Msc Thesis

THE HEAT UNDER OUR PEAT

USING TRANSFER LEARNING FOR PEAT FIRE PREDICTION IN THE ARCTIC REGION OF NORTH-EAST SIBERIA WITH SENTINEL-2 IMAGERY

By
Aafke Schreurs (s2007754)

Supervised by
Dr. Mannes Poel
Dr. Ir. Janneke Ettema

Interaction Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente

Abstract

Over 90% of peatland can be found in boreal and subarctic regions. Unfortunately, these peatland areas might be more vulnerable to burning than previously assumed. Eventually northern peatlands could turn from carbon sinks to sources. New tools are necessary to respond to fire danger in peatland areas prone to burn. Remote sensing can view remote locations such as the Arctic regions of North-East Siberia. For peat fire prediction the combination of Sentinel-2 satellite imagery and the usage of a convolutional neural network (CNN) is applied. As labelled data in the field of remote sensing is not easily found and is costly, transfer learning will be applied. This leads to the main objective to explore the viability of peat fire prediction for the remote Arctic area of North-East Siberia, using Sentinel-2 imagery and transfer learning of a CNN network. The EuroSAT dataset is used to create a pretrained network closer to the target task of peat fire prediction and ResNet50 is chosen as the network architecture. The results show that predicting peat fire in the region of North-East Siberia with the help of transfer learning is a novel approach that is still far from being able to be used as an early warning in the form of, for example, a susceptibility map. Transfer learning cannot be applied with the approach and data in this thesis; however, it is not excluded as an option for peat fire prediction and needs further investigation towards which features are important and which approaches are best applicable per region and timespan.

Acknowledgements

The topic of peat fire prediction was of great interest to me, especially considering the effects climate change has had in many different areas of the world recently and happy I was able to combine it with another interest, deep learning. Important is to acknowledge that this thesis could not have been completed without the support and guidance of several people. Firstly, my supervisors Mannes Poel and Janneke Ettema. You have given me the freedom to pursue my own topic and helping me forge my path through the many possible options. Not only do I appreciate the knowledge that you have offered me throughout the process but also the positivity and support that you gave me throughout our meetings. I also want to thank Ben Maathuis for providing me an introduction into Geographic Information Systems. I have gained much knowledge though the information given. Furthermore, I would like to thank my parents, sister, partner, and friends for supporting me and lifting me up when I needed it the most.

Contents

1 Introduction

Last year in 2020 a Nature Magazine article had an alarming message: the Arctic is burning like never before (Witze, 2020). In Siberia, the article stated, fires started as early as in May and burned much longer than normally. In the two most eastern districts of Russia the individual fires even counted 18,591. Two reasons for the large number of fires that are presented in the article are the warmer temperatures in spring and winter, and that peat fires, so called *Zombie Fires* have been smouldering beneath the snow and ice. Unfortunately, as the climate warms further, fire risk in Siberia is predicted to increase (Sherstyukov & Sherstyukov, 2014) and very likely increases the vulnerability of peat fires which in return can cause long-term environmental changes such as permafrost thawing (Turetsky et al., 2015) in the higher located peatlands.

Peatlands are a type of wetland where dead plant remains do not fully decay but are stored and build up a layer of peat as wetland area provides flooded or saturated soil to make this possible (Parish et al., 2008). Peatlands can vary highly, from blanket bogs to swamp forests and climate conditions play a role in this. As it varies highly, distinctions between peat characteristics are made between:

- Bogs, which is peatland that is raised above the ground and consist of mire (an area that is currently accumulating and forming peat). It can only be wetted through precipitation such as rainfall.
- Fens, peatland area that gathers water and nutrients from the mineral soil.

Peat can be formed from mosses, sedges, grasses, trees, shrubs, or reeds. For peat layers to grow the water level needs to be either just under, at, or a bit above the surface. Most carbon is stored in moss, litter and peat layers that could be consumed during fires (Turetsky et al., 2011).

Peat fires are mainly dominated by smouldering fires (Turetsky et al., 2015). Smouldering fires happen when oxygen comes into contact with the surface of a solid fuel (Ohlemiller, 1985). While both flaming and smouldering fires are through heat release and transfer, smouldering resolves into lower temperatures and slower spread. Smouldering fires can spread over large areas and can go deep into the soil (Rein et al., 2008). These fires can burn for long periods of time, even weeks or months, and are difficult to stop (Rein, 2016). Even fire-fighting efforts, rainfall or other weather changes cannot always stop the fires. Smouldering fires can have a great impact on the soil system (Rein et al., 2008). Eventually northern peatlands could turn from carbon sinks into sources (Hugelius et al., 2020). Even though peatlands in general account for just around 3% of the terrestrial surface, it stores around 21% of the total global carbon (C) stock (Leifeld & Menichetti, 2018) and therefore represents one of the largest carbon pools (Yu, 2012). Over 90% (4 million km^2)

of peatland can be found in boreal and subarctic regions (Yu et al., 2010). Unfortunately, Turetsky et al. (2004) presents that these boreal peatland areas might be more vulnerable to burning than previously assumed. New tools are necessary to be able to respond to fire danger in peatland areas that are likely to burn (Turetsky et al., 2015).

In countries where peatlands are widespread, satellite data that can identify hotspots in combination with spectral data, is relevant for early action tools considering peatland areas (FAO, 2020). Remote sensing has the ability to view remote locations (Lees et al., 2018), like the Arctic regions of North-East Siberia. Remote sensing also offers methods to study large wetland areas for different points in time in a cost-effective way (Rani et al., 2011). One of the newest multi-spectral satellite available for open access is the Sentinel-2. With the high-spatial resolution, wide field of view and the offering of 13 spectral bands, the Sentinel-2 is a big step forward compared to other current multi-spectral satellite options (Drusch et al., 2012). Before the Sentinel-2, open access multi-spectral imaging in high spatial and temporal resolution was difficult to come by (Sirin et al., 2020). With the topic of peatland, the use of satellites like the Sentinel-2, has mainly been focusing on mapping peatland according to an overview by the Food and Agriculture Organisation of the United Nations (FAO, 2020). The use of Sentinel-2 and the information provided from all 13 bands for peat fire prediction, however, seems to be lacking.

Deep learning is becoming more and more the preferred choice in many application fields (Zhu et al., 2017) and has been introduced in the remote sensing field for big data analysis (L. Zhang et al., 2016). From these deep learning models a Convolutional Neural Network (CNN) is arguably the most exciting in its potential to find complex non-linear feature representations from image pixels (Ban et al., 2020). CNNs are very effective in the areas such as image recognition, object detection and semantic segmentation (Zhu et al., 2017). CNNs can find complex hidden spatial patterns in image patches (C. Zhang et al., 2018). As in the case of fire, spatial characteristics are present, and applying a CNN is helpful as it can find multiple levels of representations and contextual (neighbouring) information from the input data (G. Zhang et al., 2019). According to G. Zhang et al. (2019) the CNN will be of important value for forest fire prevention.

Three deep learning methods can be described (Zhu et al., 2017). One is the use of pretrained networks such as a CNN trained on a natural image dataset (Zhu et al., 2017) for example ImageNet (Deng et al., 2009). Secondly, a pretrained model can be adapted by fine-tuning to the smaller satellite dataset (Zhu et al., 2017). Lastly, a network can be learned from scratch using satellite images. However, as these large-scale networks have a vast amount of parameters, training them on a small dataset will cause overfitting. Zhu et al. (2017) mention that some papers therefore use a small-scale network, but this shows drawbacks as the network tends to fit to the training data, which reduces the generalization. For peat fire prediction in this thesis, where the dataset will not be big enough to reach the millions of images that ImageNet consists of, one of the other two deep learning categories is needed. Both categories are approaches commonly known as transfer learning.

In the field of remote sensing, transfer learning could be important, as labelled data is not easily found, is costly and involves much effort (Marmanis et al., 2016). Transfer learning can make use of previously learned features on a specific task and transfer these to another task that is similar or even unrelated (Pires de Lima & Marfurt, 2020). Yosinski et al. (2014) found that even though the transferability gap grows when the tasks are more distant from each other, it still offers better results than initialisation with random weights. Several studies have been conducted applying transfer learning with remote sensing imagery (Marmanis et al., 2016; Petrovska et al., 2020; Pires de Lima & Marfurt, 2020) and focused on satellite imagery containing RGB valued images.

Until now, as far as is known, no studies have been conducted towards peat fire prediction using Sentinel-2 data. To explore transfer learning on the Sentinel-2 data, two larger datasets are available to create a pretrained CNN model with all thirteen available bands. These datasets are EuroSAT (Helber et al., 2019), and BigEarthNet (Sumbul et al., 2019). Both datasets contain Sentinel-2 images of all thirteen bands. EuroSAT contains 27,000 images of 64x64 pixels divided over 10 different classes. BigEarthNet (Sumbul et al., 2019) consists of 590.326 multi-labeled image patches (120 x 120 pixels for 10 m bands, 60 x 60 pixels for 20 m bands, and 20 x 20 pixels for 60 m bands). Both datasets perform well on the ResNet50 network model. This thesis focused on the use of EuroSAT.

Soil moisture can provide unique information for the susceptibility to peatland fire (Dadap et al., 2019) as reduction in soil moisture will lead to peatland being more ignitable, cause deeper burns, and longer burning durations, according to a review study by Nelson et al. (2021). Water indices are useful for moisture and can be used to detect changes in water content, or when combined with other indices, more complex information of peatland conditions (Lees et al., 2020). Therefore, not only the thirteen bands but water indices in combination with a vegetation index will be explored as well.

The main objective of this thesis is to explore the viability of peat fire prediction for the remote Arctic area of North-East Siberia, using Sentinel-2 imagery and transfer learning of a CNN network pretrained on the EuroSAT dataset. The next main research question and sub-questions follows from this objective:

RQ1. *To what extent is transfer learning using a CNN network pretrained on the EuroSAT dataset, suitable to predict peat fire at the Arctic area of North-East Siberia using Sentinel-2 data?*

RQ1.1. *To what extent can peat fire in the region be predicted using the spring months median values for the fires in the summer months?*

RQ1.2. *Which Sentinel-2 features are the most discriminative for predicting peat fire in the region?*

RQ1.2.1. *Which bands are most discriminative as features for peat fire prediction?*

RQ1.2.2. *Which vegetation and water indices are most discriminative as features for peat fire prediction?*

This thesis is hoping to contribute to the field of peat fire prediction in the Arctic area by offering the first steps towards early action tools using remote sensing and transfer learning to predict peat fires in the remote Arctic area.

The structure of this thesis is as follows, the first section hereafter contains background information about the topics such as CNNs, Sentinel-2 bands, ResNet50, etc. It is followed by the next section, the Literature Review. Then the methodology is explained and afterwards in another section the results are given. These are discussed in the Discussion section with the limitations and the section Conclusion afterwards. Lastly, the Future Work section.

2 Background

In this section the Sentinel-2 satellite bands are explained in further detail as well as the convolutional neural network (CNN), transfer learning and the evaluation metrics including the feature importance method.

2.1 Sentinel-2

The Sentinel-2 is a joint initiative of the European Commission (EC) and the European Space Agency (ESA) and is a multi-spectral imaging instrument (MSI) offering high resolution optical images (Drusch et al., 2012). It has a resolution ranging from 10-60 m with a 5-day revisit time depending on the location. It consists of thirteen different bands with each a different target wavelength. Bands B1, B9, and B10 are mainly for applying atmospheric correction and the screening for clouds (Drusch et al., 2012). B5-B7 are vegetation red-edge bands which for wetland classification add significant influence to the classification of intensive vegetated wetland classes (Kaplan & Avdan, 2019). Other bands can be found in **Table 1**, which shows an overview of the band names and central wavelength information combined from the ESA's User Guide resolution pages (ESA, n.d.-a, n.d.-c).

Table 1. The 13 bands from the Sentinel-2 and their names with the central wavelengths (µm). Adapted from ESA (n.d.-a) and (ESA, n.d.-c).

Band NR.	Name	Central wavelength (nm)
B1	Aerosols	443
B2	Blue	490
B3	Green	560
B4	Red	665
B5	Vegetation Red Edge (VNIR)	705
B6	Vegetation Red Edge (VNIR)	740
B7	Vegetation Red Edge (VNIR)	783
B8	NIR	842
B8A	Vegetation Red Edge (VNIR)	865
B9	Water Vapour	940
B10	Cirrus	1375
B11	SWIR1	1610
B12	SWIR2	2190

From these 13 bands, the three indices that will be used in this thesis can be calculated. The NDVI (**Equation 2.1**) by Rouse et al. (1973) is able to separate green vegetation from other surfaces and does this using the absorption of the red wavelengths by the chlorophyll of the vegetation in comparison with the reflection of the near infrared (NIR) wavelength caused by the internal leaf structure (Lozano et al., 2007; Tucker, 1979).

$$NDVI = \frac{NIR - Red}{NIR + Red} = \frac{B8 - B4}{B8 + B4} \tag{2.1}$$

The NDWI (**Equation 2.2**) is proposed by McFeeters (1996) and uses the NIR and visible green light to enhance the presence of open water features and eliminating the presence of soil and terrestrial vegetation features.

$$NDWI = \frac{Green - NIR}{Green + NIR} = \frac{B3 - B8}{B3 + B8} \tag{2.2}$$

The NDWI (**Equation 2.3**) proposed by Gao (1996), is a measure of water in vegetation canopy, interacting with the incoming solar radiation. In some papers this equation is revered to as the NDMI, such as with a study by Maulana et al. (2019) which is mentioned in detail in **Section 3.3**. To prevent confusion with both the water indices this NDWI proposed by Gao (1996) will be called NDMI hereafter. The NDMI uses shortwave-infrared (SWIR) and NIR as the absorption of water in tissues is related to the reduction of reflectance of the SWIR in comparison with the NIR and can therefore be used as an estimation of water content in vegetation (Lozano et al., 2007).

$$NDMI = \frac{NIR - SWIR}{NIR + SWIR} = \frac{B8A - B11}{B8A + B11} \tag{2.3}$$

2.1.1 EuroSAT dataset

EuroSAT (Helber et al., 2019) is made of Sentinel-2A images from across Europe and created with the aim of providing a dataset for Earth observation applications. Additionally, because most other datasets are not suitable for applications with the Sentinel-2 imagery. The dataset consists of 10 different classes: industrial, residential, annual crop, permanent crop, river, sea and lake, herbaceous vegetation, highway, and pasture. EuroSAT has a total of 27,000 images with a patch size of 64x64 pixels. All 13 bands of the Sentinel-2 are included in the dataset. ResNet50 pretrained on ImageNet, was used as one of the network models to evaluate the performance, scoring an accuracy of 98.57% on the EuroSAT dataset.

2.2 Deep Neural Networks

Taking the information from the Deep Learning Book (Goodfellow et al., 2016) deep neural networks (DNNs), also called feedforward neural networks or multilayer perceptrons (MLPs), are very important in the machine learning area. DNNs try to get close to a certain function f that describes the input x. This is done by evaluating x by passing it through the network that computes f to eventually an output y.

A DNN is called a network because in most cases they consist of several different functions that are added together into one network (Goodfellow et al., 2016). These different functions are called *layers* and it can have a multitude of it, having a certain *depth*, which is where the term deep learning came from. The first layer is the input layer, the last layer is called the output layer and the other layers in the network are called hidden layers. DNNs consist of *units* which represent a vector-valued layer. Each vector value can be seen as a *unit* or a *neuron* when looking at the neuroscience inspired way. For each layer these units act parallel from each other and calculates an activation function from the input of previous other units.

In the next couple of subsections, important aspects of DNNs will be explained. This includes optimization, output activation functions, regularization, and the learning rate.

2.2.1 Optimization

Most deep learning algorithms try to minimize or maximize some function $f(x)$ by changing x, which is called optimization (Goodfellow et al., 2016). This function that you want to optimize is called the loss function, or cost function. It is the criterion or objective function. This is done calculating the derivative of the function f for a certain point x to get the slope. It can show how to change x to make a small improvement for the output y. This can be done by moving x in small steps with the opposite sign of the derivative until it finds a slope with a value of 0, meaning it found a stationary point. This is called *gradient descent*. Through *back-propagation*, the weights are updated. Backpropagation starts at calculating the loss at the output layers and then goes backwards along the hidden layers to compute the gradient. Using this gradient, an optimizer such as *Stochastic Gradient Descent* (SGD) performs the learning part.

Stochastic Gradient Descent (SGD) is used in many deep networks, as datasets can become very large (Goodfellow et al., 2016). Instead of using the sum over each calculated loss value for an input datapoint, SGD uses minibatches. Instead of calculating the loss over all the input datapoints, a fixed amount of datapoints is uniformly chosen from the dataset and used to calculate the loss for each. This means that the loss function is no longer dependent on the amount of data in the dataset and is therefore computationally less expensive.

Momentum is a term that describes the method designed to speed up learning and help moving the gradients in the right direction by using moving average (Goodfellow et al., 2016). This means that the step size no longer depends solely on the learning rate but also on how large and aligned a batch of gradients are. When many gradient batches point into the same direction, the step size becomes large. Common values for momentum are 0.5, 0.9 and 0.99.

More recently other optimization algorithms have emerged making use of adaptive learning rates (Goodfellow et al., 2016). This idea of adaptive learning rates means that for each parameter there is a learning rate that adapts itself through learning. So, instead of having one learning rate, each model parameter has one and will automatically adapt while learning. Three of these algorithms mentioned by Goodfellow et. al. (2016) are AdaGrad, RMSProp and Adam. AdaGrad works well for some deep learning models, however, it can lead to too early excessive decrease of the learning rates at others. RMSProp is an adaptation to the AdaGrad algorithm and solves the problem that arises of too early decrease of the learning rates. It is one of the popular choices as an optimization method. Adam can best be seen as RMSProp in combination with momentum and is being seen as quite robust to the choice of parameters even though sometimes it needs to be adjusted from the suggested default.

Another optimization strategy is *batch normalization*. What batch normalization does is it normalizes the output from the activation function of layers (Goodfellow et al., 2016). It reparametrizes the model, solving the problem of passing on updates across many layers. It does this by batches and computing the mean and standard deviation from the activation values of that layer. Through backpropagation the mean and standard deviation are learned and normalize the activation values. Batch normalization can be applied for each input and hidden layer in a neural network.

2.2.2 Learning rate

A learning rate is a positive number that can be small and constant and determines the size of the step in which to move towards a minimal loss function value (Goodfellow et al., 2016). For SGD the learning rate is crucial and practically it is best to gradually decrease this over time as SGD creates noise that will not be gone even after arriving at a minimum. It is common to decay the learning rate linearly until a certain specified iteration (epoch).

It is important to set the learning rate to a good value, however, there are many ways to do so and it is as Goodfellow et. al. (2016) describes more of an art than it is a science way of doing it. When a learning rate is set too large, many oscillations of the loss function may occur and increases significantly. Gentle oscillations are perfectly fine. A too low of a learning rate can cause learning to proceed slowly or become stuck with a high loss value.

2.2.3 Output activation functions

The output layer is the last layer that applies transformation to the set of features that the DNN provides, to complete the task (Goodfellow et al., 2016). This can be a linear output layer, that often is used to produce the mean of a conditional Gaussian distribution. For binary classification the *sigmoid* function can be applied. This is a Bernoulli distribution that only needs to find the probability of class 1 given the input. The probability of the other class follows from subtracting the class 1 probability from 1. This does mean that the probability value needs to be within the [0,1] interval. Using a linear function would not suffice and lead to gradients of 0 when the value falls outside the interval. Gradients that are 0 means the learning algorithm is not able to know how to improve the parameters matching the unit. The *sigmoid* function maps the linear unit value to the interval (0,1). Then there is the *softmax* function. This function is used when there is a discrete variable that has several different states (or classes). The *softmax* function is in fact a generalization of the sigmoid function and is most often used as the output activation function for classification.

2.2.4 Regularization

Regularization is used for the purpose of regulating the model's parameters (Goodfellow et al., 2016). Most regularization methods use a parameter norm penalty which mostly consists of a hyperparameter and adding this to the loss function. When the hyperparameter has a value of 0, no regularization is done while large values correspond to more. For neural networks biases are left unregularized and penalizes only the weights. L^2 regularization, commonly known as weight decay, is one of these methods that use a parameter norm penalty as it adds a certain regularization term to the loss function (**Equation 2.4**). L^1 regularization is another option to penalize the weights (**Equation 2.5**).

$$L^2 = \frac{1}{2} \|w\|^2 \tag{2.4}$$

$$L^1 = \|w\| = \sum_i |w_i| \tag{2.5}$$

Another technique to regularize a DNN is through *data augmentation*. The more data the better a model can generalize (Goodfellow et al., 2016). However, often data is limited. Data augmentation can add fake data to a dataset to create a larger amount of data. With a classification model this is mostly straightforward as the model needs to be invariant to a variety of transformations anyways. This means that applying transformations such as rotation, flipping, etc. can be easily added to the dataset.

Dropout is a technique that makes use of the idea of the ensemble of different subnetworks (Goodfellow et al., 2016). Dropout creates different subnetworks by turning units on and off and therefore creating subnetworks with different units being present in each subnetwork. This can be done by multiplying a certain unit's activation value with zero to have no impact. If a unit will be present in a subnetwork, is decided by a hyperparameter for each unit, which is normally set to 0.5 for hidden layers and 0.8 for input layers. This means the chance of a unit being 'on' is 50%. *Dropout* is computationally very cheap and does not limit the type of model or the training procedure used.

2.3 Convolutional Neural Network

According to Goodfellow et al. (2016) convolutional neural networks (CNNs) are neural networks that use convolution in one or multiple of the layers. CNNs are basically a specialized sort of feedforward network. CNNs can be used with data that correspond to a certain grid-like idea. For example, images that contain $m \times n$ pixels and c channels. A convolutional layer consists of three different types of stages: the convolutional, detector, and pooling stage.

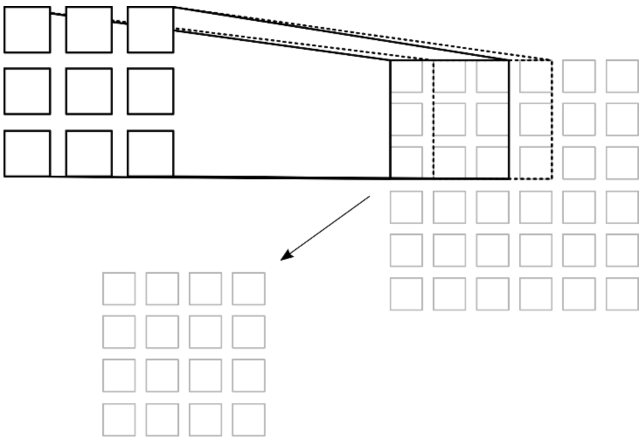


Figure 1. A kernel (top) of size 3 x 3 moving over the input of 6 x 6 using a stride of 1 outputting a 4 x 4 feature map (bottom).

2.3.1 Convolutional stage

Convolutions are linear operations using kernels to compute a weighted average over a certain area of the image (Goodfellow et al., 2016). The output is sometimes called the feature map.

With more traditional neural networks each input unit is connected to each output unit because these networks make use of matrix multiplication. This is the case for example, with fully connected layers. In the case of a CNN, the connection between input and output units are *sparse* (also called sparse interactions, sparse connectivity, or sparse weights). Meaning, that when the kernel is smaller than the input image, fewer parameters (weights) need to be stored as kernels can detect small, meaningful features (i.e., edges) using significantly less pixels.

Additionally, CNNs make use of *weight sharing* (Goodfellow et al., 2016). A neural network normally has one weight in the matrix for computing one output of the layer for only once. For a CNN, each value in the kernel, for example the 9 values in **Figure 1**, is normally used for every position of the input. Therefore, only one set of values need to be learned. This parameter sharing causes a CNN to not be affected by translation. If the input changes regarding translation, the output will also change in that way. Moving an object in the image will result in an output with a representation that moved the same amount. Eventually the convolutional stage outputs a set of linear activations that is passed on to the detector stage.

2.3.2 Detector stage

The detector stage is a non-linear function that the linear activations from the convolutional stage are passed through (Goodfellow et al., 2016). ReLU, rectified linear unit function, is an example of such a non-linear function. Deep networks can train faster and more easily with the use of ReLU in comparison with the tanh function (Krizhevsky et al., 2017). With ReLU all values under 0 will be translated to 0 and all above, will keep the same value. This can be written as **Equation 2.6**.

$$f(x) = \max(0, x) \tag{2.6}$$

After the activation function is applied, the output is sent to the pooling stage.

2.3.3 Pooling stage

The pooling stage creates a statistical summary of nearby values of the output from the detector stage (Goodfellow et al., 2016). This can be for example the max pooling, where the maximum value of neighbouring pixels is chosen as the representative value for that block. This block of neighbouring pixels consists of a rectangle. Other well-known pooling functions are average pooling where the average of the values is taken, the L2-norm, or a weighted average based on the distance from the central pixel. By using pooling functions, the representation becomes invariant to small translations of the input. Useful if it is more important to know if a certain feature is present rather than the location of this feature. Another advantage is the possibility for downsampling, when taking a stride bigger than one. This will reduce the representation size and therefore also the computational costs for the next layer.

2.3.4 ResNet50

With deeper neural networks the problem arose that at a certain point the depth of a network reached a saturation and eventually decreased the accuracy and to solve this problem He et al. (2016) created a CNN called ResNet. The idea is that a deeper neural network should not perform worse than its shallower counterpart. Instead of learning a direct function to approximate the optimal function, the authors argued that the network could also learn an underlying mapping $H(x)$. This leads to a *residual* function of $F(x) = H(x) - x$. Meaning, the original function becomes $F(x) + x$. The reasoning behind this is that if the added layers (the number of layers on top of the shallower network) are constructed as *identity* mappings, the deeper network should have no error greater than its shallower counterpart. This means that if the identity mappings are optimal, the residual value will be pushed to zero, since it should get as close as possible to the identity mapping. The identity mappings are mostly not the optimal function but the authors reason that if the optimal function is closer to an identity mapping rather than a zero mapping, it would be easier to find an approximate solution with use of the identity mapping instead of learning the function as a new one.

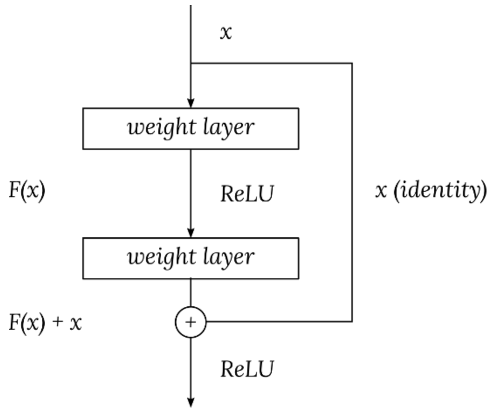


Figure 2. Residual block. Adapted from He et al. (2016).

To make this possible, He et al. (2016) introduced the concept of a residual block (**Figure 2**). These residual blocks have several stacked layers that learn the residual function $F(x)$. The input to these stacked layers is x. However, x is also the *identity* mapping that is brought over to the end of the residual block using a *shortcut connection* (the line matching 'x (identity)' in **Figure 2**). Then element-wise addition is applied of $F(x) + x$. If these values are not of the same dimension, either the shortcut connection performs identity mapping with zero entries padded to increase dimensionality or a linear projection by a 1 x 1 convolution is used. Afterwards a nonlinear activation function is applied.

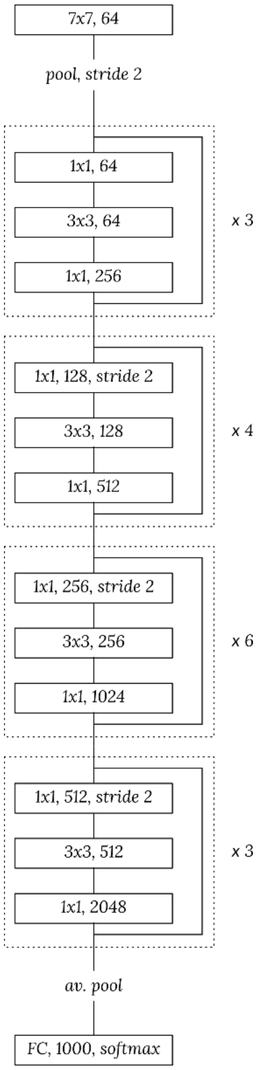


Figure 3. ResNet50 architecture. Adapted from He et al. (2016, Table 1).

The ResNet50 model, which can be seen at **Figure 3**, is named after the total number of weighted layers it consist of. In this case that is 50. Other options are 18, 34, 101, and 152. Taking ResNet50, it starts with a convolutional layer with a filter size of 7 x 7 and a stride of 2 for downsampling the output size to half. Then a pooling layer with a stride of 2 to again reduce the dimensions. In comparison to the residual block shown at **Figure 2**, these blocks consist of 3 layers instead of 2. The blocks have layers of 1 x 1, 3 x 3, and 1 x 1 convolutions. This design is called a *bottleneck* design (He et al., 2016). What happens is that the first 1 x 1 convolution decreases the input dimensions, whereafter the 3 x 3 convolution has smaller input and output dimensions, meaning this is less computationally expensive. The last 1 x 1 convolution increases the dimensions again. As can be seen in **Figure 3**, the model can be divided into four stages each with a certain number of bottleneck blocks. After these four stages a global average pooling layer is applied, followed by a 1000-way FC layer using a softmax classifier.

The ResNet architecture managed to win first place in the ILSVRC (ImageNet Large Scale Visual Recognition Challenge) 2015 contest with a top-5 error of 3.57% using an ensemble of six models with different depths (He et al., 2016).

2.4 Transfer Learning

With deep neural networks the layers close to the input data learn more general features while the layers closer to the final layers have features that become less and less general (Yosinski et al., 2014). With transfer learning learned features of the CNN model on a primary (base) task are applied to an unrelated secondary (target) task (Pires de Lima & Marfurt, 2020). Features from a number of layers of the base model, are transferred to the layers of the target model (Yosinski et al., 2014). This primary CNN is often called a pretrained model which is a model that is already trained in the domain is was intended for and therefore can save time and computing power (Alom et al., 2019).

Transfer learning consists of two common known approaches: feature extraction and fine-tuning (Peters et al., 2019; Pires de Lima & Marfurt, 2020). With feature extraction the layers extracted from the model are 'frozen'. This means that the weights stay the way they are and are not adjusted by training anymore. According to Pires de Lima and Marfurt (2020) fine-tuning starts of like feature extraction with frozen layers but afterwards unfreezes the layers and allows them to learn. Fine-tuning offers the benefit to create a general-purpose representation for many different tasks (Peters et al., 2019). Dependent on the amount of data from the target dataset and the amount of parameters (weights) in the first layers of the model, these first n layers can be fine-tuned to the new task (Yosinski et al., 2014). If the dataset is small and the number of weights is large, overfitting might occur. When the dataset is large enough or the number of weights is small, then these first n layers might also be fine-tuned. As opposed to a model trained from scratch, fine-tuning is done with a smaller learning rate and less amount of epochs (Petrovska et al., 2020).

2.5 Evaluation metrics

Peat fire prediction in this work is a binary classification where the emphasis is on the positive (fire) class. The F-score is a harmonic average of precision and recall (Radke et al., 2019). The most important aspect is that the model can predict where a fire is prone to start and that it does not overlook too many places where a fire might start (false negatives). Since the F2-score puts more weight on the recall and therefore the false negatives, this metric will be the leading factor. The AUC score from the ROC will be the second metric, which by examining the literature review (**Section 3**), is done by several studies (Maulana et al., 2019; G. Zhang et al., 2019) and gives a good idea of the performance of classification. For a more visual representation of the classification accuracy per dataset, a confusion matrix will be used as a third metric. A very common evaluation metric in classification tasks, according to Petrovska et al. (2020).

2.5.1 Permutation Feature Importance

Permutation Feature Importance (PFI) can be used to calculate the change in model error (model performance) when a certain feature's values are permutated (Molnar, 2019). This is commonly done by shuffling a certain features' values (in this case images). A feature is important if the shuffling of the images increases the error. The other way around, a feature is unimportant when it does not affect the error much. Molnar (2021) argues that using the test set to perform PFI, prevents importance on the wrong features because of possible overfitting on the training set when permutation would be applied to the training set. Therefore, the test set will be used for feature importance. For this thesis, the feature importance score FI_i for each feature i is calculated by subtracting the permutated feature model performance p_i^p from the original performance p^o (**Equation 2.7**). This is done for each feature i in the dataset. The model performance is defined by the F2-score.

$$FI_i = p^o - p_i^p \tag{2.7}$$

Molnar (2021) mentions that if features are strongly correlated the interpretation of the importance of the features should be carefully made.

3 Literature Review

This section outlines the studies done towards peat fire prediction. At the time of writing, few studies are published on the topic of predicting peat fires using machine learning. Therefore, in the first subsection several general transfer learning methods using a CNN and satellite imagery are discussed. The next subsection contains several general wildfire prediction classification approaches using CNNs. Afterwards, the few known studies that are more specific to peat fires are discussed. These studies do not focus solely on the usage of a CNN but fall into the more general category of machine learning. **Table 2** shows an overview of all the papers mentioned below.

3.1 Transfer learning in the remote sensing domain

The following studies research the effectiveness of transfer learning and the different approaches (fine-tuning and feature extraction) on remote sensing images with RGB values.

Marmanis et al. (2016) used feature extraction as a transfer learning approach. The network is Overfeat (Sermanet et al., 2014) and pretrained on ImageNet (Russakovsky et al., 2015). Marmanis et al. (2016) used ImageNet to gain information from the preactivations at the seventh and eighth layer (the two last FC layers as can be seen in **Figure 4**). These value matrixes of 1 x 1 x 4096 (8281 values) are concatenated and transformed to a 2D-array of sizes 91 x 91 pixels and passed on to the trainable network as input. This trainable network consists of two convolutional blocks and two fully connected layers, using a softmax function. SGD is applied with batch sizes of 10. Other optimizations and regularizations applied can be found in **Table 2**. The model is evaluated using the UC Merced Land dataset (Yang & Newsam, 2010) and with an accuracy of 92.4% the trainable CNN outperforms the Random Forest (RF) classifier in combination with the Overfeat pretrained network (86.9%) and both the CNN and RF without the pretrained network (44,5% and 44% respectively). The authors also observed that when no pretrained model is used beforehand, both classifiers are not able to learn appropriately and differs significantly in accuracy.

The results of this research show the potential of feature extraction as one approach to applying transfer learning, using the two last fully connected layer information from the pretrained network, and converting them to a two-dimensional input to the trainable network.

Table 2. Overview of mentioned papers.

STUDY	FEATURE EXTRACTION/ FINE-TUNING	PRETRAINED MODEL(S)	DATASETS	(TOP) MODEL/CLASSIFIER	REGULARIZATION	OPTIMISATION	PERFORMANCE	IMAGE PATCH SIZES
MARMANIS ET AL. (2016)	Feature extraction (layer 7 and 8)	Overfeat (ImageNet)	UC Merced Land dataset	CNN (2 Conv + 2 FC using Softmax classifier)	Dropout, weight decay and data augmentation (zoom, rotate, etc.)	SGD with momentum (batch size 10)	92.4% accuracy	-
PIRES DE LIMA AND MARFURT (2020)	Feature extraction and fune-tuning	VGG19 and InceptionV3 (ImageNet and another version on PatternNet)	UC Merced Land dataset, AID and PatternNet	Classifier network (average pooling, 2 FC using Softmax classifier)	Dropout	Several versions: SGD w/o momentum (0.9) and learning rate 0.001/0.001, Adam (lr 0.01), Adamax (lr 0.002)	Best accuracy (99.7%) InceptionV3 (pretrained ImageNet) fine-tuned on PatternNet	-
PETROVSKA ET AL. (2020)	Fine-tuning (first new layers then whole network), feature extraction from fine-tuned networks for SVM classifiers	ResNet50, Xception, InceptionV3, and DenseNet121 all pretrained on ImageNet	AID and NWPU-RESISC45 (train/test splits: 50/50 and 20/80, 20/80 and 10/90 respectively)	Three versions: Softmax classifier (1 FC and Dropout before classifier) and 2 SVM classifiers (feature extraction from learned features of softmax model version)	Dropout, data augmentation, label smoothing	SGD, Cyclical lr, linear decay lr scheduler	Many different combinations, overall accuracy is used.	-
JIANG ET AL. (2019)	Fine-tuning, 'fusion'	ResNet50 (ImageNet)	Arctic Wetland Dataset (AWD)	Early, middle and late fusion	Augmentation balancing, weight decay (0,00002)	SGD with momentum (0.5), poly learning rate strategy	93.12% averaged accuracy	Patches of 30x30m (either 10x10 pixels for 3m resolution, or 15x15 pixels for 2m resolution of full image). Each patch has a single label.
RADKE ET AL. (2019)	-	-	Own dataset (Red, Green, Blue and NIR from Landsat-8 satellite (to also create NDVI), DEM, athmospheric pressure, temperature, dew point, wind direction, wind speed, precipitation, and relative humidity)	CNN (2 conv, 3 dropout, 1 FC and 1 output layer)	-	SGD	Average accuracy 87.7%, recall of 91.1% and F-score of 6.4%	Each pixel has a resolution of 30x30m. Using sliding window of 30 pixels to create patches (30 x 30 pixels).
G. ZHANG ET AL. (2019)	-	-	Own dataset with 11 features (elevation, slope, aspect, average temperature, average precipitation, surface roughness, average wind speed. Forest coverage ratio, NDVI, distance to roads, and distance to rivers)	CNN (3 Conv, 2 pooling, 3 FC)	Dropout (0.5)	Adam, ReLU	Validation accuracy 87.92%, AUC of 0.86	25x25 pixel patches (pixel size of 5x5 km), sliding window for each pixel, each pixel will have a prediction between 0 and 1. Five different categories depending on probability: very low, low, moderate, high, and very high.
JANIEC AND GADAL (2020)	-	-	Fire data from FIRMS, Radiation, precipitations, temperature, max temperature, NDVI from MODIS, elevation, slope, slope direction, distance from settlements, distance from roads, distance from water lines	Random Forest, Maximum Entropy	-	-		-
LANGFORD ET AL. (2018)	-	-	Own dataset (NDVI, EVI, SAVI, Bands 1-7 from the MODIS, Daytime LST, and wildfire extent)	DNN (input layer, 3 hidden layers with ReLU and output layer with softmax)	Not mentioned	Not mentioned	0.68 precision, 0.95 recall on dataset-0, 0.61 precision and 0.96 recall on dataset-1	-
MAULANA ET AL. (2019)	-	-	Monthly precipitation, NDMI, type of peatland use and cover, road network density, peat depth, peat decomposition type, river network density, canal network density	Spatial Logistic Regression	-	-	AUC of 0.8309 and overall accuracy 85.16%	Cells of 100 x 100 m.
ROSADI ET AL. (2020)			Time of data collection, district area, LST, wind speed, humidity, height, and NDVI	AdaBoost, Random Forest, k-Nearest Neighbour, Decision Tree, Logistic Regression and more	-	-	95% overall accuracy for multiple models	-

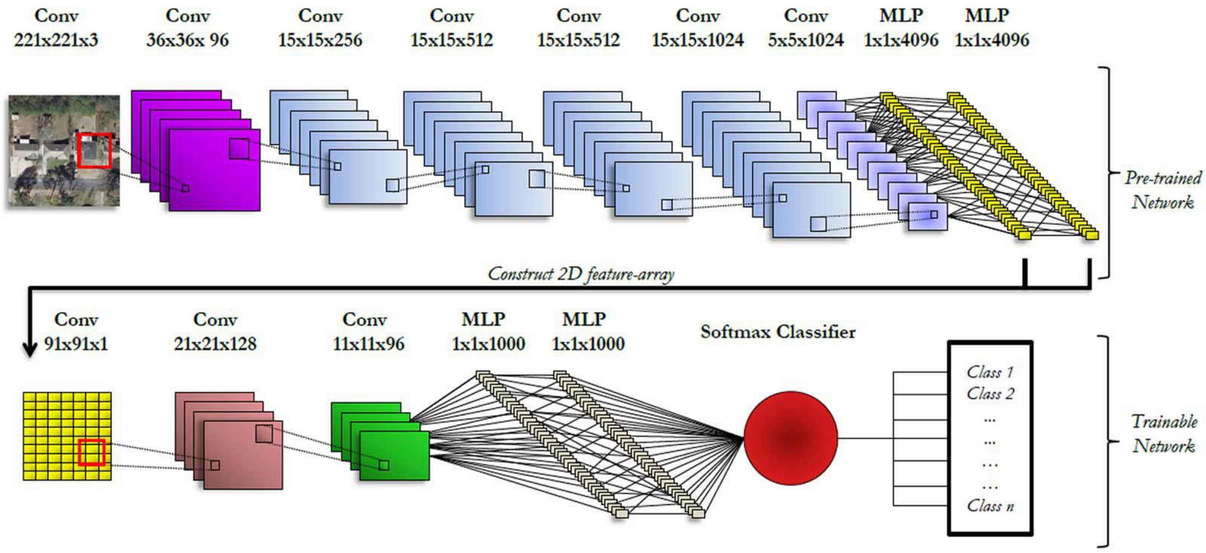


Figure 4. Workflow diagram of the feature extraction from the pretrained network and the trainable network. From Marmanis et al. (2016).

A more recent study by Pires de Lima and Marfurt (2020) focused on the transfer learning process from natural images to remote sensing images using the feature extraction as well as the fine-tuning approach. They use VGG19 (Simonyan & Zisserman, 2015) and InceptionV3 (Szegedy et al., 2016) as pretrained networks (on ImageNet), adding a classifier network (top network) with an average pooling layer, two fully connected layers, dropout in between these layers and a softmax function. Three remote sensing datasets are included in the experiments: the UC Merced Land, AID (Aerial Image Dataset) by Xia et al. (2017) and PatternNet (Zhou et al., 2018). The authors compared the performance on each dataset with randomly initialized weights, fine-tuning and feature extraction. They also evaluated the performance of a CNN pretrained with PatternNet (instead of ImageNet) on the UC Merced Land and AID datasets. More detailed information about optimizers and regularization methods can be seen in **Table 2**.

Overall, the authors showed that fine-tuning as a transfer learning approach works well and performs better than randomly initialized weights. Feature extraction can be limited by the difference in the primary task (natural image classification) and the target task (remote sensing image classification) as the initial layers are frozen and causes the model to overfit. Unfreezing the layers reduces this. Training all layers and fine-tuning after initially only training the last layer showed the highest increase in accuracy.

In the case of the pretrained network on PatternNet, the authors found that the test set accuracy after training with UC Merced Land and AID in both cases is lower than in comparison with when the datasets were trained using the ImageNet pretrained network. The authors reason this has to do with ImageNet being a complex dataset with many different examples for the same class, whereas PatternNet is meant to give clear examples of different remote sensing scenes, lacking the intra-class diversity that ImageNet does have.

Interestingly, this research found that feature extraction did not work as the layers were still all frozen. The top model only consists of fully connected layers with no convolutional layers that can still learn features. With Marmanis et al. (2016) the necessity of a pretrained network becomes clear for smaller datasets. However, this might arguably only work with a classifier on top that can still learn to some extent, as with this study, the top network has no learnable layers. Concluding, that applying transfer learning to peat fire prediction, fine-tuning is a workable approach with remote sensing imagery and that feature extraction can be another approach if a trainable classifier is added as a top network.

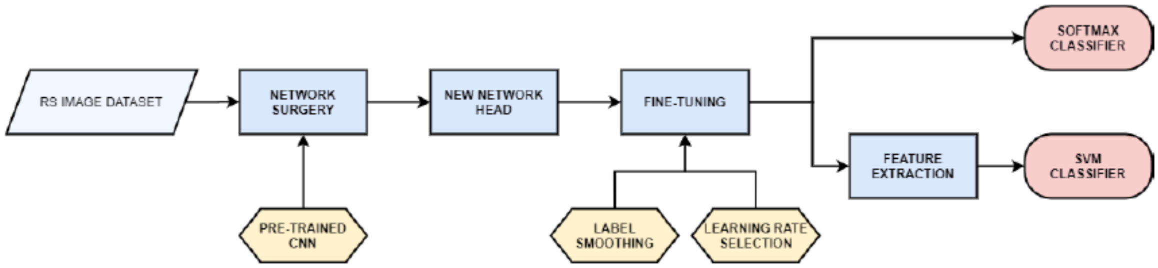


Figure 5. Flowchart of the pretrained CNN, fine-tuning and feature extraction. From Petrovska et al. (2020).

Another approach towards transfer learning is the combination of fine-tuning and feature extraction done by Petrovska et al. (2020) where the performance, measured in overall accuracy, was compared between a softmax output function and Support Vector Machine (SVM) classifier. Softmax has been used in the above two studies and the authors mention that SVM is common in other transfer learning research. Four CNN networks were tested: ResNet50 (He et al., 2016), Xception (Chollet, 2017), InceptionV3 (Szegedy et al., 2016), and DenseNet121 (Huang et al., 2017). All networks were pretrained on ImageNet. Two datasets are used: AID and NWPU-RESISC45 (Cheng et al., 2017). As can be seen in **Figure 5**, there is a pretrained CNN and through 'network surgery' the final layers after the average pooling layer are replaced by a top network (new network head) consisting of a fully connected layer, dropout, and a softmax layer. Fine-tuning is then used to train the network on the dataset. This is done for each of the two datasets, resulting in two trained networks. After the fine-tuning, feature extraction from these trained networks is applied and the features are transferred to a SVM classifier. More specific details can be found in **Table 2**. The results show that in general the SVM outperformed the softmax function, however, this also depended on the dataset used.

From the study can be concluded that transfer learning approaches can be combined using fine-tuning and feature extraction. The SVM classifier outperformed the softmax function, which agrees with previous conclusions that an additional top network that can still learn, seems to perform better than a top network that cannot.

ArcticNet (Jiang et al., 2019) is a CNN that can classify six different wetland types: water, peat bog, channel fen, dense forest, sparse forest, and wetland. From the CNN predictions a label map is created. The model uses two CNNs with the pretrained network being ResNet50. One network (called RGB Branch) inputs the RGB patches created from the full image, and the other network (NIR/DEM/NDVI branch) the patches with the NIR, DEM (Digital Elevation Model), and NDVI features (NDN). Each network is trained separately on the data. The authors researched different methods to fuse the two networks together and what influence these fusion strategies would have on the performance. The three fusion strategies are: early fusion, middle and late fusion (see **Figure 6**). Early fusion concatenates the RGB and NDN directly at the input of the network. Middle fusion concatenates the feature maps produced at the second convolutional layer of the ResNet50 and another variant at the third convolutional layer. The layers before fusion inherit the learned weights from each separate network (RGB Branch and NIR/DEM/NDVI Branch) and retrains the network's layers afterwards. For late fusion the activation vectors from each single network from the last convolutional layer are concatenated as an input for the fully connected layer afterwards. Only the last fully connected layer is retrained. Training was done using the SGD optimizer. For more details see **Table 2**.

Even though the single networks performed well, all fusion strategies showed better results with the best average accuracy of 93.12% with late fusion. Even though the fusion method shows good options, it lacks the study towards the effectiveness of directly having all the features as input. Meaning, instead of creating two different networks with a standard input of three features, there could be one network with six features as input. This research focuses on wetlands, which as mentioned before, peatlands are a certain type of. Thus, it might indicate that R, G, B, NDVI, and NIR are able to find information specific for certain wetland types and that these multi-spectral information layers might be useful for peat fire prediction as well.

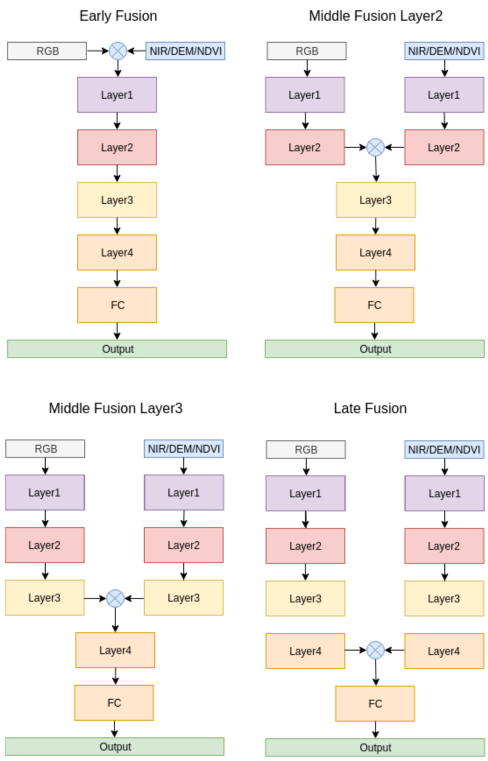


Figure 6. The three fusion strategies applied. From Jiang et al. (2019).

3.2 Wildfire prediction

Research papers towards peat fire prediction is presumably scarce, as few were found (see **Section 3.3**), therefore a more general view on wildfire prediction will be assessed in this section. The focus is on papers that use a CNN, as previously mentioned in **Section 1**, CNNs are effective for image recognition and is an exciting tool to find complex non-linear feature representations from image pixels. The following papers for wildfire prediction are observed to find a proper approach that can be applied to peat fire prediction.

FireCast (Radke et al., 2019) predicts future wildfire spread in the United States (test done at Beaver Creek, Colorado) by combining satellite imagery, a small number of location characteristics, and weather forecast. With the help of a CNN consisting of two convolutional layers, three dropout layers, one fully connected layer and an output layer. A sliding window of 30 pixels creates patches of each point-of-interest (POI). SGD is the optimizer, and no regularization is used. Landsat-8 satellite imagery is used as the visual input to the model with a resolution of 30m. FireCast uses the features shown in **Table 2**. This data includes atmospheric pressure, temperature, dew point, wind direction, wind speed, precipitation, and relative humidity for each fire location. For each fire that is in the dataset, an initial fire perimeter and the future 24h of weather data is given. With the ground truth being the fire perimeter of the next day. Using this data, FireCast outputs an image of the area for each day, showing the POIs with a colour corresponding to a certain prediction value. Evaluation is done using total prediction accuracy, recall and the F-score.

Testing FireCast, results in a total average accuracy of 87.7%, a recall of 91.1% and an F-score of 6.4%, although the authors mention that the model normally has a higher F-score on days with more fire growth. These F2-scores are however, still better than the scores found with a commonly used fire spread model. When taking a two-week period in between the pixel prediction, it shows the potential of FireCast to predict fire spread at least two weeks up ahead (F-score of 34.4%). The authors state that even though some pixels are wrongly predicted, these pixels could still provide useful information. In the end, FireCast is limited to the available data and relied on data augmentation to create enough training data. The method for data augmentation is not mentioned. Future work is aimed towards more data from different regions as FireCast is limited due to the available data. Different input resolutions are also considered.

FireCast is an example of how a small dataset can be applied in combination with a CNN. The creation of image patches is one of the approaches that can be taken from this paper, whereas the size of 30 x 30 pixels for the sliding window could also play a role for fire prediction and spatial information. Another important aspect is that it shows the potential to predict wildfires within a timespan of two weeks.

G. Zhang et al. (2019) studied the forest fire susceptibility at the Yunnan Province, China using a CNN and 11 influential features that are described in **Table 2**. The authors managed to predict which areas of the Yunnan Province had a high susceptibility of fires breaking out, outputting a susceptibility map to visually show. Susceptibility in this study, is defined as the probability estimation of a fire occurrence in a region.

To create the input data, each of the influential factors is converted to a raster map with a pixel size of 5 x 5 km, converted to WGS 1984 Web Mercator and min-max normalized. The NDVI feature was created using a mean image of the spring months (March, April, and May). Each feature map is split into 25 x 25 pixels patches (125 x 125 km in size). All the features are stacked onto each other, creating a 3-dimensional array of n x n x c to input into the CNN. Where n is the row and column of each input image and c the number of influential factors. A corresponding ground truth label was created with the corresponding XY central coordinates on an Ignition Raster Dataset (IRD). A raster where all the ignition points are stored on. To make the dataset more balanced, they added neighbouring pixels within a buffer zone of 5 km of an ignition point to the fire class. Argumentation is that following the spatial characteristics of forest fire events, areas near the fire pixel are prone to fire as well. The same amount of non-fire points as fire points were selected from the IRD. The dataset was split between 80% training data and 20% validation data for the years 2002-2009. 2010 was used as the test set. The labels for the fire points have a value of 1 and value 0 for the non-fire points label.

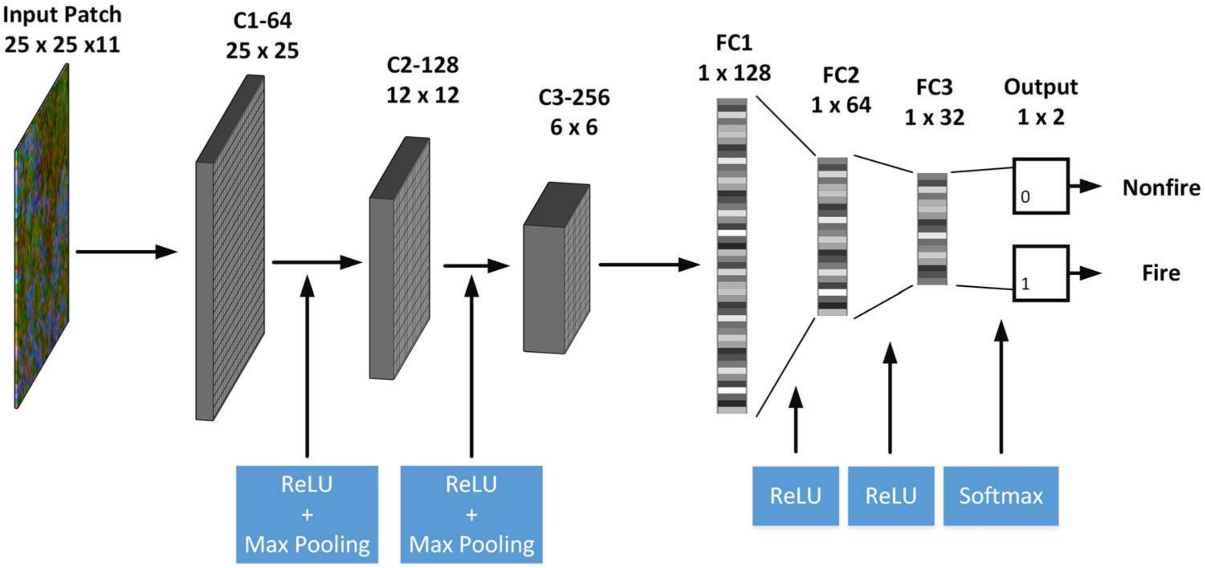


Figure 7. Architecture design of the CNN from G. Zhang et al. (2019). C1, C2, and C3 are convolutional layers. FC1, FC2, and FC3 are fully connected layers.

The CNN (**Figure 7**) consists of three convolutional layers (64, 128 and 256 kernels) with kernel sizes of 3x3, two pooling layers, and three fully connected layers (128, 64 and 32 neurons). Activation function optimizer, etc. can be found in **Table 2**. Five statistical evaluation metrics including accuracy, specificity, sensitivity, positive predictive value, and negative predictive value are used. To evaluate the global performance the AUC value of the ROC is added as well. The results give a validation accuracy of 87.92% and on the test data the CNN model has an AUC of 0.86.

This research shows a useful approach to create a dataset consisting of image patches from the feature maps and the use of creating a single image from several images over a certain timespan. Different dataset balancing approaches, mentioned above, can be considered as well. Especially, that the authors went for the selection of the same amount of non-fire points. This also shows that not the entire area needs to be included in the training data for the network to predict for all test data patches that does include the entire area. The authors use an input layer that can input eleven features, showing the possibility of a CNN with over three channels. No transfer learning was applied in this research, nor the use of an existing CNN such as ResNet50. It would, therefore, be interesting to see how these approaches can be combined to create a multi-channel CNN from a pretrained network that initially only takes three features as an input.

Janiec and Gadal (2020) studied forest fire prediction in North-East Siberia. More specific in the Republic of Sakha. The authors used two machine learning methods: Maximum Entropy and Random Forest. The fire prediction maps created were on macro level (the Republic of Sakha region) and micro level (Nyurbinsky region). Two types of models were created: presence prediction and presence probability maps. Presence prediction has two classes; absence (there is no probability of a forest fire) and presence (there is a probability of a forest fire). For training, fire data from FIRMS was used over the years 2001-2015 and a validation set of 2015-2018. Features used for the prediction of forest fire can be found in **Table 2**.

Despite the limitations of poor exploration of the area, the authors found that in general the random forest method showed better results on macro scale and the maximum entropy on micro scale. The NDVI used was highly correlated with fires in the boreal forest. The authors mainly found that fire risk in the region of Yakutia is not solved easily as boreal forests are largely different from other parts of the world. It is suggested that different types of datasets such as climatic and different remote sensing data from different sensors is used for fire risk assessment. As far as is known, no other articles can be found on this topic for the region, which would imply that there is still a lot of research that can be done towards this area and (peat) fire prediction.

3.3 Peatland fires

As noted in **Section 3.2**, as far as is known, no (peat) fire prediction research has been done in the North-East Siberia region. However, there have been some peat fire prediction/detection studies published in other areas of the world. These studies are assessed to see which features are used and which machine learning techniques are applied.

The study of Rosadi et al. (2020) uses different machine learning techniques including an AdaBoost method, Random Forest, k-Nearest Neighbour, Decision Tree, Logistic Regression and more, to predict the occurrence of fires of a peatland area in Indonesia. The features as input can be found in **Table 2**. The dataset is created by gathering data from fire hotspots of peatlands and labelling these with the value 1. The same area is then used at a different time when no fire has occurred, labelling it as the value 0.

On the test set the following accuracy with a certain train/test split ratio were obtained: SVM (95%, 90/10), kNN (95%, 80/20, k=3), Logistic Regression (90%, 90/10), Decision Tree (95%, 90/10), Naïve Bayes (90%, 90/10), and AdaBoost (95%, 90/10). The Adaboost method described in the paper outperforms the other approaches slightly, but only because the training accuracy scored the highest (100%). Other machine learning approaches also performed very well.

The authors conclude that either classical or more advanced machine learning approaches can be used for fire occurrence detection in peatlands. This means it opens a door for further research towards the use of deep learning and CNNs for peat fire prediction.

A research on wildfire detection and mapping in Alaska using only satellite data (Langford et al., 2018) includes an area that is shown on the peatland cover map by Hugelius et al. (2020) to have peatland coverage. Even though the authors main aim is in general wildfire mapping, since it involves peatland areas, it is interesting to see which features and techniques are applied involving machine learning and the use of satellite data.

Langford et al. (2018) used a deep neural network (DNN) that consists of five layers. The first layer corresponds to the different features of NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) on board the Terra and Aqua satellites, as input (see **Table 2**). The second, third and fourth layers are hidden layers with a *ReLu* activation function. The last layer is the output layer for the fire/no-fire classes and consists of a *softmax* function. One of the authors' focuses was the use of the validation-loss (VL) DNN strategy. A strategy, as explained by the authors, that splits the dataset in two (train and validation set) and uses the validation set for performance validation to select the weights. The weights that resulted in the best validation loss score, will be saved as the final network. The authors conclude that the VL strategy showed significant improvements over the standard training, especially when the training dataset is small, and the test set is large (recall of 0.00 to a score of 0.96 for the most imbalanced dataset). The results also showed that the VL strategy works well when the dataset is imbalanced.

Even though this study's focus was on the approach towards imbalanced datasets and the validation-loss strategy and consisted of fire detection and mapping with the use of pixels and not patches, it also showed that the use of solely remote sensing data (and which data is used) was able to detect fires and, therefore, might be able to apply to the prediction of peat fires as the study region did involve peatland areas.

One study that focusses on peat fire prediction in the southern hemisphere is by Maulana et al. (2019). They created spatial logistic regression models using remote sensing information (fire occurrences from MODIS burned area product images) in combination with sources of climate variables (monthly precipitation, NDMI), human activity (type of peatland use and cover, road network density), physical peat characteristics (peat depth, peat decomposition type), and physiographic variables (river network density, canal network density). The fires usually started in January. Precipitation and moisture index data were gathered from the six months prior of each of the years included. As an example, for January 2004 the months July-December of 2003 are taken.

The fire occurrence data is created using cells of 100 x 100 m, where each cell that has a burned area detected will be defined as 1. Cells with no burned area will acquire the value of 0. The created dataset is split into a training (80%) and test set (20%). Three models were created and all of them consisting of all the variables, with a difference in the approach for NDMI and precipitation. Model t-1 uses a moving average of the months December plus the three months prior for the NDMI and precipitation; t-2 uses the months November and the three months prior; and t-3 the months October and the three months prior.

The results show that NDMI and precipitation data from the months closest to the fire occurrence month January (t-1), have the highest AUC and overall accuracy, 0.8333 and 85.35% respectively. However, since climate data is mostly only available in later months (so data of December is available in January), the authors chose the t-3 model (AUC=0.8309, accuracy=85.16%) to create a peatland fire prediction map. The prediction map shows that an NDMI range of -0.06 – 0.30 is related to 96% of peatland fire. Precipitation ranges 230-246mm/month related to 99% of peatland fire occurrence. Making climate variables strongly related to peatland fires. 90% of peatland fires are predicted to take place at peat depths above the 100cm.

Interestingly, this study shows how data dating months before the actual fire occurrence can contain information for peatland fire prediction. Mainly, the use of NDMI and precipitation, which both show to be high indicators for peatland fire. That peatland fires in that study area are mostly detected on peat depths above 100cm, might carefully imply that especially these peatland areas are of need for a prediction map and that this could also apply to peatland in North-East Siberia.

Concluding, fine-tuning and feature extraction with a trainable network are both good transfer learning approaches. Studies assessed above, however, showed transfer learning only on RGB valued images. This means only three features are as input. For input of more than three features, a pretrained network on ImageNet will not suffice as it only allows three features to be inputted. One approach mentioned is to fuse two pretrained ResNet50 networks together. However, this needs several separate networks when even more features are added. Another approach is preferable. One of the approaches may be directly inputting a stacked image of all the features at the input layer (Langford et al., 2018; G. Zhang et al., 2019). The studies did not use this approach in combination with transfer learning on a CNN. A pretrained network would need to become available for peat fire prediction that allows for the input of thirteen features (all Sentinel-2 bands). Two of the options to create a pretrained network on satellite images is to either use the EuroSAT (see **Section 2.1.1**) dataset or the BigEarthNet (Sumbul et al., 2019) dataset. For the EuroSAT dataset an approach is needed to create a pretrained network of ImageNet with thirteen features allowed and then train on the EuroSAT, since the dataset is too small to train a network with random initialized weights. BigEarthNet is big enough to use random initialized weights, however since multiple models with different amount of input features are needed, this could be very time consuming. Simply because of time constraints, the EuroSAT dataset is used. ResNet50 is chosen as the network architecture, since it worked well with network fusion described by Jiang et al. (2019) and an advantage of using ResNet50 is that EuroSAT has shown to perform very well (Helber et al., 2019).

It shows that from the vegetation indices, the NDVI is used in many mentioned studies (see **Table 2**) and worth studying to see how this feature influences the performance of peat fire prediction of the CNN. Maulana et al. (2019) concluded that NDMI is a very high indicator for peat fires even from data months prior of the occurrence of the peat fire and is therefore a water index to be assessed in this thesis. Recall, that water indices could hold information about peatland conditions when combined. Another water index mentioned in **Section 2.1** is the NDWI, which is similar to the NDMI, except that is uses the green band instead of the SWIR. The NDVI, NDMI and NDWI will be studied to see their influence on the performance for peat fire prediction. Prediction follows the definition of susceptibility from G. Zhang et al. (2019) as the probability estimation of a peat fire occurrence in a region. For this thesis the Sakha Republic in North-East Siberia will be used as the region of interest, as this region has little research towards (peat) fires (Janiec & Gadal, 2020).

4 Methodology

This section describes the methodology to answer the main research question of the viability of peat fire prediction in the Arctic area of North-East Siberia using transfer learning with a CNN pretrained on the EuroSAT dataset. First up are the study areas within the region of interest. Then the creation of the datasets, of which the workflow can be seen in Figure 8. Afterwards the training of the networks with the created datasets is explained as well as the evaluation metrics used. The methods applied for further analysis on the datasets, are last.

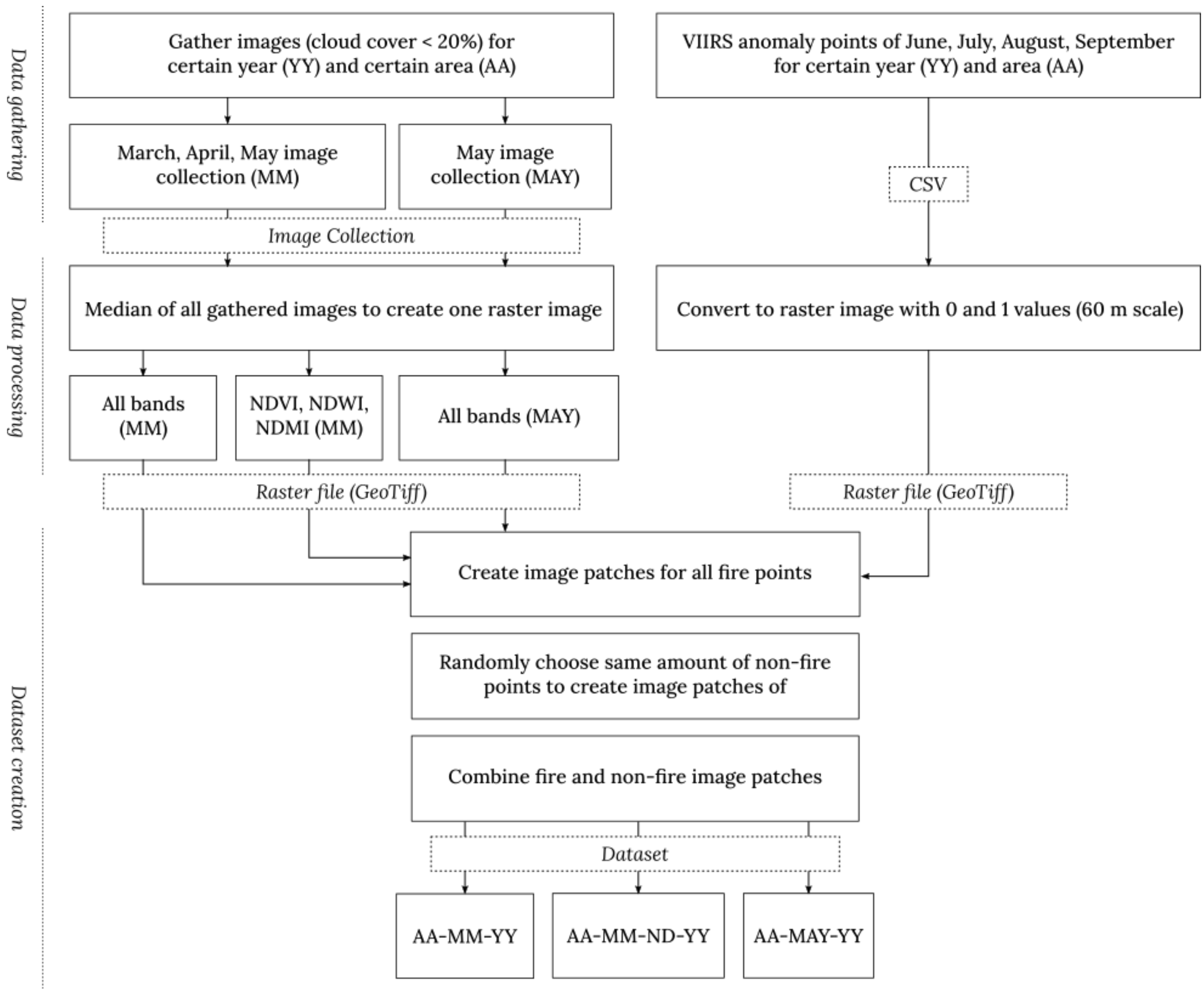


Figure 8. Workflow for creating the different datasets per year and area.

4.1 Region of Interest

The region of interest (ROI) with its two different areas A1 and A2 (Figure 9) lies in the Sakha Republic of Russia (Yakutia), which is located in North Eastern Siberia and covers around 20% of Russia (Janiec & Gadal, 2020). Yakutia is an area where it experiences for example extreme high and low temperatures and a thick layer of permafrost. The ROI selected is located above the Arctic Circle. When assessing a map by Vompersky et al. (2011) this area should contain peatlands. A recent peatland cover map by Hugelius et al. (2020), also shows the ROI containing peatland coverage. Recall that in this region large and frequent fires have been present at the Northeast of Siberia, where the Sakha Republic can be found. Additionally, when observing the FIRMS thermal anomalies, it becomes apparent that especially in 2019 and 2020 many thermal anomalies (probable fires) can be found in the Sakha Republic region above the Arctic Circle. A1 and A2 are small areas taken from this region.

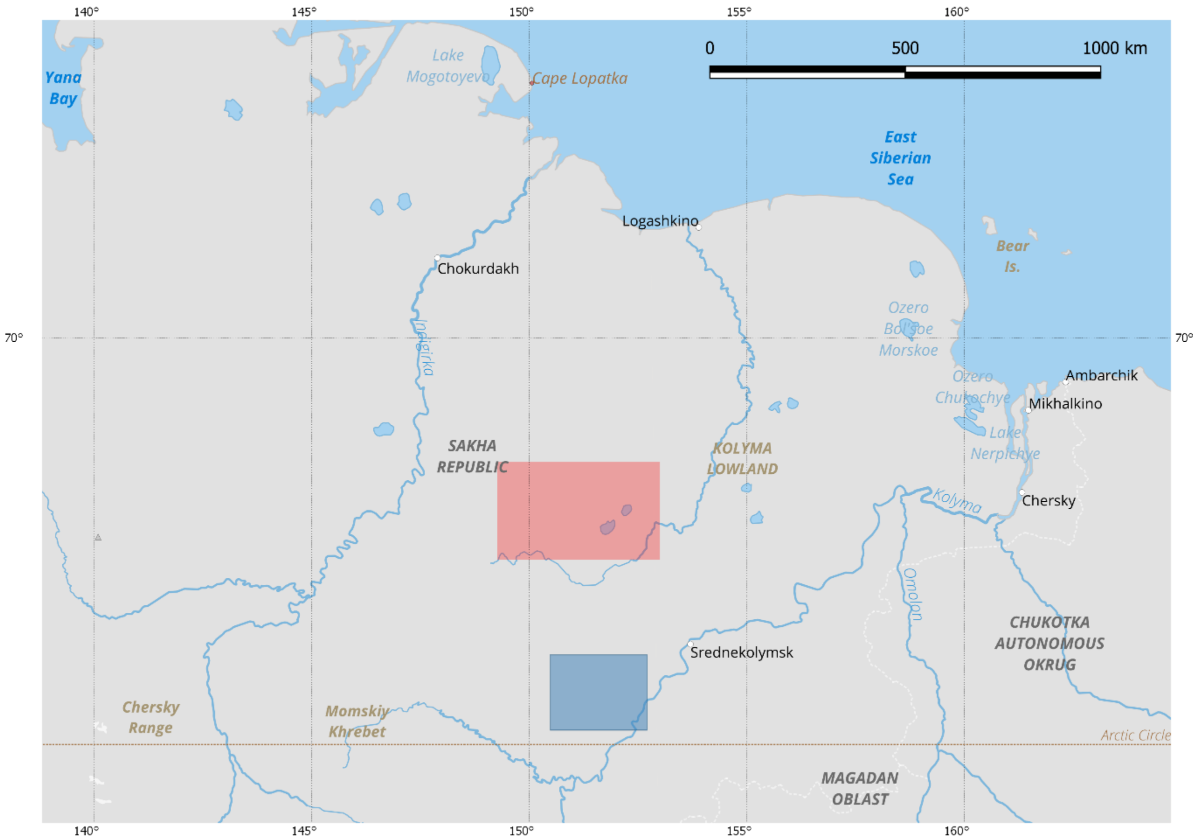


Figure 9. A1 (red) and A2 (blue) areas in the Sakha Republic of North-East Siberia.

4.2 Data gathering

Copernicus Sentinel data are gathered (see **Section 4.2.1**) and processed (see **Section 4.3.1**) using the Google Earth Engine (https://earthengine.google.com). The anomaly points are collected from National Aeronautics and Space Administration's (NASA) Fire Information for Resource Management System (FIRMS) website (https://firms.modaps.eosdis.nasa.gov/). In this thesis the anomaly points from the Visible Infrared Imaging Radiometer Suite (VIIRS) instruments aboard the Suomi National Polar-orbiting Partnership (S-NPP) are used, which was launched in

2011 and is a joint mission from the National Oceanic and Atmospheric Administration (NOAA) and NASA (Schroeder et al., 2014). The resolution of the thermal anomalies is 375 m.

4.2.1 Gather images for certain year and area

The dataset of the Sentinel-2 is from the Level-1C where Top of Atmosphere (TOA) calculations are applied. To create an image collection for March-May (MM), all images from 1st of March until the 31st of May are gathered. Using these spring months follows the approach applied by G. Zhang et al. (2019) for the NDVI feature. All 13 bands (**Table 1**) are included from each image and will function as the features. For May (MAY) the 1st of May until the 31st of May is taken. Next up, a cloud mask is applied to each image and all images with a cloud coverage above 20% are removed from the collection. Each image in the collection has a WGS 84 geospatial format and the resolution for each band is set to 30 m. This is done for the years 2019 and 2020 and study area A1 and A2.

4.2.2 Gather anomaly points for certain year and area

The fire points include all thermal anomalies detected by VIIRS from 1st of June until 30st of September for the years 2019 and 2020 and are downloaded from the FIRMS website in a CSV format. For each year a separate CSV file is created with the corresponding thermal anomalies.

4.3 Data Processing

The processing of the data is done by creating a single raster image from an image collection and a raster image of all the anomaly points.

4.3.1 Median image data

The median of the collection of images is calculated, creating a single image with 13 features (see **Table 1**), that represents the median for each pixel of the values of March until May of a certain year (MI) following G. Zhang et al. (2019). So, for each pixel location (i,j) the value of each image I in the collection M is collected and from these n values the median of that pixel location (i,j) is chosen. This is repeated for each pixel location, creating a single image MI. The MI of the whole area for each year (2019 and 2020) are exported in the GeoTiff format. Separately from this, the NDVI, NDWI and NDMI are calculated from the MI features. They are stacked together into one image being saved as a band, creating an image with 3 features. This image is also exported as a GeoTiff file.

4.3.2 Fire raster image

The CSV files for each year are imported into GEE as a (fire) point. Each (fire) point corresponds to a certain coordinate value provided in the CSV in longitude and latitude. For each area, a raster image for each year (2019 and 2020) is created where every fire point is converted to a pixel with a value of 1 and other pixels get a value of 0. These raster images have a scale of 60 m. Meaning each pixel represents a 60 x 60 m area. The fire raster image for each year is then downloaded as a GeoTiff file and has the same dimensions as the MI raster file. This means that some neighbouring pixels are also labelled as fire points, which as mentioned by G. Zhang et al. (2019) these neighbouring pixels might also be fire-prone.

4.4 Dataset creation

In this part of the workflow the actual datasets are created by using the raster files and converting them to image patches with fire or non-fire labels. All three steps within this part are combined in the next subsection.

4.4.1 Create fire and non-fire image patches

For each year and each area, image patches are created using a sliding window approach used in other studies (Radke et al., 2019; G. Zhang et al., 2019) from the GeoTiff files (images and fire points). Firstly, the x and y coordinates of all the fire pixels are stored in a list. Added to this list are the same number of non-fire pixels, roughly following the dataset balancing strategy described by (G. Zhang et al., 2019). The non-fire pixels are without any fire pixel in their patch, to make sure it completely represents a non-fire patch.

A window size n is decided on and for each of the stored coordinates in the list a patch of $n\ x\ n$ is made. With the pixel corresponding to the xy coordinate as the centre pixel. This is done for all 13 bands, resulting into a patch of $n\ x\ n\ x$ 13. In this thesis, a window size of n=21 was chosen and will result in a 21x21x13 image patch. The same xy coordinate as the centre pixel is used on the fire point raster image as the label for the matching patch. The image patch size of 21 is a rough average of the image patch sizes found in the reviewed literature, which the patch sizes can be found in **Table 2**.

The output were the following datasets that were created: A1-MM-19, A1-MM-20, A1-MAY-19, A1-MAY-20, A2-MAY-19, A2-MAY-20, A1-MM-ND-19, and A1-MM-ND-20.

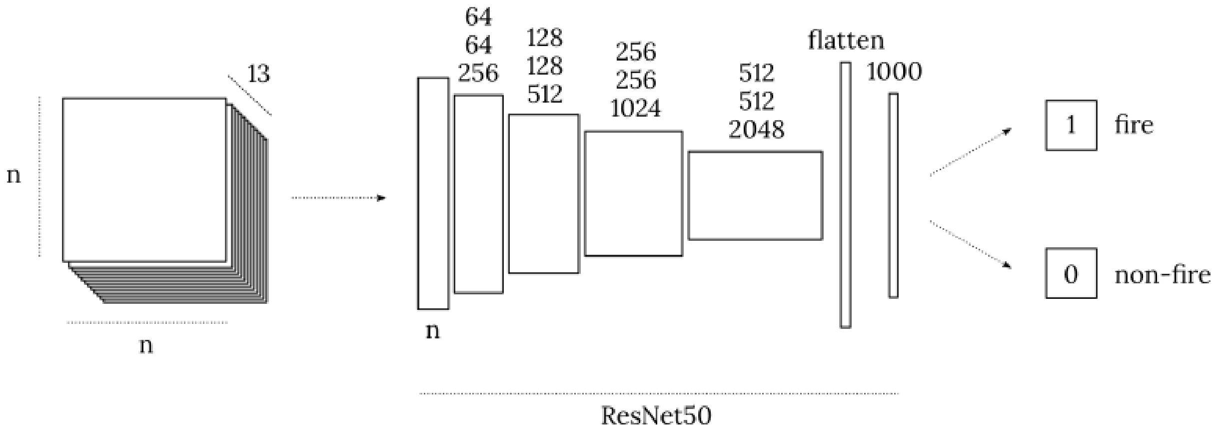


Figure 10. Model structure with the input dimensions (left), the ResNet50 architecture (middle) and the prediction outcome classes (right).

4.5 Model Structure

As is shown in **Figure 10**, the structure of the model consists of an input layer that takes in the n x n x 13 pixels image patches (or n x n x 3 in case of the indices NDVI, NDWI, and NDMI). These image patches are fed into the CNN. The output is binary; either a patch belongs to the fire or non-fire class. Binary outputs are used to calculate the scores (F2-score, AUC and confusion matrix) to evaluate the performance.

4.6 Training on EuroSAT

According to the paper connected to the EuroSAT dataset (Helber et al., 2019) the best results obtained were 98.57% accuracy. This, however, is only with training on the red, green and blue bands. The authors mention that they used a ResNet50 trained on ImageNet and first trained the last layer using a learning rate of 0.001. Afterwards they fine-tuned all the layers using a learning rate between 0.001 and 0.0001. The authors do not mention which optimizer is used. In this work the Adam optimizer is used, which is used in the research of G. Zhang et al. (2019) as well and is explained in **Section 2.2.1**. After training, the model trained on all 13 layers of the EuroSAT dataset was found to have a 95% accuracy on the test set.

The EuroSAT dataset was split in half. One half was used to train only the first and last layer and the other half to fine-tune all the layers. The first layer is included with the training as it was removed and replaced by a layer that can input 13 channels. The first initial layer had pretrained weights trained on ImageNet for the red, green, and blue channels. The weights are copied over to the other channels. This means that the fourth layer gets the weights from the red channel, the fifth from the green channel, etc. This goes on until all channels have pretrained weights. Training the first and last layer had 10 epochs, Adam optimizer, and a learning rate of

0.001. The fine-tuning used the other half of the dataset with 5 epochs, the Adam optimizer, and a learning rate of 0.0001.

The accuracy of 95% is a very good score and gets close to the result in the paper, considering it includes all 13 layers. It is therefore the pretrained network that will be used for the further training on the datasets.

4.7 Transfer learning

To explore how the spring months (March, April, and May) image composition using median values performs on peat fire prediction for the summer months (**RQ1.1**), a train and test method for transfer learning was set up. For each dataset a network model is trained. **Table 3** shows the number of train, validation, and test images for A1 and A2 over the years 2019 and 2020. As the goal is to predict for the coming years, two different test sets are available for each network. A network trained on the A1-MM-19 dataset, has a SY (same year) test set consisting of 2019 images from the same dataset as it is trained on, and an OY (other year) test set that consists of images from the 2020 year (A1-MM-20).

4.7.1 Training

Training the network models followed the approach described in **Section 4.6**, except for using ten epochs (instead of five) for the fine-tuning when unfreezing all layers.

Table 3. Number of train, validation, and test images for A1 and A2 over the years 2019 and 2020.

Area and year	*Train set*	*Validation set*	*SY test set*	*OY test set*
A1, 2019	7,097	1,972	789	25,419
A1, 2020	114,382	31,774	12,710	19,716
A2, 2019	9,693	2,693	1,078	26,928
A2, 2020	56,148	15,597	6,239	12,478

The first step is to load the pretrained network on EuroSAT. This network already has the right amount of input channels (13 for the features and 3 for the indices). The dataset is split in half and the first half is used for training the first and last layer (T1), with the other layers frozen, and the other half for fine-tuning all layers when unfreezing them (T2). Both for T1 and T2 the Adam optimizer is chosen (see **Section 2.2.1**). A small learning rate and number of epochs is chosen, recalling that a smaller learning rate and number of epochs is common for transfer learning

(**Section 2.4**). T1 has a learning rate of 0.001 and T2 a learning rate of 0.0001 with a learning rate step of 7 and gamma 0.1. Meaning that after epoch 6 the learning rate will be 0.00001 for the rest of the epochs. The total number of epochs is 10.

4.7.2 Testing

The testing was done after the fine-tuning (T2) to see how the network models would perform on data from the same year and another year. The evaluation metrics F2-score, AUC and confusion matrices described in **Section 2.5** were used to measure the performance of each model on the different test sets.

4.8 Feature Importance

To answer research question **RQ1.2** to find which features are discriminative for peat fire prediction with the two sub questions **RQ1.2.1** and **RQ1.2.2**, the features (see **Table 1**) and indices (**Section 2.1**) are tested for their importance. This will be done using the Permutation Feature Importance approach explained in **Section 2.5.1** and be applied to decide on the most discriminative features. A threshold of 0.5 was taken to find the most discriminative features over all the datasets. Additionally, for each dataset the difference between feature importance over the two years (2019 and 2020) and different areas are compared and explored.

4.8.1 Evaluation

A feature is important if the permutation leads to a decrease in the performance of the model which in this thesis is measured using the F2-score and AUC. The first step is to have a reference model. This is the network model trained on the test set without any permutations, test set X. The performance scores were derived by testing on the test set X. Each dataset is evaluated on the feature importance using the SY test set, so test set X refers to this test set. The test set was then be used for permutation on one of the features and create test set X^{perm}. The model was tested again, and the performance compared with the scores from the original set X. For each of the other features the same steps were taken until each feature has an importance score. The higher the outcome the more important the feature is assumed to be.

4.9 Dataset analysis

Analysis on the datasets, in particular the originally chosen dataset of area A1 and the months of March-May, is done applying several statistical and exploratory methods. The image patches are separately explored from the fire points. To assess the images, a correlation matrix was created, and Principal Component Analysis (PCA) applied. For the exploration of the fire points the Copernicus Global Land Cover map by Buchhorn et al. (2020) was used to create graphical and tabular information of the fire points' land coverage.

4.9.1 Analysis image patches

For the correlation matrix and PCA all image patches were converted to a single vector containing all pixels of all image patches included. As an example, a 10 x 10 x 13 image patch is converted to a 100 x 13 vector, if there are, for example, 10 image patches this results into a 1000 x 13 vector. This large vector is used to create the correlation matrix using the Pearson correlation coefficient. For PCA, after applying the standardization of the vector, the vector was used to find the two most important principal components. The choice for two is made after analysing how many components would be necessary to explain over 95% of the data.

For the 2020 datasets of A1, the same number of pixels are used as the 2019 datasets. The sole reason for this is the error it produces otherwise as it becomes a very large vector, needing a lot of computing power. The dataset is simply split where it matched this number.

4.9.2 Analysis fire points

Fire points were analysed looking at the number of fires per month for each year and for each year the land cover per month using the Copernicus Global Land Cover map (Buchhorn et al., 2020) adding the layer in GEE. The fire points from the CSV were imported in GEE. For each fire point GEE looked at the corresponding land cover type storing it as a collection. This collection was downloaded as a CSV file and processed. Each month was taken separately, and the total number of fires was calculated per month per year. Additionally, for each month every type of land cover was found with the corresponding amount of fire points matching the land cover type. A visual map with the location of each fire point per month for each area was also created.

5 Results

This section will provide the results gathered from the experiments. Several figures are provided in the text for clarification, others can be found in the Appendices.

5.1 Performance of (summer) peat fire prediction using the spring months

The results for the performance of the models using different datasets are shown in **Table 4** with the F2 and AUC scores on the SY test set and OY test set (see **Section 4.7**). The best scores are marked in bold. See **Figure 8** for the workflow of how these datasets differ and are created. The corresponding confusion matrices can be found in **Appendix A**.

Table 4. Performance scores for each dataset for the test set of the same year (for 2019 that is 2019), test set of the other year (for 2019 that is 2020) and the validation accuracy of the model after training.

	SY TEST SET		OY TEST SET		VAL. SET
	F2-score	*AUC*	*F2-score*	*AUC*	*Accuracy*
A1-MM-19	0.85	**0.90**	0.32	0.55	0.78
A1-MM-20	0.74	0.67	0.32	**0.70**	0.65
A1-MAY-19	**0.87**	0.89	0.11	0.59	0.82
A1-MAY-20	0.78	0.71	0.50	0.55	0.65
A1-MM-ND-19	0.76	0.84	0.14	0.50	0.75
A1-MM-ND-20	0.76	0.69	0.58	0.65	0.64
A2-MAY-19	0.86	0.88	0.0012	0.57	**0.84**
A2-MAY-20	0.82	0.78	**0.77**	0.51	0.73

Testing on the same year results with the best scores for the F2 and AUC found on SY test set of A1-MAY-19 (0.87 and 0.89 respectively). However, when tested on the OY test set (the A1-MAY-20 dataset), the scores are significantly lower. The AUC score of 0.58 indicates it is close to a random classifier. A score of 0.11 for the F2, indicates that many fires were falsely identified as no-fire, which can be seen with the corresponding confusion matrix in **Figure A-2** (bottom left). With the different region the A2-MAY-20 shows the highest performance on the OY test set with a F2-score of 0.77. However, as the AUC is 0.51, the classification is still random. Indicating that even though some models are better at classifying fire images, in general, as too many non-fire images are classified wrongly, it still does not properly show which areas are susceptible. This means that no model can predict peat fire from the other year. When the AUC is 0.70 in the case of A1-MM-20, the F2-score shows 0.32, showing that in this case non-fire classification was better than in other cases, but it still lacks the proper classification of fire images. Since it is desirable to have a high F2-score, as overlooking susceptible areas is not wanted, no model shows potential. Interestingly, when observing the results in **Table 4**, it seems that the accuracy on the validation set might also play a role. The highest accuracy shows the lowest performance on the OY test set. The 2020 cases show the best result with the highest accuracy and are in general lower than the 2019 cases to begin with but have higher OY test set scores than the 2019 cases.

Observing the confusion matrices (**Appendix A**) makes it clear that A1-MM-19, A1-MAY-19, and A2-MAY-19 can distinguish non-fire (class 0) from fire (class 1) image patches well when tested on the SY test set. The F2 and AUC scores for the three cases are above 0.85, with A1-MAY-19 an F2-score of 0.87. The datasets of the year 2020 (A1-MM-20, A1-MAY-20, and A2-MAY-20) have more difficulty with classifying non-fire images when tested on the SY test set which can be seen from the confusion matrices (top right). For the fire class, there are clearly more true positives than false negatives. These differences in performance scores between the two years (2019 and 2020) might be due to the amount of training examples. 2020 shows many more fires (see **Table 3**), indicating that a bigger training set lowers the performance of the models on the SY test set or 2020 shows less distinction between fire and non-fire image patches. However, as the same occurs in a different region (A2), it might suggest the first argument.

Overall, it shows that using the spring months and creating a MI is not a suitable for peat fire prediction of the summer months. Even when only one month is taken for the median values (MAY), the models are still not able to predict well for the other year as MAY shows random classifiers. Same year prediction could show potential for peat fire prediction using transfer learning as the models were able to learn, however not with the approach taken in this thesis.

5.2 Feature importance

For feature importance on all features (**Table 1**) it becomes clear that overall, in most cases B4, B5, and B7 (red and two vegetation red edge bands) show the highest importance score (**Figure 11**), indicating that vegetation red edge bands might positively influence peat fire prediction. For the A2-MAY-19 and A2-MAY-20 it shows importance at the SWIR and NIR/VNIR bands. Especially when considering the AUC scores. Band 12 (SWIR 2) is one of the most important feature for A2-MAY-19 and Band 11 (SWIR 1) for A2-MAY-20. B10 is never of any importance, due to that it is mainly meant for cirrus cloud detection as mentioned in **Section 2.1**. When only looking at the importance of bands above the 0.5 threshold for the F2-score, the features include: B1, B2, B4, B5, B6, B7, B8, and B8A. The AUC shows similar results, with B5 being in three out of six cases the highest importance. In general, no significant discriminative features were found, and differences of the most important features are visible over the years and regions. The only clear result is that B10 is of no importance in all cases and that small importance was found for vegetation red edge bands.

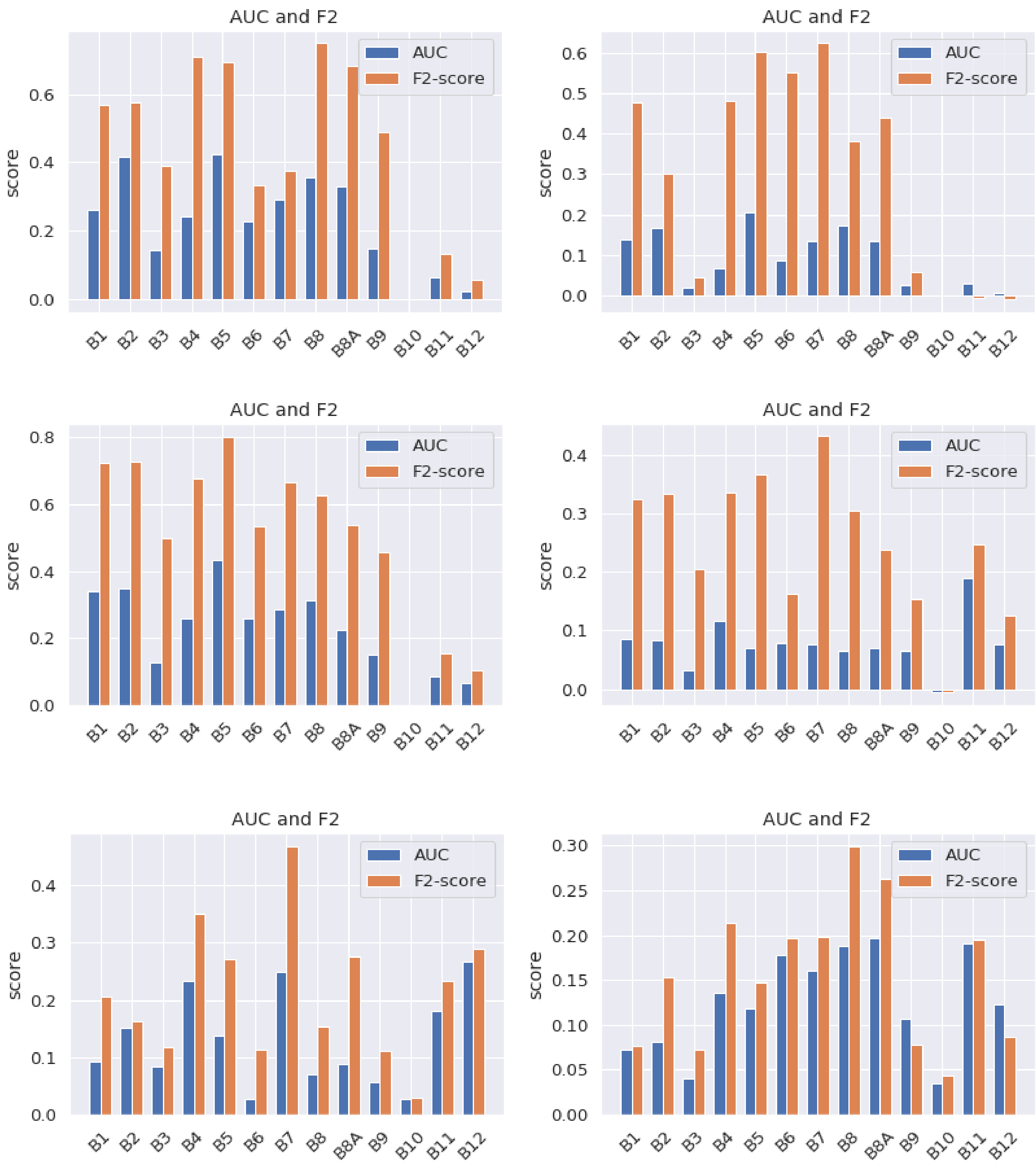


Figure 11. Feature importance scores (AUC and F2-score) for each band of the A1-MM-19 (top left) and A1-MM-20 (top right), A1-MAY-19 (middle left) and A1-MAY-20 (middle right), A2-MAY-19 (bottom left) and A2-MAY-20 (bottom right) dataset.

As permutation feature importance depends on variables being independent, a correlation matrix is presented for each of the datasets (**Appendix B**) to see how much correlation is present. The method of creating the correlation matrices can be found in **Section 4.9**. The matrices show a high correlation between B1-B9, B10 on its own and B11 and B12 for the area A1. For area A2 B1-

B6, B6-B8A, and B11-B12 are very strongly correlated. B9 could be correlated to B8 and B8A, even though it shows less correlation in 2019. Evidently showing that most features are correlated to one another. By applying PCA (see **Table C-1**) on the datasets it shows that all datasets contain the bands B1-B8A in the first principal component (PC1) with the highest values. B9 is also common in most datasets for PC1 and is also sometimes used for PC2, although it is the lowest valued number. For PC2, the bands B11 and B12 show the highest values, being supported by B10 in three of the six cases. None of the bands separately or some combined show large variance to describe the dataset. Instead, it is a linear combination of a high number of bands in the case of PC1. No clear discriminative feature can be pointed out when applying PCA, however as the eigenvector values in the first component are similar for the bands (B1-B9) scoring highest in this first component and have a high correlation, it would indicate that only one band from these would suffice to use for peat fire prediction in combination with the SWIR bands. For this one feature it may be considered using one of the vegetation red edge bands.

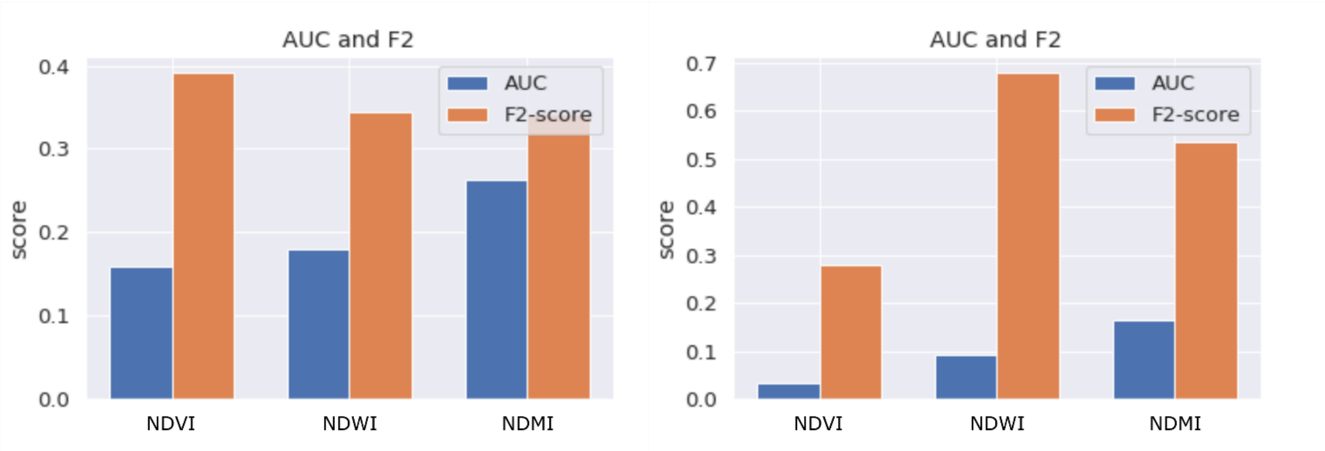


Figure 12. Feature importance scores (AUC and F2-score) for the NDVI, NDWI, and NDMI indices of the A1-MM-ND-19 (left) and A1-MM-ND-20 (right) dataset.

The feature importance over the normalized difference indices is found in **Figure 12**. With the feature importance on the indices NDVI, NDWI, and NDMI, only the A1-MM of both years has been assessed. The results show that for A1-MM-ND-19 the NDVI is the most important feature followed by both the NDMI and NDWI when looking at the F2-score. For the AUC the NDMI is highest in score. The A1-MM-ND-20 feature importance is highest with the NDWI, when assessing the F2-score and highest for the AUC score with the NDMI. There are slight similarities between the two years in feature importance over the indices however differences are also visible. The NDMI has the most importance on the AUC scores, indicating that it might be of importance to separate non-fire images from fire images. When looking at the F2-score which favours the fire class and the correct classifications of this class, the NDWI shows a very high score. Indicating that the NDWI might be of importance for classifying fire images. However, both years show differences in scores per feature, meaning that it stays unclear which indices are discriminative and can really describe the distinguishment between the fire and non-fire class.

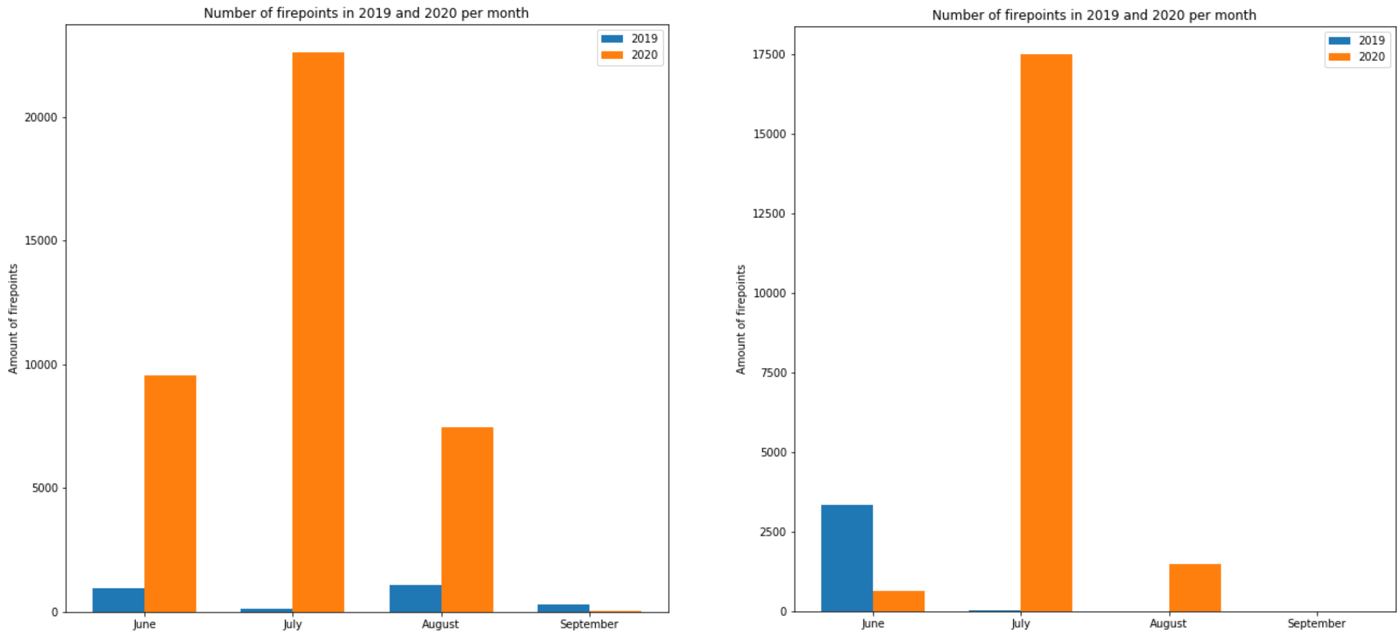


Figure 13. Number of fire points per month for 2019 and 2020 in region A1 (left) and A2 (right).

5.3 Dataset analysis

To find out what is different between the two years, the fire points acquired from VIIRS are being assessed. In 2019 there are significantly less fire points in the A1 area then there are in 2020 (see **Figure 13**). Only in September, the year 2019 had more fires. For 2019, August has the most fires. For 2020, July contains the most fire points in a month. June 2020 already starts off with a vast number of fires, indicating that the month May might also include some fires in the area that have not been considered. What mainly stands out from these observations is that within area A1, September 2019 shows many fire points, indicating many fires at the end of the summer. June 2020 starts off with many fire points early in the summer. When observing **Figure 14** it becomes visible how fires of 2020 do not happen in areas where fires have been burning in 2019. Arguing the possibility that burned areas of the last year seem to not consist of enough fuel for fires the next year as they are still recovering.

The Global Copernicus Land Cover Map (Buchhorn et al., 2020) was applied to the fire point locations using GEE to calculate the land cover type per fire point. **Table D-1** gives an overview per month and per year on the percentages of each land cover, using the same naming for land cover types as is used for the land cover map by Buchhorn et al. (2020). Only 12.8% for 2019 and 14.2% for 2020, account for herbaceous wetland land cover. Recall that peatland is a type of wetland. For area A2 (**Table D-2**) this is even lower (2.7% and 6.0% for 2019 and 2020 respectively). This could mean that image patches created from the datasets are not always corresponding with peatland but also consist of other land cover types. Do consider that the class accuracy for herbaceous wetlands is under 65% as this class has high confidence errors with mainly herbaceous vegetation having spectral similarities (Buchhorn et al., 2020). Meaning that some fire points could have been identified as the wrong land cover type. A more detailed overview of each month (for each year and area) and the number of fire points belonging to a land cover type, can be found in **Appendix E**.

After observation of the results, it shows that there is no clear indicator found from the results presented in this thesis that could explain the differences between the datasets of 2019 and 2020.

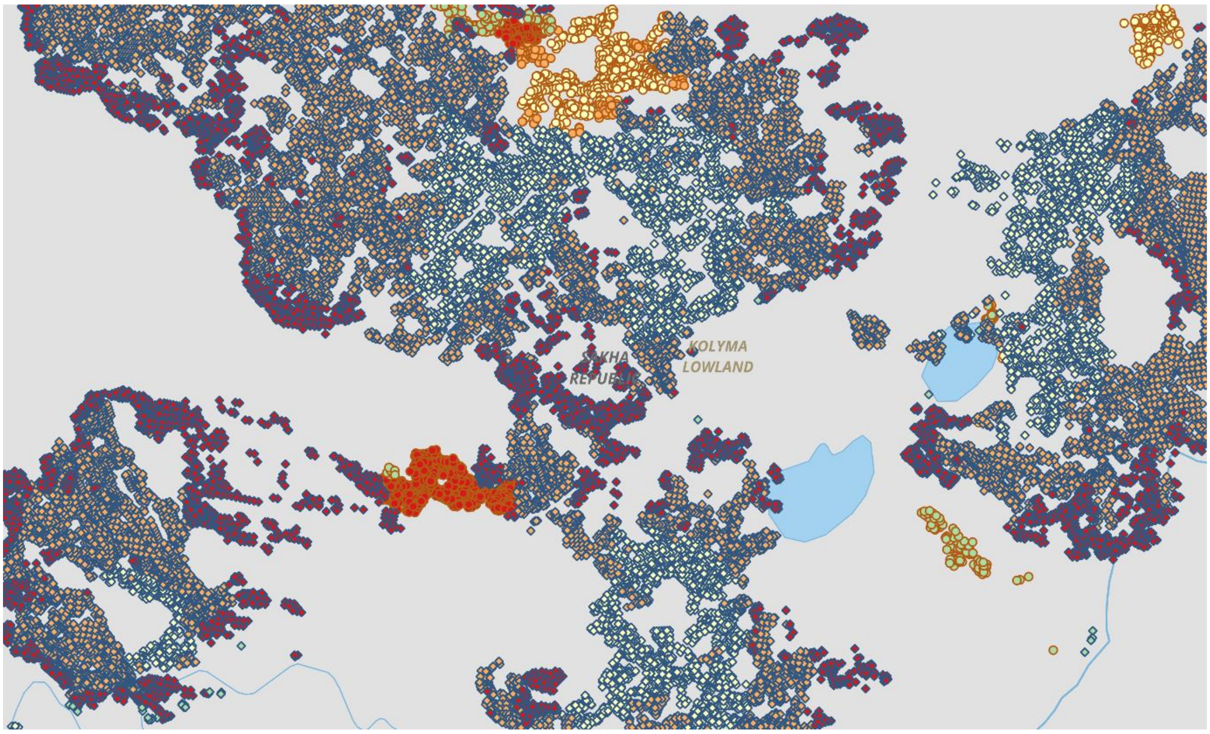


Figure 14. Map with all fire points in area A1 of 2019 (red outlined circles) and 2020 (blue outlined diamonds) per month (yellow; June, orange; July, red; August, and green; September).

6 Discussion

In this thesis the viability of peat fire prediction applying transfer learning on a CNN with the use of Sentinel-2 imagery was explored in the Arctic region of North-East Siberia (**RQ1**). The results show that peat fire prediction is a difficult task in the North-East Siberia, as Janiec and Gadal (2020) have also been pointing out for fire risk assessment. The results indicate that a model trained on one year cannot predict for the other year. This would suggest there is too much fluctuation between the datasets of 2019 and 2020. The same fluctuation is present when only the month May is used or with a different area (A2). Several reasons can cause these differences between the two years. In the first place, wetlands (which peatland is part of) are difficult to observe using remote sensing because of the diversity in vegetation content and the variability it brings at different timespans (Kaplan & Avdan, 2019). Secondly, climate change, as warmer climates cause drier soils and lowers the water table in peatland (Turetsky et al., 2015). Lower water tables may affect peatland vegetation (Breeuwer et al., 2009; Kettridge et al., 2015). This could mean that the differences observed at the datasets from 2019 and 2020 may show vegetation differences between the two years due to climate change effects. Lower water tables can also cause deeper burning which in return, also has consequences for vegetation succession and eco-system function (Turetsky et al., 2015). This indicates the third reason why differences might occur between the datasets, namely that fires are associated with changes in the vegetation (Noble et al., 2018). Due to the large number of fires observed in the two years chosen (2019 and 2020), this might have affected the landscape of the region. Turetsky et al. (2015) mention that peat fires can cause long term environmental changes such as permafrost thawing, which might have been another cause of differences between the year 2019 and 2020. According to Gibson et al. (2018) the effects of wildfires in boreal peatland could last for 30 years and results in a warmer and a deeper active layer. Even low-severity fires in peat covered wetland areas can cause complex processes and feedbacks (Ackley et al., 2021) as they found that the burned area of the region of interest showed earlier snowmelt than the unburned area. The early start of the fires in 2020 in combination with the late fires in September 2019 could mean that these changes are indeed occurring as the late fires might have caused thawing and earlier snowmelt. When observing the GEE images, less snow cover in 2020 can be roughly observed. This assumption of thawing, however, is highly speculative and needs further research as this thesis does not provide enough data to support this.

Transfer learning using the spring months for peat fire prediction (**RQ1.1**) shows a performance between 0.76-0.89 F2 and 0.84-0.90 AUC score for the 2019 datasets. The 2020 datasets have a lower performance and shows in most cases from the confusion matrices that non-fire image patches are difficult to classify correctly. The lower scores in 2020 may be due to the larger number of train data, causing fewer clear distinctions between fire and non-fire image patches and more examples for each class from different coordinates (and therefore small areas) are present. This suggests that if a good distinction between the two classes can be created with the data, what might be the case in 2019, peat fire prediction using transfer learning might have potential. Especially when taking a smaller timespan into account as with the day/two week timespan used in Radke et al. (2019). Since it is a single image (median of the values) taken from images over three months, information on certain areas might be averaged too much to still provide useful information, especially considering the variability peatlands bring over certain timespans. The MAY datasets showed the same problems, suggesting that taking an average over several images might not work well even if the timespan is one month. However, in the case of Maulana et al. (2019) where an average for the NDMI is taken, it showed good results. The 4-monthly average of the NDMI was the top indicator for peat fire prediction in the southern hemisphere. In contrary to this thesis study, other data such as peat physical characteristics was used including peat depth which may act as a key indicator for the susceptibility to high peat burn severity as shallow peat is more susceptible to it (Wilkinson et al., 2020). This is in line with Maulana et al. (2019) concluding that 90% of the fires happened at a peat depth under 100 cm. Other data types were also included in other studies towards wildfires (Radke et al., 2019; G. Zhang et al., 2019). This might incline that for peat fire prediction, the inclusion of other data is necessary, which is also suggested for wildfire assessment in the region by Janiec and Gadal (2020) and Maulana et al. (2019) mention the importance of climate data, including precipitation, as a long period of low rainfall is a risk factor for peat fire (FAO, 2020).

For the most discriminative features to apply for transfer learning on a CNN to predict peat fire (**RQ1.2**), the results show how the feature importance for the Sentinel-2 features (**RQ1.2.1**) which can be found at **Table 1**, differ from the areas (A1 and A2) and from the use of several spring months (MM) and just the month May (MAY). Difference in feature importance was also found by Räsänen et al. (2020) where they compared different remote sensing datasets (that also included NIR and NDWI) to map peatland vegetation and found that different features for different areas were most important for certain regressions and optimal performance depends on the peatland area and its structure. This could explain the difference of importance in the SWIR bands between the two areas. Räsänen et al. (2020) also conclude that a combination of multiple different remote sensing datasets should be used for peatland mapping. This might also be the case with peat fire prediction. However, as mentioned in the results there might be some similarities that can be carefully found with the bands in the infrared spectre (B5-B8A). These are slightly discriminative and could be of importance to predict peat fires. As previously mentioned by Kaplan and Avdan

(2019) the red-edge bands (B5-B7) can help identify the more intensely vegetated wetland classes. Even though the subject of mapping wetland and predicting peat fire might still be far off, it might be worth looking into further how the red edge bands of the Sentinel-2 could help with peat fire prediction.

When answering the sub question of which vegetation and water indices as features are most discriminative for peat fire prediction (**RQ1.2.2**), the lower water table depth mentioned before might indicate why the NDMI and NDWI show importance. As well as the soil moisture as an indicator for the susceptibility of peatland fires (Dadap et al., 2019). Even though Lees et al. (2020) could not find any significant relationship between the water indices (including the NDMI, mentioned as NDWI in the article) and the water table depth or soil moisture. However, in a dry month a decrease in the water table depth, soil moisture, and the NDMI was observed. The authors also found that relationships between water indices and water content is specific to certain species. The NDMI shows importance when looking at the AUC scores and being in line with Maulana et al. (2019) as an indicator for peat fire. The NDWI is the most discriminative feature in the year 2020 (A1-MM-20) and of strong importance in 2019, when considering the F2-score. Reminding that the NDWI is meant for enhancing open water features and eliminating soil and vegetation, this could indicate that the NDWI contains information on peatland areas before fire starts. However, when observing the FIRMS fire points, it becomes clear that several fire points are close to open water. Making it a possibility that the model associates open water with peat fire. This is therefore still a large unknown of why and if the NDWI is of importance and creates possibilities for further research towards water indices and peat fire prediction. The NDVI is a feature used in many of the mentioned related work and is the highest scoring in the year 2019 (A1-MM-19). It is, however, difficult to say how much of an importance the NDVI has, as it is the least important feature when assessing the AUC scores. According to Janiec and Gadal (2020) the NDVI is strongly correlated with fire. This could explain why the F2-scores are higher than the AUC scores for the NDVI, indicating the NDVI is of help classifying the fire class but less for the non-fire class. The related work is in context of forest fire prediction and not peat fires. Further research could show if the NDVI is of importance for peat fire prediction.

In the context of transfer learning, it seems that these differences in feature importance might influence the performance of the model when tested on a different year and suggests that for each region and year a model needs to be trained again when applying transfer learning using a CNN. Additionally, for each region and timespan, recalculations of feature importance scores are necessary, which can then be used to select the features most discriminant for that region and that time to be used with transfer learning.

From a technical view, one of the reasons why transfer learning might not seem to work with the approach taken in this work, is the use of the EuroSAT dataset to pretrain the models. Being in line with Marmanis et al. (2016) that found pretrained networks on ImageNet performed better than models pretrained on PatternNet. The same lack of complexity in the dataset mentioned by the authors, could be a reason for the models not being able to predict peat fire for the other year when depending on the EuroSAT pretrained network models. The performance found of ArcticNet by Jiang et al. (2019) shows the applicability of a model pretrained on ImageNet to use for remote sensing images.

6.1 Limitations

In this thesis the main limitation was the availability of precise data needed to investigate the differences in feature values found between the two years. Mainly to specify where exactly peatland in the ROI could be found and what types of vegetation. As far as is known these precise data are lacking in the northern remote peatlands of this study's ROI and therefore makes it difficult to find out what the differences are caused by.

Time constraints led to the exclusion of BigEarthNet as a dataset to pretrain the models on before training these on the peat fire datasets. This dataset contains many training samples that would suggest that the dataset is more complex and recall that the lack of complexity may have been of influence on the performance. BigEarthNet also includes wetland image patches, suggesting that the dataset is closer to the task at hand (peat fire prediction). The time constraint also led to the choice of using the NDVI, NDWI, and NDMI as water and vegetation indices. There are several other water indices, moisture indices and even other satellite types that can be applied.

It needs to be mentioned that the correlation matrices show high correlation between several bands. As the Permutation Feature Importance method relies on features being independent, the results for feature importance should not be completely dependent on this method and other feature importance methods are suggested to explore. Another limitation to mention is the use of **Equation 2.7** to calculate the feature importance score, as this score is not normalized. Meaning, that when one dataset has a higher performance score (e.g., F2-score) on the original test set than the other dataset but for both datasets the permutated test set creates the same importance score, the importance score seems the same, however when normalized the importance score might differ from each other.

Computational power proposed a limitation mainly on the calculations of the PCA components and the eigenvalues and -vectors corresponding to it. Some datasets were too large to compute and gave an error and therefore had to be made smaller, meaning that not the entire dataset is represented while calculating the values.

7 Conclusions

In this thesis the viability of predicting peat fire in the region of North-East Siberia with the help of transfer learning on a CNN using Sentinel-2 imagery was investigated. Predicting peat fire in the region of North-East Siberia is a novel approach that is still far from being able to be used as an early warning in the form of, for example, a susceptibility map. The usage of the spring months and creating an image of the median values to predict the summer fires in the region, is not a suitable approach to apply in combination with transfer learning on a CNN. Even when using only the month May, the models tested on a different year are random classifiers and can therefore not distinguish between non-fire and fire image patches. This does not exclude transfer learning as an option when different approaches are applied. In the context of transfer learning, it needs further investigation per study area and per feature to see which features are truly of importance and can be used for deep learning purposes. In this thesis it can be very carefully said that the red edge bands (especially B5), SWIR bands and the NDWI and NDMI might be discriminative for peat fire prediction using transfer learning on a CNN. However, differences are visible between different years and areas and the acquired results can only be applied for the areas used in this thesis and cannot suggest any clear discriminative features.

It has been tried in this thesis to give a first step on the viability of peat fire prediction with transfer learning on a CNN and mainly to indicate where to go further. Concluding that much research is still needed. Not only in the field of deep learning and peat fire prediction but including how vegetation changes due to peat fires and climate change and how this affects satellite data. More knowledge on what these differences between years include, is important for going into the direction of being able to know if deep learning is indeed an approach that could work for peat fire prediction.

8 Future work

Peat fire prediction in the Arctic region is a novel approach that still needs much research. Part of this research can consist of three different main focuses: the region, approach, and model choice.

Other Arctic regions such as Alaska, Canada, or North-Scandinavia where peatland coverage is present, can be considered as regions of interest. These regions might have more extended data resources available, including peat physical data, weather data or accurate data on vegetation coverage. The next step would be to explore how differences in data and feature importance can translate to vegetation coverage or other weather data that is constantly in change and how these extra data sources influence the performance of transfer learning models on peat fire prediction. With more accurate additional data available, a more precise reasoning for why differences are present can be constructed and might show how climate change and peat fires influence these changes on the data. It is interesting to see if models can be generalized and how features are of importance in other regions, to see if indeed feature importance differs over completely different areas or that certain similarities can be found.

Applying transfer learning on datasets that have a timespan of a year in between, does not show potential. This, however, could still be the case for shorter timespans, which a small potential was shown for same year predictions. Experiments were done using MI that combined several images of the spring months. This potential could be investigated further by means of using data that is closer to the actual fire-starting date. The approach could entail to average over the values and create a single image from several images again or use a single cloud-free image days or weeks pre-fire. This way it becomes interesting to study how this might be able to be generalized if enough fire-starting data is gathered over a bigger region.

In this thesis the image patches in the datasets did not entail the entire area. Interesting would be investigating how the performances are when a (smaller) area is divided into patches to make sure every part of the area is included into the training set. This way, changes for each smaller patch may be detected better as the patch will always include the same area. This also provides the opportunity to create a susceptibility map more easily as it needs a considered less amount of computing power as less image patches are used as when a sliding window includes an image patch for each pixel. Another approach by Rosadi et al. (2020) where each fire class image has a counterpart of a non-fire image taken from another time when no fire occurred, is also worth looking into.

The use of other data such as, weather data, field measurements or peat physical data, for example the peat layer depth used in Maulana et al. (2019) might need to be added to the dataset to make sure the differences and changes of vegetation, snow cover, and maybe even other factors, are also known to the network model that is trained. This also brings up another approach that might include the use of (satellite) data that is independent on vegetation changes or weather. The Sentinel-1, for example, is a Synthetic Aperture Radar (SAR) C-band satellite developed by ESA that is able to acquire images independently from what the weather is at that time (ESA, n.d.-b). Another option is the Soil Moisture Active Passive (SMAP) L-band satellite from NASA, that is able to penetrate clouds and forest cover offering the opportunity to measure soil moisture which therefore can be a useful resource for research towards peat fires (Dadap et al., 2019).

The choice of features is important as well. Other water, vegetation, and moisture indices can be researched towards peat fire prediction and its importance to performance of the model. This also includes the further research towards the importance of the indices used in this thesis and the connection with differences of timespan, area, and vegetation.

The model choice is another field of study. In the first place it is worth the further research towards different base and top models. In this thesis the focus was on ResNet50, but other studies mentioned in this work use other backbone structures. Several articles in the literature for transfer learning made use of a top model such as a SVM, which would give the option of a different learnable machine learning model on top of the CNN. Different base and top models with different backbones and configurations are all possible and might indicate the need for a review towards the best use cases for peat fire prediction in the Arctic region. Further on, investigating the potential of transfer learning using pretrained networks on more complex remote sensing datasets, such as BigEarthNet, could answer the necessity of pretraining on a dataset closer to the primary task of peat fire prediction or that pretrained networks on natural images (ImageNet) show similar results for transfer learning predicting peat fires. Additionally, completely other model choices can be investigated. In this thesis a CNN was chosen. However, other model options are available that can include, for example, the temporal dimension, think of a Long Short-Term Memory (LSTM) model (Hochreiter & Schmidhuber, 1997). A Convolutional LSTM (ConvLSTM) model such as the one by Shi et al. (2015) could offer the advantages of a CNN and LSTM for peat fire prediction. Further research towards how different models perform for peat fire prediction is appreciated.

9 References

Ackley, C., Tank, S. E., Haynes, K. M., Rezanezhad, F., McCarter, C., & Quinton, W. L. (2021). Coupled hydrological and geochemical impacts of wildfire in peatland-dominated regions of discontinuous permafrost. *Science of The Total Environment*, 782, Article 146841. https://doi.org/10.1016/j.scitotenv.2021.146841

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Hasan, M., Van Essen, B. C., Awwal, A. A. S., & Asari, V. K. (2019). A State-of-the-Art Survey on Deep Learning Theory and Architectures. *Electronics*, 8(3), Article 292. https://doi.org/10.3390/electronics8030292

Ban, Y., Zhang, P., Nascetti, A., Bevington, A. R., & Wulder, M. A. (2020). Near Real-Time Wildfire Progression Monitoring with Sentinel-1 SAR Time Series and Deep Learning. *Scientific Reports*, 10(1), Article 1322. https://doi.org/10.1038/s41598-019-56967-x

Breeuwer, A., Robroek, B. J. M., Limpens, J., Heijmans, M. M. P. D., Schouten, M. G. C., & Berendse, F. (2009). Decreased summer water table depth affects peatland vegetation. *Basic and Applied Ecology*, 10(4), 330–339. https://doi.org/10.1016/j.baae.2008.05.005

Buchhorn, M., Lesiv, M., Tsendbazar, N.-E., Herold, M., Bertels, L., & Smets, B. (2020). Copernicus Global Land Cover Layers—Collection 2. *Remote Sensing*, 12(6), Article 1044. https://doi.org/10.3390/rs12061044

Cheng, G., Han, J., & Lu, X. (2017). Remote Sensing Image Scene Classification: Benchmark and State of the Art. *Proceedings of the IEEE*, 105(10), 1865–1883. https://doi.org/10.1109/JPROC.2017.2675998

Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 1800–1807. https://doi.org/10.1109/CVPR.2017.195

Dadap, N. C., Cobb, A. R., Hoyt, A. M., Harvey, C. F., & Konings, A. G. (2019). Satellite soil moisture observations predict burned area in Southeast Asian peatlands. *Environmental Research Letters*, 14(9), Article 094014. https://doi.org/10.1088/1748-9326/ab3891

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. 2009 *IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 120, 25–36. https://doi.org/10.1016/j.rse.2011.11.026

European Space Agency. (n.d.-a). *Radiometric Resolutions*. Retrieved 23 August 2021, from https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric

European Space Agency. (n.d.-b). *Sentinel-1*. Retrieved 20 September 2021, from https://sentinel.esa.int/web/sentinel/missions/sentinel-1

European Space Agency. (n.d.-c). *Spatial Resolution.* Retrieved 11 August 2021, from https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial

FAO. (2020). *Peatland mapping and monitoring: Recommendations and technical overview*. Rome. https://doi.org/10.4060/ca8200en

Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

Gibson, C. M., Chasmer, L. E., Thompson, D. K., Quinton, W. L., Flannigan, M. D., & Olefeldt, D. (2018). Wildfire as a major driver of recent permafrost thaw in boreal peatlands. *Nature Communications*, 9(1). https://doi.org/10.1038/s41467-018-05457-1

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. https://www.deeplearningbook.org/

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 770–778. https://doi.org/10.1109/CVPR.2016.90

Helber, P., Bischke, B., Dengel, A., & Borth, D. (2019). 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. https://doi.org/10.1109/JSTARS.2019.2918242

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 2261–2269. https://doi.org/10.1109/CVPR.2017.243

Hugelius, G., Loisel, J., Chadburn, S., Jackson, R. B., Jones, M., MacDonald, G., Marushchak, M., Olefeldt, D., Packalen, M., Siewert, M. B., Treat, C., Turetsky, M., Voigt, C., & Yu, Z. (2020). Large stocks of peatland carbon and nitrogen are vulnerable to permafrost thaw. *Proceedings of the National Academy of Sciences*, 117(34), 20438–20446. https://doi.org/10.1073/pnas.1916387117

Janiec, P., & Gadal, S. (2020). A Comparison of Two Machine Learning Classification Methods for Remote Sensing Predictive Modeling of the Forest Fire in the North-Eastern Siberia. *Remote Sensing*, 12(24), Article 4157. https://doi.org/10.3390/rs12244157

Jiang, Z., Von Ness, K., Loisel, J., & Wang, Z. (2019). ArcticNet: A Deep Learning Solution to Classify Arctic Wetlands. *ArXiv Preprint ArXiv*:1906.00133. http://arxiv.org/abs/1906.00133

Kaplan, G., & Avdan, U. (2019). Evaluating the utilization of the red edge and radar bands from sentinel sensors for wetland classification. *CATENA*, 178, 109–119. https://doi.org/10.1016/j.catena.2019.03.011

Kettridge, N., Turetsky, M. R., Sherwood, J. H., Thompson, D. K., Miller, C. A., Benscoter, B. W., Flannigan, M. D., Wotton, B. M., & Waddington, J. M. (2015). Moderate drop in water table increases peatland vulnerability to post-fire regime shift. *Scientific Reports*, 5(1), Article 8063. https://doi.org/10.1038/srep08063

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. https://doi.org/10.1145/3065386

Langford, Z., Kumar, J., & Hoffman, F. (2018). Wildfire Mapping in Interior Alaska Using Deep Neural Networks on Imbalanced Datasets. *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 770–778. https://doi.org/10.1109/ICDMW.2018.00116

Lees, K. J., Artz, R. R. E., Khomik, M., Clark, J. M., Ritson, J., Hancock, M. H., Cowie, N. R., & Quaife, T. (2020). Using Spectral Indices to Estimate Water Content and GPP in Sphagnum Moss and Other Peatland Vegetation. *IEEE Transactions on Geoscience and Remote Sensing*, 58(7), 4547–4557. https://doi.org/10.1109/TGRS.2019.2961479

Lees, K. J., Quaife, T., Artz, R. R. E., Khomik, M., & Clark, J. M. (2018). Potential for using remote sensing to estimate carbon fluxes across northern peatlands – A review. *Science of The Total Environment*, 615, 857–874. https://doi.org/10.1016/j.scitotenv.2017.09.103

Leifeld, J., & Menichetti, L. (2018). The underappreciated potential of peatlands in global climate change mitigation strategies. *Nature Communications*, 9(1), Article 1071. https://doi.org/10.1038/s41467-018-03406-6

Lozano, F. J., Suárez-Seoane, S., & de Luis, E. (2007). Assessment of several spectral indices derived from multi-temporal Landsat data for fire occurrence probability modelling. *Remote Sensing of Environment*, 107(4), 533–544. https://doi.org/10.1016/j.rse.2006.10.001

Marmanis, D., Datcu, M., Esch, T., & Stilla, U. (2016). Deep Learning Earth Observation Classification Using ImageNet Pretrained Networks. *IEEE Geoscience and Remote Sensing Letters*, 13(1), 105–109. https://doi.org/10.1109/LGRS.2015.2499239

Maulana, S. I., Syaufina, L., Prasetyo, L. B., & Aidi, M. N. (2019). Spatial logistic regression models for predicting peatland fire in Bengkalis Regency, Indonesia. *Journal of Sustainability Science and Management*, 14(3), 55–66.

McFeeters, S. K. (1996). 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. https://doi.org/10.1080/01431169608948714

Molnar, C. (2019). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. https://christophm.github.io/interpretable-ml-book/

Nelson, K., Thompson, D., Hopkinson, C., Petrone, R., & Chasmer, L. (2021). Peatland-fire interactions: A review of wildland fire feedbacks and interactions in Canadian boreal peatlands. *Science of The Total Environment*, 769, Article 145212. https://doi.org/10.1016/j.scitotenv.2021.145212

Noble, A., Palmer, S. M., Glaves, D. J., Crowle, A., Brown, L. E., & Holden, J. (2018). Prescribed burning, atmospheric pollution and grazing effects on peatland vegetation composition. *Journal of Applied Ecology*, 55(2), 559–569. https://doi.org/10.1111/1365-2664.12994

Ohlemiller, T. J. (1985). Modeling of smoldering combustion propagation. *Progress in Energy and Combustion Science*, 11(4), 277–310. https://doi.org/10.1016/0360-1285(85)90004-8

Parish, F., Sirin, A., Charman, D., Joosten, H., Minayeva, T., Silvius, M., & Stringer, L. (Eds.). (2008). *Assessment on Peatlands, Biodiversity and Climate Change: Main Report*. Global Environment

Centre, Kuala Lumpur and Wetlands International, Wageningen. http://www.imcg.net/media/download_gallery/books/assessment_peatland.pdf

Peters, M. E., Ruder, S., & Smith, N. A. (2019). To Tune or Not to Tune? Adapting Pretrained Representations to Diverse Tasks. *ArXiv:1903.05987 [Cs]*. http://arxiv.org/abs/1903.05987

Petrovska, B., Atanasova-Pacemska, T., Corizzo, R., Mignone, P., Lameski, P., & Zdravevski, E. (2020). Aerial Scene Classification through Fine-Tuning with Adaptive Learning Rates and Label Smoothing. *Applied Sciences*, *10*(17), Article 5792. https://doi.org/10.3390/app10175792

Pires de Lima, R., & Marfurt, K. (2020). Convolutional Neural Network for Remote-Sensing Scene Classification: Transfer Learning Analysis. *Remote Sensing*, *12*(1), Article 86. https://doi.org/10.3390/rs12010086

Radke, D., Hessler, A., & Ellsworth, D. (2019). FireCast: Leveraging Deep Learning to Predict Wildfire Spread. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, IJCAI-19, 4575–4581. https://doi.org/10.24963/ijcai.2019/636

Rani, M., Kumar, P., Yadav, M., & Hooda, R. S. (2011). Wetland Assessment and Monitoring Using Image Processing Techniques: A Case Study of Ranchi, India. *Journal of Geographic Information System*, *3*(4), 345–350. https://doi.org/10.4236/jgis.2011.34032

Räsänen, A., Aurela, M., Juutinen, S., Kumpula, T., Lohila, A., Penttilä, T., & Virtanen, T. (2020). Detecting northern peatland vegetation patterns at ultra-high spatial resolution. *Remote Sensing in Ecology and Conservation*, *6*(4), 457–471. https://doi.org/10.1002/rse2.140

Rein, G. (2016). Smoldering Combustion. In M. J. Hurley, D. Gottuk, J. R. Hall, K. Harada, E. Kuligowski, M. Puchovsky, J. Torero, J. M. Watts, & C. Wieczorek (Eds.), *SFPE Handbook of Fire Protection Engineering* (pp. 581–603). Springer. https://doi.org/10.1007/978-1-4939-2565-0_19

Rein, G., Cleaver, N., Ashton, C., Pironi, P., & Torero, J. L. (2008). The severity of smouldering peat fires and damage to the forest soil. *CATENA*, *74*(3), 304–309. https://doi.org/10.1016/j.catena.2008.05.008

Rosadi, D., Andriyani, W., Arisanty, D., & Agustina, D. (2020). Prediction of Forest Fire Occurrence in Peatlands using Machine Learning Approaches. *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 48–51. https://doi.org/10.1109/ISRITI51436.2020.9315359

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. In *Proceedings of the third ERTS Symposium* (Vol. 1, pp. 309–317). US Government Printing Office, NASA.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, *115*(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y

Schroeder, W., Oliva, P., Giglio, L., & Csiszar, I. A. (2014). The New VIIRS 375m active fire detection data product: Algorithm description and initial assessment. *Remote Sensing of Environment*, *143*, 85–96. https://doi.org/10.1016/j.rse.2013.12.008

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. (2014). OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. *ArXiv:1312.6229 [Cs]*. http://arxiv.org/abs/1312.6229

Sherstyukov, B. G., & Sherstyukov, A. B. (2014). Assessment of increase in forest fire risk in Russia till the late 21st century based on scenario experiments with fifth-generation climate models. *Russian Meteorology and Hydrology*, *39*(5), 292–301. https://doi.org/10.3103/S1068373914050021

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W., & Woo, W. (2015). Convolutional LSTM Network: A machine learning approach for precipitation nowcasting. *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, 802–810.

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv:1409.1556 [Cs]*. http://arxiv.org/abs/1409.1556

Sirin, A. A., Medvedeva, M. A., Makarov, D. A., Maslov, A. A., & Joosten, H. (2020). Multispectral satellite based monitoring of land cover change and associated fire reduction after large-scale peatland rewetting following the 2010 peat fires in Moscow Region (Russia). *Ecological Engineering*, *158*, Article 106044. https://doi.org/10.1016/j.ecoleng.2020.106044

Sumbul, G., Charfuelan, M., Demir, B., & Markl, V. (2019). Bigearthnet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding. *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 5901–5904. https://doi.org/10.1109/IGARSS.2019.8900532

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 2818–2826. https://doi.org/10.1109/CVPR.2016.308

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

Turetsky, M. R., Amiro, B. D., Bosch, E., & Bhatti, J. S. (2004). Historical burn area in western Canadian peatlands and its relationship to fire weather indices. *Global Biogeochemical Cycles*, *18*(4). https://doi.org/10.1029/2004GB002222

Turetsky, M. R., Benscoter, B., Page, S., Rein, G., van der Werf, G. R., & Watts, A. (2015). Global vulnerability of peatlands to fire and carbon loss. *Nature Geoscience*, *8*(1), 11–14. https://doi.org/10.1038/ngeo2325

Turetsky, M. R., Kane, E. S., Harden, J. W., Ottmar, R. D., Manies, K. L., Hoy, E., & Kasischke, E. S. (2011). Recent acceleration of biomass burning and carbon losses in Alaskan forests and peatlands. *Nature Geoscience*, *4*(1), 27–31. https://doi.org/10.1038/ngeo1027

Vompersky, S. E., Sirin, A. A., Sal'nikov, A. A., Tsyganova, O. P., & Valyaeva, N. A. (2011). Estimation of forest cover extent over peatlands and paludified shallow-peat lands in Russia. *Contemporary Problems of Ecology*, *4*(7), 734–741. https://doi.org/10.1134/S1995425511070058

Wilkinson, S. L., Tekatch, A. M., Markle, C. E., Moore, P. A., & Waddington, J. M. (2020). Shallow peat is most vulnerable to high peat burn severity during wildfire. *Environmental Research Letters*, *15*(10), Article 104032. https://doi.org/10.1088/1748-9326/aba7e8

Witze, A. (2020). The Arctic is burning like never before—And that's bad news for climate change. *Nature*, *585*, 336–337. https://doi.org/10.1038/d41586-020-02568-y

Xia, G.-S., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., Zhang, L., & Lu, X. (2017). AID: A Benchmark Data Set for Performance Evaluation of Aerial Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7), 3965–3981. https://doi.org/10.1109/TGRS.2017.2685945

Yang, Y., & Newsam, S. (2010). Bag-of-visual-words and spatial extensions for land-use classification. *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 270–279. https://doi.org/10.1145/1869790.1869829

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How Transferable Are Features in Deep Neural Networks? *Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2, 3320–3328.

Yu, Z. C. (2012). Northern peatland carbon stocks and dynamics: A review. *Biogeosciences*, 9(10), 4071–4085. https://doi.org/10.5194/bg-9-4071-2012

Yu, Z., Loisel, J., Brosseau, D. P., Beilman, D. W., & Hunt, S. J. (2010). Global peatland dynamics since the Last Glacial Maximum. *Geophysical Research Letters*, 37(13). https://doi.org/10.1029/2010GL043584

Zhang, C., Pan, X., Li, H., Gardiner, A., Sargent, I., Hare, J., & Atkinson, P. M. (2018). A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 140, 133–144. https://doi.org/10.1016/j.isprsjprs.2017.07.014

Zhang, G., Wang, M., & Liu, K. (2019). Forest Fire Susceptibility Modeling Using a Convolutional Neural Network for Yunnan Province of China. *International Journal of Disaster Risk Science*, 10(3), 386–403. https://doi.org/10.1007/s13753-019-00233-1

Zhang, L., Zhang, L., & Du, B. (2016). Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2), 22–40. https://doi.org/10.1109/MGRS.2016.2540798

Zhou, W., Newsam, S., Li, C., & Shao, Z. (2018). PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145, 197–209. https://doi.org/10.1016/j.isprsjprs.2018.01.004

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8–36. https://doi.org/10.1109/MGRS.2017.2762307

We acknowledge the use of data and/or imagery from NASA's *Fire Information for Resource Management System* (FIRMS) (*https://earthdata.nasa.gov/firms*), *part of* NASA's *Earth Observing System Data and Information System* (EOSDIS).

Appendix A: Confusion Matrices

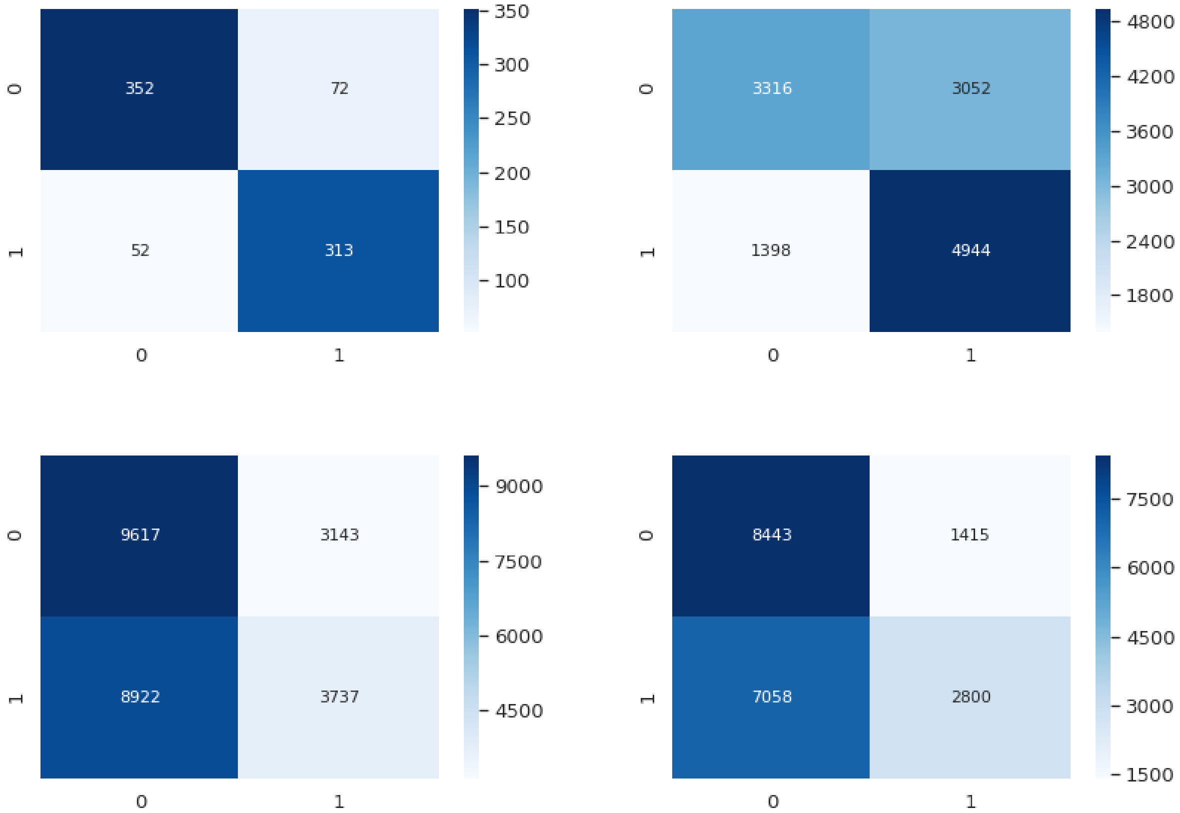


Figure A-1. Confusion matrices of same year test set A1-MM-19 (top left), same year test A1-MM-20 (top right), other year test A1-MM-19 (bottom left), and other year test A1-MM-20 (bottom right).

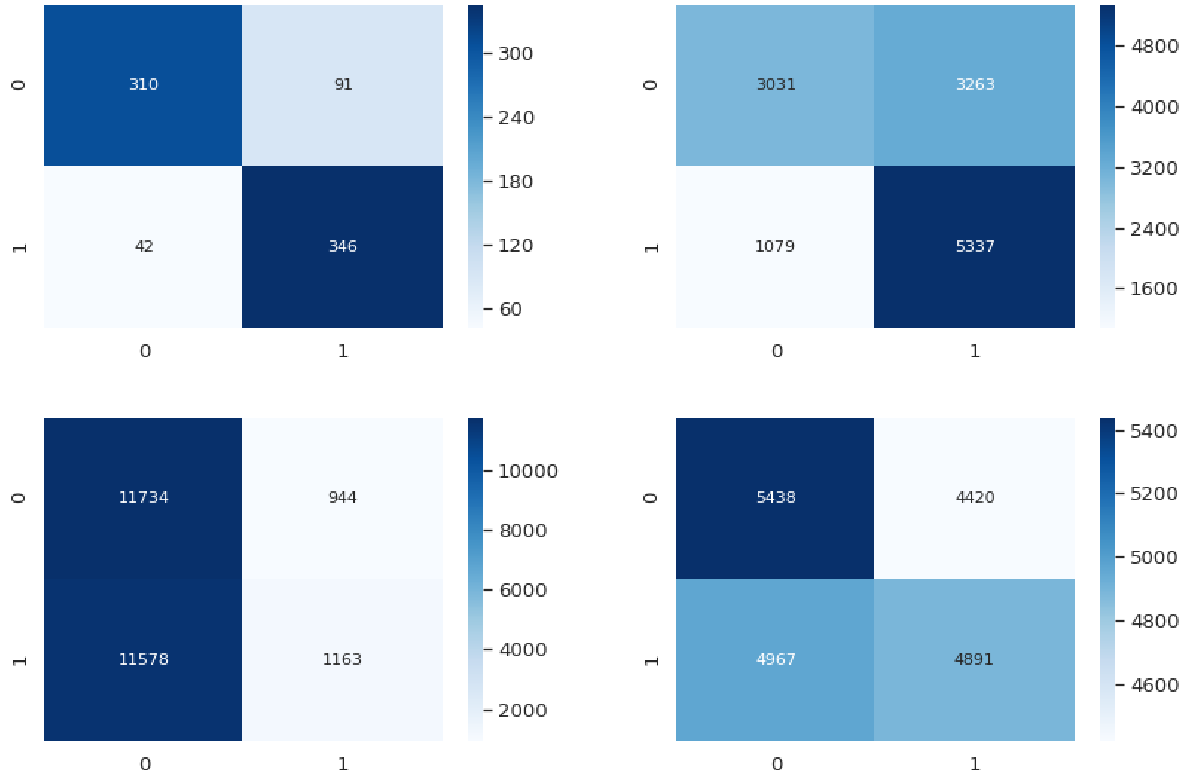


Figure A-2. Confusion matrices of same year test set A1-MAY-19 (top left), same year test A1-MAY-20 (top right), other year test A1-MAY-19 (bottom left), and other year test A1-MAY-20 (bottom right).

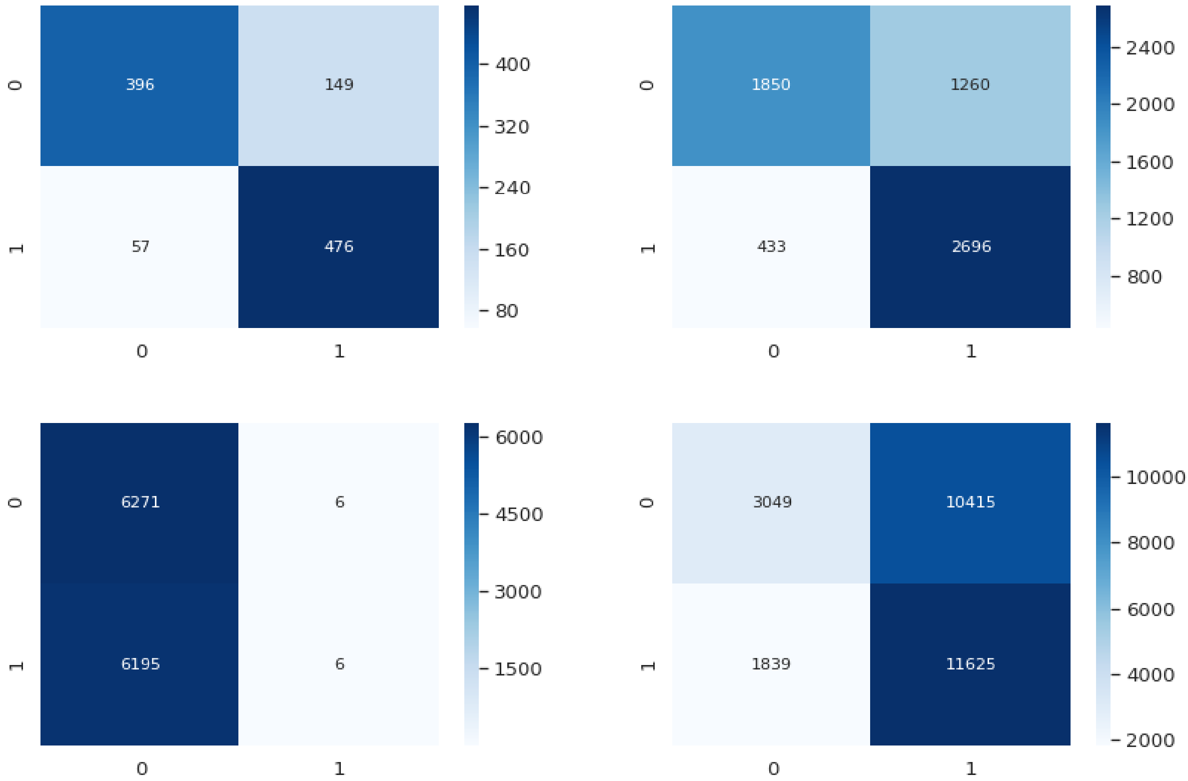


Figure A-3. Confusion matrices of same year test set A2-MAY-19 (top left), same year test A2-MAY-20 (top right), other year test A2-MAY-19 (bottom left), and other year test A2-MAY-20 (bottom right).

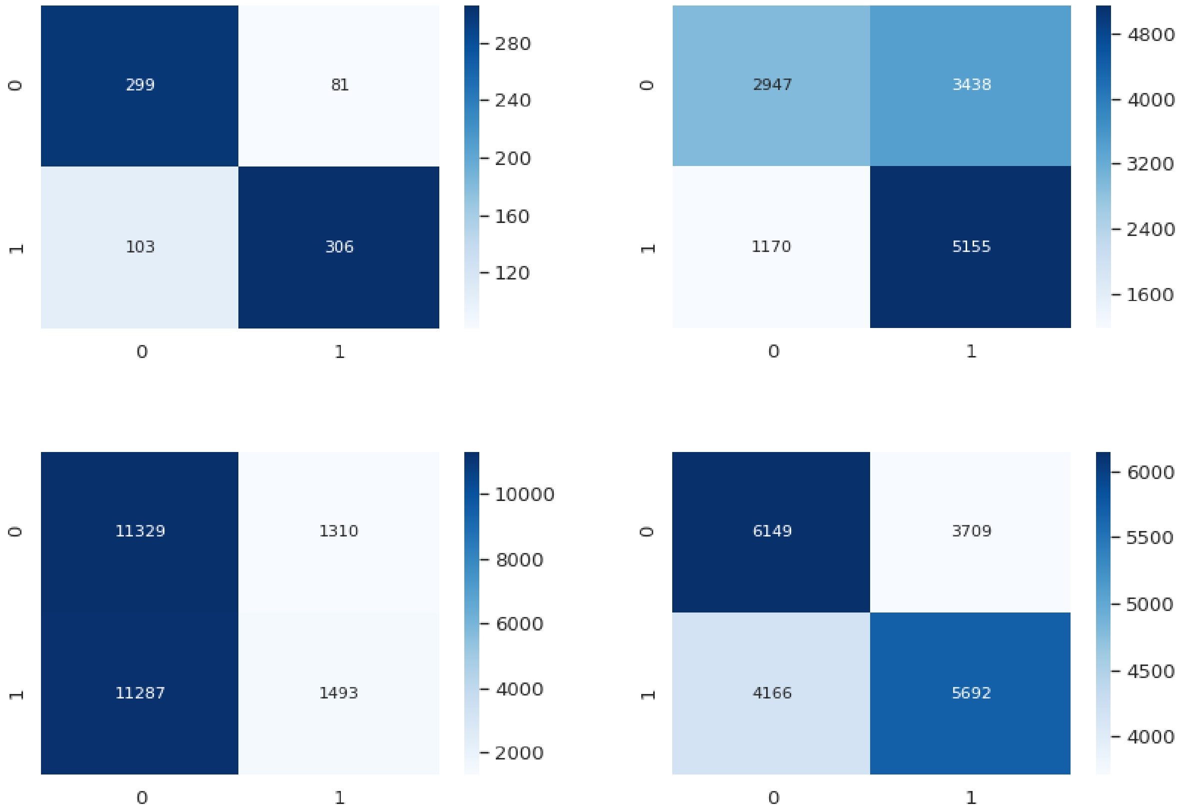


Figure A-4. Confusion matrices of same year test set A1-MM-ND-19 (top left), same year test A1-MM-ND-20 (top right), other year test A1-MM-ND-19 (bottom left), and other year test A1-MM-ND-20 (bottom right).

Appendix B: Correlation Matrices

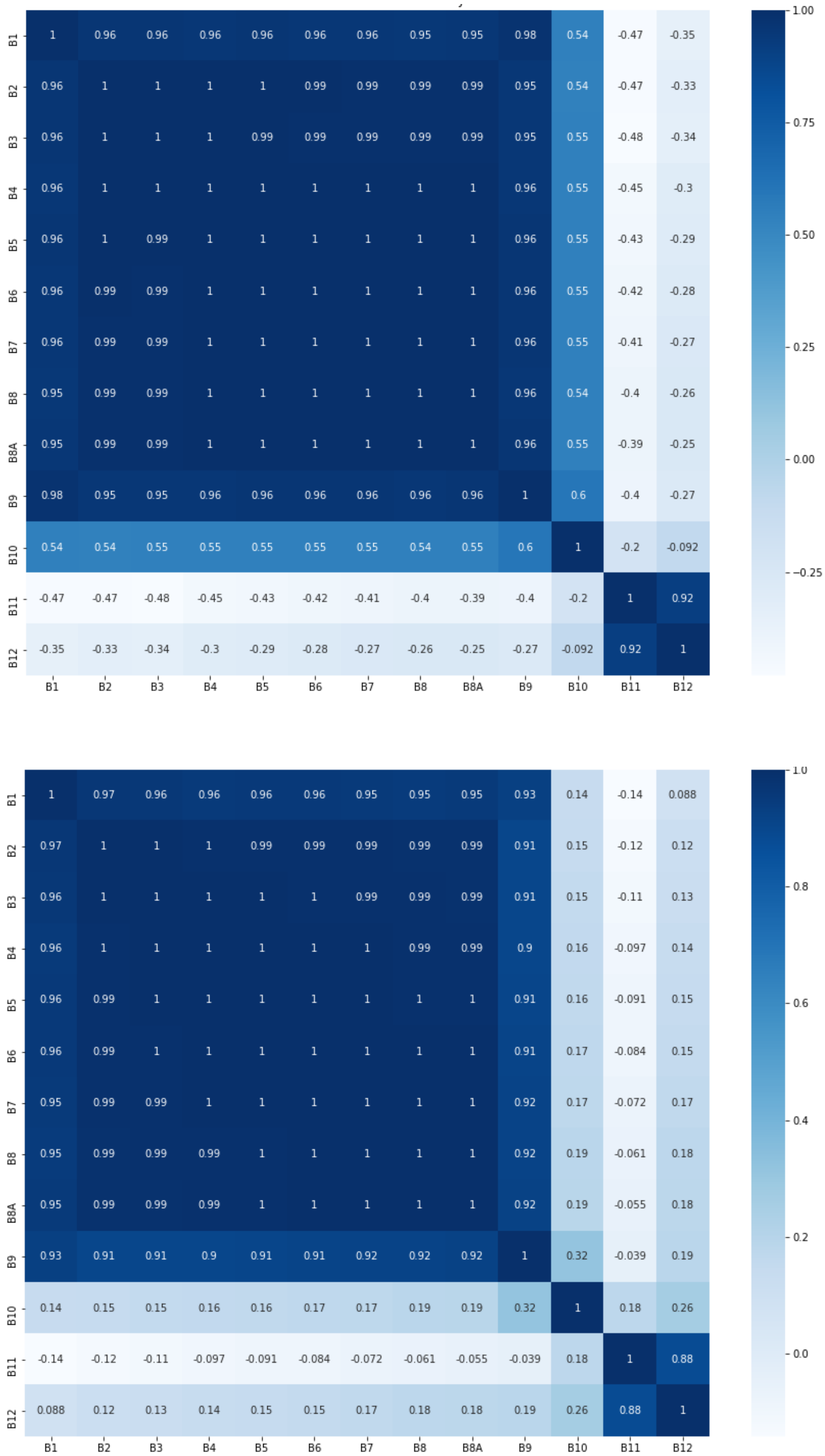


Figure B-1. Correlation matrix of the dataset A1-MM-19 (top) and A1-MM-20 (bottom) for each band.

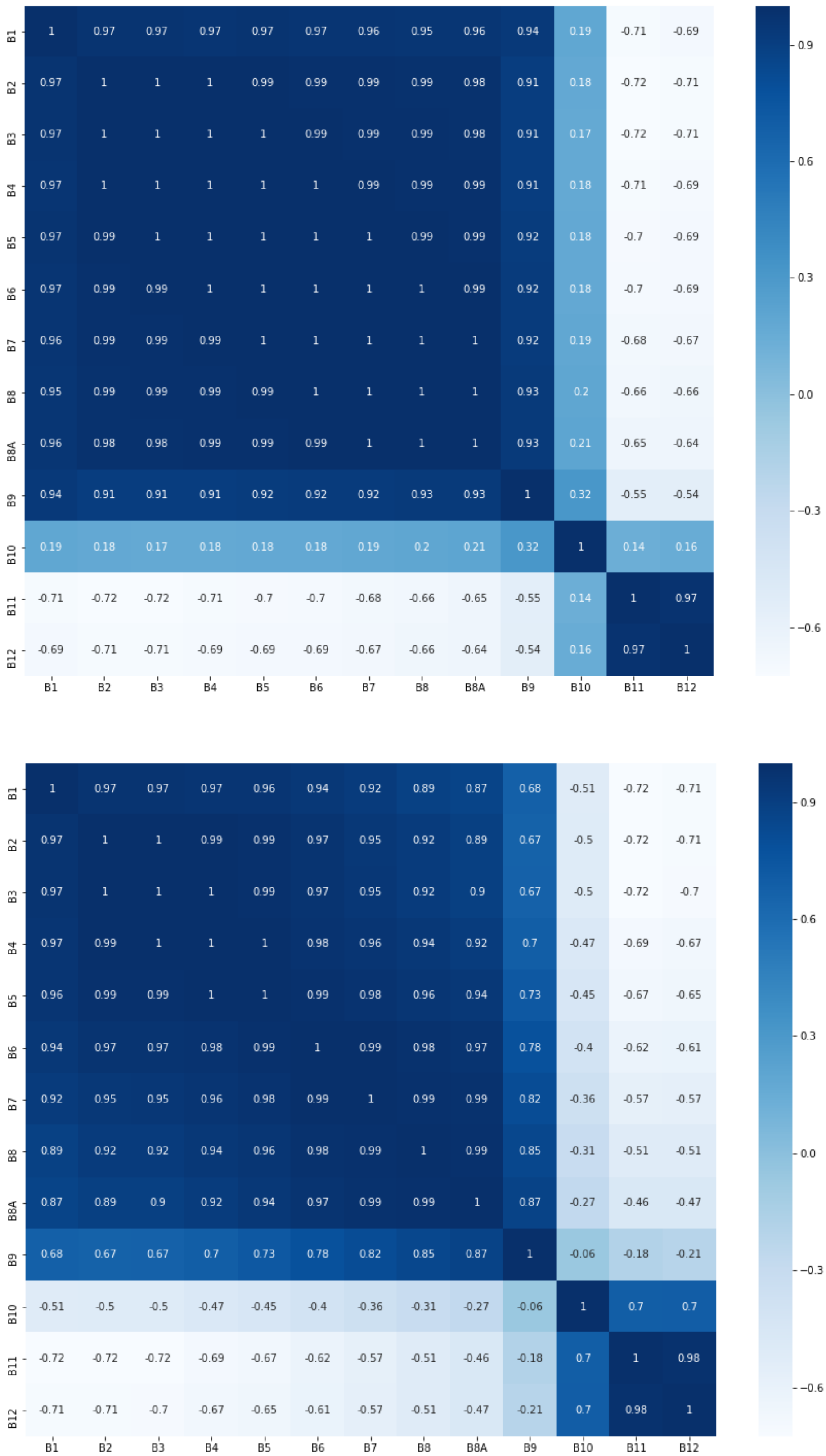
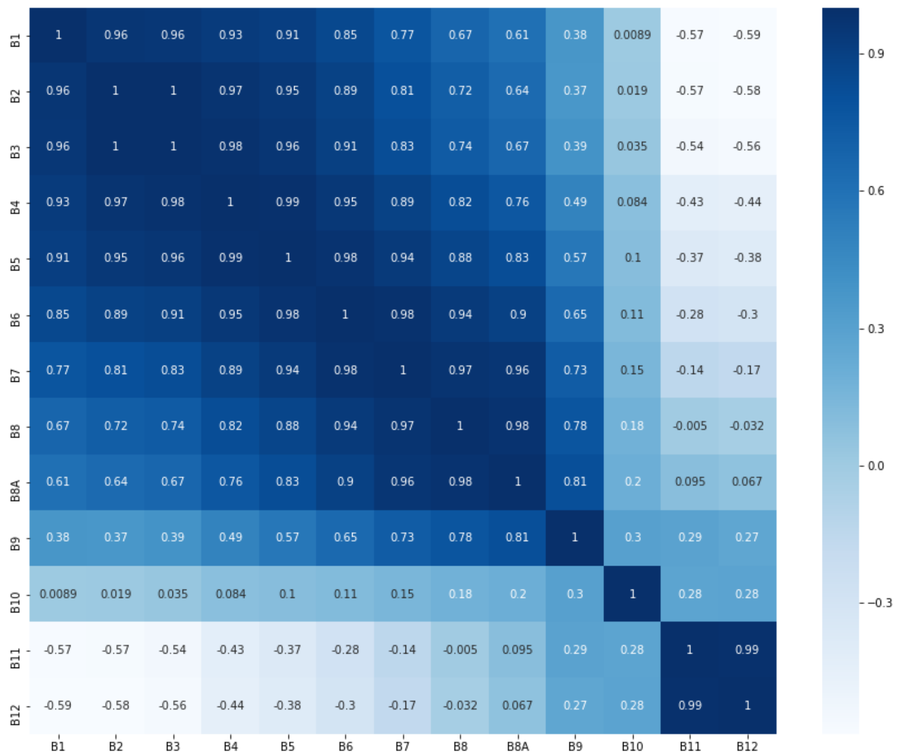


Figure B-2. Correlation matrix of the dataset A1-MAY-19 (top) and A1-MAY-20 (bottom) for each band.



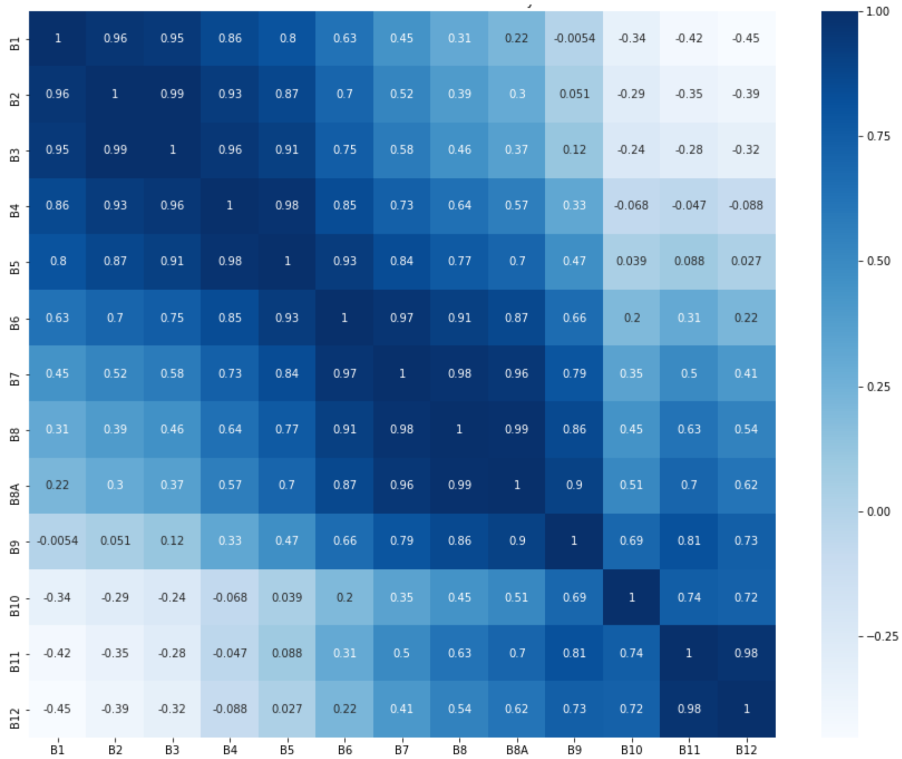


Figure B-3. Correlation matrix of the dataset A2-MAY-19 (top) and A2-MAY-20 (bottom) for each band.

Appendix C: PCA

Table C-1. Principal components 1 and 2 of the datasets with the eigenvector values for all bands.

	BANDS												
	B1	B2	B3	B4	B5	B6	B7	B8	B8A	B9	B10	B11	B12
A1-MM-19 (PC1)	**-0.30**	**-0.31**	**-0.31**	**-0.31**	**-0.31**	**-0.31**	**-0.31**	**-0.31**	**-0.31**	**-0.30**	-0.18	0.15	0.11
A1-MM-19 (PC2)	0.00	0.02	0.01	0.04	0.06	0.07	0.08	0.08	0.09	0.07	0.13	**0.66**	**0.72**
A1-MM-20 (PC1)	**0.30**	**0.31**	**0.31**	**0.31**	**0.31**	**0.31**	**0.32**	**0.32**	**0.32**	**0.31**	0.10	0.02	0.10
A1-MM-20 (PC2)	-0.09	-0.07	-0.05	-0.04	-0.03	-0.03	-0.02	-0.01	-0.01	0.01	0.25	**0.69**	**0.66**
A1-MAY-19 (PC1)	**0.30**	**0.30**	**0.30**	**0.30**	**0.30**	**0.30**	**0.30**	**0.30**	**0.30**	**0.28**	0.05	-0.23	-0.23
A1-MAY-19 (PC2)	0.04	0.02	0.02	0.03	0.04	0.04	0.06	0.07	0.08	0.20	**0.71**	**0.46**	**0.48**
A1-MAY-20 (PC1)	**0.30**	**0.31**	**0.30**	**0.31**	**0.31**	**0.32**	**0.32**	**0.32**	**0.31**	0.23	-0.09	-0.20	-0.19
A1-MAY-20 (PC2)	-0.06	-0.05	-0.05	-0.01	0.02	0.07	0.12	0.17	0.21	**0.41**	**0.59**	**0.46**	**0.43**
A2-MAY-19 (PC1)	**0.31**	**0.32**	**0.32**	**0.33**	**0.34**	**0.34**	**0.32**	**0.30**	**0.29**	0.21	0.05	-0.13	-0.13
A2-MAY-19 (PC2)	-0.17	-0.16	-0.14	-0.06	-0.01	0.06	0.15	0.23	0.28	**0.38**	0.28	**0.52**	**0.52**
A2-MAY-20 (PC1)	0.23	**0.26**	**0.28**	**0.32**	**0.35**	**0.36**	**0.35**	**0.34**	**0.32**	**0.26**	0.10	0.13	0.11
A2-MAY-20 (PC2)	**0.34**	**0.32**	**0.30**	0.20	0.14	0.02	-0.09	-0.16	-0.20	**-0.30**	**-0.36**	**-0.42**	**-0.41**

Appendix D: Land Cover Type tables

Table D-1. Number of fire points per month (June-September) and per land cover class with the total percentage for each class for the years 2019 and 2020 for region A1.

	2019					2020				
	June	*July*	*August*	*September*	**Total (%)**	*June*	*July*	*August*	*September*	**Total (%)**
Shrubs (S)	0	0	0	0	**0,0**	5	9	5	0	**0,0**
Herbaceous vegetation (HV)	89	5	131	63	**11,6**	1653	3582	895	4	**15,5**
Permanent water bodies (PWB)	36	14	51	14	**4,6**	196	567	285	1	**2,6**
Herbaceous wetland (HW)	120	22	116	60	**12,8**	1941	2900	798	3	**14,2**
Closed forest, deciduous needle leaf (CF:NL)	0	0	5	0	**0,2**	27	157	60	1	**0,6**
Closed forest, not matching any of the other definitions (CF:Other)	534	59	594	108	**52,2**	4354	11900	4264	15	**51,8**
Open forest, deciduous needle leaf (OF:NL)	0	0	1	0	**0,0**	0	1	2	0	**0,0**
Open forest, not matching any of the other definitions (OF:Other)	168	23	213	56	**18,5**	1383	3483	1173	11	**15,2**

Table D-2. Number of fire points per month (June-September) and per land cover class with the total percentage for each class for the years 2019 and 2020 for region A2.

	2019					2020				
	June	*July*	*August*	*September*	**Total (%)**	*June*	*July*	*August*	*September*	**Total (%)**
Shrubs (S)	2	1	0	0	**0,1**	2	32	2	0	**0,2**
Herbaceous vegetation (HV)	171	7	0	0	**5,3**	115	2710	298	0	**15,9**
Permanent water bodies (PWB)	141	1	0	0	**4,2**	6	252	28	0	**1,5**
Herbaceous wetland (HW)	86	7	0	0	**2,7**	45	1057	72	0	**6,0**
Closed forest, deciduous needle leaf (CF:NL)	533	4	0	0	**15,9**	298	6580	460	0	**37,4**
Closed forest, not matching any of the other definitions (CF:Other)	2135	4	1	0	**63,2**	78	4687	477	0	**26,7**
Open forest, deciduous needle leaf (OF:NL)	6	0	0	0	**0,2**	7	99	6	0	**0,6**
Open forest, not matching any of the other definitions (OF:Other)	277	8	0	0	**8,4**	92	2086	156	0	**11,9**

Appendix E: Land Cover Type per Month

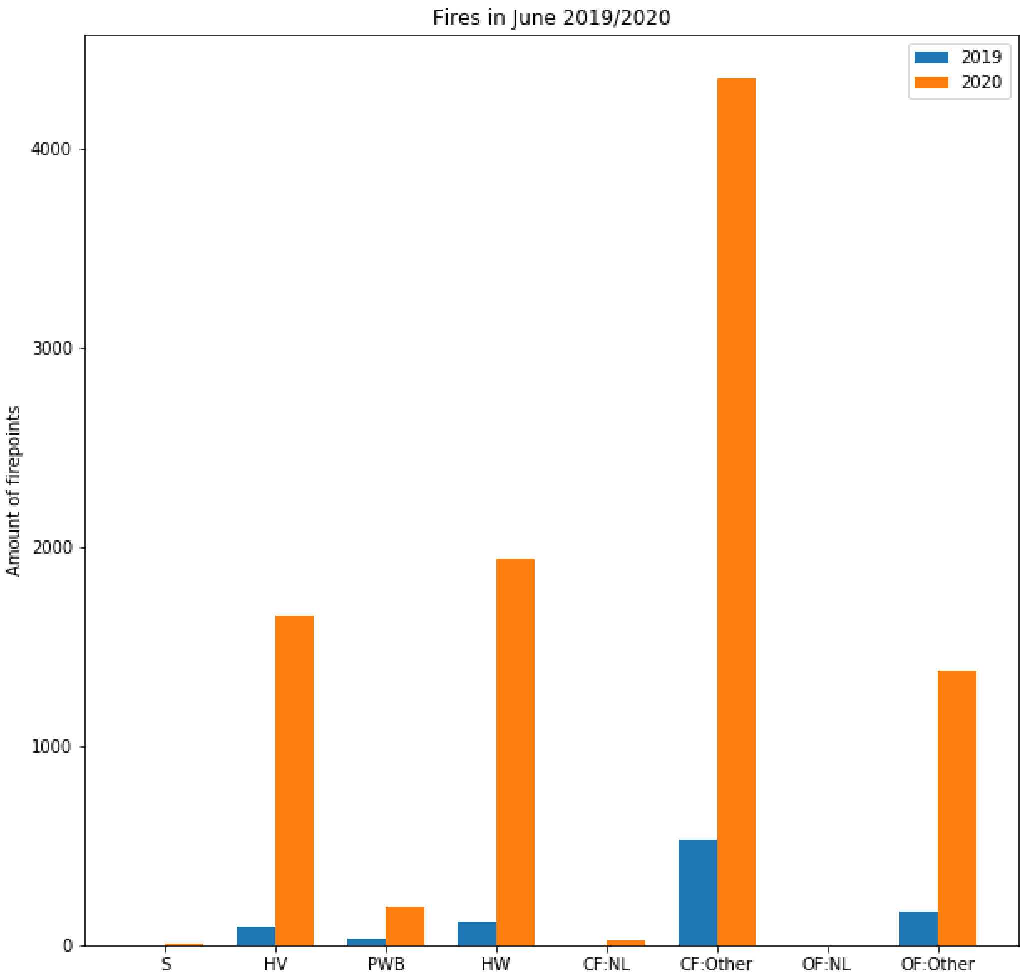


Figure E-1. Number of fire points per land cover class in the month June for 2019 and 2020 in region A1.

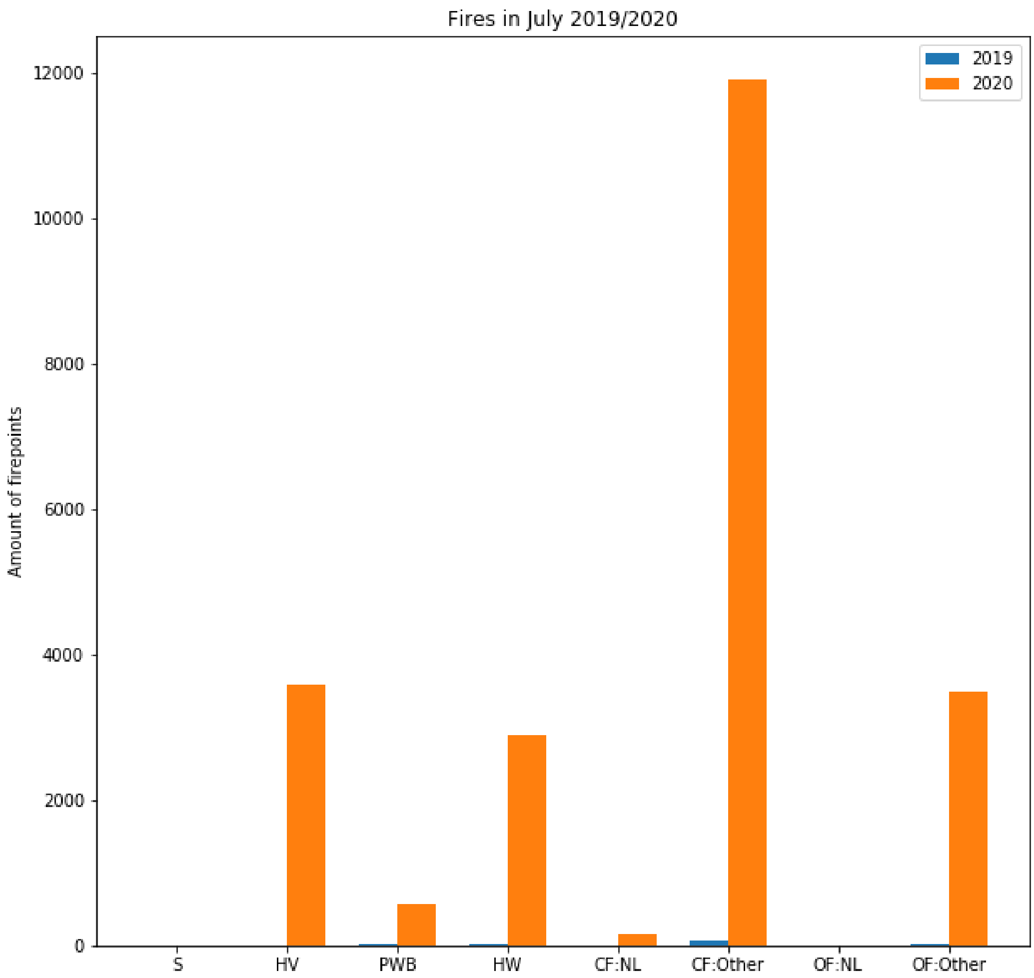


Figure E-2. Number of fire points per land cover class in the month July for 2019 and 2020 in region A1.

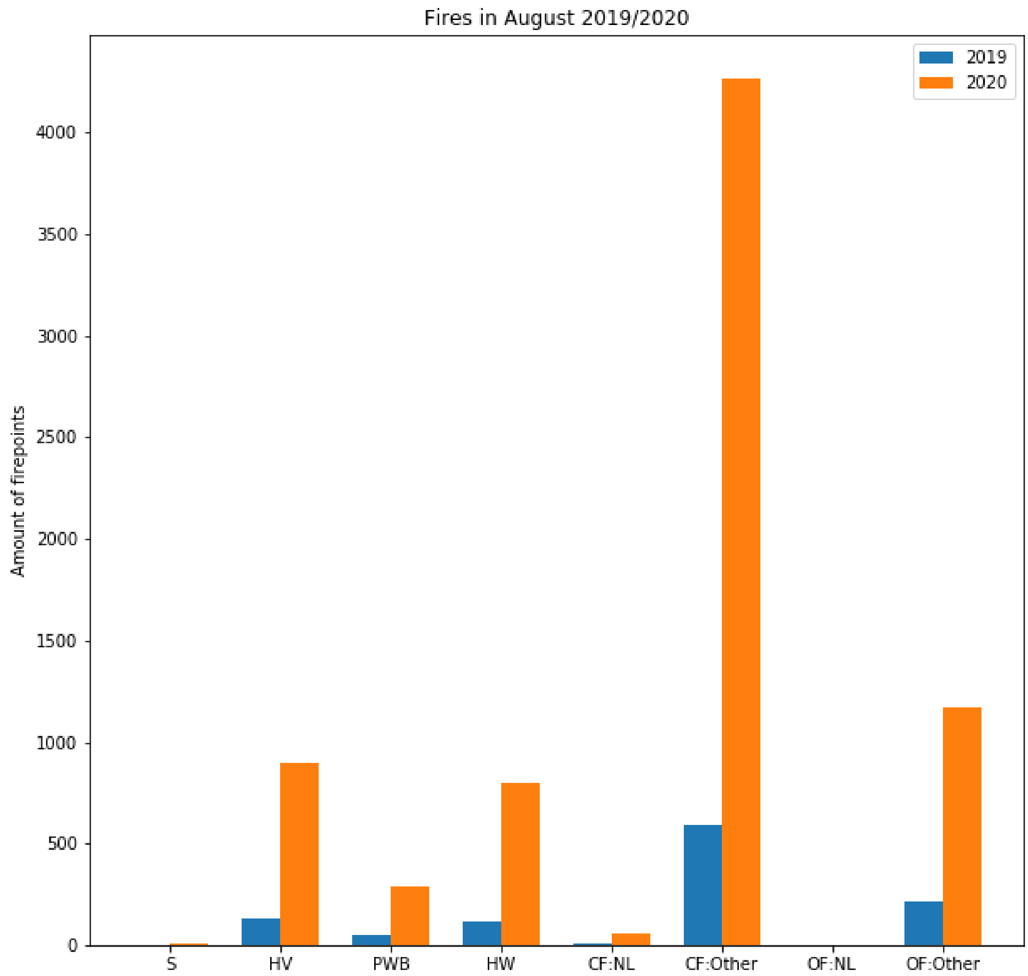


Figure E-3. Number of fire points per land cover class in the month August for 2019 and 2020 in region A1.

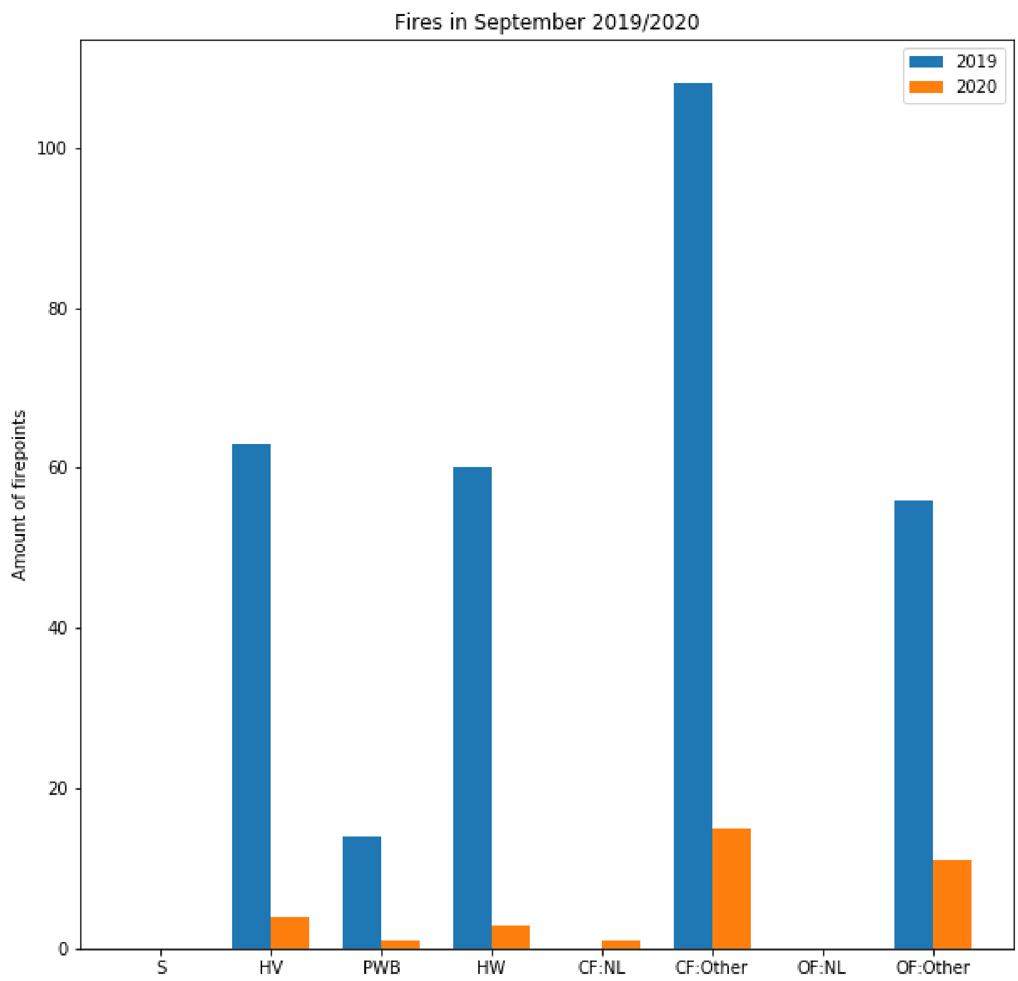


Figure E-4. Number of fire points per land cover class in the month September for 2019 and 2020 in region A1.

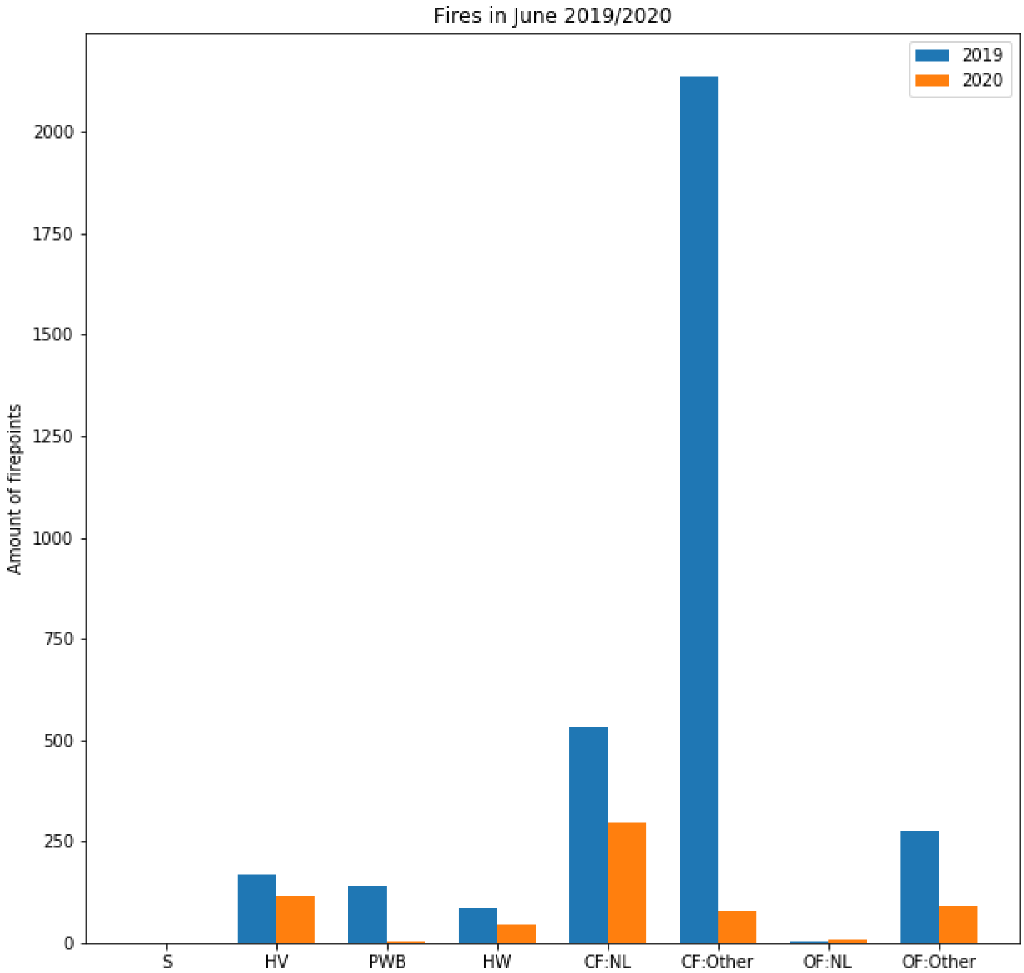


Figure E-5. Number of fire points per land cover class in the month June for 2019 and 2020 in region A2.

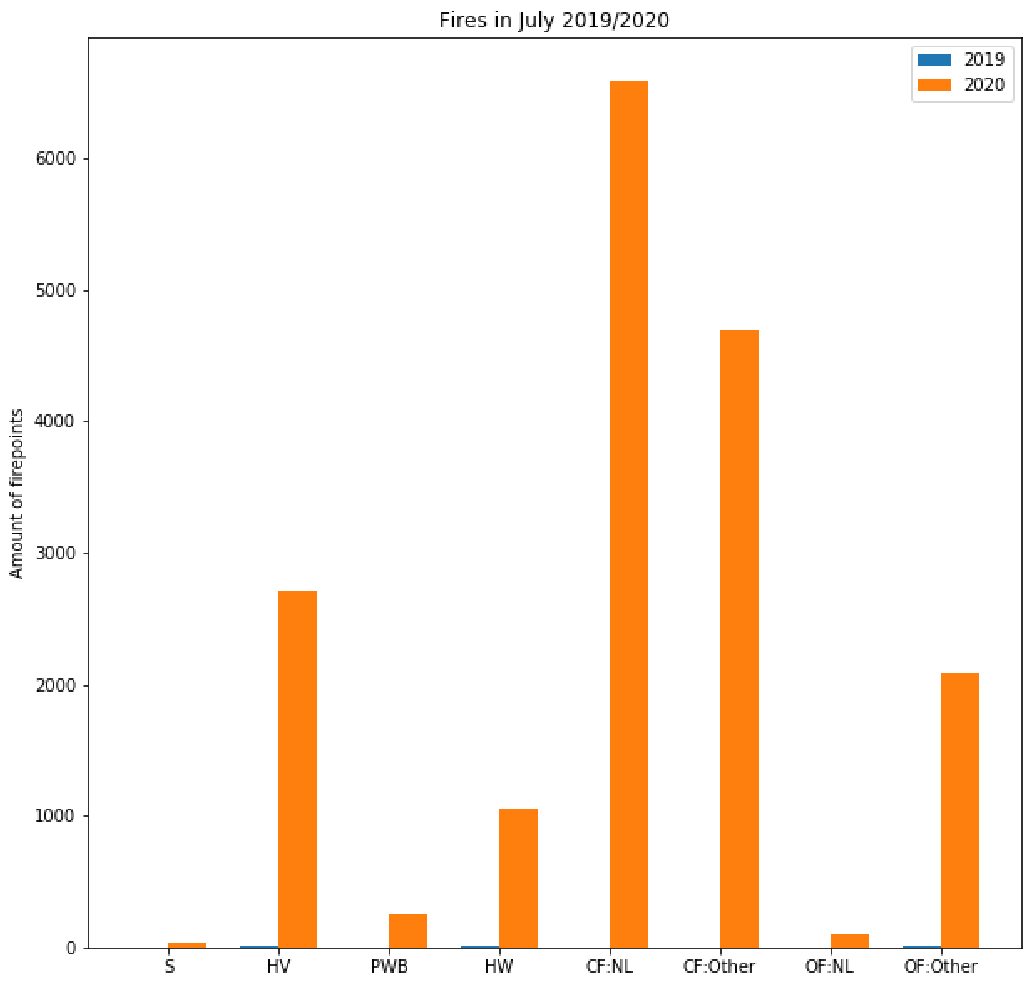


Figure E-6. Number of fire points per land cover class in the month July for 2019 and 2020 in region A2.

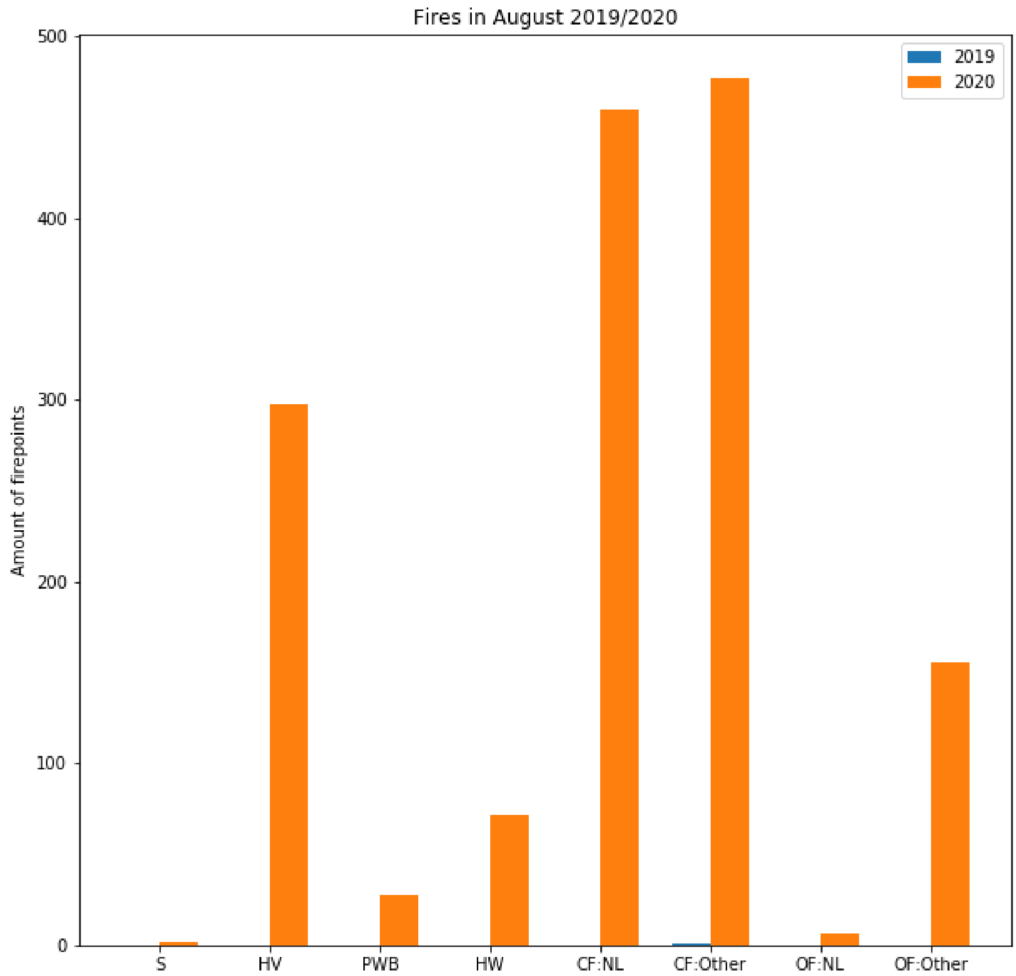


Figure E-7. Number of fire points per land cover class in the month August for 2019 and 2020 in region A2.