **RQ 2: Given the multi-aspect and specifically multi-spectral nature of satellite imagery data, what is the overall benefit of applying traditional transfer learning methods to the satellite data processing challenge?** Over time, transfer learning has benefited several domains by applying pre-learned features to task-specific model creation and fine-tuning. Considering the richness of information content in satellite data and the scarcity of labelled data in the EO domain, it is essential to investigate the traditional transfer learning models, i.e. pre-trained model on ImageNet data, in order to tackle disaster analysis tasks.
3. **RQ 3: How can feature representations learned from one type of spatio-spectral data be applied to data with other spatio-spectral**

properties, and indeed modalities, for downstream tasks to assist in disaster analysis? Spatio-spectral information works wonders in distinguishing several geographical features in traditional handcrafted techniques. Hence learning spatio-spectral representations can be highly advantageous for EO domain tasks. However, a problem exists in the size of data required to learn those representations. Although pre-trained models have shown great performance in several domains including EO, they are not optimised for the EO domain since they consist of only optical data, i.e. RGB images. Considering this, I hypothesise that having features learned from spatio-spectral data specific to the EO domain can easily be transferred to other aspects of satellite data to solve multiple EO domain problems including disaster analysis.

4. **RQ 4: How can semi-supervised and self-supervised learning be used to reduce the large labelled dataset dependency and provide for robust performance compared to supervised learning for the EO domain and especially for different disasters?** It is commonly understood, that a lack of data can reduce the performance of supervised learning models significantly whereas transfer learning improves performance. But as mentioned, pre-trained models are generally limited to RGB and hence non-EO domain, which motivates our investigation of semi-supervised learning to learn spatio-spectral features for disaster analysis. By this, I also hypothesis that semi-supervised and self-supervised learning methods will improve the robustness of models due to active learning as compared to supervised learning.
5. **RQ 5: How can generalisation across different disasters be achieved through transfer learning from EO trained models?** So far models are limited to one disaster domain and there is no single model for different disasters. One promising direction is to investigate the application of spatio-

spectral features learned from one disaster to other disasters as downstream tasks.

To address these research questions, I believe that deep learning is an appropriate computational framework due to its ability to self-learn complex patterns within data without requiring hand-crafted features. Deep learning methods also show the state of the art performance in several satellite imaging and fusion tasks due to their robustness, scalability, ease in fine-tuning hyper-parameters, and knowledge transfer properties. Fortunately as also seen there exists a wide number of EO datasets. While these are not all labeled, they do provide a firm foundation for studies.

4 Work Done To Date

Given the research questions above, my initial research focused on the applicability of the multi-spectral aspect of satellite imaging for flood detection along with the comparison between transfer learning and model training from scratch [51], [52]. I also investigated several deep learning architectures to explore the impact of supervised learning with multi-spectral data for disaster analysis in floods. In the following, these initial research activities are expanded upon.

4.1 Multi-Spectral Satellite Images and Flood Detection

To address research question 1 & 2, my initial work carried out an investigation of the benefit of indexing techniques in multi-spectral satellite images; along with traditional transfer learning methods i.e., pre-trained deep learning models trained on RGB images from the ImageNet dataset, to use with multi-spectral and multi-resolution data.

From this work I looked at data from a flood detection task, which was sourced from the MediaEval 2019 competition– the competition provided the labelled multi-spectral data as mentioned in Section 4.1.1. Multi-spectral satellite data as introduced earlier varies in terms of its reflectance and absorption properties at different frequencies, eg., NIR absorbs water and reflects vegetation. Such properties of multi-spectral data help to separate different geographical features and land use types. In this work I considered in particular the challenge of identifying water bodies and the detection of floods. This is one particular challenge type, but in general the detection of land usage type and major changes of use is applicable across a range of disaster management scenarios. Although there have been many works which have shown the potential of identifying water bodies from multi-spectral images, floods are still difficult to identify due to shallow water, clouds, or building shadows and mixed pixels [74], [9], [4].

With the focus on the automatic flood detection task, and given the two research questions in MS satellite imagery, I proposed two methods. Firstly, I proposed an index for flood detection, and secondly, I explored different combinations of spectral bands, which can be applied with pre-trained deep CNN models to learn features for floods. For water indices, the final image obtained is grayscale whereas with spectral band combination the ideal image should consist of 12 spectral image channels. However, due to the fact that I utilise pre-trained models, which are limited to three channel information only, grayscale channels were tripled and for raw bands, I utilised three spectral bands at a time. This led me to identify the best tri-band combination in order to detect floods in images. Also, with the many variants of deep learning models, it is necessary to choose the right model for the tasks. For that, in this work I explored the popular VGG16, ResNets, and EfficientNets architectures.

With this as the overall goal, I further worked with three sub-questions:

1) Can classic flood detection methods benefit from enhancement with deep learning based image processing?

2) Can existing pre-trained models be leveraged for multi-spectral flood damage analysis, and if so, which three spectral bands combinations are most suitable for flood detection?
3) How do pre-trained models perform as compared to models trained from scratch for spectral band combination based detection?

In the following, I first give an overview of the dataset used for this work, before expanding on the analysis and results with respect to these three sub-questions.

4.1.1 Dataset Overview

SENTINEL-2 is a satellite platform that provides multi-spectral instrument (MSI) data with 13 spectral bands. Among the 13 spectral bands, four bands are at 10 metres, six bands at 20 metres and three bands at 60 metres spatial resolution. For flood detection, I leveraged the annotated dataset provided by the MediaEval[1] 2019 competition [6]. This dataset consists of 335 image sets with 267 identified as development sets and 68 as test sets. Each set consists of between 1 to 24 day time-series images of before and after flood events; this provides a total of 2,770 images. The data has 12 bands as shown in appendix A.1, which comes in three different sets of resolutions: 10 metres, 20 metres and 60 metres. Each 10-metre resolution image is 512 x 512 pixels in size, 20-metre resolution images are 256 x 256 pixels, and 60-metre images are 128 x 128 pixels in size. The provided dataset includes ground truth only for the development dataset. Therefore, in this work I utilised only the development dataset which is subsequently split into three parts, i.e. training, validation, and test, in the ratio of 80:10:10. Additionally, I pre-processed the data with the following steps:

– Upscale the low-resolution bands to 10m resolution using nearest neighbour interpolation.

[1] http://www.multimediaeval.org/mediaeval2019/

– Normalise each band's pixel values between 0 and 255.
– Augment images by shifting, rotating, and flipping the images with batch sizes of between 8 and 16 in order to increase the size of the training dataset.

4.1.2 Water Index Driven CNN Analysis

While the primary focus is on deep learning driven solutions, the reality is that hand crafted functions have been used in multispectral data analysis for the last 40 years. Rather than ignoring such research, I ask whether it is advantageous to pre-process image data to generate such index features prior to deep image analysis rather than simply processing the raw spectral information. This work was presented in the Symposium of Applied Computing (SAC 2020) conference [51].

Over time many water indices have been proposed for water bodies detection. For example, normalised difference water index (NDWI) [71] showed good results in mapping water but suffered from giving water bodies and built-up areas similar values. Xu meanwhile proposed modified NDWI, which used SWIR and the Green band to improve mapping and overcome the built-up area problem [103], while Mishra & Prasad [74] used the combination of NDWI with Blue/NIR spectral indices to detect shallow water [74]. Similarly, the AWEI technique by Fyesia et al. [31] helps in overcoming the cloud shadow problem by calculating indices using coefficient values for different bands [31]. However, all these indices are primarily designed for water bodies instead of flood detection. Floods requires different approaches to be map from or detected in satellite images [9], [65].

With my focus on flood detection, in this work I proposed an index (PI), motivated by the work of Mishra & Prasad, i.e. MI, by integrating two indices, MNDWI and an index based on Blue/NIR. The reason for using MNDWI instead of the originally used NDWI is because SWIR is good at separating built-up areas from water, and has a capability to capture moisture in the soil, which can be an important factor with the flooded region. Another property of SWIR is that it can pass through thin clouds, which is important as during floods the region is typically covered with clouds due to weather conditions. Moreover, as Mishra & Prasad showed that combining NDWI with Blue/NIR improves the mapping of the shallow water body, it is appropriate to utilise Blue/NIR with MNDWI. The proposed index (PI) is described by the following functions:

$$PI = \frac{Green - SWIR}{Green + SWIR} + \frac{Blue - NIR}{Blue + NIR} \tag{1}$$

Thus, I pre-process four input data channels to provide a single PI channel which is used in subsequent modelling.

For the image processing backbone, I utilised the popular VGG16 model with pre-trained weights based on ImageNet and froze the first four blocks in the VGG16 model while leaving block 5 trainable to allow task specific features to be learned. Global average pooling was used to reduce the overfitting of the model by reducing the total parameters. After the VGG16 blocks, I used a fully connected layer of 128 units followed by a dropout layer with a dropout parameter of 0.5, and finally a softmax layer [2]. Rectified linear unit (ReLU) were used as the activation function in all but the final layer, and I made use of the Adam optimiser to guide the training process. As the problem was a binary classification problem I used the binary cross entropy loss function. This architecture is depicted in Figure 1.

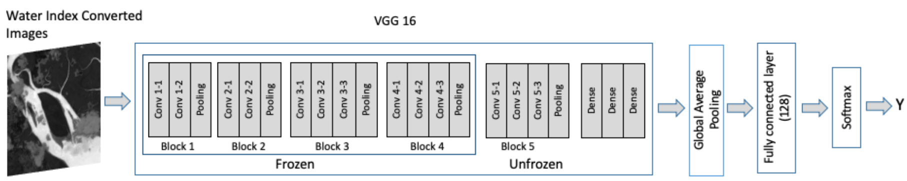


Fig. 1: Model Architecture For Flood Detection with Water Index Technique

[2] Since this was a binary classification problem, a single logistic unit could have been used here and was indeed used for all future studies.

As these pre-trained models were designed for RGB images (i.e, three channel images) whereas index images are single channel, as part of pre-processing I duplicated the PI image to each channel in order to feed them to the model architecture.

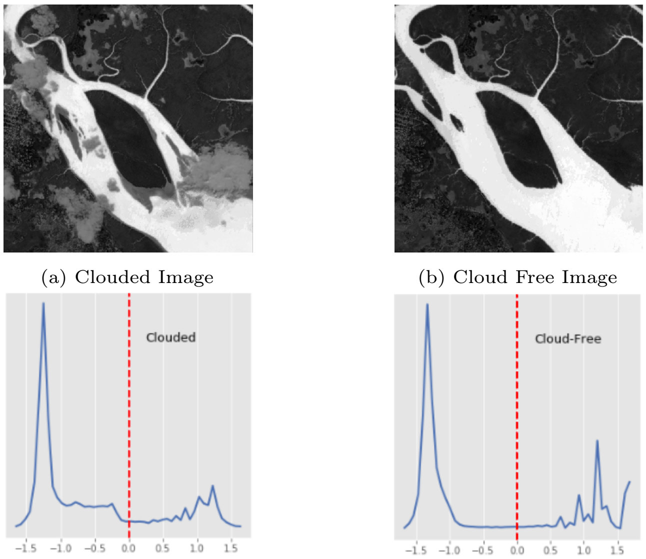
Results showed that the water index flood detection model outperformed all prior work which had aimed to make use of the water index model directly as shown in Table 1. To provide some insights into the potential causes of improved performance, I further analysed the histograms of indices values in Figure 2 for different geographical features. I found that the proposed model approach reduces the noise in the images and clearly distinguish between clouded and cloud free water images. The approach causes the cloud index count to move near to zero with water values remaining positive, while everything else remaining negative. I argue that this reduces the likelihood of misclassification of water areas due to either built-up areas or cloud shadows.

Index Type	TP	TN	F1	Kappa
NDWI [71]	0.84	0.78	0.80	0.59
MNDWI [103]	0.76	0.96	0.88	0.74
AWEI [31]	0.74	0.90	0.83	0.66
Mishra et al. [74]	0.83	0.94	0.90	0.79
Li et al. [63]	0.92	0.84	0.87	0.71
Proposed	**0.93**	**0.98**	**0.96**	**0.92**

Table 1: Evaluation of Water Indices Techniques with VGG16 Model

4.1.3 Tri-Band Estimators for Flood Detection

Whereas the first study considered the benefit of applying a deep image processing architecture to pre-processed water index data, the second study instead broadened the approach to allow the model to determine the best possible combination of input channels while taking advantage of pre-trained CNN features. This work



(a) Clouded Image (b) Cloud Free Image

(c) Clouded Histogram (d) Cloud Free Histogram

Fig. 2: Proposed Index. The index makes water values positive while non-water region negative. Red line is a threshold for water and non water region.

was presented in Machine Learning for Earth Observation (MACLEAN 2020) [52].

In order to utilise these pre-trained models, I stacked three different bands together to form three channel combinations. By stacking three bands together, I got 33 different band combination out of 10 base bands. The selection of 33 combinations out of 120 was made by evaluating the performance of all combinations and selecting those combinations which had an F1 score greater than 0.75.

I proposed the architecture as shown in Figure 3, where I fed the 3 band combination to different CNN models, i.e. VGG16, ResNet18, ResNet50, and EfficientNetB0 (Baseline). Global average pooling was used at the output of each pre-trained model, which was then fed to a fully connected layer of 512 units with a ReLU activation function. In order to avoid over-fitting during training, a dropout of 0.5 was used. The final output layer used a sigmoid function for binary classification. The model architecture remains the same for models trained from scratch or with models that made use of pre-trained weights.

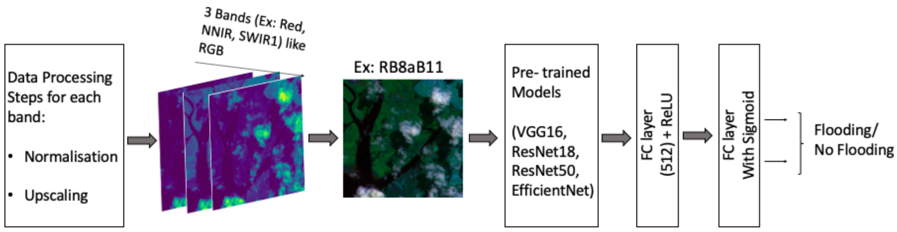


Fig. 3: Model Architecture For Flood Detection with 3 Band Combination

Again, the rectified linear unit (ReLU) was used as an activation function for all but the final layer and the Adam optimiser was applied to guide the training process. Binary cross entropy function was used to calculate the loss.

Figure 4 shows the F1 score spread of 33 band combinations across the four pre-trained models for each architecture variant. Each model had its own band combination, that outperformed all other combinations but a few combinations showed an overall better score across all models. Such combinations are RB11B, RB8aB11, and B7B11B, while combinations like RGB, RB12B and B8B11B showed the overall worst performance across all four models.

Since SWIR bands are known for good performance on water identification, the best performance of all three combinations (RB11B, RB8aB11, and B7B11B) are justifiable. But that doesn't make all combinations with SWIR good performers, evaluation results showed that RB12B and B8B11B were the worst performers among 33 band combinations, which highlights the need for meaningful band combinations to detect flooding. Similar to SWIR, combination with NIR were expected to perform well, but this was not seen in practice as we look at the top 5 combinations in the boxplots. Instead, narrow-band NIR (NNIR) shows better

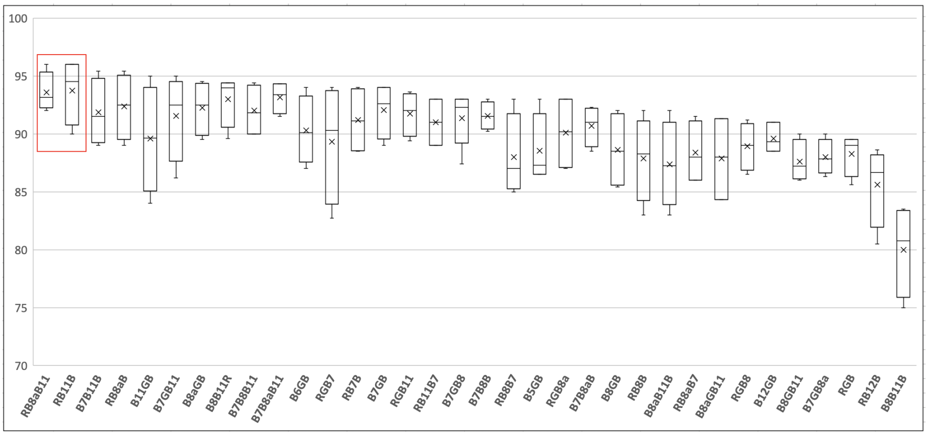


Fig. 4: Boxplot of 33 band combinations showing F1 score from all four models (VGG16, ResNet18, ResNet50, and EfficientNetB0)

performance. Individual bands are known to have specific abilities to identify types and hence to be useful in object detection, but single bands in themselves are insufficient for useful detection, and instead requires useful combinations. With that, I argue that to identify shallow water bodies such as floods, RB11B, RB8aB11, and B7B11B combinations are better identifiers with deep learning models.

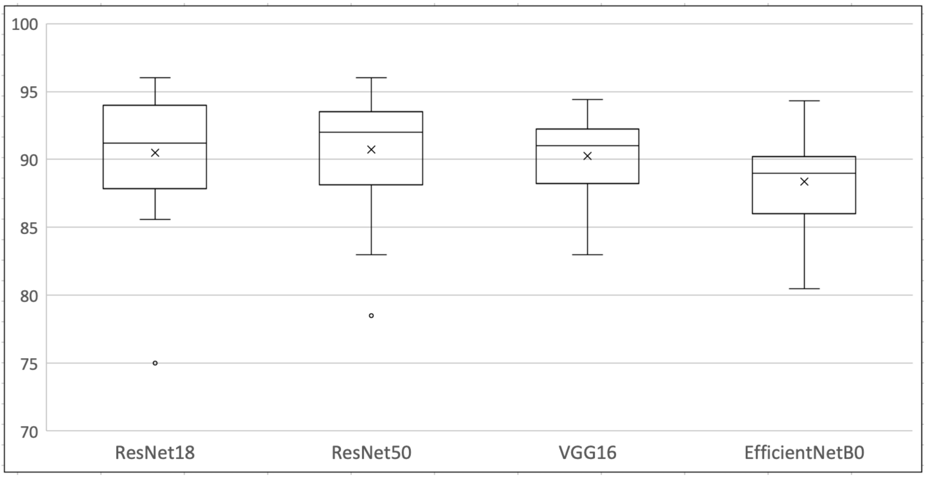


Fig. 5: Model Performance Across all 33 Combinations

Figure 5 shows the overall performance of the four pre-trained architectures across all 33 band combinations. This shows that ResNet18 and ResNet50 had similar results but by examining the top-10 best performing combination in Table 2, I see that the top 3 results are from ResNet18 and those are also the best combinations according to the boxplot comparison in Figure 4. One potential reason for ResNet18 outperforming ResNet50 is that ResNet18 better generalised with the relatively small training sets.

Combination	**Model**	**F1**	**Kappa**	**TN**	**TP**
RB8aB11	**ResNet18**	**0.96**	**0.913**	**0.95**	**0.97**
RB11B	**ResNet18**	**0.96**	**0.912**	**0.97**	**0.94**
B7B11B	**ResNet18**	**0.954**	**0.905**	**0.95**	**0.95**
RB8aB	ResNet50	0.954	0.90	0.96	0.94
B11GB	ResNet50	0.95	0.89	0.95	0.94
B7GB11	ResNet18	0.95	0.89	0.96	0.93
B8aGB	ResNet50	0.945	0.884	0.95	0.94
RB8B11	ResNet18	0.944	0.882	0.92	0.98
B7B8B11	VGG16	0.944	0.88	0.97	0.90
B7B8aB11	EfficientNetB0	0.943	0.88	0.94	0.95

Table 2: Top-10 Best Performing Combinations in Terms of F1 and Kappa with Pre-Trained Models

4.1.4 Usefulness Of RGB Based Pre-Trained Models

Considering the overall best performance of RB8aB11 combinations, I trained VGG16, ResNet18 and ResNet50 from scratch on RB8aB11 and compared results with similar models derived from pre-trained weights.

The models ran with varied epoch lengths in order to find the best performance. The results are presented in Figure 6. Here it can be seen that among the three models, ResNet50 showed the best performance, but it could not compete with the pre-trained models with the performance of 0.96 F1 score that was shown for the previous study in Table 2. Also, it can be noted that among pre-trained models, ResNet18 performed better than ResNet50 when trained for 100 epochs. Whereas in the case of 'from scratch' models ResNet50 performed better when trained for 500 epochs. By that, I argue that the pre-trained model requires less deep models, while training from scratch requires deeper models to perform at their optimum level.

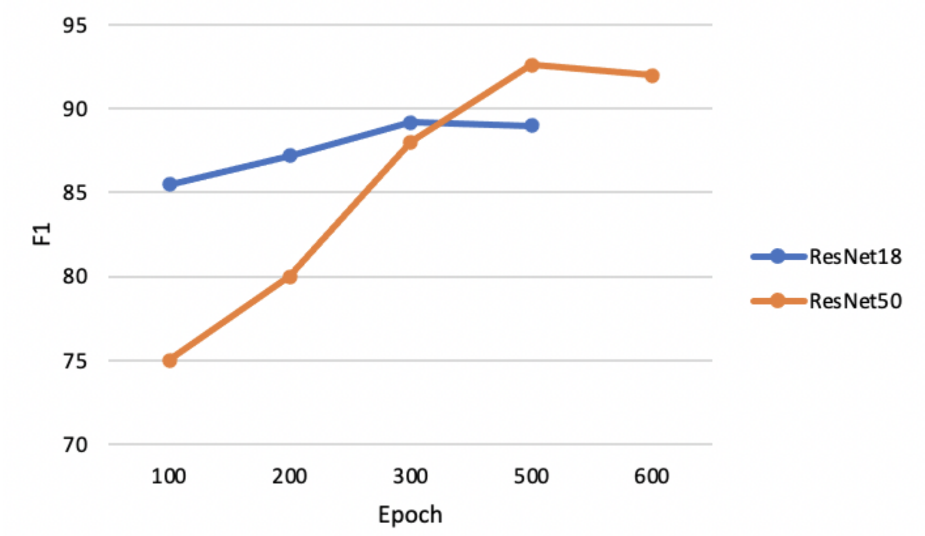


Fig. 6: Validation Results for Models Trained from Scratch on RB8aB11

4.2 Social Media Disaster Tweet Classification

My second distinct branch of research activity took the multi-modality of data for disaster analysis to perhaps its most extreme case and investigated issues relevant to representation use when applied to social media content that encodes information about disasters. It is worth noting that this work was performed prior to the identification of the main research questions that are now being proposed in this work.

As social media has grown tremendously, it has also become one of the major sources of information in disaster management. Previous work has shown the importance of social media in disaster management due to high uses of social media to post disaster related updates, or to ask for or to provide help. Even rescue agencies use social media platforms to spread information. However, social media can also be a source of fake or non-informative information that can obscure the truly valuable content. For this reason my initial research focused on an investigation of disaster related social media content analysis to determine whether the level of informativeness could easily be estimated and hence factored into a complete framework for multi-modal disaster data integration.

Specifically, my initial work focused on the classification of Twitter disaster related tweet data into a two level classification, i.e., informativeness and type of information, where the type of information can be further categorised into *affected individual*, *caution and advice*, *donation and volunteering*, and *sympathy and support*. This particularly classification approach had already been proposed by Imran et al., [46]. Hence from my perspective, this study was mainly carried out to understand the impact of different textual feature representation techniques for disaster related social media data. Given the potential uniqueness of disaster related social media with novel terminology and a lack of time for composing grammatically well-formed tweets, I believe that the approach to analysing tweet data in this context is vital.

This study analysed two aspects of textual representations, that is: 1) comparison of the feature representation techniques Bag-of-Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF), pre-trained Word2Vec model trained on Google news data, and Doc2Vec; and 2) comparison of pre-trained distributed representations, i.e., Word2Vec, GloVe, ELMo and BERT. ELMo and BERT are recent state-of-art transformer techniques, which have shown great performance in several domains.

For this work I utilised the benchmark CrisisLex [78] and CrisisNLP [46] datasets. These datasets consist of tweets relating to earthquakes, floods, and storms over a three years time period (2012 - 2015). I used a total of 15 Twitter datasets of different disasters (earthquakes, floods, and storms) where 6 datasets were taken from CrisisLex [78] and 9 datasets were taken from Crisis NLP [46]. Each dataset consists of data labelled according to the informativeness of the tweet on the particular event, and according to the type of information in each tweet. For transparency in results, I split the data in two ways: 1) Leave one out (LOO), and 2) Cross Disaster. In the case of LOO, I trained the model on 14 datasets out of 15 and kept back one dataset as a test dataset. Consequently, the model trained 15 times, and I obtain 15 test scores from which I calculated the average weighted F1 score of all test results. In the case of Cross Disaster, I trained the model on one type of disaster data at a time and tested the model on different disaster datasets individually. From this, I subsequently calculated the average of weighted F1 score over all test results.

4.2.1 Comparison of Feature Representation Techniques

For the first study, that is the comparison of the feature representation techniques BoW, TF-IDF, Word2Vec, and Doc2Vec, I also compared the performance based on unigram, hybrid unigram and bigram features to add a word-to-word relation feature and part of speech (POS) tagging, which capture the syntactic behaviour of words. However for Word2Vec, POS tagging was not used, since the pre-trained model does not support this. The actual classification model was based on logistic regression which was fed with data processed through the different feature representations.

Word2Vec with Unigram encodings outperformed all other representations as shown in Table 3. This I believe was due to the fact that a pre-trained model is trained on a comparatively large corpus, which creates a much improved context similarity representation. However, Word2Vec struggles with out of vocabulary words, which are more prevalent in text from a Twitter feed, as such data contains human created hashtags and slang words. Meanwhile, there is very little difference between hybrid n-gram and unigram. This could be due to the use of a pre-trained model, which does not consist of words related to hashtags and bigrams. This work was presented as a poster in the International Systems for Crisis Response and Management (ISCRAM 2019) [49]

Data Trained On:	Earthquake		Flood		Storm		LOO Avg F1	
Feature	I	IT	I	IT	I	IT	I	IT
BoW Hybrid	0.66	0.63	0.7	0.69	0.71	0.66	0.76	0.73
BoW Unigram	0.67	0.64	0.68	0.71	0.67	0.66	0.76	0.74
BoW POS	0.67	0.63	0.69	0.71	0.75	0.67	0.77	0.74
Doc2Vec Hybrid	0.72	0.61	0.68	0.66	0.67	0.66	0.79	0.71
Doc2Vec Unigram	0.71	0.62	0.68	0.66	0.65	0.66	0.78	0.71
Doc2Vec POS	0.72	0.61	0.67	0.67	0.63	0.66	0.78	0.71
TF-IDF Hybrid	0.69	0.54	0.72	0.62	0.62	0.6	0.79	0.74
TF-IDF Unigram	0.72	0.58	0.73	0.69	0.62	0.65	0.79	0.75
TF-IDF POS	0.72	0.58	0.72	0.7	0.72	0.65	0.79	0.75
Word2Vec Hybrid	0.76	0.68	**0.75**	**0.73**	**0.81**	0.7	0.8	**0.76**
Word2Vec Unigram	**0.78**	**0.7**	**0.75**	**0.73**	**0.81**	**0.71**	**0.81**	**0.76**
Word2Vec POS	NA	NA	NA	NA	NA	NA	NA	NA
Average F1 score for I = Informativeness Classification, and IT = Information Type Classification								

Table 3: Comparative Results for Different Feature Representations with Cross Disaster and Leave One Out

4.2.2 Comparison Of The Pre-Trained Distributed Representations

For the second study, I utilised a number of pre-trained distributed representations, namely Word2Vec, GloVe, ELMo and BERT, to perform a similar study with two level classification into informativeness and information type. More concretely, I made use of the following embedding specifications: (a) GloVe, which was trained on 2 billion tweets, 27 billion tokens, 1.2 million words and with 200 dimension vectors; (b) Word2Vec, which is a pre-trained model which includes 300 dimension word vectors for a vocabulary of 3 million words and phrases, and has been trained on 100 billion words from a Google News dataset; (c) ELMo (Small), which was trained on a raw 1 Billion Word Benchmark [12] and which has a 1024 dimension output vector; and (d) BERT (Base), which is trained on the concatenation of BooksCorpus (800M words) [111] and English Wikipedia (2,500M words) and uses 12 transformer blocks, a hidden layer of size 768 with a filter size of 3,072, and 12 self-attention heads.

With the objective of evaluating the strength of embedding models for disaster tweet data, I utilised a vanilla feed-forward neural network (FFNN) for the actual classification task. The FFNN is chosen due to a desire to reduce the complexity as ELMo and BERT themselves are computationally expensive, and to provide transparency of the model. I used the model architecture as shown in figure 7, but I fed each pre-trained embedding layer to two fully connected (FC) layers of 128 units. As the activation function, I utilised Leaky ReLU with alpha=0.1 after each layer to overcome the dying ReLU problem i.e. instead of having zero slope for each $x<0$, Leaky ReLU uses a small negative slope. In order to avoid over-fitting, I applied both a dropout of 0.5 after each FC layer, and L2 regularisation of 0.001 in each layer. For the output layer, a softmax function was used with an output of 2 or 5 dimension logit, depending on the classification task. The Adam optimiser and categorical cross-entropy loss function were used for training. Since ELMo and BERT require the additional parameter of maximum sequence length to process the input sentence at once, I set the maximum sequence length to 128 characters, which is based on the Twitter word limit.

The results from the LOO training approach showed that ELMO outperformed in the informativeness classification task, while Word2Vec outperformed in information type classification. However, there was not too much difference in

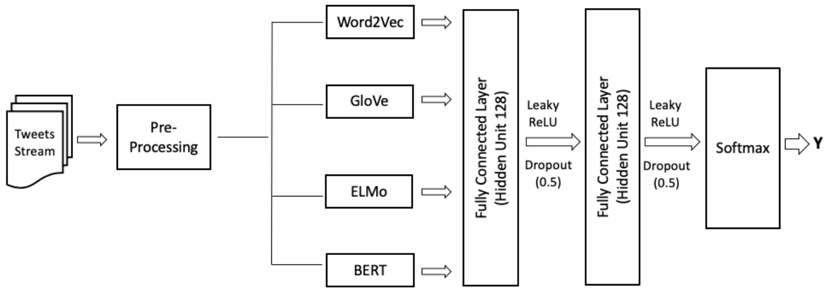


Fig. 7: Study 1.2 - Model Architecture

the results among different embeddings. Meanwhile, results from cross disaster training showed similar outcomes with mixed results between Word2Vec, GloVe and ELMo for both classification tasks.

	Embedding	P	R	A	F1
Informativeness	Word2Vec	0.83	0.80	0.80	0.80
	GloVe	0.83	0.80	0.80	0.80
	ELMo	0.84	0.80	0.80	**0.81**
	BERT	0.83	0.77	0.77	0.78
Information Type	Word2Vec	0.78	0.76	0.76	**0.76**
	GloVe	0.77	0.75	0.75	0.75
	ELMo	0.76	0.76	0.75	0.75
	BERT	0.77	0.75	0.75	0.75

Table 4: Classification Result of LOO Training Approach for Distributed Representations Comparison

The advantage of ELMo and BERT is that they capture the context of word in a sentence rather than generalising the word representation as in Word2Vec and GloVe representations. This overcomes the out of vocabulary (OOV) issue, where BERT utilises the WordPiece tokenisation embeddings and ELMo uses a character-based approach. These results clearly highlighted that this advantage of ELMo and BERT did not benefit us in the disaster related data. This is perhaps due to the fact that Twitter messages are short and highly informal messages

Trained on	Embedding	P		R		A		F1	
		I	IT	I	IT	I	IT	I	IT
Earthquake Data	Word2Vec	0.83	0.74	0.75	0.7	0.75	0.7	0.78	**0.69**
	GloVe	0.83	0.72	0.77	0.69	0.77	0.69	0.79	0.67
	ELMo	0.84	0.73	0.79	0.69	0.79	0.69	**0.81**	0.68
	BERT	0.83	0.67	0.76	0.73	0.76	0.68	0.78	0.68
Flood Data	Word2Vec	0.79	0.76	0.73	0.72	0.73	0.72	0.73	**0.72**
	GloVe	0.8	0.73	0.77	0.69	0.77	0.69	**0.76**	0.68
	ELMo	0.8	0.76	0.75	0.7	0.75	0.7	0.75	0.70
	BERT	0.8	0.72	0.76	0.67	0.76	0.67	0.75	0.66
Storm Data	Word2Vec	0.83	0.74	0.76	0.71	0.76	0.71	0.78	0.71
	GloVe	0.84	0.74	0.74	0.72	0.74	0.72	0.77	**0.72**
	ELMo	0.84	0.75	0.78	0.71	0.78	0.71	**0.80**	0.71
	BERT	0.83	0.74	0.75	0.69	0.75	0.69	0.77	0.69

Table 5: Cross Disaster Classification Result for Distributed Representations Comparison, I = Informativeness and IT = Information Type

with significant amounts of unknown or rare words which cannot be captured by pre-trained models.

Further analysis was performed on wrongly predicted tweets and found that tweets with the following issues commonly were interpreted wrongly: 1) messages with a lot of hashtags like #ineedwater, or 2) where the pre-processing step reduced some tweets to 2-3 words whose meaning was difficult to interpret.

With this, I argue that for data like Twitter, Word2Vec and GloVe are still better representation techniques. Also that there is a need for more fine-grain pre-processing approach in the case of Twitter data. Apart from that, the final conclusion from both studies showed that in cross-disaster training, disaster related tweets appears to be independent of the actual type of disaster. This could be due to similar vocabulary shared across the type of disaster in terms of donation, sympathy, needs, etc. This work was presented in Advances in Social Network Analysis and Mining (ASONAM 2019) [48].

4.2.3 Summary

My results to date have covered both textual and multi-spectral analysis. My focus moving forward is however on the multi-spectral data. From this work, it can be seen that indices and band combinations (i.e., the first and third studies) provide similar results in terms of performance, that is 0.96 F1 score. Also, I conclude that combining the right bands can enhance the performance of shallow water or flood water detection, which in this case was RB8aB11, and RB11B. The reason for their best performance could be that SWIR bands are sensitive towards the water, soil moisture and clouds, whereas the Red band is sensitive to built-up areas and vegetation, which makes them ideal for flood detection. The major limitation in this work remains the use of the existing pre-trained model, which can only train three-band combinations, and has limited us from seeing the potential impact of all 10 base band on the flood detection task. This limitation motivates future work to obtain transfer learning modelling approaches specific to the EO domain.

5 Planned Work

In the main branch of our initial work, I explored the concept of multi-spectral satellite imaging with transfer learning. The analysis showed that correct spectral combinations or indices enhance the specific information content in multi-spectral images. Work also showed that deep learning is an excellent approach to address the underlying multi-spectral data complexity. Unfortunately, it also shows that training from scratch requires a lot of labelled data and pre-trained models are designed for only three band information. This underscores the limitations of conventional supervised learning and transfer learning approaches to completely explore multiple band data.

Considering the limitation of pre-trained models and supervised learning, in the next phase of this research I plan to extend my previous work with respect to two aspects: firstly, learning the spatio-spectral features of the EO domain by utilising semi-supervised learning techniques to provide a more robust and generalised framework. Secondly, exploring transfer learning by applying already learned weights on various satellite data sources varying in their spectral, spatial and even their mode (i.e., radar versus imaging). With that goal I can further divide my future work into the following high level tasks for disaster analysis from satellite data:

- Learning the spatio-spectral representations in semi-supervised and self-supervised settings to improve information content.
- Evaluating transfer learning with already learned spatio-spectral features across different dimensions of satellite data.
- Testing the model for cross-domain application across different disasters.

In the following sub-sections, I expand on these plans.

5.1 Learning spatio-spectral representations in semi-supervised and self-supervised settings

Multi-spectral imagery provides a great level of information based on different absorption and reflectance properties for different geographical features. I believe that this property of multi-spectral data can be one of great benefit and can provide better representations for EO tasks than normal RGB images with features that are more relevant to the given data types. Apart from normal spectral variation, multi-spectral data commonly have various spatial resolution and temporal properties, which can increase its usefulness while also providing challenges.

So far I have worked with multi-spectral information that has low and high-resolution spectral bands to identify the floods. I observed that multi-spectral information provides distinguishable geographical features, which help in mapping floods in the images. But our work was limited to three bands or indexing techniques, which was due to the lack of labelled data to utilise deep learning model for all spectral bands.

To eradicate the data size issue, pre-trained models are commonly applied in several domains including satellite imagery, but they are limited to three channels information only, i.e. RGB, which is not ideal for the multi-spectral type data. As per our current research into this area, there is very limited available pre-trained models for remote sensing that specifically incorporate multi-spectral information. Our goal here therefore is to build on transfer learning methods with representations learned from spectral data specific to the EO domain.

In order to build a pre-trained model in a supervised learning setting, a large amount of labelled multi-spectral satellite data is required, which unfortunately is the major issue across the EO domain. For this reason, I focus on a semi-supervised and self-supervised approach, which utilises largely unlabelled data in order to learn representations actively.

As described in Section 2.3.2, the good performance of S4L [106], and SimCLRv2 [14], BYOL [33], SimSiam [16] was based on the applicability of smoothness assumption. I believe that it will be beneficial for satellite data, as multi-spectral and SAR data itself is a variation of an image and should have the same labels across different bands and modality. Most of the methods utilise cropping, rotation, Gaussian blur or colour distortion of images as augmentation. Among them, colour distortion and Gaussian blur can be considered as variations within multi-spectral and SAR data. I believe that these can help in learning better

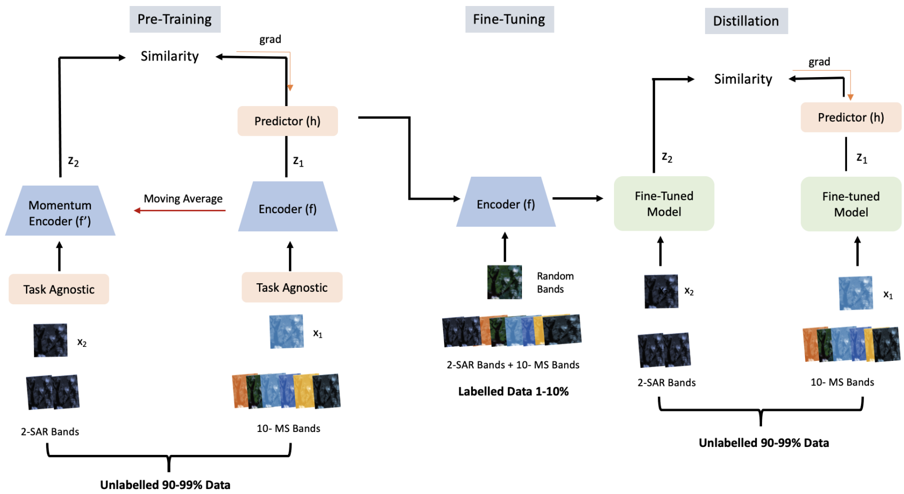


Fig. 8: Model architecture motivated from SimCLRv2 [14] and BYOL [33]

spatio-spectral representations and can help in generalising spectral and SAR data with geographical features.

5.1.1 Experimental Plan

For training such a model, various benchmark datasets are already available, these include EuroSat [40], BigEarthNet [93], and Sen12MS [87]. The BigEarthNet and Sen12MS datasets consist of images from Sentinel-1 (SAR) and Sentinel-2 (MS) for land coverage classifications and segmentation. EuroSat consists of 64x64 pixel chip of MS with single labels of land coverage, BigEarthNet dataset consists of 120x120 images with multi-labels for land coverage, and Sen12MS data consists of 256x256 images with labels for scene classification and semantic segmentation. Considering theses datasets, this study will utilise the Sen12MS dataset for training the model as it is the most diverse and consists of a huge volume (180K) of Sentinel-1 and Sentinel-2, which makes it more relevant for learning spatio-spectral information for different geographical aspects.

The model architecture will be based on a combination of self-supervised and semi-supervised learning as shown in Figure 8, which is motivated from recent work in SimCLRv2 [14], BYOL [33], and SimSiam [16]. This work will utilise the semi-supervised approach by leveraging a large amount of unlabelled data and train it in a self-supervised task agnostic way. This then can be utilised for fine-tuning with a supervised mechanism on a small amount of labelled data. For the task agnostic model, I will leverage the random MS band and SAR images, along with basic augmentation methods such as rotation, cropping, flipping etc. This way, the self-supervised model will be trained to get latent representations of the unlabelled data, which then can be fine-tuned further with supervised learning. The added advantage of this training approach is that the model learns latent representation from unlabelled data once. And then supervised learning part can leverage both classification and segmentation tasks, as the labelled data requirement is reduced in such a training method. The last step is self-distillation for task-specific learning from the fine-tuned model, which in our case is random spectral and radar bands. This will allow us to have a model, which can be highly advantageous in other EO domain tasks.

Such training requires an unsupervised loss in order to optimise the performance of pre-training. For my model, the unsupervised loss will be the contrastive loss [13] as in Eq 2, which is calculated based on the similarities within two variations of an image. The goal is to maximise the agreement between two variations of a single image and minimise it with respect to the other set. For supervised training loss, traditional cross-entropy loss will be used. I believe that this approach will provide us with the model, which can learn the representations that can be transferred to a more task-specific model including for example our central flood prediction task. The contrastive loss is defined as:

$$Loss = 2 - 2 \cdot \frac{h}{\|h\|_2} \cdot \frac{z_2}{\|z_2\|_2} \quad (2)$$

where $\|\cdot\|_2$ is l_2 normalisation, h is the multi-layer perceptron (MLP) prediction output and z_2 is the MLP projection output.

5.2 An investigation of transfer learning with already learned spatio-spectral features across different dimensions of satellite data.

While spatio-spectral data can be considered to be the baseline of satellite data, I believe that by achieving a model with learned geographical features, it can be used for several other EO domain tasks. As mentioned earlier, MS, MR, MT or radar data might initially be thought of as different data types, but they are highly interdependent. Though optical and radar sensors are different in terms of data format, they do provide similar information in terms of geographical features along with different resolution and temporal information. Since transfer learning in general has shown outstanding performance in several domains, I believe the applicability of transfer learning in EO specific domains can be improved.

5.2.1 Experimental Plan

The next task will be to evaluate the applicability of the model trained on spatio-spectral data with self-supervised settings as mentioned in section 5.1.1. These learned representations then can be used to transfer for more task-specific supervised training for different EO domain tasks including disaster analysis with different data such as radar, or different spectral information. For the evaluation of knowledge transfer success, I will be using benchmark datasets such as EuroSat [40] for EO domain land cover classification, MediaEval2019 [6] and Multi3Net

[85] for floods related tasks and xView2 [35] for various disaster mapping tasks. The performance of same will also be evaluated against state-of-art pre-trained models trained on ImageNet [21] data and recently available BigEarthNet pre-trained models, which is based on supervised methods for the same EO domain tasks.

5.3 Generalising the model across different disasters

Different disasters naturally lead to different types of damages such as flooding mostly consisting of water mapping around areas or buildings, whereas earthquake leads to more infrastructure damage and hence involves changes to roads, bridges or building. Meanwhile, hurricanes or typhoons result in both infrastructure damage and flooded regions; similarly, wildfires and volcano require different mapping. Considering the different type of damage from different disasters, individual task-specific models are required. This increases the requirement of large labelled data which are scarce in the remote sensing domain.

5.3.1 Experiment Plan

Considering the challenge of having task specific models, this work will evaluate the pre-trained models on two tasks, namely flood and building damage segmentation. I believe these two tasks can be combined in order to have a single model for damage assessment due to floods, typhoon, hurricane etc. With this, I identified a number of datasets such as MediaEval2017 [10], Multi3Net [85] and xView2 [35], which are different in terms of resolution, spectral information, and sensors. This is where our pre-trained model will be helpful in transferring the EO related features and train on a specific task irrespective of the resolution, spectral information and sensors. Although another challenge remains in dealing with the different types of labels available for each dataset, such as MediaEval2017 consisting of flood segmentation, Multi3Net consisting of buildings and flooded buildings segmentation. Whereas xView2 consist of different disasters (i.e. floods, earthquake, hurricane, volcano, wildfires) images with labels as disaster, building detection and their intensity of the damage. With that, this work will leverage these datasets as multi-label segmentation problems by fine-tuning the pre-trained models for task-specific learning. I believe by learning disaster related features, this work can be a step towards better generalisation for different disaster damage types. In result, this work can produce an end-to-end system to analyse floods and building damage from satellite images irrespective of resolution and spectral information.

5.4 Planned Timeline

The two years (2021-2022) research plan shown in Figure 9 is split into eight quarters. The two years of work is divided into four achievable tasks as follows:

- *Implementation of Self-Supervised and Semi-Supervised learning methods:* The task will be carried out throughout four quarters of 2021 to the first quarter of 2022 with different settings of data processing, hyperparameters, and learning algorithms. Meanwhile, I plan to submit this research work in at least two top tier-1 computer vision and remote sensing conferences.
- *Evaluation of transfer learning on benchmark datasets:* Simultaneously, evaluation of learned features from Self-Supervised and Semi-Supervised learning methods will be performed on EO domain benchmark datasets such as BigEarthNet, Sen12MS etc.
- *Evaluate the application of learned features in disaster analysis tasks:* From later quarters of 2021 until mid of Q2 in 2022, I will begin evaluating the usefulness of the learned features in disaster analysis tasks. The major focus of this task will remain on methods to obtain a model to generalise across different disasters.

- *Thesis writing:* The last two and half quarters of 2022 will be dedicated to thesis writing.

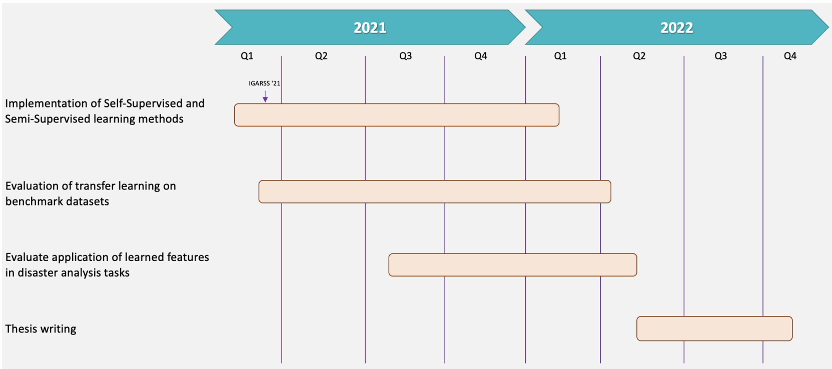


Fig. 9: Research Timeline

6 Concluding Remarks

In the domain of multi-spectral satellite data, the impact of spectral information was analysed to identify floods in images applying transfer learning. From this work, it was evident that traditional pre-trained models (i.e. models trained on ImageNet) provide better performance than the model trained from scratch, even though ImageNet data are natural images rather than multi-spectral images. Based on this, my future work focuses on attaining an EO based pre-trained model with a semi-supervised learning approach to provide robust performance. The efficiency of learned representations will be evaluated on several benchmark EO domain tasks including disaster analysis.

Our achievements in this research to date include one poster presentation, one short and two long papers along with participation in the MediaEval2019 competition and various academic events.

Long Paper

[51] Jain, P., Schoen-Phelan, B., Ross, R.: Automatic flood detection in sentinel-2images using deep convolutional neural networks. In: Proceedings of the 35th Annual ACM Symposium on Applied Computing. pp. 617–623 (2020)

[52] Jain, P., Schoen-Phelan, B., Ross, R.: Tri-Band Assessment of Multi-Spectral Satellite Data for Flood Detection. In: Proceedings of the Machine Learning for Earth Observation. (2020) (Accepted)

Short Paper

[48] Jain, P., Ross, R., Schoen-Phelan, B.: Estimating distributed representation performance in disaster-related social media classification. In: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. pp. 723–727 (2019)

[50] Jain, P., Ross, R., Schoen-Phelan, B.: MediaEval2019: Flood Detection in Time Sequence Satellite Images. In: Proceedings of the MediaEval 2019 Workshop (2020)

Poster

[49] Jain, P., Schoen-Phelan, B., Ross, R.: Comparative study of feature representations for disaster tweet classification. In: Proceedings of the 16th International Conference on Information Systems for Crisis Response and Management. pp. 1371-1372 (2019)

Acknowledgments

References

1. Akter, S., Wamba, S.F.: Big data and disaster management: a systematic review and agenda for future research. Annals of Operations Research **283**(1-2), 939–959 (2019)
2. Anthony, G., Gregg, H., Tshilidzi, M.: Image classification using svms: one-against-one vs one-against-all. arXiv preprint arXiv:0711.2914 (2007)

3. Ayush, K., Uzkent, B., Meng, C., Burke, M., Lobell, D., Ermon, S.: Geography-aware self-supervised learning. arXiv preprint arXiv:2011.09980 (2020)
4. Bangira, T., Alfieri, S.M., Menenti, M., van Niekerk, A.: Comparing thresholding with machine learning classifiers for mapping complex water. Remote Sensing **11**(11), 1351 (2019)
5. Benedetti, P., Ienco, D., Gaetano, R., Ose, K., Pensa, R.G., Dupuy, S.: m^3Fusion: A deep learning architecture for multiscale multimodal multitemporal satellite data fusion. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **11**(12), 4939–4949 (2018)
6. Benjamin Bischke, Patrick Helber, Erkan Basar, Simon Brugman, Zhengyu Zhao and Konstantin Pogorelov: The Multimedia Satellite Task at MediaEval 2019: Flood Severity Estimation. In: Proc. of the MediaEval 2019 Workshop. Sophia-Antipolis, France (2019)
7. Berthelot, D., Carlini, N., Cubuk, E.D., Kurakin, A., Sohn, K., Zhang, H., Raffel, C.: Remixmatch: Semi-supervised learning with distribution matching and augmentation anchoring. In: International Conference on Learning Representations (2019)
8. Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C.A.: Mixmatch: A holistic approach to semi-supervised learning. In: Advances in Neural Information Processing Systems. pp. 5049–5059 (2019)
9. Bischke, B., Helber, P., Folz, J., Borth, D., Dengel, A.: Multi-task learning for segmentation of building footprints with deep neural networks. In: 2019 IEEE International Conference on Image Processing (ICIP). pp. 1480–1484 (2019)
10. Bischke, B., Helber, P., Schulze, C., Venkat, S., Dengel, A., Borth, D.: The multimedia satellite task at mediaeval 2017. In: Proc. of the MediaEval 2017 Workshop. Dublin, Ireland
11. Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., Joulin, A.: Unsupervised learning of visual features by contrasting cluster assignments. arXiv preprint arXiv:2006.09882 (2020)

12. Chelba, C., Mikolov, T., Schuster, M., Ge, Q., Brants, T., Koehn, P., Robinson, T.: One billion word benchmark for measuring progress in statistical language modeling. arXiv preprint arXiv:1312.3005 (2013)
13. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. arXiv preprint arXiv:2002.05709 (2020)
14. Chen, T., Kornblith, S., Swersky, K., Norouzi, M., Hinton, G.: Big self-supervised models are strong semi-supervised learners. arXiv preprint arXiv:2006.10029 (2020)
15. Chen, X., Fan, H., Girshick, R., He, K.: Improved baselines with momentum contrastive learning. arXiv preprint arXiv:2003.04297 (2020)
16. Chen, X., He, K.: Exploring simple siamese representation learning. arXiv preprint arXiv:2011.10566 (2020)
17. Chen, Y., Jiang, H., Li, C., Jia, X., Ghamisi, P.: Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Transactions on Geoscience and Remote Sensing **54**(10), 6232–6251 (2016)
18. Cheng, G., Han, J., Lu, X.: Remote sensing image scene classification: Benchmark and state of the art. Proceedings of the IEEE **105**(10), 1865–1883 (2017)
19. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: Randaugment: Practical automated data augmentation with a reduced search space. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 702–703 (2020)
20. Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., Raska, R.: Deepglobe 2018: A challenge to parse the earth through satellite images. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 172–17209. IEEE (2018)
21. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
22. DeVries, T., Taylor, G.W.: Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552 (2017)
23. Dian, R., Li, S., Fang, L., Wei, Q.: Multispectral and hyperspectral image fusion with spatial-spectral sparse representation. Information Fusion **49**, 262–270 (2019)

24. Dickey, T., Lewis, M., Chang, G.: Optical oceanography: recent advances and future directions using global remote sensing and in situ observations. Reviews of geophysics **44**(1) (2006)
25. Dong, L., Shan, J.: A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. ISPRS Journal of Photogrammetry and Remote Sensing **84**, 85–99 (2013)
26. Doshi, J., Basu, S., Pang, G.: From satellite imagery to disaster insights. arXiv preprint arXiv:1812.07033 (2018)
27. Dosovitskiy, A., Fischer, P., Springenberg, J.T., Riedmiller, M., Brox, T.: Discriminative unsupervised feature learning with exemplar convolutional neural networks. IEEE transactions on pattern analysis and machine intelligence **38**(9), 1734–1747 (2015)
28. Duarte, D., Nex, F., Kerle, N., Vosselman, G.: Satellite image classification of building damages using airborne and satellite image samples in a deep learning approach. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences **4**(2) (2018)
29. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. International journal of computer vision **88**(2), 303–338 (2010)
30. Fathollahi, F., Zhang, Y.: Adaptive band selection for pan-sharpening of hyperspectral images. International Journal of Remote Sensing **41**(10), 3924–3947 (2020)
31. Feyisa, G.L., Meilby, H., Fensholt, R., Proud, S.R.: Automated water extraction index: A new technique for surface water mapping using landsat imagery. Remote Sensing of Environment **140**, 23–35 (2014)
32. Gidaris, S., Singh, P., Komodakis, N.: Unsupervised representation learning by predicting image rotations. arXiv preprint arXiv:1803.07728 (2018)
33. Grill, J.B., Strub, F., Altché, F., Tallec, C., Richemond, P.H., Buchatskaya, E., Doersch, C., Pires, B.A., Guo, Z.D., Azar, M.G., et al.: Bootstrap your own latent: A new approach to self-supervised learning. arXiv preprint arXiv:2006.07733 (2020)

34. Guo, Y., Liu, Y., Georgiou, T., Lew, M.S.: A review of semantic segmentation using deep neural networks. International journal of multimedia information retrieval **7**(2), 87–93 (2018)
35. Gupta, R., Goodman, B., Patel, N., Hosfelt, R., Sajeev, S., Heim, E., Doshi, J., Lucas, K., Choset, H., Gaston, M.: Creating xbd: A dataset for assessing building damage from satellite imagery. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 10–17 (2019)
36. Han, X.H., Shi, B., Zheng, Y.: Residual hsrcnn: Residual hyper-spectral reconstruction cnn from an rgb image. In: 2018 24th International Conference on Pattern Recognition (ICPR). pp. 2664–2669. IEEE (2018)
37. Haq, M., Akhtar, M., Muhammad, S., Paras, S., Rahmatullah, J.: Techniques of remote sensing and gis for flood monitoring and damage assessment: A case study of sindh province, pakistan. The Egyptian Journal of Remote Sensing and Space Science **15**(2), 135–141 (2012)
38. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
39. Hegazy, I.R., Kaloop, M.R.: Monitoring urban growth and land use change detection with gis and remote sensing techniques in daqahlia governorate egypt. International Journal of Sustainable Built Environment **4**(1), 117–124 (2015)
40. Helber, P., Bischke, B., Dengel, A., Borth, D.: Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **12**(7), 2217–2226 (2019)
41. Hu, F., Xia, G.S., Hu, J., Zhang, L.: Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sensing **7**(11), 14680–14707 (2015)
42. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708 (2017)

43. Hughes, L.H., Merkle, N., Bürgmann, T., Auer, S., Schmitt, M.: Deep learning for sar-optical image matching. In: IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. pp. 4877–4880. IEEE (2019)
44. Hunt, E.R., Hively, W.D., Fujikawa, S.J., Linden, D.S., Daughtry, C.S., McCarty, G.W.: Acquisition of nir-green-blue digital photographs from unmanned aircraft for crop monitoring. Remote Sensing **2**(1), 290–305 (2010)
45. Ienco, D., Interdonato, R., Gaetano, R., Minh, D.H.T.: Combining sentinel-1 and sentinel-2 satellite image time series for land cover mapping via a multi-source deep learning architecture. ISPRS Journal of Photogrammetry and Remote Sensing **158**, 11–22 (2019)
46. Imran, M., Mitra, P., Castillo, C.: Twitter as a Lifeline: Human-annotated Twitter Corpora for NLP of Crisis-related Messages (may 2016). https://doi.org/10.1006/bbrc.1995.2431, `http://arxiv.org/abs/1605.05894`
47. Iqbal, S., Ghani, M.U., Saba, T., Rehman, A.: Brain tumor segmentation in multi-spectral mri using convolutional neural networks (cnn). Microscopy research and technique **81**(4), 419–427 (2018)
48. Jain, P., Ross, R., Schoen-Phelan, B.: Estimating distributed representation performance in disaster-related social media classification. In: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. pp. 723–727 (2019)
49. Jain, P., Schoen-Phelan, B., Ross, R.: Comparative study of feature representations for disaster tweet classification pp. 1371–1372 (2019)
50. Jain, P., Schoen-Phelan, B., Ross, R.: Mediaeval2019: Flood detection in time sequence satellite images (2019)
51. Jain, P., Schoen-Phelan, B., Ross, R.: Automatic flood detection in sentinei-2 images using deep convolutional neural networks. In: Proceedings of the 35th Annual ACM Symposium on Applied Computing. pp. 617–623 (2020)
52. Jain, P., Schoen-Phelan, B., Ross, R.: Tri-band assessment of multi-spectral satellite data for flood detection. In: Proceedings of the Machine Learning for Earth Observation (2020)

53. Jiang, J., Liu, F., Xu, Y., Huang, H., et al.: Multi-spectral rgb-nir image classification using double-channel cnn. IEEE Access **7**, 20607–20613 (2019)
54. Jiang, Y., Wu, D., Deng, Z., Qian, P., Wang, J., Wang, G., Chung, F.L., Choi, K.S., Wang, S.: Seizure classification from eeg signals using transfer learning, semi-supervised learning and tsk fuzzy system. IEEE Transactions on Neural Systems and Rehabilitation Engineering **25**(12), 2270–2284 (2017)
55. Joshi, A.R., Tarte, I., Suresh, S., Koolagudi, S.G.: Damage identification and assessment using image processing on post-disaster satellite imagery. In: 2017 IEEE Global Humanitarian Technology Conference (GHTC). pp. 1–7. IEEE (2017)
56. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. Medical image analysis **36**, 61–78 (2017)
57. Khan, S.H., He, X., Porikli, F., Bennamoun, M.: Forest change detection in incomplete satellite images with deep neural networks. IEEE Transactions on Geoscience and Remote Sensing **55**(9), 5407–5423 (2017)
58. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
59. Kulkarni, S.C., Rege, P.P.: Pixel level fusion techniques for sar and optical images: A review. Information Fusion **59**, 13–29 (2020)
60. Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A.: Deep learning classification of land cover and crop types using remote sensing data. IEEE Geoscience and Remote Sensing Letters **14**(5), 778–782 (2017)
61. LeCun, Y., Bengio, Y., et al.: Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks **3361**(10), 1995 (1995)
62. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998)
63. Li, N., Martin, A., Estival, R.: An automatic water detection approach based on dempster-shafer theory for multi-spectral images. In: 2017 20th International Conference on Information Fusion (Fusion). pp. 1–8. IEEE (2017)

64. Li, W., Zhang, X., Peng, Y., Dong, M.: Dmnet: A network architecture using dilated convolution and multiscale mechanisms for spatiotemporal fusion of remote sensing images. IEEE Sensors Journal (2020)
65. Li, Y., Martinis, S., Wieland, M.: Urban flood mapping with an active self-learning convolutional neural network based on terrasar-x intensity and interferometric coherence. ISPRS Journal of Photogrammetry and Remote Sensing **152**, 178–191 (2019)
66. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision. pp. 740–755. Springer (2014)
67. Liu, X., Liu, Q., Wang, Y.: Remote sensing image fusion based on two-stream fusion network. Information Fusion **55**, 1–15 (2020)
68. Liu, X., Chi, M., Zhang, Y., Qin, Y.: Classifying high resolution remote sensing images by fine-tuned vgg deep networks. In: IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. pp. 7137–7140. IEEE (2018)
69. Liu, X., Deng, C., Chanussot, J., Hong, D., Zhao, B.: Stfnet: A two-stream convolutional neural network for spatiotemporal image fusion. IEEE Transactions on Geoscience and Remote Sensing **57**(9), 6552–6564 (2019)
70. Mahdianpari, M., Salehi, B., Rezaee, M., Mohammadimanesh, F., Zhang, Y.: Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. Remote Sensing **10**(7), 1119 (2018)
71. McFeeters, S.K.: The use of the normalized difference water index (ndwi) in the delineation of open water features. International journal of remote sensing **17**(7), 1425–1432 (1996)
72. Meng, X., Shen, H., Li, H., Zhang, L., Fu, R.: Review of the pansharpening methods for remote sensing images based on the idea of meta-analysis: Practical discussion and challenges. Information Fusion **46**, 102–113 (2019)
73. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). pp. 565–571. IEEE (2016)

74. Mishra, K., Prasad, P.: Automatic extraction of water bodies from landsat imagery using perceptron model. Journal of Computational Environmental Sciences **2015** (2015)

75. Moumtzidou, A., Bakratsas, M., Andreadis, S., Gialampoukidis, I., Vrochidis, S., Kompatsiaris, I.: Road passability estimation using deep neural networks and satellite image patches. In: Proceedings of the 2019 International Conference on Big Data from Space (BiDS'19), Publications Office of the European Union. pp. 1–4 (2019)

76. Mountrakis, G., Im, J., Ogole, C.: Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing **66**(3), 247–259 (2011)

77. Noroozi, M., Favaro, P.: Unsupervised learning of visual representations by solving jigsaw puzzles. In: European conference on computer vision. pp. 69–84. Springer (2016)

78. Olteanu, A., Castillo, C., Diaz, F., Vieweg, S.: CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises. In Proceedings of the 8th International AAAI Conference on Weblogs and Social Media (ICWSM'14) (January 2014) (2014)

79. Papadomanolaki, M., Vakalopoulou, M., Zagoruyko, S., Karantzalos, K.: Benchmarking deep learning frameworks for the classification of very high resolution satellite multispectral data. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences **3**(7) (2016)

80. Peng, Y., Li, W., Luo, X., Du, J., Gan, Y., Gao, X.: Integrated fusion framework based on semicoupled sparse tensor factorization for spatio–temporal–spectral fusion of remote sensing images. Information Fusion (2020)

81. Qiu, C., Schmitt, M., Geiß, C., Chen, T.H.K., Zhu, X.X.: A framework for large-scale mapping of human settlement extent from sentinel-2 images via fully convolutional neural networks. ISPRS Journal of Photogrammetry and Remote Sensing **163**, 152–170 (2020)

82. Quang, N.H., Tuan, V.A., Hao, N.T.P., Hang, L.T.T., Hung, N.M., Anh, V.L., Phuong, L.T.M., Carrie, R.: Synthetic aperture radar and optical remote sensing image fusion for flood monitoring in the vietnam lower mekong basin: a prototype application for the vietnam open data cube. European Journal of Remote Sensing **52**(1), 599–612 (2019)
83. Rau, J.Y., Chen, L.C., Liu, J.K., Wu, T.H.: Dynamics monitoring and disaster assessment for watershed management using time-series satellite images. IEEE transactions on geoscience and remote sensing **45**(6), 1641–1649 (2007)
84. Rokni, K., Ahmad, A., Solaimani, K., Hazini, S.: A new approach for surface water change detection: Integration of pixel level image fusion and image classification techniques. International Journal of Applied Earth Observation and Geoinformation **34**, 226–234 (2015)
85. Rudner, T.G., Rußwurm, M., Fil, J., Pelich, R., Bischke, B., Kopačková, V., Biliński, P.: Multi3net: segmenting flooded buildings via fusion of multiresolution, multisensor, and multitemporal satellite imagery. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 702–709 (2019)
86. Sa, I., Chen, Z., Popović, M., Khanna, R., Liebisch, F., Nieto, J., Siegwart, R.: weednet: Dense semantic weed classification using multispectral images and mav for smart farming. IEEE Robotics and Automation Letters **3**(1), 588–595 (2017)
87. Schmitt, M., Hughes, L.H., Qiu, C., Zhu, X.X.: Sen12ms–a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion. arXiv preprint arXiv:1906.07789 (2019)
88. Schmitt, M., Hughes, L.H., Zhu, X.X.: The sen1-2 dataset for deep learning in sar-optical data fusion. arXiv preprint arXiv:1807.01569 (2018)
89. Sharma, Y., Ross, R.: Less is more when applying transfer learning to multi-spectral data
90. Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M.: Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. IEEE transactions on medical imaging **35**(5), 1285–1298 (2016)

91. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

92. Sohn, K., Berthelot, D., Li, C.L., Zhang, Z., Carlini, N., Cubuk, E.D., Kurakin, A., Zhang, H., Raffel, C.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. arXiv preprint arXiv:2001.07685 (2020)

93. Sumbul, G., Charfuelan, M., Demir, B., Markl, V.: Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. In: IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. pp. 5901–5904. IEEE (2019)

94. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. pp. 6105–6114 (2019)

95. Unnikrishnan, A., Sowmya, V., Soman, K.: Deep learning architectures for land cover classification using red and near-infrared satellite images. Multimedia Tools and Applications **78**(13), 18379–18394 (2019)

96. Van Engelen, J.E., Hoos, H.H.: A survey on semi-supervised learning. Machine Learning **109**(2), 373–440 (2020)

97. Vincenzi, S., Porrello, A., Buzzega, P., Cipriano, M., Fronte, P., Cuccu, R., Ippoliti, C., Conte, A., Calderara, S.: The color out of space: learning self-supervised representations for earth observation imagery. arXiv preprint arXiv:2006.12119 (2020)

98. Wang, Q., Yuan, Z., Du, Q., Li, X.: Getnet: A general end-to-end 2-d cnn framework for hyperspectral image change detection. IEEE Transactions on Geoscience and Remote Sensing **57**(1), 3–13 (2018)

99. Wang, Q., Zhang, X., Chen, G., Dai, F., Gong, Y., Zhu, K.: Change detection based on faster r-cnn for high-resolution remote sensing images. Remote sensing letters **9**(10), 923–932 (2018)

100. Wei, Y., Yuan, Q., Shen, H., Zhang, L.: Boosting the accuracy of multispectral image pansharpening by learning a deep residual network. IEEE Geoscience and Remote Sensing Letters **14**(10), 1795–1799 (2017)

101. Wu, H., Prasad, S.: Semi-supervised deep learning using pseudo labels for hyperspectral image classification. IEEE Transactions on Image Processing **27**(3), 1259–1270 (2017)

102. Xie, Q., Zhou, M., Zhao, Q., Meng, D., Zuo, W., Xu, Z.: Multispectral and hyperspectral image fusion by ms/hs fusion net. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1585–1594 (2019)

103. Xu, H.: Modification of normalised difference water index (ndwi) to enhance open water features in remotely sensed imagery. International journal of remote sensing **27**(14), 3025–3033 (2006)

104. Yu, M., Yang, C., Li, Y.: Big data in natural disaster management: a review. Geosciences **8**(5), 165 (2018)

105. Yuan, Q., Wei, Y., Meng, X., Shen, H., Zhang, L.: A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **11**(3), 978–989 (2018)

106. Zhai, X., Oliver, A., Kolesnikov, A., Beyer, L.: S4l: Self-supervised semi-supervised learning. In: Proceedings of the IEEE international conference on computer vision. pp. 1476–1485 (2019)

107. Zhang, C., Pan, X., Li, H., Gardiner, A., Sargent, I., Hare, J., Atkinson, P.M.: A hybrid mlp-cnn classifier for very fine resolution remotely sensed image classification. ISPRS Journal of Photogrammetry and Remote Sensing **140**, 133–144 (2018)

108. Zhang, R., Isola, P., Efros, A.A.: Colorful image colorization. In: European conference on computer vision. pp. 649–666. Springer (2016)

109. Zhao, Y., Huang, B., Song, H.: A robust adaptive spatial and temporal image fusion model for complex land surface changes. Remote sensing of environment **208**, 42–62 (2018)

110. Zhu, Q., Zhang, J., Ding, Y., Liu, M., Li, Y., Feng, B., Miao, S., Yang, W., He, H., Zhu, J.: Semantics-constrained advantageous information selection of multimodal

spatiotemporal data for landslide disaster assessment. ISPRS International Journal of Geo-Information **8**(2), 68 (2019)

111. Zhu, Y., Kiros, R., Zemel, R., Salakhutdinov, R., Urtasun, R., Torralba, A., Fidler, S.: Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In: arXiv preprint arXiv:1506.06724 (2015)

A Satellite Imaging

A.1 Satellite Bands- Sentinel 2

Sentinel-2 bands	**Central wavelength (nm)**	**Bandwidth (nm)**	**Spatial resolution (m)**
Band 1 – Coastal aerosol	442	21	60
Band 2 – Blue	492	66	10
Band 3 – Green	559	36	10
Band 4 – Red	664	31	10
Band 5 – Vegetation red edge	704	15	20
Band 6 – Vegetation red edge	740	15	20
Band 7 – Vegetation red edge	782	20	20
Band 8 – NIR	832	106	10
Band 8A – Narrow NIR	864	21	20
Band 9 – Water vapour	945	20	60
Band 11 – SWIR	1613	91	20
Band 12 – SWIR	2202	175	20

Table 6: Sentinel-2 Bands Detail

A.2 Water Indexing Techniques

Water Indexing Techniques	
NDWI [71]	Green-NIR/Green+NIR
MNDWI [103]	Green-SWIR1/Green+SWIR1
AWEI_S [31]	Blue + 2.5 x Green - 1.5 * (NIR+SWIR1) - 0.25 x (SWIR2)
AWEI_NS [31]	4 x (Green - SWIR2) - (0.25 x NIR + 2.75 x SWIR1)
MI [74]	NDWI + Blue-NIR/Blue+NIR

Table 7: Water Indexing Techniques