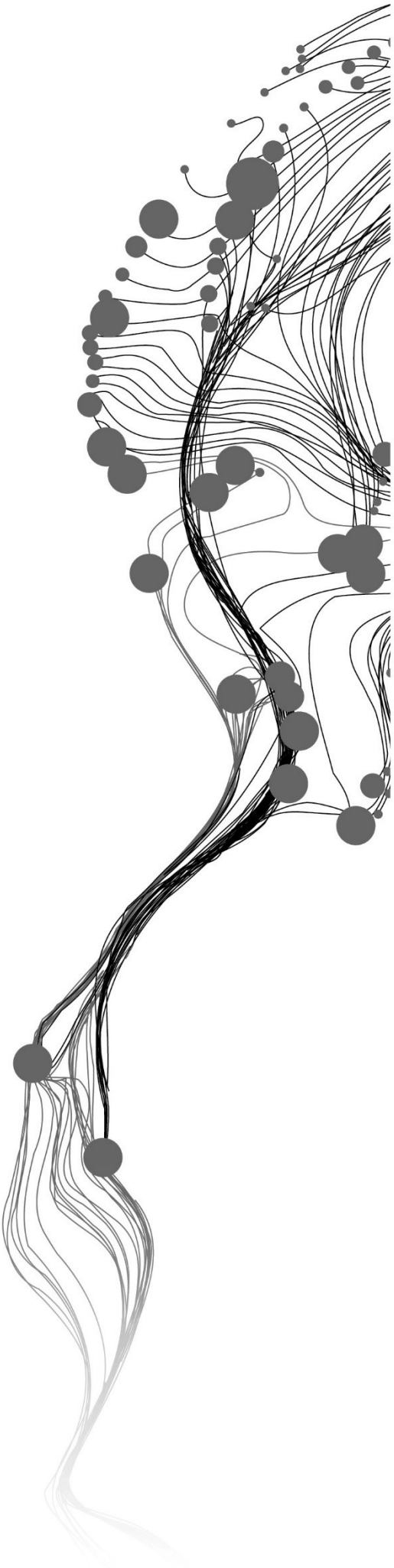
CONVOLUTIONAL NETWORKS FOR THE CLASSIFICATION OF MULTI-TEMPORAL SATELLITE IMAGES

RATNA MAYASARI
February 2019

SUPERVISORS:
dr. C. Persello
dr. M. Belgiu



CONVOLUTIONAL NETWORKS FOR THE CLASSIFICATION OF MULTI-TEMPORAL SATELLITE IMAGES

RATNA MAYASARI
Enschede, The Netherlands, February 2019

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialisation: Geoinformatics

SUPERVISORS:
dr. C. Persello
dr. M. Belgiu

THESIS ASSESSMENT BOARD:
prof. dr. ir. A. Stein (Chair)
dr. D. Tiede (External Examiner, University of Salzburg)

ABSTRACT

Satellite images have been widely used to produce classification maps which are further used for various applications. Nowadays, many satellite missions have been launched and provide images with a high spatial, spectral and temporal resolution. Many studies have been conducted to investigate the methods that are capable of utilising all the available information simultaneously especially for classifying objects that spectrally changes throughout the time, i.e., crops. Multi-temporal satellite images (MTSI) provides additional information in the temporal domain to discriminate crops classes. We study the neural network approach, especially fully convolutional network (FCN) architecture to produce accurate land cover maps of agricultural areas by using MTSI. We design and investigate the use of FCN architecture by adopting the dilated convolution layer (FCN-SNet architecture) and concatenating network (FCN-SubNet Architecture). We apply these networks to Sentinel-2 images where the two study areas are located, Romania and California. We perform several experiments for selecting the appropriate hyper-parameter values for the FCN. In addition, we identify several errors in the reference data which caused the accuracy of the classification results is relatively low. Therefore, we make a refinement for the datasets to improve the classification result. Based on the results, FCN-SNet as the proposed technique outperforms Support Vector Machine (SVM), Dynamic Time Warping (DTW), and FCN-SubNet approach. It also offers a more efficient computation.

Keywords: Fully Convolutional Network (FCN), multi-temporal satellite images, classification

ACKNOWLEDGEMENTS

Praises and thanks to Allah for giving me blessing, opportunity, strength and good health to go through and complete my study and research in the Faculty of ITC in Enschede.

My special gratitude and thanks to my supervisors, dr. C. Persello and dr. M. Belgiu, for the time, technical and non-technical advice, discussion and continuous support in the successful completion of this research. I learn a lot from both of you.

My deep gratitude to prof. dr. ir. A. Stein for his critical feedback to my research.

I would like to thank drs. J.P.G. Bakx (course director of GFM) and dr. D. Tiede (External Examiner, University of Salzburg) for their insightful feedback.

I thank the Ministry of Research, Technology and Higher Education, and Geospatial Information Agency of Indonesia for giving me the opportunity and financial support to study. Especially for dr. W. Ambarwulan, ir. I. Herliningsih, m.si, dr. A. K. Mulyana, deceased ir. E Hendrayana, and my colleagues in BIG for encouraging and supporting me to pursue this MSc.

My sincere thanks to Ratna Sari Dewi, Aji Putra Perdana, and Aldino Rizaldy for the discussion and valuable advice to my research. Many thanks to Yibo Zhou for allowing me to use his DTW script codes for my research. I also would like to thank the Indonesian student community in Enschede, my fellows in ITC, my Geoinformatics (GFM) classmates and ITC staff for sharing the experience and providing support in the academic and non-academic matter during my study.

There are also a large number of people who are not possible to mention them all here and giving me valuable support for this research. I would like to extend my sincere thanks to all of them.

Finally, my utmost gratitude and love for my parents and my family who are always understanding and supporting me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Motivation and Problem Statement

Land cover classification (LCC) is a fundamental part of geospatial information's provision commonly displayed on a map. Geospatial information can be used for many applications, e.g., agricultural production, urban planning, land development, land cover and crops monitoring. In line with the global goals that have been set by the United Nations, sustainable agricultural production supports the achievement of the Sustainable Development Goals (SDGs) number 2, "end hunger, achieve food security and improved nutrition and promote sustainable agriculture" (United Nations, 2015).

Monitoring in the agricultural sector is needed because agricultural fields are influenced by climate change more than any other sectors. Providing information about crop types through time by considering the phenology is one of the activities in crops monitoring. Phenology describes the vegetation cycle according to a natural seasonal growth pattern that is useful for distinguishing the type of vegetation (Rußwurm & Körner, 2017).

Remote sensing data, i.e., aerial photo or satellite images, have been used as a primary source to generate the LCC map. Satellite images become a suitable choice of data sources for large monitoring areas. Furthermore, current satellite missions provide a huge volume of images with a short revisit time and various bands which are useful in giving spatial and spectral information for mapping LCC. The visible, near-infrared, and middle infrared channels are commonly used for vegetation detection purposes (EUMeTrain, 2010; Xue & Su, 2017). Sentinel-2 also provides these commonly used channels. The Sentinel-2 mission has two satellites, Sentinel-2A and Sentinel-2B, and provides 13 channels of images with five days revisit time (European Space Agency, 2018b). It provides multi-temporal satellite images (MTSI), a collection of satellite images acquired at different times over the same location. MTSI is useful for both mapping and monitoring purposes by providing information over a period of time.

In the quick development of the computation technology, automation of the LCC mapping becomes an imperious need. It helps to optimise the time factor for data analysis and brings a possibility to use large datasets as input. These advantages overcome the problem of manual interpretation method that requires more human intervention. Various supervised algorithms are used to perform the automatic land cover classification, i.e., non-parametric classification algorithms such as a neural network (Jensen, 2015). Neural network (NN) algorithm has many advantages including no assumption of data distribution, ability to learn from examples and to model non-linear and complex data, can generalise a model and predict data, and can automatically extract information by generating intermediate features. Furthermore, Gómez, White, & Wulder (2016) state that for a large area with unknown data distribution, non-parametric classifier, i.e., NN, is proven as being more capable than the parametric classifier.

Deep Learning (DL) is different from normal NN because it uses hidden layers. These hidden layers construct an architecture that does the computation to learn the information from the data hierarchically and gradually (Lecun, Bengio, & Hinton, 2015). Kamilaris & Prenafeta-Boldú (2018) provide a review of the DL application in agriculture, including the used methods. From their review, Convolution Neural Network (CNN) frequently appears as the used technique in agricultural applications, for example in crop type classification, crop detection and plant recognition. In the component of methods comparison, CNN is superior to the other approaches, i.e., Support Vector Machine (SVM), Artificial Neural Network (ANN), and Random Forest (RF) in most of the studies.

CNN does not apply a pixel-wise classification because it is designed for image recognition aimed to predict the label of the whole input image, not for every pixel in the input image (Guo et al., 2018). Fully

Convolutional Network (FCN) applies the end-to-end pixel-wise classification by predicting the label of every pixel in the input image. FCN is applied for classification based on CNN architecture by replacing the fully connected layer with a convolution layer (Guo et al., 2018; Shelhamer, Long, & Darrell, 2017). FCN has successfully applied for various purposes by using different datasets, e.g., lidar point clouds (Rizaldy, Persello, Gevaert, & Oude Elberink, 2018), synthetic aperture radar (SAR) images (Gao, Zhang, & Xue, 2017; Li et al., 2018), aerial images (Bergado, Persello, & Stein, 2018; Persello & Stein, 2017; Yang et al., 2018), mono-temporal satellite images (Bittner, Cui, & Reinartz, 2017; Maggiori, Tarabalka, Charpiat, & Alliez, 2016) and DTM extraction (Gevaert, Persello, Nex, & Vosselman, 2018).

When we use MTSI as a data source in the mapping process, developing an approach that fully incorporates the temporal dimension for mapping remains a potential research area and indicates the primary problem for operational mapping (Gómez et al., 2016). Additionally, we also need to tackle the classification problem with low inter-class spectral variability from MTSI that produces confusion among the targeted class (Kamilaris & Prenafeta-Boldú, 2018; Rußwurm & Körner, 2017).

1.2. Research Identification

The availability of multi-temporal data brings opportunities and challenges in deriving LCC map. Although multi-temporal data can be useful for capturing the phenology of particular crops, it also potentially brings higher intra-class spectral variability because the observations are repeated in the same location (with same objects) in a different time (Landgrebe, 1978). Along with it, the spatial information is also an important part to determine the classes by recognising the spatial appearance of the objects, such as shape, size, and pattern. Approaches that incorporate spatial, spectral and temporal data are needed to meet the various needs of information from the targeted classes in the LCC map (Gómez et al., 2016). LCC map combined with other thematic layers as the base map provides information that supports various applications such as urban planning, land management, and agricultural monitoring. As mentioned earlier, FCN offers an end-to-end pixel-wised classification that is useful in the mapping process that targeted LCC map as an output.

Addressing limitation in the high computational cost of CNN models during the testing period, Persello & Stein (2017) propose to use FCN for detecting informal settlements by using remote sensing image in a single time acquisition. The authors conclude that FCN performs better than patch-based CNN. The authors observe that FCN has an advantage in applying classification for any size of input images that can be different from the size of training patches. FCN has a less computational time than patch-based CNN because it removes the process of splitting and re-joining input image to fit the patch size. Other studies show the advantages of FCN over CNN for building detection (Maggiori et al., 2016). While, Fu, Liu, Zhou, Sun, & Zhang (2017) use FCN to classify land cover types, and Guo et al. (2018) use FCN to distinguish car and tree from other classes.

Therefore, in this thesis we investigate and design FCN to classify land cover in agricultural areas using MTSI. The proposed FCN network learns and extracts discriminative features automatically from a dataset that contains spatial, spectral and temporal information. A comparison with other approaches, i.e., SVM and Dynamic Time Warping (DTW) is necessary to measure the performance of the proposed approach. SVM is well known as a traditional approach for classification (Bruzzone & Persello, 2009; Hsu, Chang, & Lin, 2003; Rußwurm & Körner, 2017). DTW for remote sensing data is introduced by Petitjean, Inglada, & Gançarski (2012) for addressing particular problems raised when classifying MTSI such as difficulties in providing up to date reference data, unequal temporal spacing of input images, and irregular behaviour of targeted objects from a time perspective for instance due to the weather condition.

1.2.1. Research Objectives

The general objective of this research is to investigate a network that exploits spatial, spectral and temporal information simultaneously from MTSI and produces the LCC map that provides information about the crops. The following sub-objectives support the aforementioned general objective:

1. To design a network for crops classification using MTSI
2. To implement and investigate the performance of the proposed network for crops classification
3. To compare the performance of the proposed network with other classification methods

1.2.2. Research Questions

Each of the sub-objectives can be achieved by answering the following questions:

Questions for sub-objective 1:

a. What are the existing NN approaches that have been applied for crops classification using MTSI?
b. What is the most suitable design for crops classification using MTSI that exploits spatial, spectral and temporal information simultaneously?

Questions for sub-objective 2:

a. What is the suitable structure of an input file for performing classification using the proposed network?
b. What are the optimal hyper-parameters values for the proposed network to be used for performing crops classification using MTSI?
c. How significant are the contributions of the spatial, spectral and temporal information for the classification result?
d. What is the relevant assessment and evaluation to measure the performance of the proposed network?

Questions for sub-objective 3:

a. Which method performs better based on the performance assessment?
b. What aspect of the method contributes to the classification result?

1.3. Innovation

This research investigates the use of FCN by adopting the dilated convolution layer (FCN-SNet architecture) and concatenating network (FCN-SubNet Architecture) to be able to extract spatial, spectral and temporal information from MTSI automatically and simultaneously. The extraction is performed by utilising Sentinel-2 images that have spectral information through the time and applying convolution operations that continuously learn the spatial information.

A network that implements the FCN approach to produce the LCC map as output by incorporating the available spatial, spectral and temporal information from the MTSI in an end-to-end manner is a breakthrough. This approach is expected to overcome drawbacks of the methods and the classification itself, e.g., computational time, utilising spatial, spectral and temporal simultaneously, distinguishing crops that have low intra-class variability.

1.4. Thesis Structure

Structure of this thesis includes the following chapters:

1) Introduction: introduces the background and aims of the research.
2) Literature Review: provides an overview of the related research and a brief overview of the methodology.
3) Methods: explain the used methodology for the research in detail.
4) Datasets: describe the list of datasets and the processing for preparing input of the experiments.

5) Experiments Setting: provides information about the conducted experiments to answer the research questions mentioned in Chapter 1.
6) Result and Discussion: present findings and results of the experiments and provide a discussion related to the results.
7) Conclusion: concludes the research according to the result and discussion. This chapter also provides answers to the research questions.

2. LITERATURE REVIEW

2.1. Related Work on Crops Classification using MTSI

Rußwurm & Körner (2017) use the Long-Short Term Memory (LSTM) model to classify the crop vegetation using MTSI by considering the phenology. LSTM is a variant of Recurrent Neural Network (RNN) that uses loop connection for analysing sequential data. LSTM is initially designed for speech recognition and achieves better accuracy compared to SVM with mono-temporal images as input. The authors successfully classify Landsat and Sentinel-2 for crop vegetation. However, some classes, such as meadow and fallow, cannot be distinguished precisely. Hybrid vegetation, such as triticale (a hybrid of wheat and rye), is also difficult to differentiate because it shares the spectral and temporal features with the wheat and rye crops.

Crops classification with MTSI shows a better result than classification with a mono-temporal satellite image. Although we need to address some challenges such as the availability of training samples, providing a complete series of cloud-free image, and annual changes of a cultivated area caused by weather or agricultural practice variation (Belgiu & Csillik, 2018). Pointing out the input for the classification, the authors successfully use the Normalised Difference Vegetation Index (NDVI) from Sentinel-2 for classifying crops. The authors apply the Time-Weighted DTW method and recommend some further works on how to reduce computational time and to use more spectral channels to classify crop vegetation.

Mou, Bruzzone, & Zhu (2018) use Recurrent CNN that combines CNN and RNN to learn spectral, spatial and temporal features for change detection. The authors classify binary classes (change and unchanged region) and multiple classes (unchanged region, city expansion, soil change, and water change). Recurrent CNN with LSTM model performs better than a combination of CNN and Fully Connected RNN or Gated Recurrent Unit. Recurrent CNN-LSTM reduces the noisy scattered results of wrongly detected classes when using RNN solely.

Ji, Zhang, Xu, Shi, & Duan (2018) experiment the CNN to classify crops by using multi-temporal data of Gaofen satellite images. The authors introduce the use of three-dimensional convolution to utilise the temporal information from MTSI. This approach increases the classification accuracy especially for the crops that have similar spectral value representation in almost every time. The authors also point out the use of an active learning strategy to refine the training dataset by adding a more random sample in each iteration of CNN.

Choosing the classification algorithm needs multiple considerations, such as the type of data, target accuracy, and class distribution to make a balance between the optimal use of the resource and the acceptable accuracy. There are different strategies to perform classification for a specific application. Comparing various studies using deep learning techniques on agricultural and food production has been conducted by Kamilaris & Prenafeta-Boldú (2018). The authors provide a comprehensive review and summarise it based on some common criteria, e.g., what type of data, what architecture of deep learning, how well is the performance, how to apply the methods, and what problems are needed to be addressed. The authors mention some popular deep learning architectures. Each of them has different advantages and makes the architecture suitable for a specific problem. The authors summarise the advantages of deep learning, i.e., a faster method in term of the testing period compared to the traditional approach, e.g., SVM, RF, and ANN; performing automatic feature extraction and better generalisation of classification compared to the other approaches that need to extract feature manually. Despite these advantages, some of the already known problems still need to be addressed, i.e., generally longer training time and needs of a large dataset for training the network, optimisation issue, and how to optimally differentiate the two crop classes that have low inter-class variability.

2.2. Overview of Support Vector Machine

SVM is a non-parametric classifier that becomes popular due to its empirical performance to solve various problems and practically used in many applications (Bruzzone & Persello, 2009; Wang & Zhong, 2003). The basic concept of SVM is to find the optimal hyperplane that separates classes with a maximum margin between the classes and minimises the misclassification on test data. Hsu et al. (2003) state that SVM aims to generate a model from training data and predicts the label of the test data. In practice, it needs to extend this definition for non-linearly separable data where the perfect separation is hard to get. Figure 1 represents the basic idea of a support vector machine. Data (symbolise in point) lie on the dashed line are called as support vectors which determine the hyperplane (the solid black line between the dashed line).

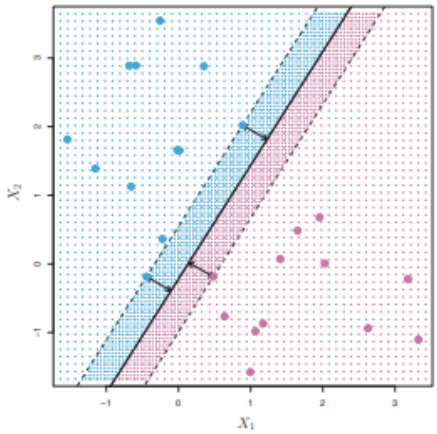


Figure 1. The basic idea of support vector machine in separating two classes by defining the hyperplane and the maximum margin
Adapted from : James, Witten, Hastie, & Tibshirani (2013)

According to Hsu et al. (2008), SVM with RBF (Radial Basis Function) kernel is a good initial choice of model selection for data classification. It has two parameters, i.e., penalty parameter (C) and kernel parameter (gamma, γ). Gamma represents the width of a kernel function (Ndikumana, Minh, Baghdadi, Courault, & Hossard, 2018). Penalty parameter controls the balance between generalisation of decision boundary and classifying the training data correctly. Higher gamma leads to overfitting because the classifier tries to generate perfect boundaries that fit the training data. C parameter takes a role to avoid the worst condition where the classifier uses many points of training data as support vectors (overfitting), so the classifier creates more general boundaries but still classifying the data optimally. Both parameters, C and gamma, are identified from the training data and used to predict the label of test data. Selection of the best value of this parameter determines the computational time of the SVM implementation.

2.3. Overview of Dynamic Time Warping

Successful implementation of DTW is reported by Baumann, Ozdogan, Richardson, & Radeloff (2017). The authors use multi-temporal data of MODIS in vegetation index format to apply DTW approach in order to generate the annual phenological curve. Another implementation is reported by Guan, Huang, Liu, Meng, & Liu (2016). The authors map the rice cropping system. Then, Maus et al. (2016) report that land cover and land use classification of MTSI using a time-weighted version of DTW.

According to Petitjean et al., (2012), DTW is a parameter-free approach that exploits the temporal information when time sampling of the input is irregular. DTW compares two radiometric profiles over time, reference and targeted profile, by measuring the similarity between them and analyse the temporal information of MTSI (Zhai, Qu, & Hao, 2018). Since DTW is originally designed for 1-Dimensional (1D) data, e.g., a speech signals (Sakoe & Chiba, 1978), in remote sensing application, it needs some options of modification from the original definition such as to handle the multi-dimensional time series images (multi-temporal and multi-spectral) by providing a single radiometric profile. Using the 1D data as input is an

appropriate solution because the sequence of all bands is dependent (Petitjean et al., 2012). Although it requires an additional step to prepare the 1D data from the remote sensing image that originally has dimension more than one.

2.4. Overview of Fully Convolutional Network

FCN is a variation of CNN that consists of a set of layers with learnable parameters of weights and biases. FCN classifies each pixel of the input and generates the output which is labelled every input pixel to a specific class.

2.4.1. Layers of FCN

The layers of an FCN architecture can be:

- Input Layer

  The input layer is the input image that has dimension $W \times H \times D$. Where $W \times D$ represents the spatial dimension of the image in Width and Height. For the training stage, the dimension of the input layer is equal to the dimension of patches, while for the prediction stage, it is equal to the test image dimension. D is the depth of the input that typically equal to the number of bands (spectral information).

- Convolution Layer

  Convolution layer is the main block of the FCN. This layer consists of a certain number of filters with a small value in the spatial dimension (commonly being used in practice) and extends through the full depth of the input dimension (Stanford University, 2018). Even though the dimension of the filter is set in the three dimensions, this type of filter is called as 2-Dimensional (2D) convolution, because the filter convolves only on the two dimensions (width and height) of the input. A convolutional layer has a dimension of $F \times F \times D \times K$, where $F \times F$ is the spatial dimension of the filter, D is the depth of the filter. Depth of 2D filter is equal to the depth of the input image, and K is the number of filters.

  Dilated convolution is a version of the convolution layer with a dilation factor as the parameter. The dilation factor represents the space between cells inside the filter. Standard convolutional layer uses dilation factor equal to 1 (no dilation). Increasing dilation factor makes the space between filter elements increases (red dot) and expands the receptive field (blue colour cells) more significantly than using convolution layer with no dilation as shown in Figure 2. The receptive field defines the number of considered pixels for the training process.

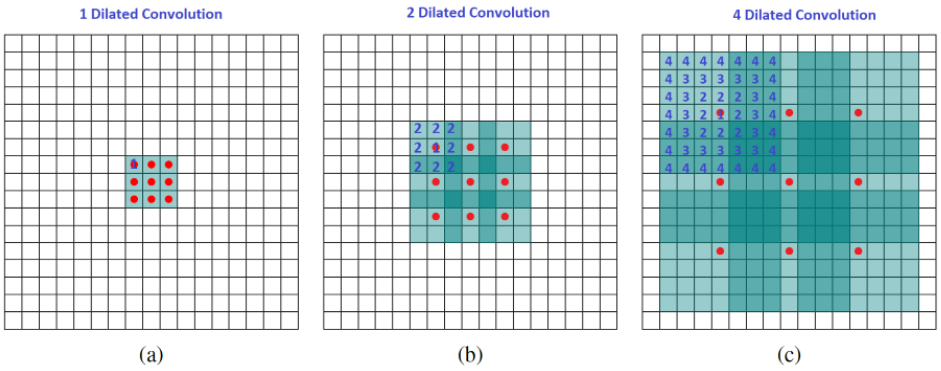


Figure 2. Dilated convolution with filter size 3 and dilated factor of 1, 2 and 4
Adapted from: Yu & Koltun (2016)

If we have 3x3 filters with dilation factor 1 in the first layer, this layer has a 3x3 view of the input image. When we stack 3x3 filters with dilation factor 1 in the second layer, this layer has a 3x3 view of the output of the first layer which means a 5x5 view of the input dimension. This type of network has an effective receptive field of 5x5. However, it is different if we stack 3x3 filters with dilation factor 2 in the second layer. This type of network has a 7x7 view of the input dimension (receptive field).

Dilated convolution or dilated kernel (DK) layer has hyper-parameters, i.e., size of the filter (F), stride (S), padding (P), dilation factor (D), and the number of filters (K). It is essential to pay attention to these parameters for controlling the size of output feature maps.

- Batch Normalisation (BN) layer

  In practice, the BN layer is commonly used after the convolutional layer. Batch normalisation is used to handle the issue of vanishing gradients when the network uses too high leaning rate (Ioffe & Szegedy, 2015).
- ReLU (Rectified Linear Unit) layer

  It is one of the activation function types commonly used in practice. It is supported by Krizhevsky, Sutskever, & Hinton (2012) who examined that training CNN with ReLU takes less time than another activation function, such as tanh units.
- Dropout layer

  It is one of the regularisation unit types that control the network so overfitting can be prevented (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014).
- Softmax layer (output)

  It is one of the loss function types that performs classification by calculating the score of every class in each pixel.

2.4.2. Hyper-Parameters of The Network

Besides learnable parameter, weight and bias calculated during the training process, FCN has hyper-parameters to be defined by the user. The considered hyper-parameters to design the network are categorised into two part as follow:

- Architectural Parameters

  The selection of architectural parameters influences the performance of the classification more than the selection of the training parameters (Rußwurm & Körner, 2017). Architectural parameters construct the network by providing value setting for the size of patches, layer's architecture, and number of filters.
  - **Patch size**. It refers to the dimension of the training image used in the network as input. The filters only look at this given patch size, not the entire input image.
  - **Layer's architecture** is the structure of layers in the network. Layer's architecture is important to be defined by considering the available data. Different dataset might need a different layer's architecture.
  - **Number of filters** represents the number of expected feature maps generated from the convolution layer. A larger number of filters means that more feature maps are generated, and it increases the number of learnable parameters. Thus, it increases the computational time.
- Training Parameters

  Training parameters consist of:
  - **Learning rate** defines how much we adjust the weights of the network. Small learning rate means a slow movement of gradient descent to seek for the global minimum. A too small gradient makes the network take a long time to converge, while a too large value of learning rate might make the network fails to convergence because it skips the global minimum.
  - **Number of epochs** expresses the time needed by the network to converge during the training stage. This parameter interacts with the other training parameters.
  - **Mini-batch size** determines how many samples are executed into memory for each iteration during the training process. When we have 2000 samples and use batch size 100, it means that the network takes 20 iterations for each epoch. This mini-batch size is also dependent on hardware capacity.

3. METHODS

A general overview of the methods is shown in Figure 3. By applying the workflow, we can evaluate how the selected methods perform classification using multi-temporal images of Sentinel-2 to provide crops information.

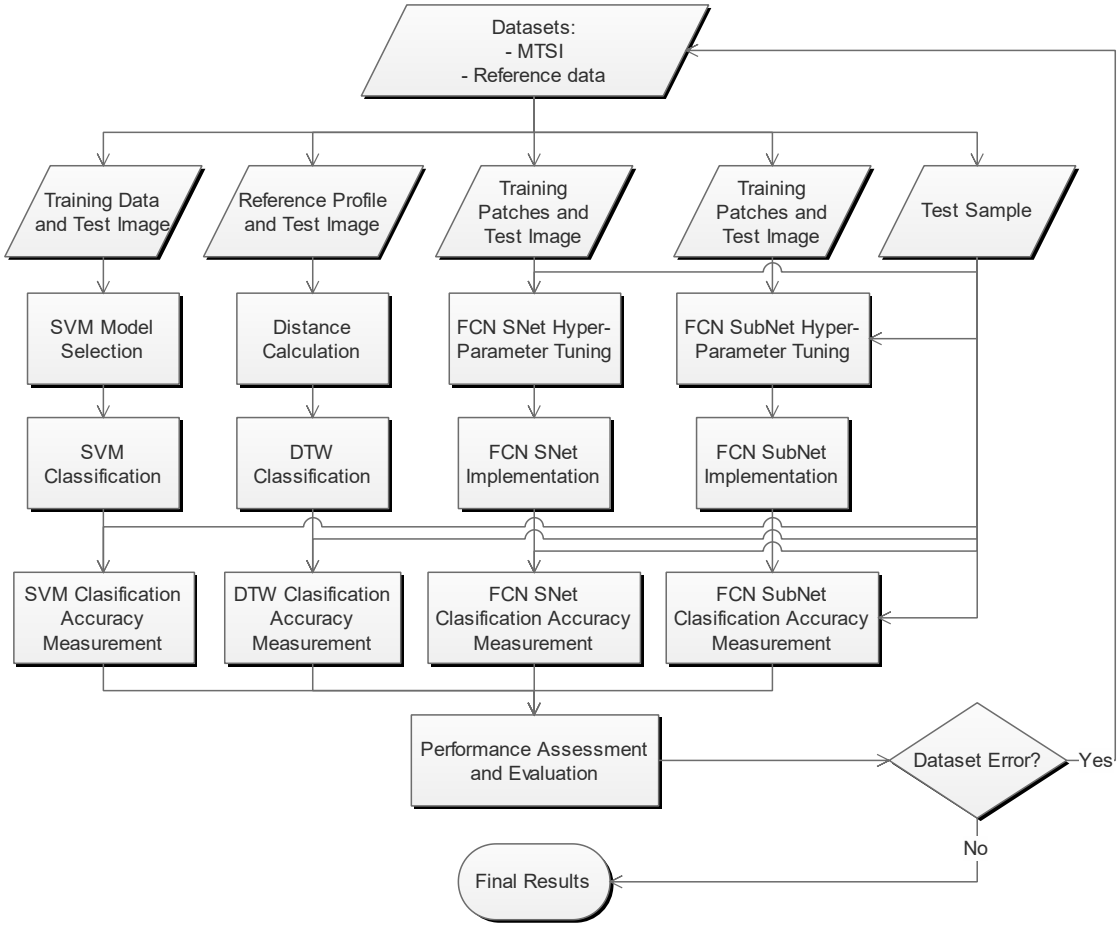


Figure 3. Workflow of applied methods

3.1. Baseline Methods: SVM and DTW

We use SVM as a standard classification strategy to produce the LCC map from MTSI. In this research, we use an RBF kernel with 400 pairs of OAs and SVM parameters (C, gamma). The implementation of SVM use LIBSVM (library for SVM) tool for MATLAB extension (Chang & Lin, 2011).

Besides applying SVM, we apply a standard DTW by measuring the spectral value similarity of the input image to the reference. The reference spectral value is a series of NDVI value along the time dimension of the input image for each of the targeted classes derived by averaging all NDVI profile of training samples.

3.2. FCN

To meet the research objective, we design two FCN architectures, FCN-SNet and FCN-SubNet, that treat the spectral information in a different way. These architectures are implemented by using a library of MatConvNet for MATLAB (Vedaldi & Lenc, 2015).

3.2.1. FCN-SNet

FCN-SNet is adopted from Persello & Stein (2017). The authors use FCN-DK architecture for detecting informal settlement using satellite images. DK means dilated kernel that refers to the dilated convolution. Instead of using down-sampling and up-sampling technique combined with a standard convolution, convolution layer with dilated filters (dilated convolution) is used to capture a larger spatial pattern and maintain the size of every layer to be the same as the input layer. By using dilated convolution, the number of parameters increases with respect to the receptive field increase. However, the number of parameters is not exponentially increased as we use the standard convolution (no dilation factor).

We adopt the FCN-DK architecture to avoid the unnecessary interpolation in the convolution-deconvolution network because we aim to produce the classification map with the same size to the input image. Table 1 presents the proposed initial architecture for this research. We use the dilated convolution without the pooling layer. We design this architecture to be able to process the input of multi-temporal images. The initial setting for the number of filters is expected to maintain the variation of the extracted features from the temporal and spectral dimensions.

Table 1. Initial architecture for FCN-SNet4.2 configuration

Block	Layer Type	Dimension				dilation	stride	pad
		width	height	depth	number of filters			
1	Convolution	3	3	40	40	1	1	1
	Batch Normalisation							
	lReLU							
2	Convolution	3	3	40	40	1	1	1
	Batch Normalisation							
	lReLU							
3	Convolution	3	3	40	40	2	1	2
	Batch Normalisation							
	lReLU							
4	Convolution	3	3	40	40	2	1	2
	Batch Normalisation							
	lReLU							
class	Convolution	1	1	40	5	1	1	0
	Batch Normalisation							
	Dropout							
	Softmax							

FCN-SNet means a single and straight network of FCN. The initial configuration network consists of four blocks dilated convolution layer by dilation factor 1 (block 1 and 2) and 2 (block 3 and 4), so we named it FCN-SNet4.2 where 4 represents the number of layers, and 2 indicates the largest dilation factor being used (starting from 1). Each block consists of three-layer types, dilated convolution, Batch Normalisation, and lRelu. Each convolution has a small size of the filter by 3x3. Larger filter size makes the network loss the detail and leads into underfitting. Stride 1 is used for all convolution layer to maintain the size of feature maps be equal to the size of the input image. The pad size is equal to the dilation factor.

For experiments, we refer to the different structures to test as FCN-SNet<a>.<b>, where a refers to the number of blocks, and b refers to the dilation factor of the last blocks. FCN-SNet6.2 means the network consists of six blocks of convolution layer by dilation factor 1,1,1,2,2,2 in sequence.

3.2.2. FCN-SubNet

FCN-SubNet is adopted from the FCN-FuseNet developed by Bergado, Persello, & Stein (2018) to utilise the multi-resolution images for classification. FuseNet is designed for panchromatic and multi-spectral bands input of very high-resolution satellite images by using two separate streams with different spatial resolution at the beginning of the network. The network then fused those two to produce a single output gradually.

We prepare the network with modification and apply it for MTSI. For the initial configuration, we use ten separate streams (sub-networks) at the beginning of the network then combined the output of the sub-network into one stream to produce the classification map. Therefore, we named it FCN-SubNet that indicates the structure of the network which contains more than one stream. The number of sub-networks represents the temporal information of multi-temporal images. We design the initial structure of the FCN-SubNet as presented in Table 2. This structure also adopts the dilated convolution layer as being used in FCN-SNet architecture.

Table 2. Initial architecture for FCN-SubNet10.2.2 configuration

Sub-network 1	Sub-network …	Sub-network n
Convolution 3x3 dilation 1 Batch Normalisation lRelu Convolution 3x3 dilation 1 Batch Normalisation lRelu Convolution 3x3 dilation 2 Batch Normalisation lRelu Convolution 3x3 dilation 2 Batch Normalisation lRelu	Convolution 3x3 dilation 1 Batch Normalisation lRelu Convolution 3x3 dilation 1 Batch Normalisation lRelu Convolution 3x3 dilation 2 Batch Normalisation lRelu Convolution 3x3 dilation 2 Batch Normalisation lRelu	Convolution 3x3 dilation 1 Batch Normalisation lRelu Convolution 3x3 dilation 1 Batch Normalisation lRelu Convolution 3x3 dilation 2 Batch Normalisation lRelu Convolution 3x3 dilation 2 Batch Normalisation lRelu
Concatenate Network		
Convolution 3x3 dilation 1 Batch Normalisation lRelu Convolution 3x3 dilation 1 Batch Normalisation lRelu Convolution 3x3 dilation 2 Batch Normalisation lRelu Convolution 3x3 dilation 2 Batch Normalisation lRelu Convolution Batch Normalisation Softmax		

Sub-Network 1-to-n uses the same architecture, where n indicates the number of available dates of acquisition. We use index '10.2.2' to indicate ten sub-networks with dilation 2 and concatenate network also uses dilation 2.

3.2.3. Design Implementation

After building the design of the proposed network, we carry out some experiments to tune the hyper-parameters value. The selected values from hyper-parameters tuning are used for final implementation.

3.3. Performance Assessment and Evaluation

Assessing the accuracy of the classification map is an important activity to provide information how good is the map to the user, besides it also exhibits the potential source of errors to improve the quality of the map to provide a reliable result (Congalton & Green, 2010). Classification accuracy represents the degree of the correctness of the LCC map (Foody, 2002). We compare the classification result to the reference data that are assumed to be true (ground reference data).

To perform the evaluation and accuracy assessment, we use the measures derived from the confusion matrix. Confusion matrix has been commonly used in practice and becomes the main point of the classification accuracy assessment (Foody, 2002). Confusion matrix shows the relation between the reference data and the corresponding classified data in cross-tabulation data. To assess the classification performance, we use measures as follows:

- Overall Accuracy

  Overall Accuracy (OA) is derived from the confusion matrix and indicates the total number of correctly classified pixels in all classes compare to the reference data (test sample). OA indicates the correctness of the classification map in percentage.

  Besides the OA, we also assess the accuracies of individual classes by calculating user's accuracy (UA), producer's accuracy (PA), and F-Measure. To provide general information for all classes, we calculate the average of UA (AUA), PA (APA) and F-Measure (AFM).
- User's Accuracy

  UA provides information from the perspective of the user, how accurate is the classification map in percentage. The user's accuracy indicates how many pixels of a particular class correctly portray that class on the ground.
- Producer's Accuracy

  PA provides information from the perspective of the producer, how accurate is the classification map in percentage. The producer's accuracy indicates how many pixels of the reference data in a certain class are correctly classified.
- F-Measure

  It provides information about the precision and robustness of the classifier in percentage. It is derived from the UA (precision) and the PA (recall).

$$FMeasure = 2 \cdot \frac{UA \cdot PA}{UA + PA}$$

- Visual Inspection on The Classification Map

  Besides quantitative evaluation, we assess the classification result qualitatively by inspecting the classification map.

4. DATASETS

In this chapter, we present the activities of image collection, pre-processing data, training, and test samples generation and creation of the input file for the experiments. We have two datasets in Romania (Dataset 1) and California (Dataset 2) explained in detail in Section 4.2 and 4.3. Section 4.1 explains the common activities in processing Sentinel-2 images for both datasets.

4.1. Image Pre-Processing

For this research, we use multi-temporal images of Sentinel-2A and 2B year 2017. Sentinel images are free and openly accessible through the website of Copernicus Open Access Hub. We collect the images by defining search criteria, i.e., cloud cover not more than 10% with sensing period during January-December 2017 in the selected study area. The data can be downloaded after login. These multi-temporal images are collected to represent the growing stages of the crops adequately.

Sentinel-2 mission has five days revisit time and 13 channels in which each of the channels has a different objective. Table 3 provides the channel's resolution and objective of Sentinel-2 (Earth Observation Portal, 2014).

Table 3. Spectral resolution and objective of Sentinel-2

Band	Spatial Resolution (m)	Mission Objective
1	60	Aerosols correction
2	10	Aerosols correction, land measurement band
3	10	Land measurement band
4	10	Land measurement band
5	20	Land measurement band
6	20	Land measurement band
7	20	Land measurement band
8	10	Water vapour correction, Land measurement band
8a	20	Water vapour correction, Land measurement band
9	60	Water vapour correction
10	60	Cirrus detection
11	20	Land measurement band
12	20	Aerosols correction, Land measurement band

We pre-process the images through the following operations:

a. Image Correction and Resampling

We use the Sen2Cor plugin for Sentinel Application Platform (SNAP) for correcting and resampling Sentinel-2 images. Sen2Cor performs atmospheric, terrain and cirrus correction and creates new images for each band with Bottom of Atmosphere (BOA) value except for Band-10. These new images are equivalent to the level 2A of Sentinel-2 product (European Space Agency, 2018a).

After performing image correction, we resample the images to obtain 10m resolution images for the 13 bands. This resampling is needed to set the spatial resolution in the same size. Resampled images are used for the experiments part. For Band-10, as it does not contain the surface information (Müller-Wilm, 2018) so we directly resample the image of Band-10 from level 1A images. Images from both locations are projected in WGS 1984 UTM Zone 35N (Dataset 1) and 11N (Dataset 2).

c. NVDI Calculation

We also prepare the images in NDVI value, so we apply NDVI calculation using bands 4 (red) and 8 (near infra-red).

$$NDVI = \frac{near\ infrared - red}{near\ infrared + red}$$

d. Image Stacking

We stack the images based on structure as in Figure 4. We stack four commonly used bands for classification, i.e., Band 2, 3, 4 and 8. For experimental purpose, besides four bands stacking, we also prepare dataset in full bands stacking (13 bands), ten bands stacking and NDVI stacking. Ten bands stacking contains images from the bands that originally have 10m and 20m resolution in Table 3.

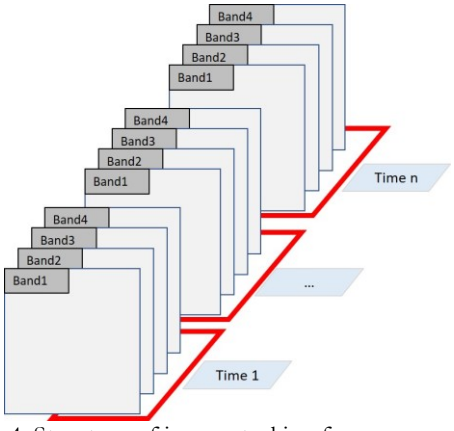


Figure 4. Structure of image stacking for a year.

4.2. Dataset 1: Romania

The first dataset is located in the agricultural site in Romania as displayed in Figure 5. Romania allocates one-third of the land for agricultural that provides a majority of the agricultural products in Europe (Encyclopædia Britannica, 2018). Romania is the number five of the largest utilised agricultural area in Europe (European Union, 2018)

Figure 5. Romania boundary and the extent of Dataset 1

We collect ten images of Sentinel-2 (the year 2017) for Dataset 1. The time acquisition of those images are as follows: 07 March, 03 April, 03 May, 05, 22 and 30 June, 22 July, 01 and 19 August, also 30 September. We also prepare the reference data provided by The National Agency for Payment and Intervention of Agricultural (APIA) of Romania. The reference data is available in shapefile format. The data has already been split into test and training and contains 1250 points in 5 classes. Figure 6 shows one of the images in Dataset Romania and the available reference points located over the study area. The image dimension of Dataset 1 is 4460x5716 pixels.

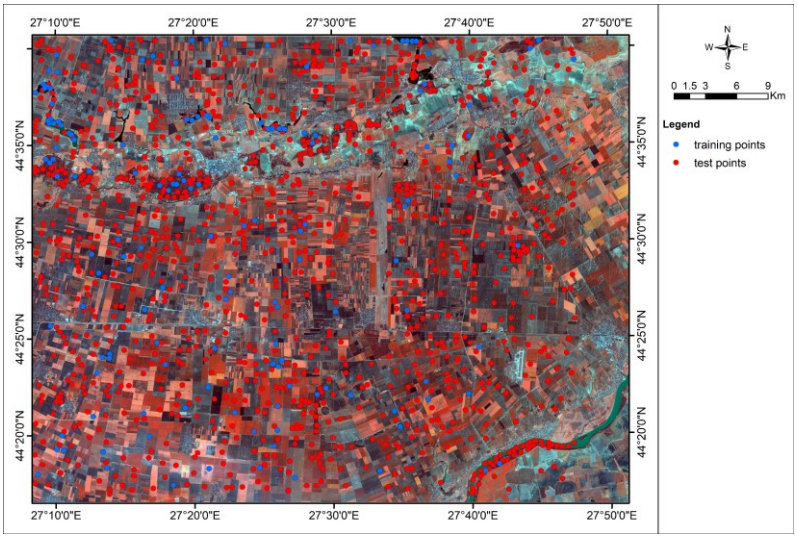


Figure 6. A preview of reference points in Dataset 1 overlay with the Sentinel-2 image on 7 March 2017 (RGB:832)

Since we need data in raster representation for the input of the network, we generate the data from the available reference points. We automatically create a buffer of 75m around the points and reshape it to a square polygon with the size of 150mx150m or equal to 155x15 pixels. We put a label with code in number (see Table 4) and convert the polygon into raster.

Table 4. Training and test area composition – Dataset 1

Code	Class Name	Number of Points		Number of Generated Polygons		Number of Generated Pixels	
		Training	Test	Training	Test	Training	Test
1	Wheat	30	400	30	395	6750	88749
2	Maize	30	250	30	235	6750	52785
3	Sunflower	30	250	30	250	6750	56163
4	Forest	30	150	30	150	6735	33634
5	Water	30	50	30	50	6750	11235
	Total	150	1100	150	1080	33735	242566

Table 4 shows the composition of training and test samples in point, polygon and raster format. Carefully look, there is a reduction in the number of test samples after it converts into a polygon format. The reason is the removal of some unnecessary polygons to make sure that the generated polygons are located inside the objects and disjoined from each other. Figure 7 presents the example of the samples in point and polygon format which is overlaid with the image.

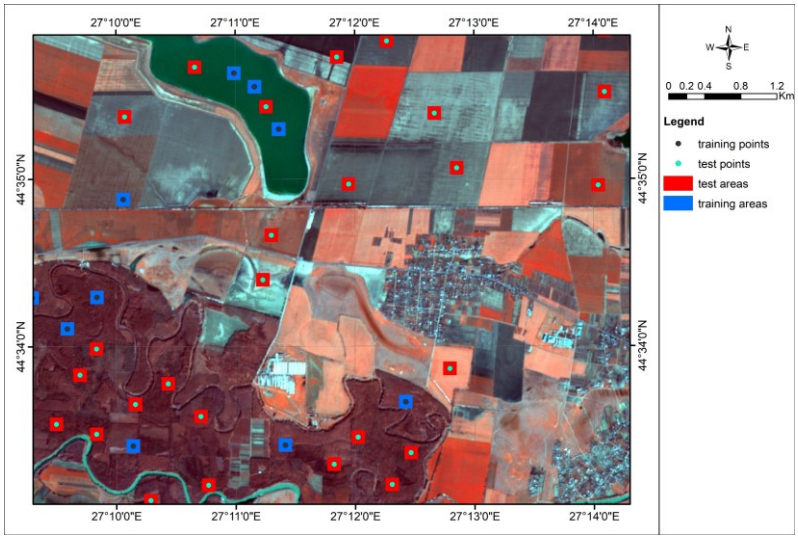


Figure 7. A preview of sample polygons, over a subset of Dataset 1. Image using Sentinel-2 on 7 March 2017 (RGB:832)

4.3. Dataset 2: California

California has a large coverage of agricultural, more than 38,000km² from the total 1,567,900.73 km² of the agricultural in the United States (World Bank, 2018). The coverage of the agricultural area in California is above the average area of agricultural land in a state. Location of the Dataset 2 in the agricultural site in California as presented in Figure 8.

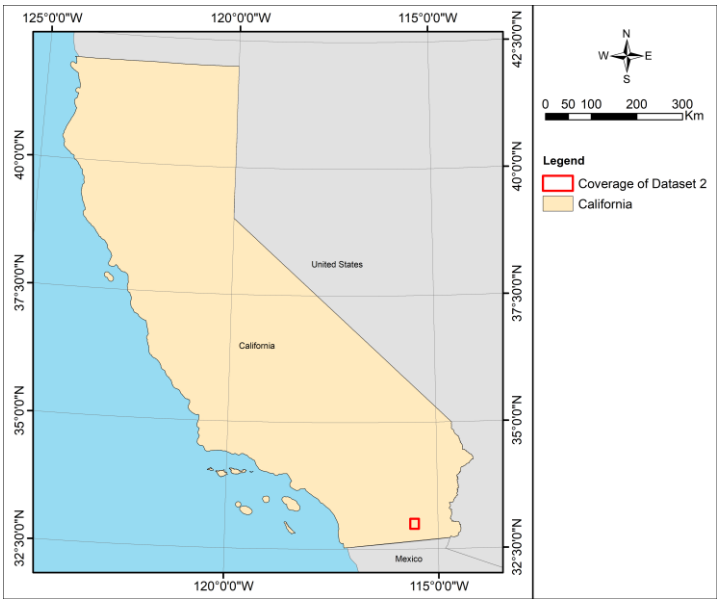


Figure 8. California boundary and the extent of Dataset 2

We collect 12 images of Sentinel-2 (the year 2017) for Dataset 2. The time acquisitions of those images are as follow 01 January, 20 February, 02 March, 21 April, 21 July, 20 June, 10 July, 19 August, 18 September, 23 October, 22 November, and 22 December. As an addition to the images, we prepare the reference data obtained from the website of the United States Department of Agriculture, National Agricultural Statistics Service. This department provides annual Cropland Data Layer (CDL) of the United States.

Different from the Dataset 1, we had a reference map as reference data. Therefore, we can estimate the coverage area of each class. We reclassify the available classes by selecting classes covered by less than

2% of the study area. After reclassification, we had 12 classes, i.e. Alfalfa, Carrots, Developed Area, Fallow/Idle Cropland, Lettuce, Onions, Open Water, Other Hay/Non-Alfalfa, Shrubland, Sod/Grass Seed, Sugar beets, and Winter Wheat. Compare to the Dataset 1, Dataset 2 has more complex classes, and the assignment of the training and test samples are not provided yet. We need to define the training and test sample by ourselves.

To generate the samples, we make a selection by taking only the objects covered by more than or equal to 225 pixels as samples and had homogeneous classes in a boundary field. From these samples, since we had more flexibility to select the samples from available reference data, we use a different setting for training and test sample by proportion about 50:50. We create the same size of polygons as Dataset 1 for training and test samples. Table 5 describes the composition of training and test of Dataset 2 in vector (polygon) and raster format.

Table 5. Training and test area composition – Dataset 2

Code	Class Name	Number of Generated Polygons		Number of Generated Pixels	
		Training	Test	Training	Test
1	Alfalfa	99	98	22125	21990
2	Carrots	17	16	3734	3570
3	Developed Area	49	48	7860	7152
4	Fallow/Idle Cropland	53	52	11649	10999
5	Lettuce	22	21	4935	4695
6	Onions	43	43	9375	9405
7	Open Water	6	6	1143	1350
8	Other Hay/Non-Alfalfa	42	41	9180	9000
9	Shrubland	25	25	3924	3393
10	Sod/Grass Seed	27	26	6075	5835
11	Sugar beets	31	31	6855	6930
12	Winter Wheat	18	17	3945	3615
	Total	432	424	90800	87934

Figure 9 shows the preview of the training and test samples in a subset of Dataset 2 randomly placed over the study area (simple random sampling). The image dimension of Dataset 2 is 2192x1899 pixels.

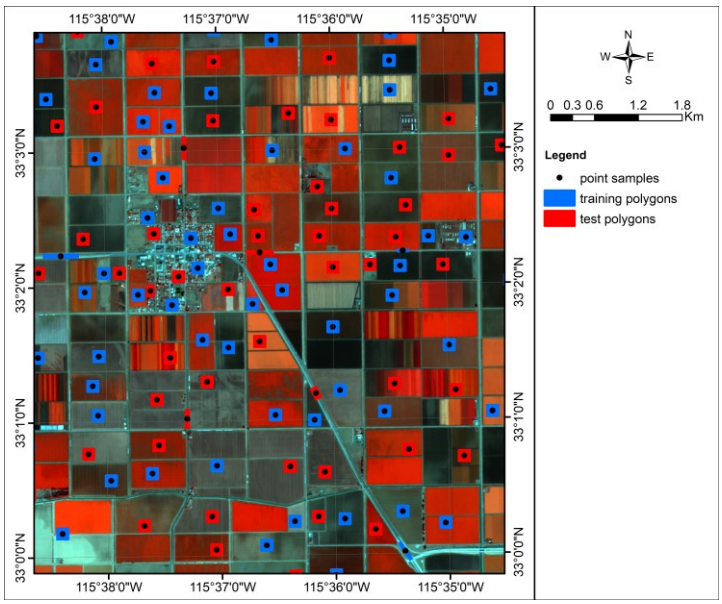


Figure 9. A preview of sample polygons, over a subset of Dataset 2 using Sentinel-2 on 01 December 2017 (RGB:832)

4.4. Structuring Input File for The Network

We prepare input file for the network by creating a set of training input that contains the image patches, class or label and attribute of the patches (training or validation). We randomly generate 2000 patches for training and 1000 patches for validation from the available training pixels mentioned in Table 4 and Table 5. For consistency in all experiments, we use the same training samples by using the same central patches indicated by its indexes. For the initial size of patches, we use 13x13 pixels which represent the 130x130 m^2 area on the ground. It considers the effective receptive field of the initial architecture. Since the objects of interest do not have a high spatial dependency to the neighbourhood pixels, it is not necessary to have a large patch to cover the neighbourhood objects in determining a label for a specific pixel.

5. EXPERIMENTS SETTING

We experimentally evaluate the optimal parameters to design the proposed architecture of FCN then implement the design to produce a classification map. We compare the result to other methods: SVM and DTW.

5.1. Initial Experiments

For a starting point, we apply the initial experiments for Dataset 1 and Dataset 2 with the input and methods as mentioned in Table 6. We start with the baseline methods, SVM and DTW, against the proposed method. The results of these experiments are presented in Section 6.1.

Table 6. Initial experiments setting

Methods	Input	Architecture
SVM	4 bands (2,3,4,8); NDVI	-
DTW	NDVI	-
FCN	4 bands (2,3,4,8); NDVI	FCN-SNet4.2

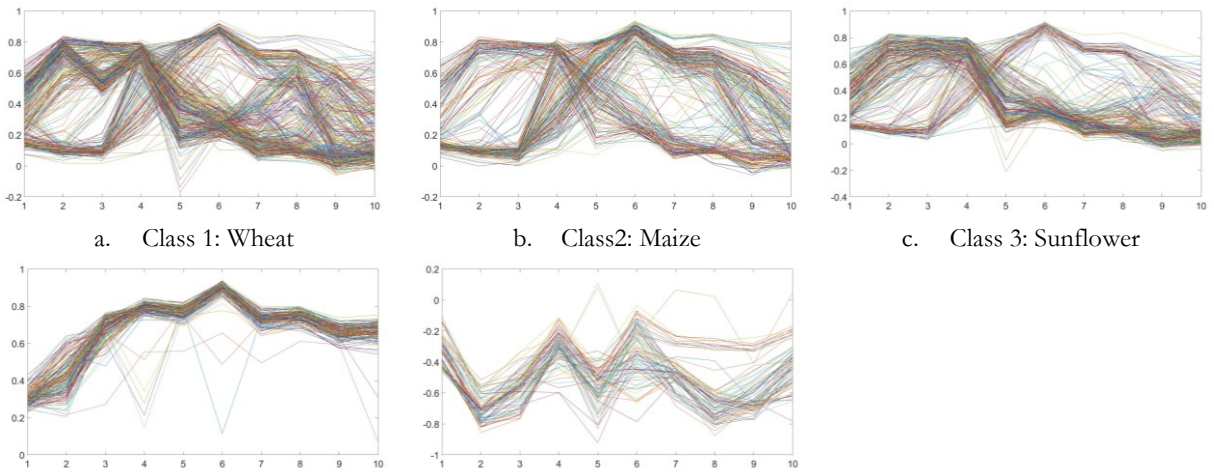
5.2. Datasets Refinement

We evaluate the quality of reference data based on NDVI value and spatial distribution of the training and test samples (spatial sampling strategies) of Dataset 1 and Dataset 2. We use the NDVI value from the available images to estimate the phenology pattern of the crops (Gómez et al., 2016). The NDVI value in a year provides an insight into the individual crop types and indicate the crops cycle. Evaluation based on NDVI value is expected to reduce the confusion among classes. The spatial sampling strategies are applied to measure the influence of the spatial distribution of samples to the classification result.

5.2.1. Dataset 1: Romania

5.2.1.1. Evaluation Based on NDVI Value

To refine Dataset 1, we check on the NDVI value of the training samples and plot the variation over time as presented in Figure 10. The NDVI value is generated from the centre of training polygons.



a. Class 1: Wheat b. Class2: Maize c. Class 3: Sunflower

d. Class 4: Forest e. Class 5: Water

Figure 10. The plot of NDVI value of the samples on Dataset 1

For comparison, we refer to the pattern generated by Belgiu & Csillik (2018) on location near to Dataset 1 as displayed in Figure 11.

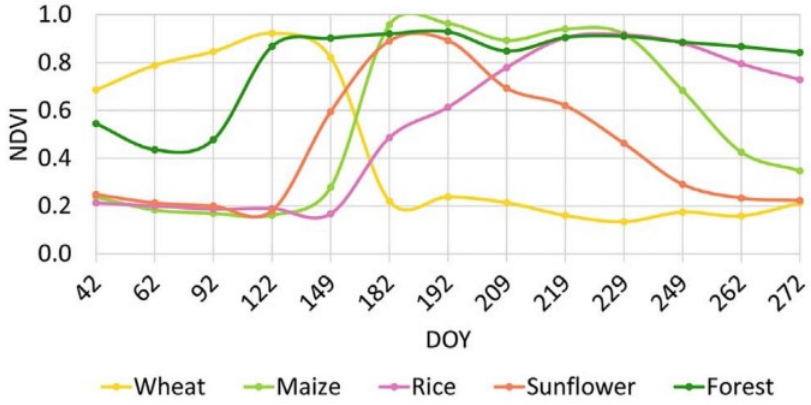
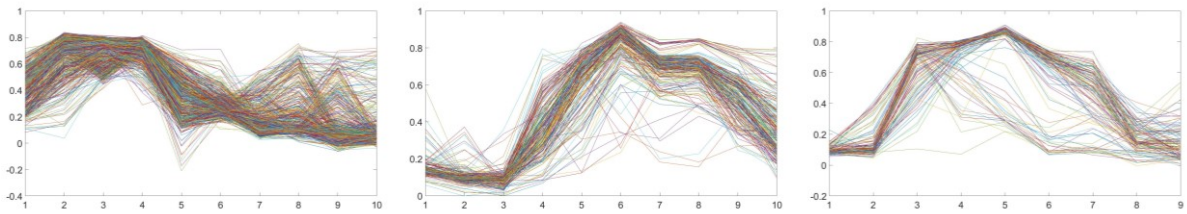


Figure 11. Temporal pattern of NDVI value on Dataset 1 area DOY = Day of Year in 2017. Adapted from: Belgiu & Csillik (2018).

We observe in Figure 11 that each of the classes has a single pattern that indicates crops with a single growing period (plantation, growing, and harvesting) during a year. We observe the pattern generated from the available samples in Figure 12, and we conclude that there is a potential error in the samples available for our study, especially in Wheat, Maize, and Sunflower classes. This problem might be the reason why the classification accuracy is low, and why the confusions exist in those three classes.

Therefore, by using this assumption and information, we evaluate the samples of those three classes to refine the dataset. We evaluate the samples by reselecting samples for each of these three classes by considering the NDVI value over time and compare the similarity with the reference pattern in Figure 11. We check further by inspecting the samples visually. We need visual interpretation because maize and sunflower are difficult to distinguish only by looking at the pattern based on the NDVI value. Both have a similar pattern along a year. Besides that, the number of classes and samples makes it feasible to perform the visual interpretation.

We try to maintain a variety of samples in a class by keeping some samples with a similar pattern and only omit the samples that had a completely different pattern (See Figure 12). Based on the pattern in Figure 11, we use the image on date 20170503 (DOY 123) to categorise the class based on the NDVI value for wheat. Meanwhile, we use an image on date 20170819 (DOY 231) to distinguish between maize and sunflower.



a. Class 1: Wheat b. Class2: Maize c. Class 3: Sunflower

Figure 12. The plot of NDVI value of the refined samples for three classes in Dataset 1

Figure 13, Figure 14 and Figure 15 show example of the samples that are displayed together with the images to help visual interpretation activity in order to reselect the samples for Class1, Class2, and Class3 respectively.

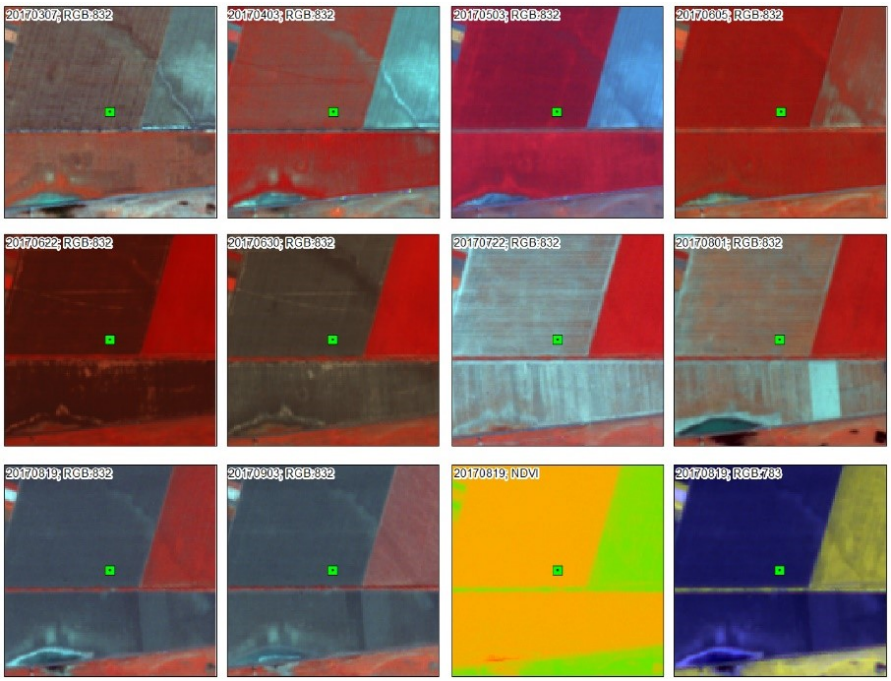


Figure 13. A sample of Class 1 appearance on RGB:832 for 10-time step, NDVI value and RGB: 783 on 19 August 2017

Figure 13 shows one of the samples representations for class 1. Wheat is represented by reddish colour in-band RGB: 832 on date1-date6 (2070307, 20170403, 20170503, 20170605, 20170622 and 2070630) and tends to darker in date5 and date6. Starting from date7 to date10 (20170722, 20170801, 20170819, and 20170903), wheat is represented by slate grey colour in-band RGB: 832 and tends to darker on date9 and date10. On date9, wheat is represented by yellow colour in NDVI image and displayed in blue colour on the image of RGB: 783. The samples with similar characteristic to this example are categorised as Class 1.

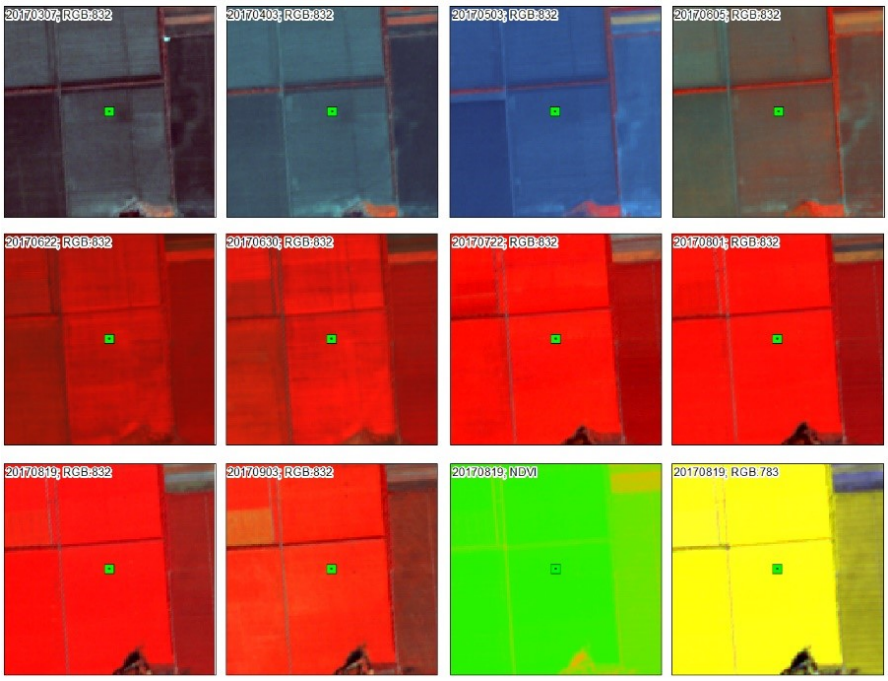


Figure 14. A sample of Class 2 appearance on RGB:832 for 10-time step, NDVI value and RGB: 783 on 19 August 2017

Figure 14 shows one of the samples representations for class 2. Maize is represented by slate grey colour in-band RGB: 832 on date1 and date2 (2070307 and 20170403). On date3 (20170503), maize is represented by bluish colour in-band RGB: 832. On date4 (20170605), maize is represented by green colour. Starting from date5 to date10 (201706022, 2070630, 20170722, 20170801, 20170819, and 20170903), maize is represented by red colour. On date9, maize is represented by lawn green colour in NDVI image and displayed in yellow colour on the image of RGB: 783. When a sample has a similar characteristic to this example are categorised as Class 2.

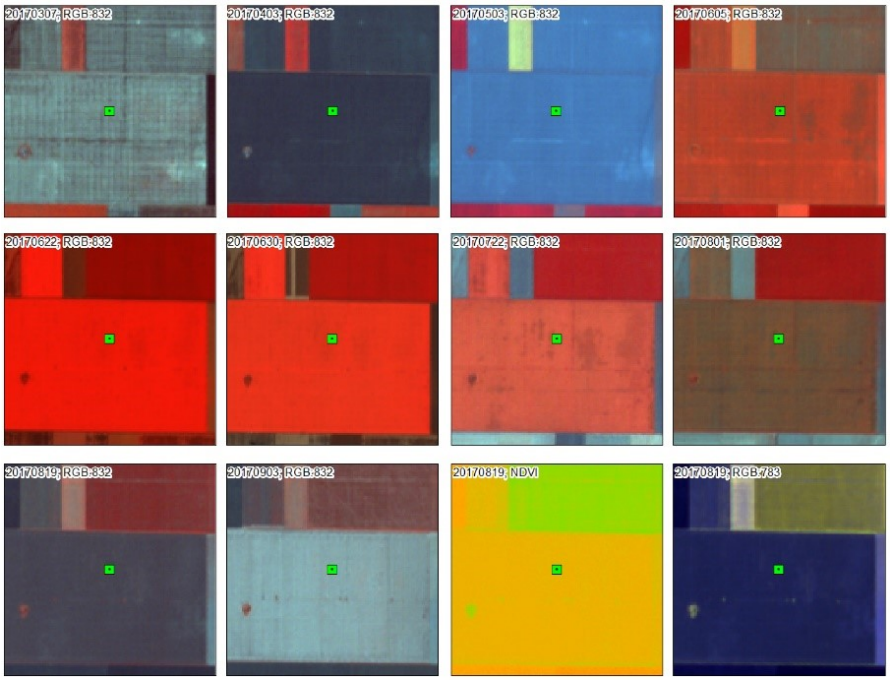


Figure 15. A sample of Class 3 appearance on RGB:832 for 10-time step, NDVI value and RGB: 783 on 19 August 2017

Figure 15 shows one of the samples representations for class 3. Sunflower is represented by light slate grey colour in-band RGB: 832 on date1 (20170307) and dark slate grey on date2 (20170403.) On date3 (20170503), sunflower is represented by bluish colour in-band RGB: 832. Starting from date4 to date8 (20170605, 2070630, 20170722, and 20170801), sunflower is represented by a red colour and tend to dark on the last date. Later, on date9 and date10 (20170819 and 20170903) sunflower is represented by slate grey colour. On date9, sunflower is represented by yellow colour in NDVI image and displayed in blue colour on the image of RGB: 783. Date 9 and 10 in RGB: 832 or 783, and NDVI image clearly distinguish maize and sunflower. The samples that have similar characteristic to this example are categorised as Class 3.

5.2.1.2. Evaluation Based on Spatial Sampling Strategies

From the allocated distribution of reference data, the training and test samples are distributed randomly over the study area. Based on the evaluation, the training number of polygons are also very low compared to the test by ratio 12:88. Therefore we test whether the spatial distribution of the reference data influences the accuracy of the classification result or ratio between training and test which might impact the accuracy of the classification results.

We evaluate this assumption by making a regular grid to systematically split the training and test area with the same coverage. By doing this, we also expect the ratio of the training and test sample increases to 50:50. We make several strategies to get spatially distributed sampling as displayed in Figure 16. We apply these strategies to reference data before and after applying evaluation based on spectral value.

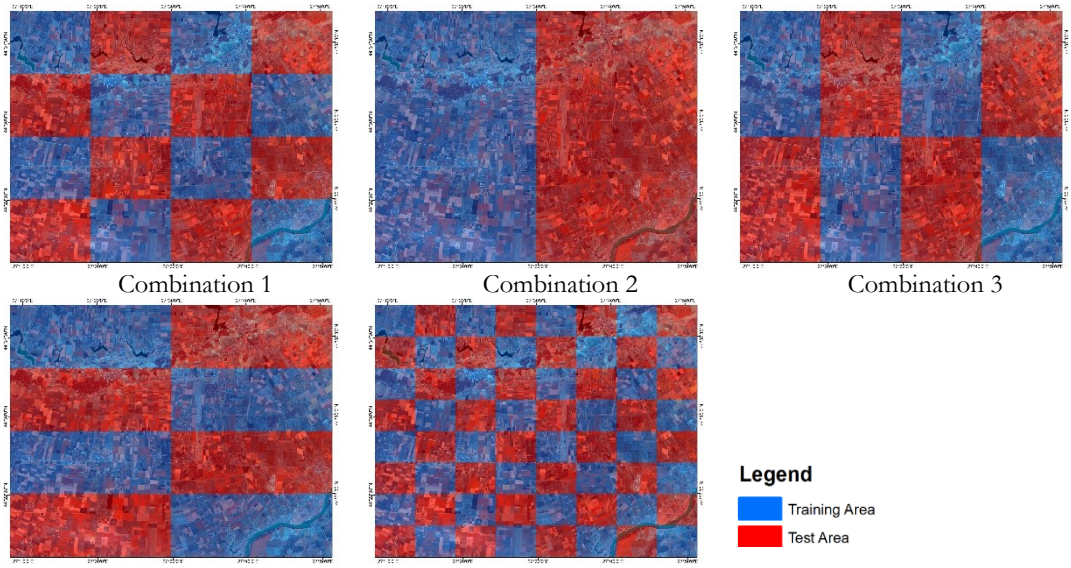


Figure 16. Set of combination for applying spatial sampling strategies

5.2.1.3. Applied Refined Dataset for FCN-SNet Initial Experiments Setting

For testing the refined dataset, we use FCN-SNet architecture. We set the experiments as Table 7 and Table 8.

Table 7. Refinement experiment set up for Dataset 1

Set	Training
1	Refined based on NDVI value only
2	Refined based on spatial location only
3	Full refined: based on NDVI value and spatial location

The results of these experiments are presented in Section 6.2 after optimising the last three hyper-parameters from Table 8.

Table 8. Detail setting for the FCN-SNet4.2 – Dataset 1

Parameter	Value
Layer architecture	FCN-SNet4.2
Size of Filter	3
Number of Filters	40
Patch Size	13
Size of Mini Batch	16, 32, 64, 100, 128
Learning Rate	1e-8, 1e-7
Number of Epochs	10, 150, 200, 250

5.2.2. Dataset 2: California

5.2.2.1. Evaluation Based on NDVI Value

For Dataset 2, we had more classes and more complex pattern to distinguish. Figure 17 shows the NDVI pattern for a year for seven classes out of 12 classes used in Dataset 2. From this pattern, we see that differentiating target classes is a challenge because some of them have a similar pattern to each other during a year. It is difficult to find the time interval, i.e., the growing phase when one class can be completely distinct from other classes. In this case, we only remove the samples that completely different from the dominant pattern in the same class. We do not apply visual interpretation since it is hard to recognise the

appearance of the sample by looking at the spectral representation in the RGB layer. It is also a time-consuming task due to the number of classes and samples.

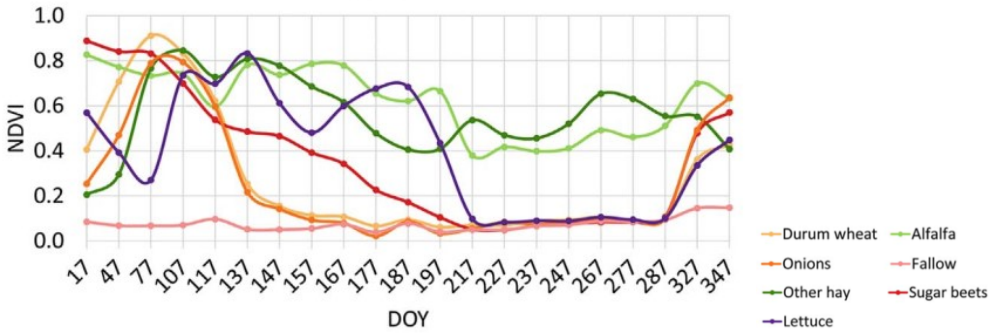
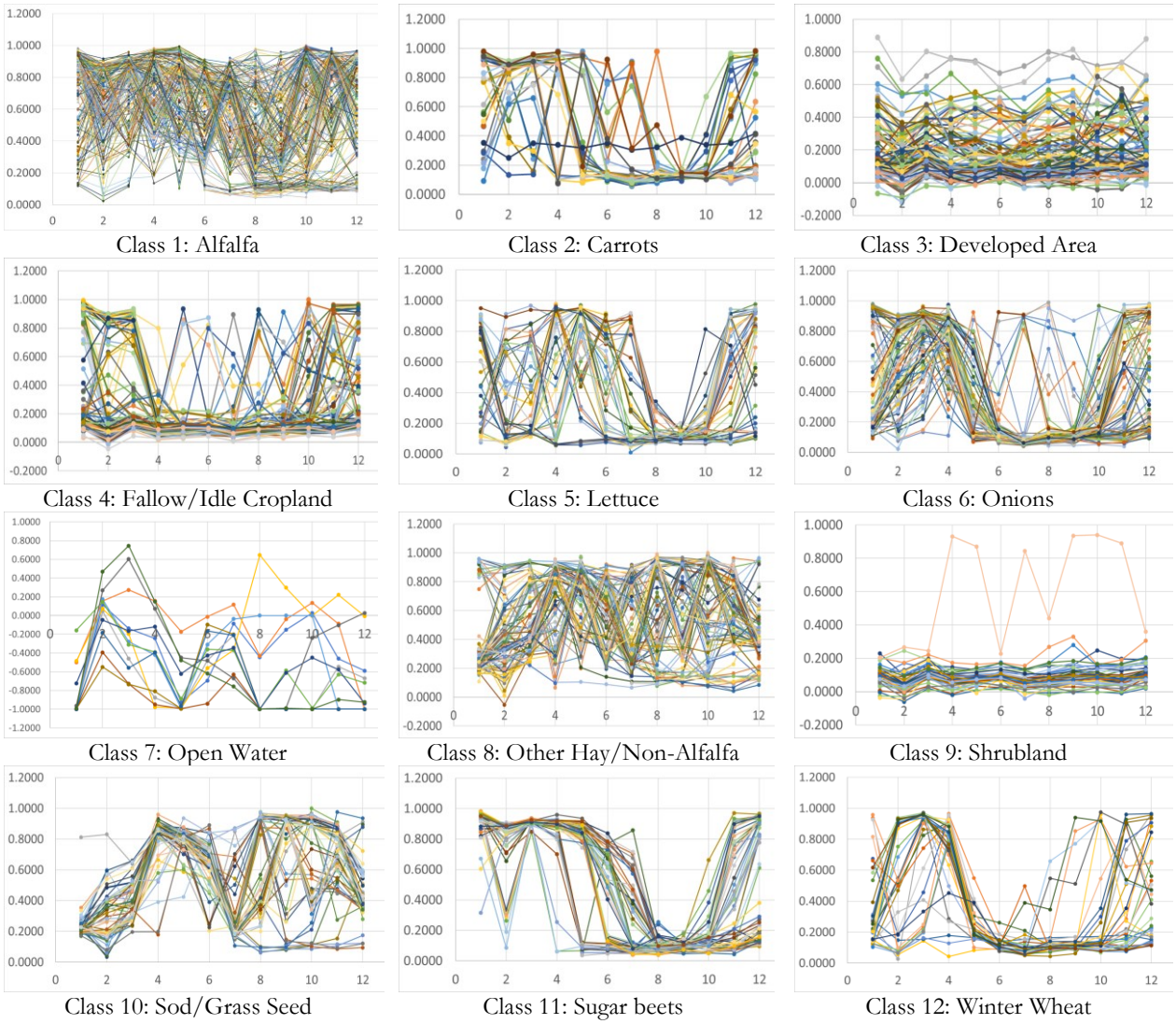


Figure 17. Temporal pattern of NDVI value on Dataset 2 area
Adapted from: Belgiu & Csillik (2018)

From the available training and test points in Dataset 2, we plot the spectral value using NDVI value. Figure 18 shows plots of NDVI value for 12 classes in the centre of training polygons.



Class 1: Alfalfa | Class 2: Carrots | Class 3: Developed Area
Class 4: Fallow/Idle Cropland | Class 5: Lettuce | Class 6: Onions
Class 7: Open Water | Class 8: Other Hay/Non-Alfalfa | Class 9: Shrubland
Class 10: Sod/Grass Seed | Class 11: Sugar beets | Class 12: Winter Wheat

Figure 18. The plot of NDVI value of the samples for 13 classes on Dataset 2

We observe the dominant pattern from the available samples to refine the selected samples. The plot of NDVI value after refinement is available in Figure 19.

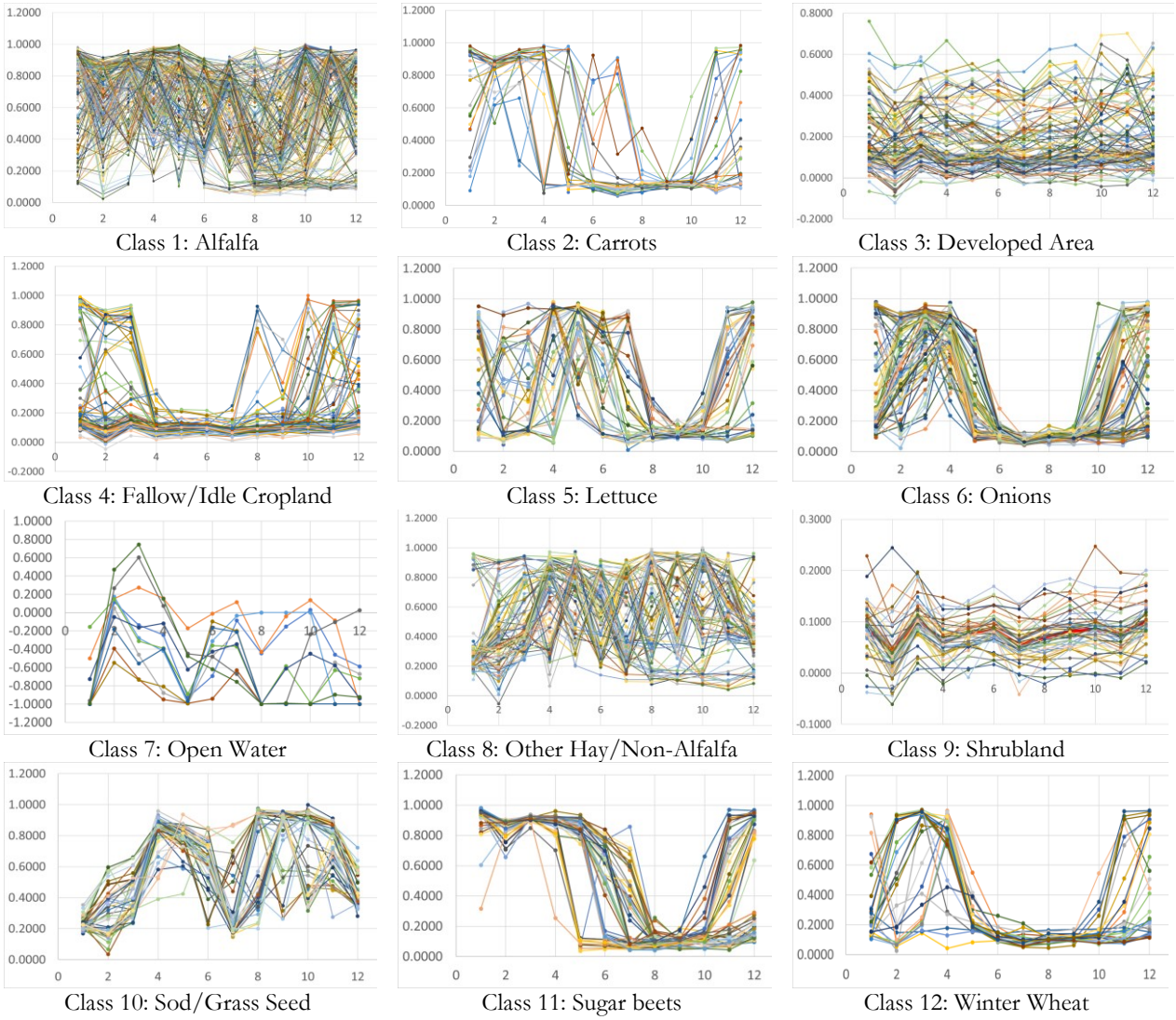


Figure 19. The plot of NDVI value of the samples for 13 classes on Dataset 2

5.2.2.2. Evaluation Based on Spatial Sampling Strategies

From the allocated distribution of reference data, the training and test data are distributed randomly over the study area by ratio 51:49. We test whether the spatial distribution of the reference data influences the accuracy of the classification result by modifying the spatial distribution of the samples.

We evaluate the spatial sampling distribution by creating a regular grid to systematically split the training and test area which maintain the ratio of about 50:50. We prepare several strategies to spatially distributed the samples as displayed in Figure 20. We apply these strategies to reference data before and after applying evaluation based on spectral value. For Dataset 2, we only use four combinations because the coverage area is smaller than the Dataset 1. We assume that combination 5 applied on Dataset 1 is already represented by combination 1 in Dataset 2.

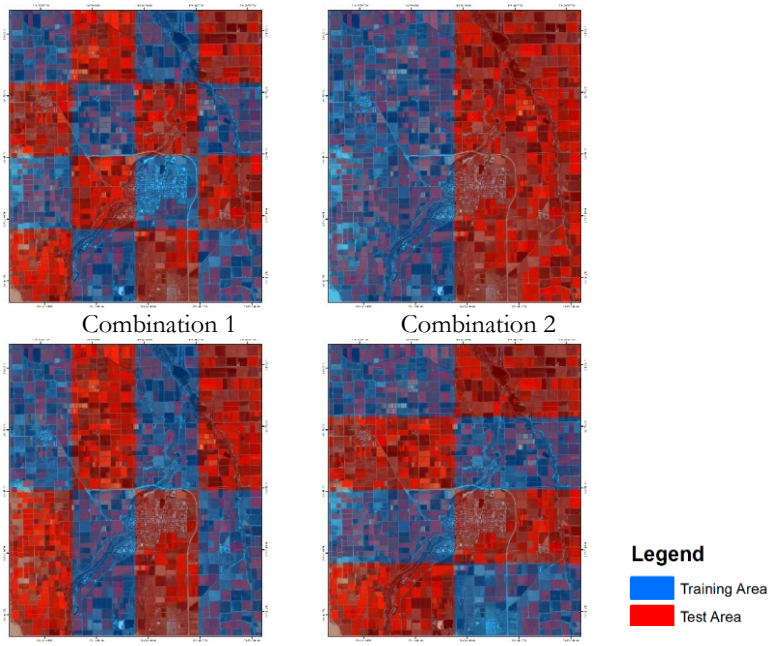


Figure 20. Set of combination for applying spatial sampling strategies

5.2.2.3. Applied Refined Dataset for FCN-SNet Initial Experiments Setting

For testing the refined dataset, we use FCN-SNet architecture. We set the experiments as shown in Table 9 and Table 10.

Table 9. Refinement experiment set up for Dataset 2

Set	Training
1	Refined based on NDVI value only
2	Refined based on spatial location only
3	Full refined: based on NDVI value and spatial location

The results of these experiments are presented in Section 6.2 after optimising the last three hyper-parameters from Table 10.

Table 10. Detail setting for the FCN-SNet4.2 – Dataset 2

Parameter	Value
Layer Architecture	FCN-SNet4.2
Size of Filter	3
Number of Filters	40
Patch Size	13
Size of Mini Batch	16, 32, 64, 100, 128
Learning Rate	1e-8, 1e-7, 1e-6
Number of Epochs	250, 500, 750, 1000

5.3. SVM Parameter Tuning

We tune in C and gamma as the parameters of SVM with RBF kernel. We select the optimal C and gamma by providing a range of values for each parameter and select the optimal value that produces the maximum OA. We set 20 values for each C and gamma, so in total it generates 400 possible combinations (see Table 11). We record the OA for each possible combination and select the best combination.

Table 11. SVM parameter experiments setting

Parameter	Value
C	20 value of 1e2 to 1e5
Gamma	20 value in range 0.1 to 10

5.4. DTW Implementation

In order to implement the DTW approach, we need to provide a temporal pattern for reference data and compare this pattern with every pixel of test data. For consistency, we generate a reference pattern for each class by averaging the NDVI value of central patches from FCN input. We generate five reference patterns for Dataset 1 and 12 reference patterns for Dataset 2. The reference pattern is an essential part of DTW implementation.

5.5. FCN Hyper-Parameters Optimisation

Hyper-parameter of the FCN is categorised into two components, architectural parameters and training parameters. The architectural parameters define the structure or design of the network, while the training parameters are used during the training. The optimum hyper-parameter values are used in the final implementation. Each parameter is tuned by varying the value of a single parameter and keeping the other parameters using the same value. Selection for candidate values of structural parameters as follow:

- **Patch size**. We start with value 13, then make it 2, 3, 4, and 5 times larger (take the odd number).
- **Layer architecture**. We test some combinations of layer architecture. The combinations aim to test whether adding more layers (deeper network) affect the classification accuracy positively. It is important to note that the addition of the layer increases the receptive field.
- **Number of filters**. We start from 40 by an assumption that the first convolutional layer produces feature maps in the same number as the input image when the network uses 10 images with 4 bands as input. Then we increase the number to 80, 120 and 160. By using this number, we expect that we could maintain the variability of features that are hierarchically produced to classify the targeted classes.

Training parameters depend on the used dataset and the architecture of the network. Experiments for training parameters as follow:

- **Learning rate**. We observe the training curve to estimate the suitable learning rate. Training curve with a low gradient to converge indicates that we need to increase the learning rate (Zulkifli, 2018).
- **Number of epochs**. We set a large number of epochs in the beginning to learn the trend of the training curve then decide when to stop the training.
- **Size of mini batch**. To see the relation of the mini batch size to the accuracy of the classification result, we vary the value of the batch. Based on the practical recommendation, we use a value as a power of two that fits the hardware capacity (size of memory). We start from 16 to 128 to test. Another suggestion is to use the value in multiple factors from the number of samples, so we put 100 as the candidate value.

5.5.1. FCN-SNet Experiments Setting

We set candidate values for each of hyper-parameters for FCN-SNet architecture as described in Table 12.

Table 12. FCN-SNet experiments setting

Hyper-parameter	Candidate Value
Layer's Architecture	2.2, 4.2, 6.2, 6.3, 9.3, 8.4, 12.4
Size of Filter	3
Number of Filters	40, 80, 120, 160

Hyper-parameter	Candidate Value
Patch Size	13, 25, 39, 51,65
Size of Mini Batch	16, 32, 64, 100, 128
Learning Rate	1e-6, 1e-7, 1e-8, 1e-9
Number of Epochs	Dataset 1: 100,150,200,250 Dataset 2: 250,500,750,1000
Input Band Type	4b, 10b, 13b, NDVI

5.5.2. FCN-SubNet Experiments Setting

We set candidate values for each of hyper-parameters for FCN-SubNet architecture as described in Table 12.

Table 13. FCN-SubNet experiments setting

Hyper-parameter	Candidate Value
Layer's Architecture	1.2,2.2,3.2,2.1,2.3
Size of Filter	3
Number of Filters	40, 80, 120, 160
Patch Size	13, 25, 39
Size of Mini Batch	16, 32, 64, 100, 128
Learning Rate	1e-6, 1e-7, 1e-8, 1e-9
Number of Epochs	Dataset 1: 500,750,1000,1250 Dataset 2: 750,1000,1250,1500
Input Band Type	4b, 10b, 13b, NDVI

5.5.3. Final Implementation

The final implementation uses the value out of experimented candidate values explained in section 5.5.3 and 5.5.4. The result of the final implementation is presented in section 6.7. Table 14 shows the final configuration of the proposed method to utilise the MTSI for generating the LCC maps that provide information about crop types.

Table 14. FCN final implementation setting

Parameter	Selected Value Dataset 1		Selected Value Dataset 2	
Network	FCN-SNet4.2	FCN-SubNet10.2.1	FCN-SNet4.2	FCN-SubNet10.2.1
Size of Filter	3	3	3	3
Number of Filters	40	40	40	80
Patch Size	39	25	13	13
Size of Mini Batch	100	16	128	16
Learning Rate	1e-7	1e-9	1e-7	1e-6
Number of Epochs	100	750	500	1500
Band Input	4b, NDVI	4b, NDVI	4b, NDVI	4b, NDVI

6. RESULTS AND DISCUSSION

This chapter presents the findings and discussion about the results derived from the experiments.

6.1. Initial Experiments

We apply classification using the standard approach, SVM, and we use the standard implementation of DTW approach as a comparison. We then apply the proposed network of FCN-SNet. Classification accuracies presented in Table 15 are generated by implementing SVM parameter setting as mentioned in Section 5.3 and FCN-SNet4.2 architecture as mentioned in Table 1.

Table 15. The classification accuracies of the initial experiments

Input	Methods	OA Dataset 1	OA Dataset 2
4 bands (2,3,4,8)	SVM	59.2	66.9
	FCN-SNet4.2	65.9	68.8
NDVI	SVM	57.4	56.7
	DTW	47.5	25.9
	FCN-SNet4.2	63.5	61.1

According to Table 15, the overall accuracies of all methods are less than 70%. We evaluate and examine the possible source of error contributed to the classification result. Confusion matrix for applying FCN-SNet4.2 on Dataset 1 and Dataset 2 are presented in Table 16 and Table 17. Individual class accuracies in UA and PA are provided in Table 18 and Table 19 (indicating by heading 'before').

Table 16. Confusion matrix Dataset 1 – FCN-SNet4.2

Class	Wheat	Maize	Sunflower	Forest	Water
Wheat	56707	16107	11751	183	167
Maize	12493	25246	10775	135	0
Sunflower	18341	11327	33629	2	0
Forest	1140	105	8	33306	0
Water	68	0	0	8	11068

Table 17. Confusion matrix Dataset 2 – FCN-SNet4.2

Class	AF	CR	DP	FA	LC	ON	OW	OH	SR	GS	SB	WW
AF	16339	52	51	192	205	0	0	292	0	293	27	210
CR	792	1718	10	250	0	1166	0	732	1	8	94	300
DP	85	9	6396	541	0	47	0	29	803	0	0	0
FA	186	809	123	7712	553	106	0	225	651	0	210	4
LC	153	463	0	651	2918	559	113	393	0	225	26	346
ON	151	165	28	394	388	5172	11	40	0	98	811	498
OW	0	0	78	178	0	0	1226	0	0	0	0	0
OH	1467	0	1	101	225	233	0	5333	0	1463	0	0
SR	7	3	460	922	0	1	0	225	1922	6	0	16
GS	1099	0	0	0	134	80	0	1154	0	3742	0	0
SB	1711	351	0	1	0	881	0	267	0	0	5762	0
WW	0	0	5	57	272	1160	0	310	16	0	0	2241

AF : Alfalfa
CR : Carrots
DP : Developed Area
FA : Fallow/Idle Cropland
LC : Lettuce
ON : Onions
OW : Open Water
OH : Other Hay/Non-Alfalfa
SR : Shrubland
GS : Sod/Grass Seed
SB : Sugar beets
WW : Winter Wheat

6.2. Dataset Refinement

We examine the confusion matrix of the classification results provided in Table 16 and Table 17. We notice that there is a larger confusion among three classes, i.e., wheat, maize and sunflower, than the other two classes, forest and water. We expect these three classes can be distinguished well as the other two classes. So, we decide to investigate various possibilities to refine the dataset before continuing further experiments and analysis. Because we do not have the ancillary data, such as additional ground measurement, available base map, or aerial photo with higher resolution, to increase the number of samples by adding polygons in the different location, we make a refinement for the reference dataset by evaluating existing samples and reselecting training and test samples. A detailed explanation of the dataset's refinement is presented in Section 5.2. Classification accuracies generated for this section are calculated from the result of applying the configuration of the FCN-SNet approach.

6.2.1. Refined Based on NDVI Value

Table 18 compares the individual class accuracies in UA and PA before and after refining samples for Dataset 1 by applying refinement explained in section 5.2.1.1. In Dataset 1, refining the samples based on NDVI values and reselecting the samples means that the number of training and test polygons are changing for classes wheat, maize and sunflower. However, the total polygons and the ratio of training and test samples are not changing.

Table 18 indicates that providing the refined samples by examining the NDVI value in a year influences the classification accuracies in Dataset 1, especially if we carefully look at the class of Wheat, Maize and Sunflower. These three classes have a significant increase of the UA in a range of 19.2% – 32.9%. Refinement in these three classes also affects the class Forest and Water which the UA increase by 3.5% and 0.7%. In general, the dataset refinement generates enhancement of the overall accuracies by 29.8%, from 65.9% to 95.7%.

Table 18. The result of before and after dataset refinement for the initial experiments – Dataset 1

Class	UA		PA	
	Before	After	Before	After
Wheat	66.8	99.7	63.9	97.5
Maize	51.9	96.3	47.8	89.8
Sunflower	53.1	72.3	59.9	93.7
Forest	96.4	99.9	99.0	99.2
Water	99.3	100	98.5	95.8

Table 19. The result of before and after dataset refinement for the initial experiments – Dataset 2

Class	UA		PA	
	Before	After	Before	After
Alfalfa	92.5	89.6	74.3	75.5
Carrots	33.9	44.4	48.1	66.4
Developed Area	80.9	84.7	89.4	87.6
Fallow/Idle Cropland	72.9	93.4	70.1	77.4
Lettuce	49.9	71.8	62.2	80.0
Onions	66.7	65.6	55.0	62.5
Open Water	82.7	83.9	90.8	97.4
Other Hay/Non-Alfalfa	60.4	51.6	59.3	53.8
Shrubland	54.0	58.0	56.6	67.0
Sod/Grass Seed	60.3	51.3	64.1	53.4
Sugar beets	64.2	65.5	83.1	83.1
Winter Wheat	55.2	50.9	62.0	61.0

For Dataset 2, there is an increase of OA from 68.8% to 71.7%. This improvement is reasonable since the plot of NDVI values after refining the Dataset 2 as displayed in Figure 19 and before refining as presented in Figure 18, does not show a clear distinction between them. However, Table 18 and Table 19 show an increase of UA for almost all classes after refinement. It means that samples refinement based on the NDVI value become a potentially effective way to improve the classification accuracies with no ancillary data.

In the proposed methods, the number of training samples is determined by the number of patches that are generated randomly. Furthermore, we also test the impact of an increasing number of training patches in Dataset 2 upon the classification results as presented in Table 20. It is not necessary to increase the number of patches since there is no difference in the classification results when different numbers of patches are used for training (e.g. 2000 patches versus 6000 patches). The number of test patches is half of the training patches, so the number of test patches increases when the training patches increase.

Table 20. Additional training patches experiments for Dataset 2

Number of training patches	2000 for 13 classes	4000 for 13 classes	6000 for 13 classes	@600 x 13 classes
OA	71.7	71.4	71.1	71.7
AUA	67.6	66.4	65.5	68.1
APA	72.1	70.9	69.1	68.8

6.2.2. Refined Based on Spatial Sampling Strategies

We perform the experiments to refine the samples based on the spatial locations to eliminate the potential error caused by the spatial distribution of samples (Congalton, 1991). We prepare several combinations as explained in Section 5.2.1.2 and 5.2.2.2.

Table 21 exposes the classification accuracies in different combinations of sampling strategies by varying the selection of spatial location. The overall accuracies for the ten different combinations of spatial sampling are in a range of 64.5% to 71.7% with average value is 68.0%. Since the difference of the accuracies after and before refined is relatively low in a range 1.4% to 5.8% in Dataset 1, it indicates the spatial distribution of the samples is not a significant source of the classification errors. It is also supported by the individual class accuracies (UA) from the five classes in ten combinations that all of them have a range of minimum and maximum UA below 25% (column Max-Min). In Dataset 1, these results also indicate the changing of training and test polygons ratio from about 12:88 to 50:50 does not significantly affect the classification accuracies.

Table 21. Classification accuracy of Dataset 1 by applying refinement in spatial location

Items		Before Refined	After Refined										
Reference data		C0	C01	C01s	C02	C02s	C03	C03s	C04	C04s	C05	C05s	Max-Min
OA		65.9	65.4	71.7	68.7	64.5	66.4	67.8	68.9	67.2	69.7	70.0	7.2
UA	Wheat	66.8	75.4	69.7	79.3	57.9	71.1	75.6	76.4	67.8	73.8	78.5	21.4
	Maize	51.9	45.3	57.8	52.0	51.6	57.8	46.9	52.2	55.8	55.1	58.5	13.2
	Sunflower	53.1	49.1	56.6	47.4	60.8	49.9	51.0	50.8	54.3	51.2	54.0	13.4
	Forest	96.4	99.8	94.2	94.8	97.0	94.7	97.7	97.1	98.1	98.1	96.0	5.6
	Water	99.3	96.5	97.3	92.3	98.4	93.1	92.7	98.2	93.6	95.3	96.3	6.1

Table 22 presents the individual class accuracies (UA) of Dataset 2 when varying the spatial selection of the samples. Alfalfa, Fallow, and Other Hay have a range of maximum and minimum UA below 25%. The other classes have differences more than 25% by varying the spatial distribution of the samples. Spatial distribution of samples in Dataset 2 implies that it has more influence on the classification accuracies compare to Dataset 1. However, both datasets imply similar behaviour in overall accuracy. Although there

is an influence on the classification accuracies in varying the spatial distribution of the samples, especially in certain classes in Dataset 2, the overall accuracies deviation is relatively low, 11.2% in Dataset 1 and 7.2% in Dataset 2. We infer that the spatial selection of the samples does not significantly contribute to the overall accuracies, but indeed influences the individual class accuracies.

Table 22. Classification accuracy of Dataset 2 by applying refinement in spatial location

Items		Before Refined	After Refined								
Reference data		C0	C01	C01s	C02	C02s	C03	C03s	C04	C04s	Max-Min
OA		68.8	61.9	65.6	63.7	55.4	64.6	64.9	67.1	59.1	11.7
UA	Alfalfa	92.5	86.3	81.7	89.7	78.8	82.8	86.3	79.0	84.9	10.9
	Carrots	33.9	39.2	32.0	0.0	5.5	10.4	13.3	45.5	18.1	45.5
	Developed Area	80.9	84.8	80.1	78.5	73.3	93.1	82.1	93.1	54.3	38.8
	Fallow/Idle Cropland	72.9	78.9	67.2	80.4	61.8	74.3	73.4	83.4	64.6	21.6
	Lettuce	49.9	48.7	56.8	27.4	35.4	64.3	31.6	34.3	50.9	36.9
	Onions	66.7	40.4	78.3	57.8	67.5	55.9	65.7	68.3	59.7	37.9
	Open Water	82.7	71.1	89.6	74.3	96.1	74.8	83.9	83.9	93.2	25.0
	Other Hay/Non-Alfalfa	60.4	49.0	49.2	36.9	39.3	35.0	54.1	48.3	45.3	19.1
	Shrubland	54.0	37.0	60.0	44.7	33.2	53.5	47.4	52.7	37.7	26.8
	Sod/Grass Seed	60.3	58.1	43.1	40.5	27.0	54.7	50.8	50.4	40.4	31.1
	Sugar beets	64.2	75.2	52.8	82.0	43.9	53.4	72.0	64.3	70.5	38.1
	Winter Wheat	55.2	49.9	40.6	55.5	31.1	49.7	40.1	27.7	62.7	35.0

6.2.3. Full Refined

We check further by applying both refinements based on NDVI value and spatial sampling strategies to test whether the previous finding remains valid. Table 23 and Table 24 provide individual class accuracies of the full refined dataset. We observe the maximum and minimum overall accuracies and individual class accuracies in both datasets.

Table 23. The result of the initial experiments applied to a full refined dataset – Dataset 1

Items		Before refined	After Refined										
Reference data		C0	C1	C1s	C2	C2s	C3	C3s	C4	C4s	C5	C5s	Max-Min
OA		65.9	96.1	97.2	95.5	97.3	96.5	96.9	96.9	96.7	97.2	96.5	1.8
UA	Wheat	66.8	99.5	98.7	98.1	100.0	99.1	99.4	98.7	100.0	99.6	98.6	1.9
	Maize	51.9	95.2	97.2	94.6	97.4	95.8	97.8	98.9	95.3	97.4	95.6	4.3
	Sunflower	53.1	77.5	87.3	80.8	81.8	85.3	79.9	85.4	80.3	80.2	85.5	9.8
	Forest	96.4	98.8	98.5	96.9	99.5	98.7	98.0	97.5	99.6	98.3	99.7	2.8
	Water	99.3	93.2	98.3	95.4	98.7	93.2	97.5	95.4	96.9	97.0	96.5	5.5

Table 23 provides records of overall accuracies on Dataset 1 that are not significantly different among the combinations. The difference is 1.8%. However, sunflower is the most affected class because the difference among combinations shows the biggest value of 9.8%. Table 24 provides the results of Dataset 2. It shows that the difference among combinations is 11.2%. This condition indicates that there is a higher influence of varying spatial distribution to Dataset 2 than Dataset 1. The biggest influence is shown for the class Carrots, indicated by the highest difference of its individual class accuracy 44.1%. This finding agrees with the aforementioned finding that spatial selection of the samples does not significantly contribute to the overall accuracies, but indeed influences the individual class accuracies.

Table 24. The result of the initial experiments applied to a full refined dataset – Dataset 2

Items		Before refined	After Refined								
Reference data		C0	C1	C1s	C2	C2s	C3	C3s	C4	C4s	Max-Min
OA		68.8	66.2	70.6	71.7	60.5	66.5	67.4	68.8	65.1	11.2
UA	Alfalfa	92.5	86.3	86.8	87.7	75.9	84.3	83.3	88.7	86.2	12.8
	Carrots	33.9	40.2	35.7	0.0	44.1	20.5	9.3	29.2	27.8	44.1
	Developed Area	80.9	74.3	86.0	84.2	75.1	84.7	90.0	86.7	66.1	23.9
	Fallow/Idle Cropland	72.9	84.0	74.2	83.9	64.2	78.2	84.6	88.6	82.5	24.4
	Lettuce	49.9	47.2	69.9	58.6	54.8	77.8	34.8	61.5	61.9	43.0
	Onions	66.7	52.3	85.1	78.2	64.7	54.1	73.0	64.8	70.2	31.0
	Open Water	82.7	50.5	88.3	84.1	96.4	76.9	92.6	87.4	83.4	19.5
	Other Hay/Non-Alfalfa	60.4	57.0	48.5	57.1	40.3	40.3	60.3	48.1	51.3	20.0
	Shrubland	54.0	42.4	55.9	45.6	30.4	49.9	41.7	45.2	37.6	25.5
	Sod/Grass Seed	60.3	62.0	42.8	51.9	38.6	60.1	56.5	60.8	43.9	22.2
	Sugar beets	64.2	75.4	56.2	79.1	53.7	61.4	65.7	68.4	62.4	25.4
	Winter Wheat	55.2	58.4	47.8	49.5	33.9	50.1	41.6	33.1	56.9	23.8

Since we do not have a special interest in a particular crop type, so we select final dataset based on the OA. For further experiments, Combination 2 is selected for a more practical reason. For example, if we need to process a new dataset with a similar characteristic to Dataset 1, we can use this trained network with Dataset 1 and directly apply classification for the new dataset that located separately to generate the classification map (pre-trained network approach). The composition of sample polygons for the selected dataset are described in Table 25 and Table 26.

Table 25. Number of polygons in Dataset 1 after dataset refinement in Combination 2

Class	Training	Test	Training (%)	Test (%)
Wheat	377	234	62	38
Maize	168	73	70	30
Sunflower	72	44	62	38
Forest	99	81	55	45
Water	43	37	54	46
Total	759	469	-	-

Table 26. Number of polygons in Dataset 2 after dataset refinement in Combination 2

Class	Training	Test	Training (%)	Test (%)
Alfalfa	103	92	53	47
Carrots	26	2	93	7
Developed Area	35	60	37	63
Fallow/Idle Cropland	50	47	52	48
Lettuce	30	9	77	23
Onions	47	28	63	37
Open Water	3	8	27	73
Other Hay/Non-Alfalfa	31	50	38	62
Shrubland	14	33	30	70
Sod/Grass Seed	16	28	36	64
Sugar beets	30	26	54	46
Winter Wheat	20	9	69	31
Total	405	392	-	-

6.3. Hyper-Parameter Tuning

After defining the final dataset for further experiments, we tune the parameter of SVM and FCN. DTW has no parameter to tune so it can directly be implemented.

6.3.1. SVM Parameters because those parameters are dependent on the used dataset.

Table 27 shows the combinations of parameter C and gamma that produce the best classification accuracy by applying SVM for Dataset 1 and Dataset 2. Both datasets used different parameter value out of the same range of candidate values because those parameters are dependent on the used dataset.

Table 27. Combination of SVM parameter that generates the best result

Parameter	Dataset 1		Dataset 2	
Input type	4b	NDVI	4b	NDVI
C	48329.3024	69519.2796	143.8450	1274.2750
gamma	0.6952	0.1000	6.1585	10.0000

We observe that Dataset 2 requires a higher gamma value to be able to discriminate the classes. We can infer that Dataset 2 is more difficult to classify because it has more classes compared to Dataset 1. Having more classes increases the possibility of having a similar spectral pattern along the temporal dimension, see Figure 12 and Figure 19. The higher value of C on NDVI input indicates that the dataset is generalised more than applied on four-bands input classification.

6.3.2. FCN-SNet Hyper-Parameters

This section reports the findings during the experiments to design FCN-SNet architecture. The OAs are recorded to measure the classification result when varying the hyper-parameter values.

6.3.2.1. Patch Size

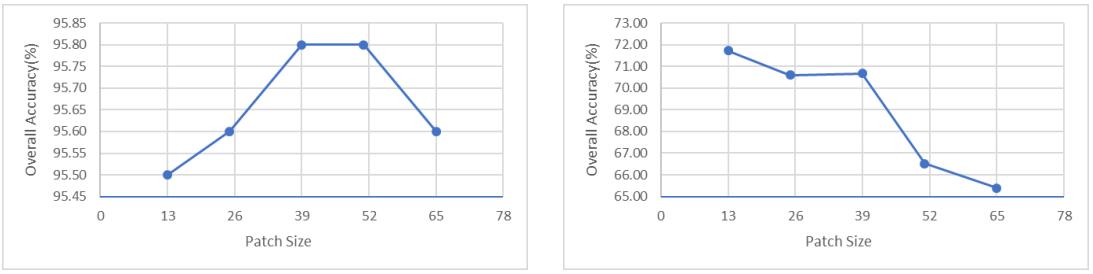
Patch size represents the spatial dimension of the considered training samples to assign the label for every pixel inside the patch (see section 4.4) for a given central pixel. Table 28 presents the influence of varying patch size on classification accuracy. We use size 13x13 as an initial value and make it larger by 2,3,4, and five times larger (use the odd number) to feed into FCN-SNet4.2 architecture. Size 13x13 is the smallest size of an effective receptive field on the architecture tested in the layer depth experiments.

Table 28. FCN-SNet experiments results – patch size

Patch Size	13x13	25x25	39x39	51x51	65x65
OA Dataset 1	95.5	95.6	95.8	95.8	95.6
OA Dataset 2	71.7	70.6	70.7	66.5	65.4

Figure 21 (a) shows that, for Dataset 1, increasing the patch size do not always lead to an increase of classification accuracies, despite the expectation that it makes an increase. Although the network considers larger area (dimension of the patch) to extract the information for predicting the label, it does not positively influence the classification results. Increasing patch size from 51x51 to 65x65 lead to degrading the classification accuracies. Patch size 39x39 and 51x51 are the obvious choices by considering the OA.

Figure 21 (b) shows a different trend in Dataset 2. The increase of patch size results in a negative impact on the classification accuracies. From this plot, size 13x13 is an obvious choice for implementation on Dataset 2.



(a) Dataset 1 (b) Dataset 2

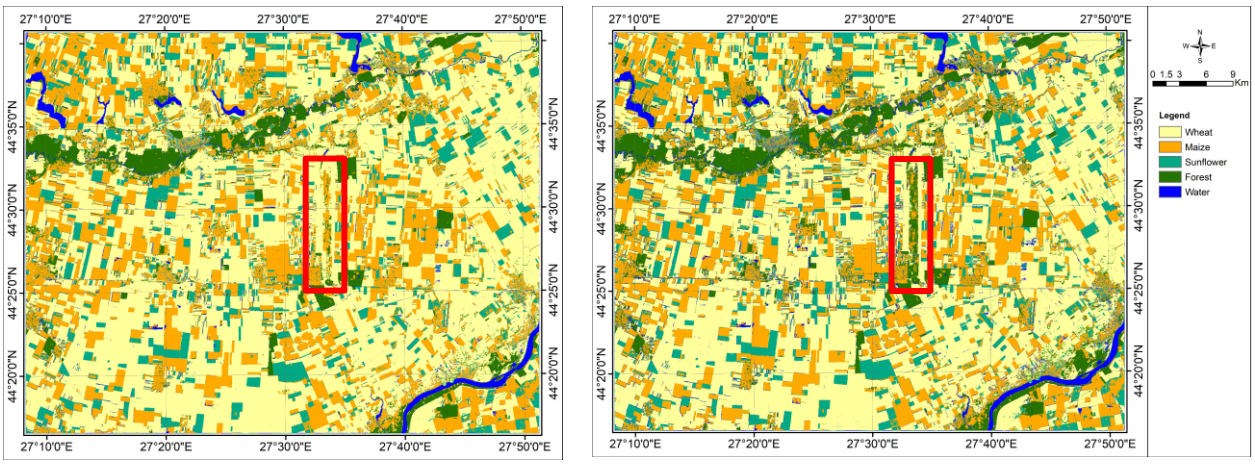
Figure 21. Effect of varying patch size on FCN-SNet

Later, we compare the average of UA (AUA), PA (APA) and F-Measure (AFM) for both patches of size to select the patch size of Dataset 1 for implementation.

Table 29. Classification accuracy comparison of patch size 39x39 and 51x51 of FCN-SNet – Dataset 1

Patch Size	OA	AUA	APA	AFM
39x39	95.8	94.1	93.2	93.7
51x51	95.8	93.9	93.4	93.6

By observing Table 29, the classification result of patch size 39x39 has a slightly higher value on AUA and AFM. Furthermore, we inspect the classification map for both patch size presented in Figure 22. See a detailed red box of both maps. It shows that the classification result of patch size 39x39 is more suitable because this area is not a forest (dark green colour). We compare it to the input image, and it seems that Figure 22 (b) is misclassified in the mentioned location.



(a) Classification map for patch size 39x39 (b) Classification map for patch size 51x51

Figure 22. A comparison for the classification map for patch size 39x39 and 51x51

Accordingly, we deduce that patch size 39x39 and 13x13 are the selected value for implementation on Dataset 1 and Dataset 2 respectively. This condition reveals that with the same spatial resolution data (10 m), the network possibly uses different patch size to solve the problems in classification. As mentioned before, Dataset 1 has five classes, and Dataset 2 has 12 classes. This might be the reason that there is no agreement in both datasets in term of patch size.

Although the crop field characteristics in both locations are different, where Dataset 2 has a smaller size of a crop field, we infer that it does not contribute to the different selection of the patch size. It happens because it does not need to consider the size of the crop fields to determine the label for a given pixel. For example, a pixel is labelled as a maize because of its spectral value in all time-step of images (temporal information), not because of its neighbourhood (if all its neighbourhood pixels are maize, then it must be a maize), nor the size of the crop field meet some specific size (if a pixel and its two pixels neighbourhood

are maize, it is considered as maize). Although, the network still needs more than one pixel to determine targeted classes because it is not sufficient to determine classes by looking at one pixel solely.

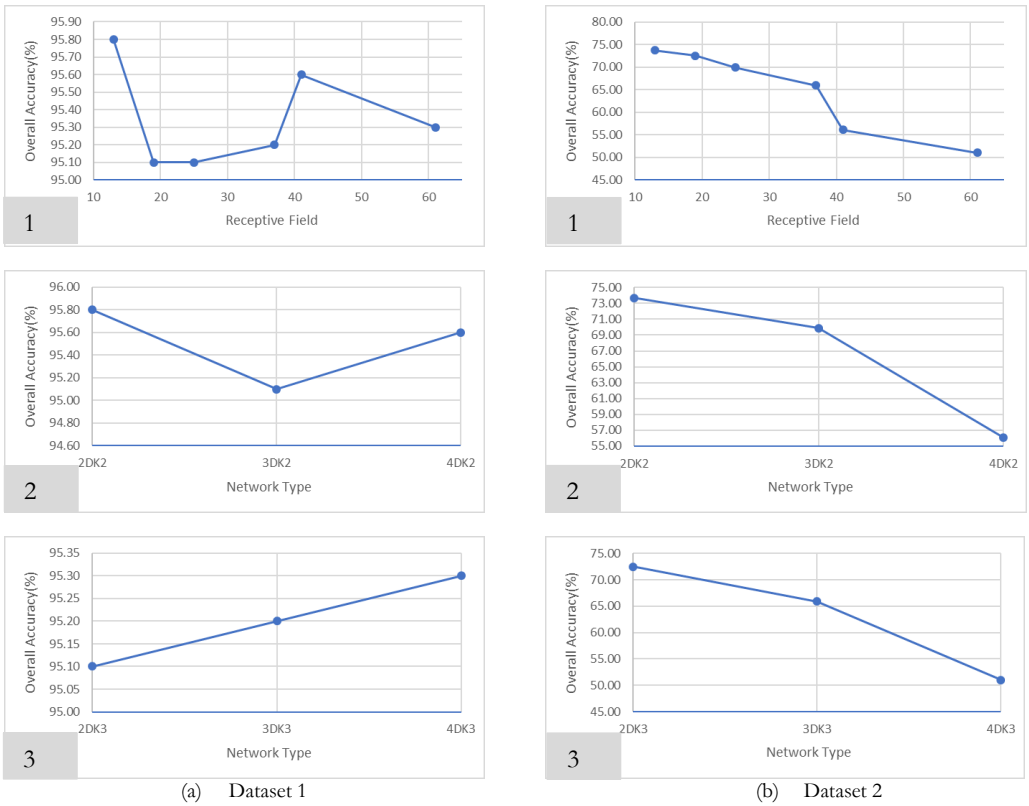
6.3.2.2. Layer Depth

Layer depth refers to the number and type of layers for processing the input to the output layer. We design six different architectures by varying the depth and a combination type of the layer. According to Table 30, the shallowest architecture network, FCN-SNet4.2, generates the highest accuracy for Dataset 1 and Dataset 2 compare to another architecture. The architecture of the layer is related to the size of the receptive field in a network that extracts the information from a given patch in the training process. Theoretically, a network with a receptive field that closes to or equals with the size of the patch means that the network includes the all pixels of a patch to learn during the training. An effective receptive field that larger than the size of the patch is not be useful in the learning process because the input does not provide information outside the patch. If we look back at the previous section, it is expected that the network FCN-SNet9.3 (RF = 37x37) would suit and produce highest classification accuracy for Dataset 1 (patch 39x39) and network FCN-SNet4.2 (RF = 13x13) suits Dataset 2 (patch 13x13). However, this result shows that network FCN-SNet4.2 (RF = 13x13) suits for Dataset 1 (patch size 39x39) and Dataset 2 (patch size 13x13).

Table 30. FCN-SNet experiments results – layer depth

Layer Architecture	4.2	6.3	6.2	9.3	8.2	12.3
Receptive Fields (RF)	13x13	19x19	25x25	37x37	41x41	61x61
OA Dataset 1	95.8	95.1	95.1	95.2	95.6	95.3
OA Dataset 2	73.7	72.5	69.9	65.9	56.1	51.0

If we look from the receptive field point of view, Figure 23-a.1 and b.1 show that there is no linear correlation between the increasing of receptive field and the classification accuracy. However, both datasets show a general trend of decreasing the classification accuracy if we change the receptive field form a small to a larger value.



(a) Dataset 1 (b) Dataset 2

Figure 23. Effect of the varying architecture of layer depth on FCN-SNet

Figure 23-a.2, a.3, b.2 and b.3 show that adding more layers tends to degrade the classification accuracy. Although we expect that layer addition (adding more parameters in the calculation) makes a more accurate calculation in predicting the label output, this assumption is only proven by the experiment of adding a block from 2 to 3 and 4 (in Figure 23-a.3). The other block additions are negatively correlated with the classification accuracy escalation. Based on this finding, we select FCN-SNet4.2 to implement for both datasets. The architecture of FCN-SNet4.2 is sufficient for the available datasets to produce classification maps with accuracy performance superior to the other architecture FCN-SubNet that has more complex architecture than FCN-SNet.

6.3.2.3. Number of Filters

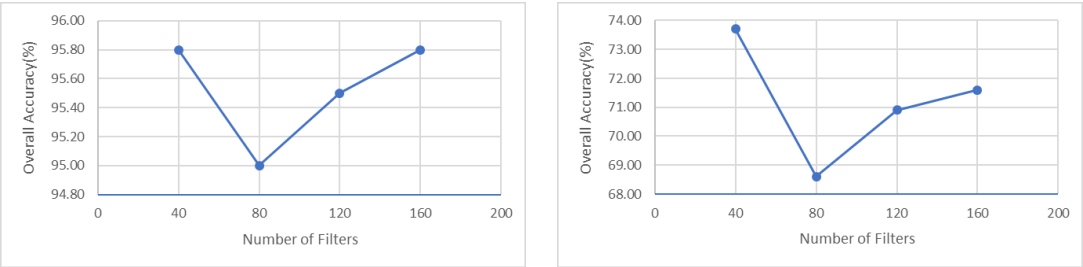
The number of filters defines the number of feature maps that are generated form convolution layers and further learned by the next layer in the network to predict the label of the input image. A larger number of filters means more feature maps that can be extracted. We start with the number of filters 40 because we expect that the generated feature maps could accommodate the depth of input images (stacking four bands with images in ten different time acquisition). So, the network produces adequate feature maps to be learned.

Table 31 shows the change of the classification accuracies by varying the number of filters which is started from 40, then makes it 2, 3, and 4 times larger.

Table 31. FCN-SNet experiments results – the number of filters

Number of Filters	40	80	120	160
OA Dataset 1	95.8	95.0	95.5	95.8
OA Dataset 2	73.7	68.6	70.9	71.6

Figure 24 shows that, on Dataset 1, there is a reduction from 95.8% to 95.0% of the OA when we increase the number of filters from 40 to 80, then OA starts to increase in a marginal value from 95.0%, for the number of filters 8, and it increases to 95.8% for the number of filters 120. While on Dataset 2, there is a similar trend where the OA decreases from 73.7% to 68.6% when the number of filters increases from 40 to 80, then continue to increase to 71.6% when the number of filters changes from 80 to 160. If we observe the trend in Figure 24, it might be possible for increasing OA for both datasets when we further increase the number of filters. However, an increasing number of filters lead to an increase in the number of parameters that affects computational time. Therefore we stop the experiments on the number of filters 160.



(a) Dataset 1 (b) Dataset 2

Figure 24. Effect of varying number of filters on FCN-SNet

We select the number of filters 40 to implement in Dataset 2. While we further check on other measures to select the implementation for the number of filters for Dataset 1. We compare the OA, AUA, APA and AFM of the number of filter 40 and 160 of Dataset 1 as presented in Table 32. We notice that AUA and AFM produced from the classification with 40 number of filters generates a slightly higher result than classification with 160 number of filters. Hence, we use 40 as the number of filters for the implementation of Dataset 1. This value is assumed as optimum for the network to learn and produce

comparable classification accuracies than using a larger number of filters. Network with 40 number of filters also has a fewer number of parameters to estimate during the training compared to a network with 160 number of filters.

Table 32. Classification accuracy comparison of the number of filters 40 and 160

Number of Filters	OA	AUA	APA	AFM
40	95.8	94.1	93.2	93.7
160	95.8	93.1	94.2	93.6

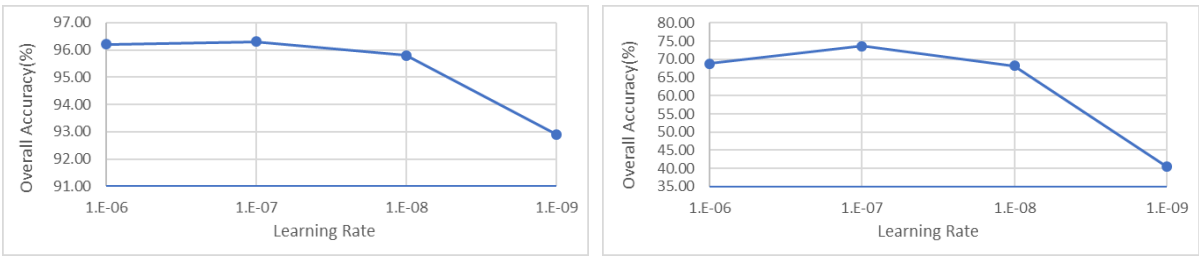
6.3.2.4. Learning Rate

We train the network with a fixed value of learning rate. We estimate the learning rate form the initial configuration and set from big to a low value for experiments. Selecting the optimum learning rate is important because we need to find a balance between the computational cost and the targeted accuracy. Learning rate is categorised as a training parameter which means that it depends on the trained dataset. The different dataset has a different variation value to treat so that it needs a different strategy to find out the optimal learning rate.

Table 33 shows the variation of classification accuracies by varying learning rate for Dataset 1 and Dataset 2. Varying learning rate from 1e-6 to 1e-9 generates a classification map with an accuracy range from 92.9% to 96.3% for Dataset 1 and from 40.5% to 73.7% for Dataset 2. According to Figure 25, both datasets show a similar trend where the OA starts to increase, peaks at learning rate 1e-7 then tends to decrease after that. Hence, we select the same learning rate 1e-7 for implementation to Dataset 1 and Dataset 2.

Table 33. FCN-SNet experiment results – learning rate

Learning Rate	1e-6	1e-7	1e-8	1e-9
OA Dataset 1	96.2	96.3	95.8	92.9
OA Dataset 2	68.8	73.7	68.2	40.5



(a) Dataset 1 (b) Dataset 2

Figure 25. Effect of varying learning rate on FCN-SNet

6.3.2.5. Number of Epochs

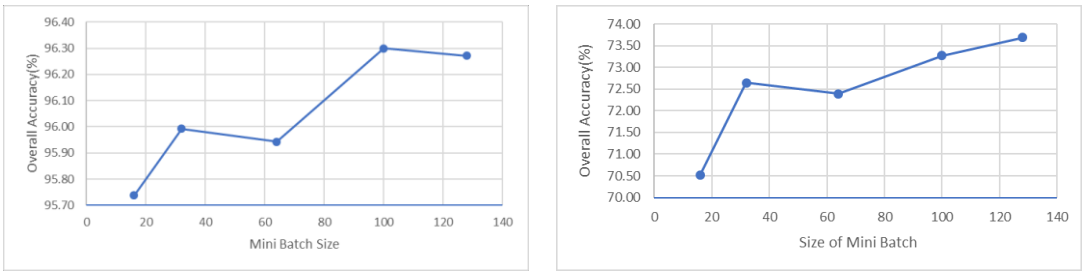
The number of epochs also depends on the used dataset. It is essential to select the adequate number of epochs along with the other training parameters to prevent under or overtraining. We can monitor the flattening of the error curve to assess the training process and decide when to stop the training. By observing the training curve, we set the number of epochs 100 for Dataset 1 and 500 for Dataset 2 to implement.

6.3.2.6. Size of Mini Batch

Table 34 presents the classification accuracies by varying the mini-batch size. According to Figure 26, the plot shows a positive correlation between the size of the mini batch and the overall accuracy. By considering the achieved overall accuracies, we select 100 for Dataset 1 and 128 for Dataset 2. The interaction between the learning rate and mini-batch size determines the number of epochs needed for the network to converge. Besides that, it defines training time per epoch and model quality.

Table 34. FCN-SNet experiments results – the size of a mini batch

Size of Mini Batch	16	32	64	100	128
OA Dataset 1	95.7	96.0	95.9	96.3	96.3
OA Dataset	70.5	72.6	72.4	73.3	73.7



(a) Dataset 1 (b) Dataset 2

Figure 26. Effect of varying size of mini batch FCN-SNet

6.3.2.7. Input Band Type

Based on Table 35, classification using NDVI input on Dataset 1 outperforms the classifications obtained using the reflectance value for these experiments. Even though it does not agree with the result on Dataset 2 that using four bands input generates the highest classification accuracy from all tested input type.

Table 35. FCN-SNet experiments results – the type of input band

Input Band Type	NDVI	4 bands	7 bands (MI)	10 bands	13 bands
OA Dataset 1	97.9	96.3	95.4	97.0	96.6
OA Dataset 2	63.5	73.7	70.7	71.4	70.9

The use of 10 bands (band 2, 3, 4, 5, 6, 7, 8, 8a, 11 and 12) of Sentinel-2 on Dataset 1 improves the classification accuracy from 96.3% to 97.0% compared to only use four bands (band 2,3,4 and 8) as input. By using all 13 bands of Sentinel-2 for classification, the results show a positive impact compared to classification with four bands. The additional nine bands increase the classification accuracies in a marginal amount from 96.3% to 96.6%. This result confirms the conclusion by Zhang, Su, Liu, & Chen (2019) who reveal that incorporating more related band information, in their case 13 bands of Sentinel-2 compare to mutual information (MI) based bands (band 3, 5, 6, 7, 8, 8a, and 9), can escalate the performance of classification in a marginal amount than using less spectral bands. However, it is surprising that using 10 bands generates better performance compared to classification with 13 bands. This result implies that three additional bands of Sentinel-2 (band 1,9 and 10) do not provide additional information to determine the targeted classes. Meanwhile, in Dataset 1, NDVI input generates the highest overall accuracy. This fact provides an indication that the refinement based on NDVI value gives a significant contribution to the classification result of Dataset 1.

Figure 27 and Figure 28 present the spectral plot of the targeted classes in 13 available bands of Sentinel-2. Both imply that band 10 (b10) do not provide any additional information because the spectral value of all classes is almost at the same value (low-class separability). Thereby band 10 is not a recommended band to use in determining the classes.

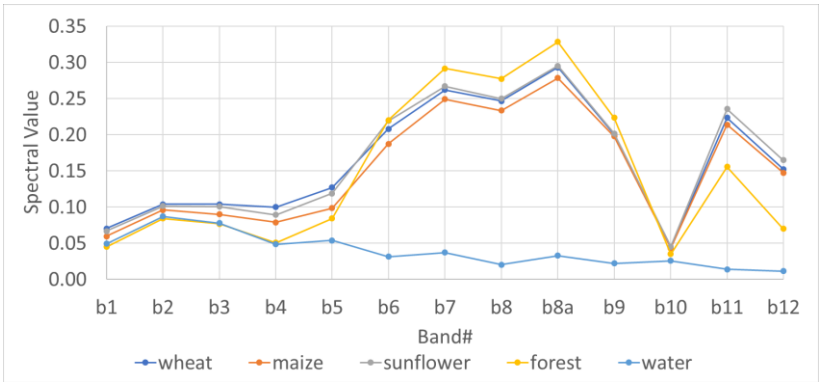


Figure 27. Spectral plot of samples from Dataset 1

Figure 27 shows that band 4, 5, 7, 8, 8a and 11 have a good separability indicated by the line separation on the plot for Dataset 1. While in Figure 28, we hardly see the separation of 12 classes in one of the bands. It indicates why the classification accuracies of Dataset 2 are generally lower than Dataset 1.

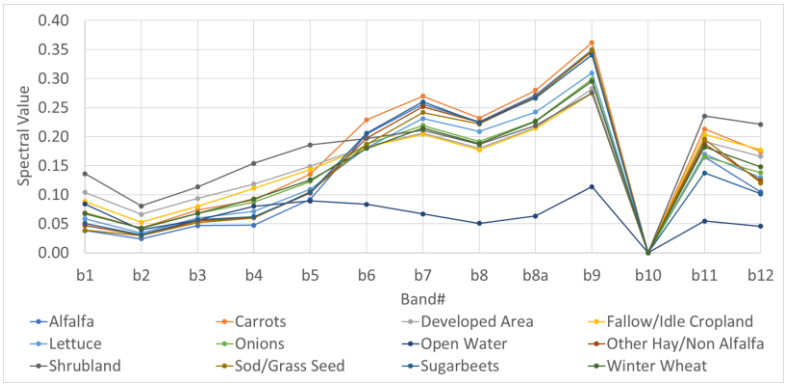


Figure 28. Spectral plot of samples from Dataset 2

6.3.3. FCN-SubNet Hyper-Parameters

This section reports the findings during the experiments to design FCN-SubNet architecture and the implementation. The OAs are recorded to measure the classification results when varying the hyper-parameter values. The selection of the FCN-SubNet hyper-parameter values are made by concerning the experiments of FCN-SNet hyper-parameter tuning

6.3.3.1. Patch Size

In FCN-SNet, both datasets are suited with a relatively small value of patch size and the OA tends to decrease at the moment the patch size is increased. Since FCN-SubNet use the same Dataset (same problems to solve) for the experiments, according to the result of applying the patch size 13x13, 25x25 and 39x39 as presented in Table 36, for efficiency, we stop increasing the patch size at the third candidate value. In the implementation, we select value 25x25 for Dataset 1 and 13x13 for Dataset 2.

Table 36. FCN-SubNet experiments results – patch size

Patch Size	13x13	25x25	39x39
OA Dataset 1 (%)	89.9	90.3	90.2
OA Dataset 2 (%)	61.7	56.3	57.2

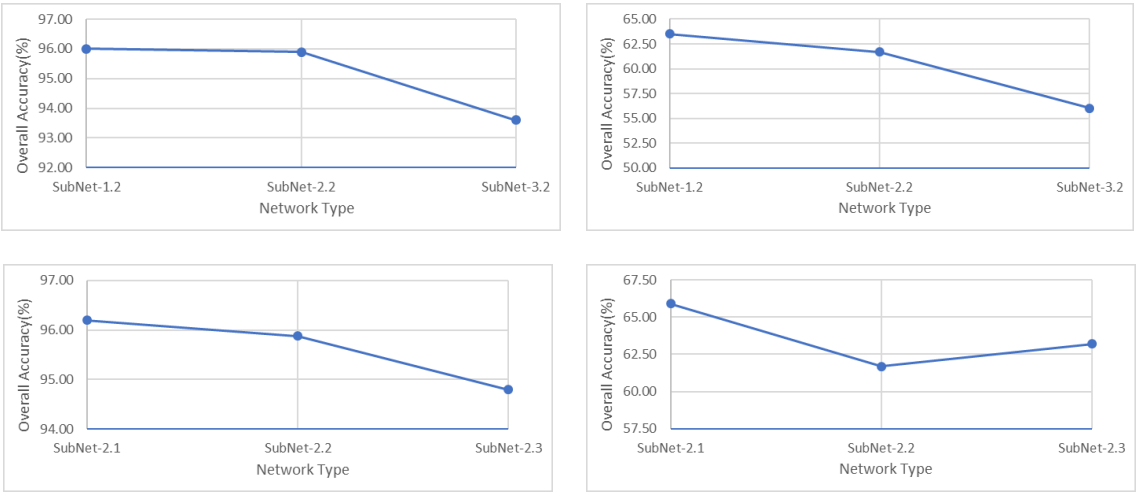
6.3.3.2. Layer Depth

Table 37 shows the tested candidate networks and the classification results. Architecture SubNet10.2.1 generates the highest classification accuracy than the other architectures for both datasets.

Table 37. FCN-SubNet experiments results – layer depth

Layer Architecture	SubNet10.1.2	SubNet10.2.2	SubNet10.3.2	SubNet10.2.1	SubNet10.2.3
OA Dataset 1 (%)	96.0	95.9	93.6	96.2	94.8
OA Dataset 2 (%)	63.5	61.7	56.0	65.9	63.2

From Figure 29 and recalling Figure 23, in designing the network for our dataset, both imply that additional layers tend to degrade the classification accuracies. This result is also in line with the result of the layer depth experiment of FCN-SNet as explained in Section 6.3.2.2. FCN-SubNet10.2.1 is a selected architecture to implement for both datasets.



(a) Dataset 1 (b) Dataset 2

Figure 29. Effect of the varying architecture of layer depth on FCN-SubNet

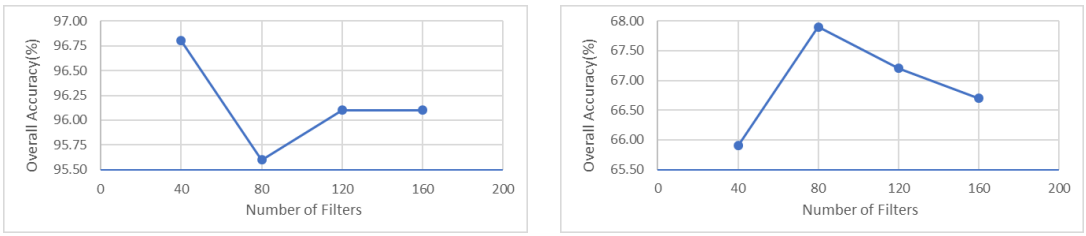
6.3.3.3. Number of Filters

Table 38 shows the classification results by varying the number of filters. Number of filters 40 has the highest OA for Dataset 1, while on Dataset 2, number of filters 80 generates the highest OA.

Table 38. FCN-SubNet experiments results – the number of filters

Number of Filters	40	80	120	160
OA Dataset 1 (%)	96.8	95.6	96.1	96.1
OA Dataset 2 (%)	65.9	67.9	67.2	66.7

Figure 30 indicates the behaviour of Dataset 1 and Dataset 2 in a different number of filters setting. Each of them indicates a different trend. Figure 30 (a) has the highest accuracy in the first candidate value. While in Figure 30 (b) second candidate value, number of filters 80 has the highest accuracy. These values are used to implement the FCN-SubNet on Dataset 1 and Dataset 2. These values are assumed as an adequate number of required feature maps to distinguish classes.



(a) Dataset 1 (b) Dataset 2

Figure 30. Effect of varying number of filters on FCN-SubNet

6.3.3.4. Learning Rate

By examining the training curve while applying the candidate values of the learning rate. We use 1e-9 for Dataset 1 and 1e-6 for Dataset 2 to implement. These values are different from the selected value of FCN-SNet architecture. This condition points out dependency of the learning parameter on the architecture of the network and the used dataset.

6.3.3.5. Number of Epochs

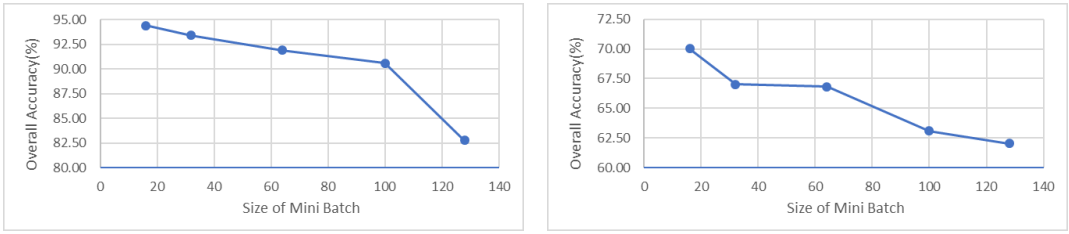
By examining the training curve, we set the number of epochs 750 for Dataset 1 and 1500 for Dataset 2 to implement. These values show that different architecture which is applied to the same dataset require a different number of epochs because it requires different time to convergence during the learning process. More complex architecture, FCN-SubNet, requires more time to converge compare to the simpler architecture as FCN-SNet architecture.

6.3.3.6. Size of Mini Batch

Table 39 describes the result in OA by varying the mini-batch size. For both datasets, size 16 has the highest OA. Figure 31 shows a completely different behaviour of increasing mini-batch size for classification accuracies compare with Figure 26. This condition implies that the more complex architecture, FCN-SubNet is more suitable with a smaller mini-batch size.

Table 39. FCN-SubNet experiments results – the size of the mini batch

Size of Mini Batch	16	32	64	100	128
OA Dataset 1 (%)	94.4	93.4	91.9	90.6	82.8
OA Dataset 2 (%)	70.0	67.0	66.8	63.1	62.0



(a) Dataset 1 (b) Dataset 2

Figure 31. Effect of varying size of mini batch FCN-SubNet

6.3.3.7. Input Band Type

Table 40 provides information about varying input bands to feed into the networks. Comparing to Table 35, we cannot see the agreement from these experiments on both datasets. The result shows a different behaviour. In these experiments, four bands input on both datasets produces higher OA that using NDVI input. However, it shows the different effect of using 10-bands and 13 bands input on both datasets. On Dataset 1, incorporating more bands to feed into the network tends to degrade the OA. While on Dataset 2, incorporating more bands gives a positive impact on the classification result indicated by the increasing OA by using 10 bands and slightly drops on 13 bands usage.

Table 40. FCN-SubNet experiments results – the type of input band

Input Band Type	NDVI	4 bands	10 bands	13 bands
OA Dataset 1	90.8	93.8	90.0	89.2
OA Dataset 2	56.4	70.0	70.4	70.1

6.4. Comparison of Final Implementation

Comparison for classification accuracies of final implementation to Dataset 1 and Dataset 2 are provided in Table 41 and Table 42. From those tables, the FCN-SNet4.2 confirms that this architecture can deal with a complex problem represented by Dataset 2. We also observe that DTW has the lowest OA compared to other methods. Although DTW considers the spectral and temporal information from the dataset, this method does not consider the spatial information. The standard implementation of DTW is applied to 1-Dimensional input, so we only used NDVI input. SVM, which also utilise the spectral and temporal information as DTW, produces 1% higher classification accuracy compares to DTW result on Dataset 2. FCN-SNet utilises the spatial, spectral and temporal information and produces a classification map with the highest classification accuracy compared to other methods. FCN-SubNet which also deals with that three information becomes a potential choice in the perspective of classification accuracies.

Table 41. Classification accuracies on the final implementation of Dataset 1

Input	Methods	OA	AUA	APA	F-Measure
4bands	SVM	90.6	86.0	82.6	83.4
	FCN-SNet4.2	96.3	94.9	94.1	94.5
	FCN-SubNet 10.2.1	93.8	89.4	92.3	90.7
NDVI	SVM	93.7	89.6	87.1	88.2
	DTW	92.7	89.4	92.4	90.6
	FCN-SNet4.2	97.9	97.5	96.7	97.1
	FCN-SubNet 10.2.1	90.8	86.1	91.9	88.2

Table 42. Classification accuracies on the final implementation of Dataset 2

Input	Methods	OA	AUA	APA	F-Measure
4bands	SVM	64.9	56.7	59.8	57.3
	FCN-SNet4.2	73.7	64.2	67.0	65.1
	FCN-SubNet10.2.1	70.0	61.0	62.7	61.3
NDVI	SVM	61.3	50.3	48.5	47.2
	DTW	29.3	34.0	38.4	25.8
	FCN-SNet4.2	63.5	57.2	51.9	51.4
	FCN-SubNet10.2.1	56.4	49.0	51.0	48.9

From the perspective of the SVM method, all the classification result confirms this method works well in a simple or more complex classification problem, as on Dataset 2 where several classes have an overlapping profile. Although SVM uses fewer hyper-parameter to tune than FCN methods, it is able to generate a baseline accuracy comparable to the FCN. This result provides additional fact to use SVM as baseline methods.

Table 42 provides classification accuracies of Dataset 2 that are generally lower from Dataset 1. Although Dataset 2 use the same images of Sentinel-2, it had different classes and use a different source of reference data. Reference data of Dataset 1 is a ground measurement without information of the accuracies, and for Dataset 2 we derive the reference data from reference map with accuracies about 85% and spatial resolution of the reference map is 30m. To minimise the issue of the reference data accuracy, we derive the reference pixels as explained in Chapter 4 through polygon generation based on the images and attributing the label based on reference data.

DTW implementation depends on the reference pattern that we need to provide. If we look at the plot of the reference pattern, Dataset 2 indicates that it is hard to separate the classes because many classes have similar NDVI value for every time-step (Figure 33). While for Dataset 1, it can be presumed that information from time 6-10 positively contributes to the class separation (Figure 32). Although the DTW does not consider the spatial information and only use limited spectral information, it can achieve 92.7% OA in Dataset 1 because the temporal information supports additional information to classify the targeted classes. However, the provided temporal information on Dataset 2 seems to be not sufficient to support

the classification for 12 classes. Based on Table 42, on the same condition where the classification only relies on the NDVI value and temporal information, SVM shows better performance.

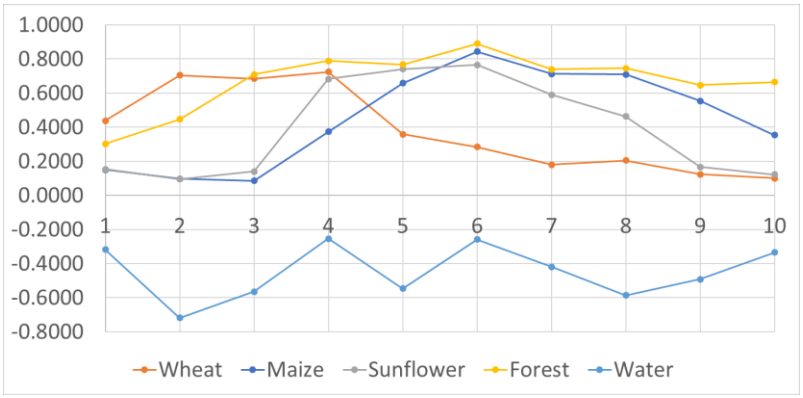


Figure 32. The plot of reference pattern for Dataset 1

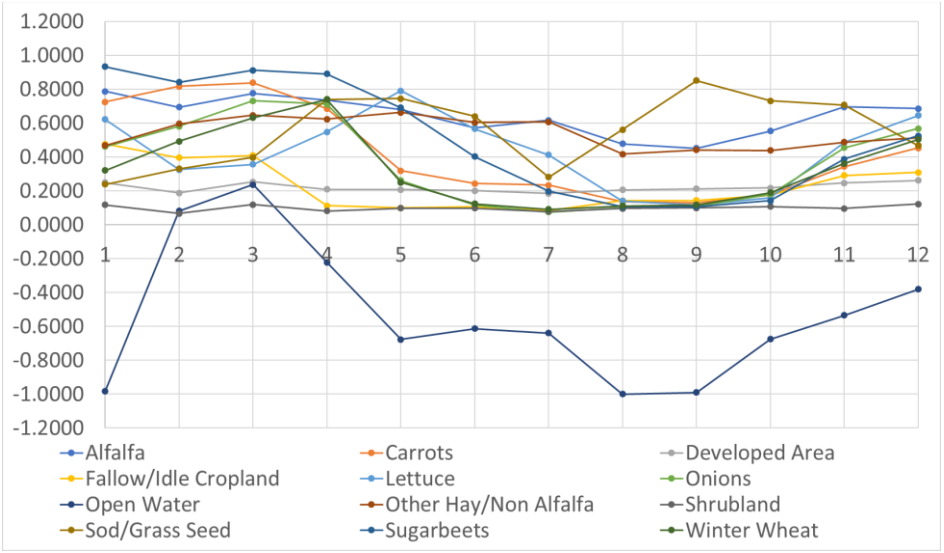


Figure 33. The plot of reference pattern for Dataset 2

Table 43 provides the individual class accuracies for Dataset 1. From the user's accuracies, all methods show that sunflower has the lowest accuracy. SVM only have about 55.4% of sunflower from the classification map is real sunflower in the ground. In the perspective of producer's accuracy, SVM has only 69.4% of sunflower from reference data correctly classified as sunflower. However, FCN-SNet and FCN-SubNet are able to perform better by achieving 81.2% and 80.0%.

FCN-SNet has the highest UA of all classes except for wheat, where FCN-SubNet achieves a better accuracy. Meanwhile, it indicates that the classification map of FCN-SNet has more misclassified pixels in forest and water compared to FCN-SubNet.

Table 43. The accuracies of individual classes of four bands input -- Dataset 1

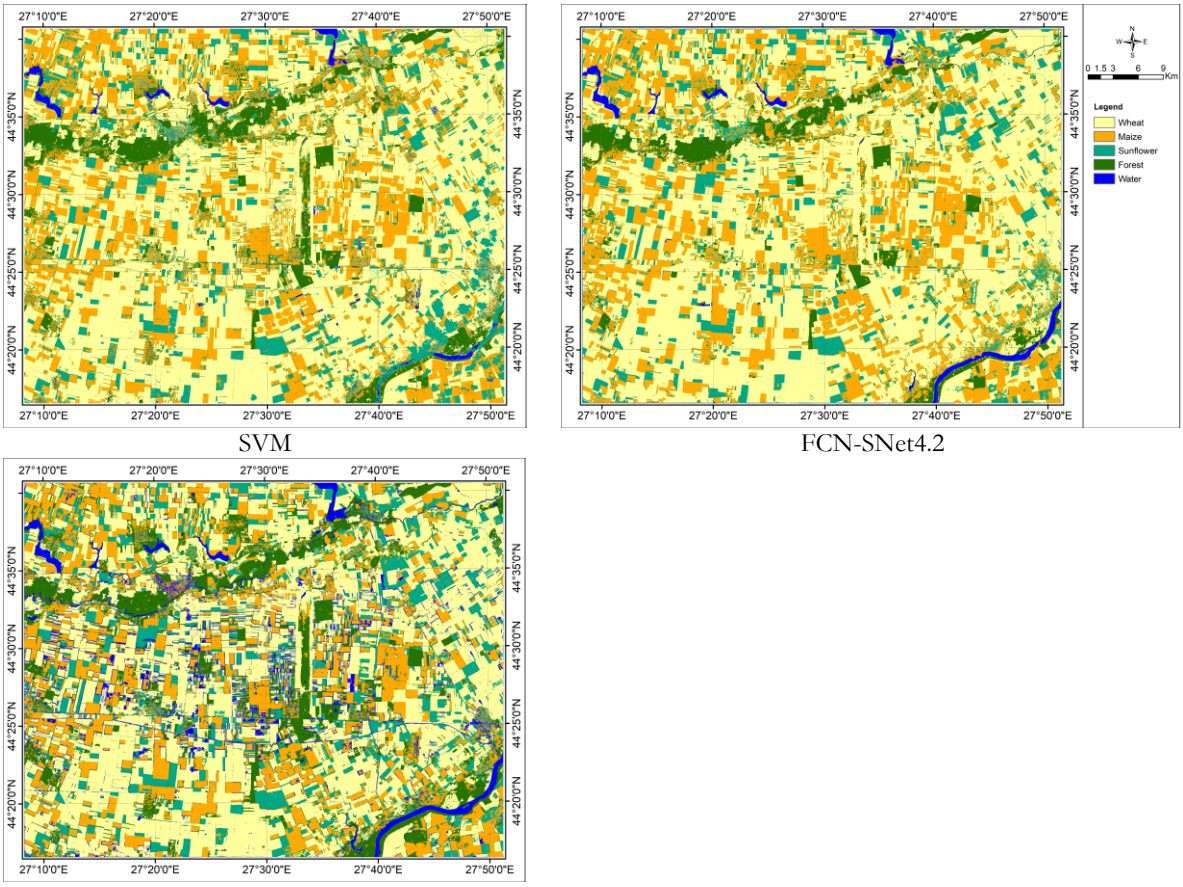
Class	UA			PA		
	SVM	FCN-SNet4.2	FCN-SubNet10.2.1	SVM	FCN-SNet4.2	FCN-SubNet10.2.1
Wheat	97.8	98.1	100.0	97.6	99.1	95.6
Maize	85.0	91.7	88.7	89.0	91.9	88.2
Sunflower	55.4	86.0	80.8	69.4	81.2	78.5
Forest	98.5	99.0	93.9	97.8	98.9	99.3
Water	93.1	99.8	83.5	59.2	99.5	100.0

Table 44 provides the individual class accuracies for Dataset 2. All methods are failed to classify Carrots. SVM has Other Hay as the most misclassified class, while FCN-SNet has shrubland and FCN-SubNet has winter wheat. The available samples from datasets are not sufficient to classify Carrots correctly (see Table 24 and Table 26).

Table 44. The accuracies of individual classes of four bands input -- Dataset 2

Class	UA			PA		
	SVM	FCN-SNet4.2	FCN-SubNet10.2.1	SVM	FCN-SNet4.2	FCN-SubNet10.2.1
Alfalfa	83.3	89.5	86.0	78.2	78.0	77.0
Carrots	0.0	0.0	0.0	0.0	0.0	0.0
Developed Area	73.3	86.0	92.1	85.5	94.4	78.3
Fallow/Idle Cropland	72.4	80.8	76.9	52.9	80.7	75.9
Lettuce	59.2	45.6	39.1	66.8	60.5	66.7
Onions	66.0	77.8	63.8	71.6	79.3	82.7
Open Water	86.6	96.0	96.5	99.6	99.2	87.0
Other Hay/Non-Alfalfa	50.0	61.1	54.5	28.7	47.8	51.5
Shrubland	32.3	53.7	61.8	45.5	46.9	60.3
Sod/Grass Seed	45.4	52.2	55.0	74.3	81.1	56.3
Sugar beets	74.9	77.8	75.1	70.7	73.5	76.1
Winter Wheat	36.7	50.4	31.5	44.0	62.2	40.6

Aside from the quantitative measure for the classification result, we do the qualitative measure by inspecting the classification maps. Figure 34 shows that FCN-SubNet generates more misclassified area with water coverage compare to the other two maps.



SVM FCN-SNet4.2

FCN-SubNet

Figure 34. Classification map of SVM, FCN-SNet4.2, FCN-SubNet for 4 bands input on Dataset 1

Figure 35 displays the reference map and classification maps of Dataset 2 by using four bands input. Square red boxes indicate some locations of misclassified pixels compared to the reference map.

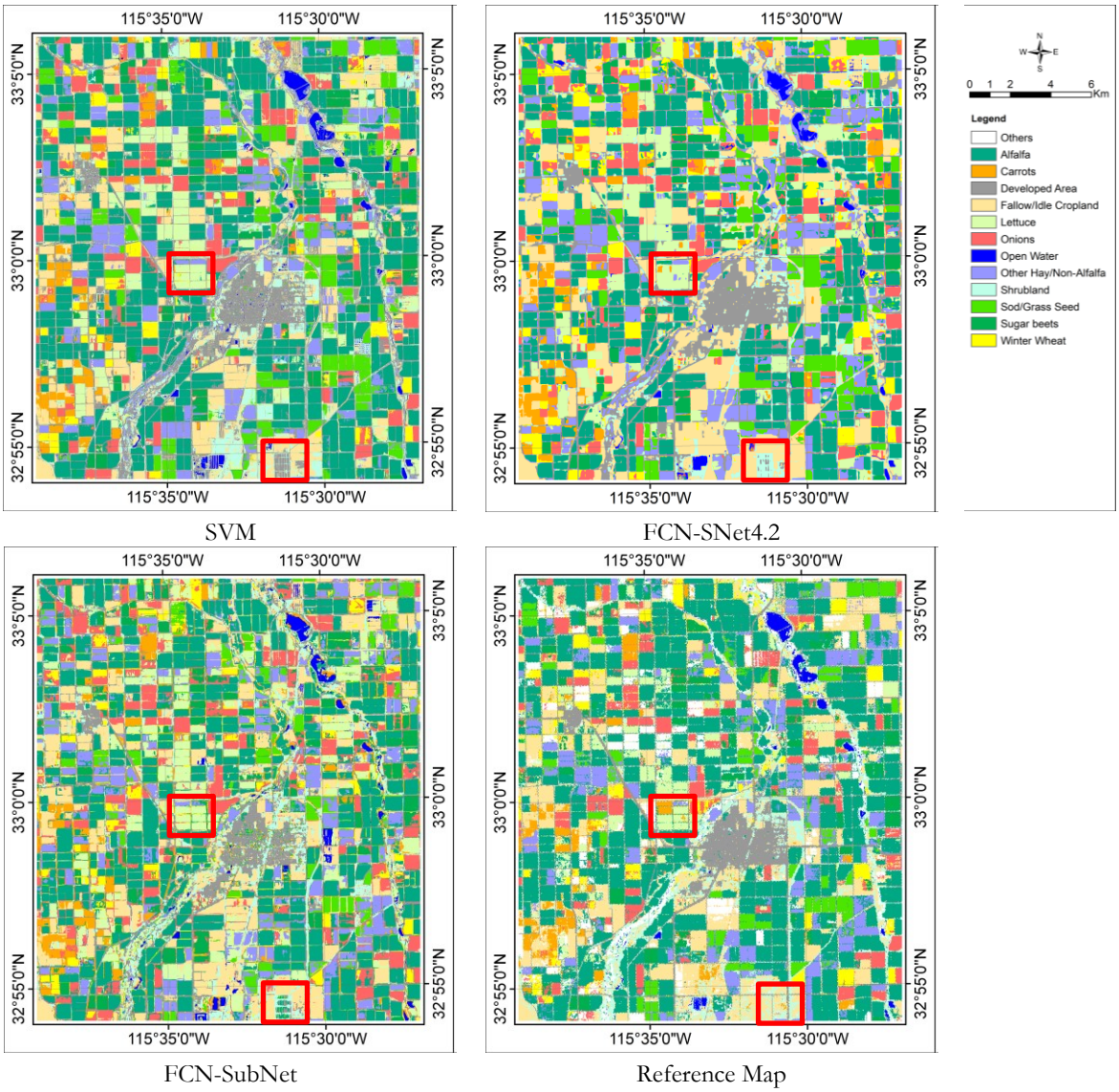


Figure 35. Classification maps of SVM, FCN-SNet4.2, FCN-SubNet of four bands input on Dataset 2

In term of computational time, as presented in Table 45, FCN-SubNet 10.2.1 can be an optional method to FCN-SNet4.2 than using DTW. Although it uses more complex architecture and needs longer processing time compared to FCN-SNet4.2, on Dataset 2, FCN-SubNet 10.2.1 produces higher classification results than DTW with NDVI input and also higher OA than SVM with four-bands input. The experiments are performed on a Laptop with an Intel Core i7-7700HQ CPU 2.80GHz, NVIDIA Quadro M1200, and 48GB of RAM.

Table 45. Estimation of processing time on Dataset 1

Methods	Input	Estimation of Processing Time
SVM	Dataset 1, 4 bands	125 minutes
DTW	Dataset 1, NDVI	4 days
FCN-SNet4.2	Dataset 1, 4 bands	40 minutes
FCN-SubNet10.2.1	Dataset 1, 4 bands	22 hours

6.5. Information Extractor

Recalling the research objective to investigate a network that exploits spatial, spectral and temporal information simultaneously, the applied methods utilise the available information in different ways. For the proposed method of FCN-SNet, it extracts the available information by using the following components:

- Spatial-information extractor
  - Patch size
  - The spatial dimension of the filter
  - Effective receptive field
- Spectral and temporal information extractor
  - Number of filters
  - The depth of the input image
  - Input band type

We carry out experiments that indicate the effect of removing some information from the Dataset.

1. Spatial information

   The effect of removing the use of spatial information can be inferred from Table 41 and Table 42 where FCN is the only method that exploits the spatial information by defining the value for the spatial-information extractor, i.e., patch size, size of the filter and effective receptive field. The OA calculated for SVM and DTW are generated from the classification by utilising only the spectral and temporal information. Table 41 denotes that the OA of Dataset 1, with NDVI input, drops to 92.7% and 93.7% from 97.9% if we remove the spatial information extraction. While for Dataset 2, with NDVI input, Table 42 shows that there is a reduction from 63.5% to 61.3% and 29.3%.

2. Spectral information

   Table 46 provides classification accuracies, in OA, by applying the proposed architecture FCN-SNet4.2 with single spectral information and full temporal information (images from all different acquisition time).

Table 46. Classification accuracies by using single spectral information

Band number	Dataset 1	Dataset 2
1	87.7	60.8
2	88.9	62.5
3	85.9	60.6
4	90.6	63.4
5	92.1	61.5
6	94.0	61.8
7	94.9	61.4
8	94.9	60.9
8a	94.9	61.7
9	92.3	57.7
10	90.7	25.6
11	94.7	63.6
12	97.3	65.3

3. Temporal information

   Effect of reducing some temporal information from the input is tested by performing classification using a single time acquisition image. We experiment by using the proposed architecture FCN-SNet4.2.

   From Table 47, the OAs on Dataset 1 vary from 81.0% to 89.3%. The maximum overall accuracies that can be obtained by a single time image for classification is 89.3%. While on Dataset 2, the OAs vary from 43.2% to 58.9%. The maximum OA achieved by a single time image for classification is 58.9%.

Table 47. Classification accuracies by using single time acquisition image

Input Image Dataset 1	OA	Input Image Dataset 2	OA
Date 1: 20170307	86.2	Date 1: 20170101	55.1
Date 2: 20170403	86.2	Date 2: 20170220	50.6
Date 3: 20170503	89.3	Date 3: 20170302	55.1
Date 4: 20170605	85.2	Date 4: 20170421	52.5
Date 5: 20170622	82.7	Date 5: 20170521	58.9
Date 6: 20170630	89.2	Date 6: 20170620	58.4
Date 7: 20170722	84.9	Date 7: 20170710	49.9
Date 8: 20170801	81.0	Date 8: 20170819	43.2
Date 9: 20170819	85.4	Date 9: 20170918	48.5
Date 10: 20170903	86.5	Date 10: 20171023	44.0
-	-	Date 11: 20171122	44.6
-	-	Date 12: 20171222	51.5

From these three experiments, we summarise the classification results in Table 48. Both datasets indicate the same behaviour where:

- including all information: it provides the highest accuracy,
- excluding temporal information: it is most influential on the classification results (producing the lowest accuracy),
- excluding spatial information extraction: it affects classification results in second place, and
- using only single spectral information: it is least influential on the classification result (OA drops but still higher than the result of excluding temporal or spatial information extraction).

Table 48. Classification accuracies by varying the use of spatial, spectral and temporal information

Parameters	To be included?			
Extract spatial information: applying FCN-SNet4.2	No	Yes	Yes	Yes
Extract spectral information: using NDVI or 4 bands	Yes	No	Yes	Yes
Extract temporal information: use all image from all time acquisition	Yes	Yes	No	Yes
Maximum generated OA (%) – Dataset 1	93.7	97.3	89.3	97.9
Maximum generated OA (%) – Dataset 2	64.9	65.3	58.9	73.7

These results confirm that applying land cover classification by using multi-temporal images classification provides an improvement in the classification accuracy compared to only use mono-temporal image. It is also in line with the conclusion by Sharma, Liu, & Yang (2018) when applying classification to Landsat-8 using a patch-based recurrent neural network.

7. CONCLUSION

7.1. Concluding Remarks

We develop two main architectures of FCN to classify land cover that contains crops information using MTSI. These architectures are expected to deal with the classification problems coming from the targeted objects, i.e., agricultural areas, and the multi-temporal input images. The optimal classification results are achieved by FCN-SNet4.2 architecture with four blocks of dilated convolutional layers and stacks the spectral and temporal information in the third dimension as the input images. According to the qualitative and quantitative measures applied to the classification results, FCN-SNet4.2 architecture performs better than the other two popular classification algorithms such as SVM and DTW. FCN-SNet4.2 architecture is also superior to the FCN-SubNet10.2.1 architecture that deals with the temporal information of MTSI in different ways. This result also points out the importance of designing the proper and adequate architecture to deal with the dataset.

In term of pre-processing, FCN has an advantage over the DTW approach. FCN approach directly uses the NDVI or spectral image to feed into the network and extracts the learning features automatically. However, DTW needs an additional step to provide a reference pattern by defining a single profile for each class to be further used in the classification. Moreover, FCN-SNet4.2 architecture significantly outperforms to SVM and DTW in terms of computational time.

7.2. Answers for The Research Questions

Each of the sub-objectives has been achieved by answering the following questions:
Questions for sub-objective 1:

a. What are the existing NN approaches that have been applied for crops classification using MTSI?

   Section 2.1 explains the related work on crops classification using MTSI. In the neural network methods, CNN is widely used for various purposes, including crops classification. There are variations of CNN such as FCN and LSTM that are also used for MTSI.

b. What is the most suitable design for crops classification using MTSI that exploits spatial, spectral and temporal information simultaneously?

   We studied the existing NN approaches and developed the methods that are expected to exploit better the MTSI in providing accurate LCC map. We experimented with the proposed methods to achieve the optimal design to be implemented. Based on the explanation in Section 6.4, FCN-SNet4.2 is an optimal design to deal with Dataset 1 and Dataset 2 used in our study. This architecture is designed by carrying out design experiments to vary the hyper-parameter value with candidate value. The setting for the experiments are explained in Section 5.5.1, and selected value for the hyper-parameter is presented in Section 6.3.2. Layer construction of this design is presented in Table 1.

   Based on Section 6.4, FCN-SNet4.2 generates higher classification accuracy compared to the other tested methods. FCN-SNet4.2 exploits the spectral and temporal information by concatenating the input image as Figure 4 and spatial information extracted by using a convolutional layer.

   Different architectures, i.e., FCN-SubNet also exploits the spatial information by using the convolution layer, but it uses a different way to handle the temporal information by providing sub-network and concatenating extracted feature maps from each sub-network to gradually generate the classification map. Spectral information is extracted in a standard way in sub-network design. However, if we look further to the computational time as shown in Table 45, FCN-SNet4.2 become the best choice. In addition, we explain the correlation of the FCN components to the information extraction in Section 6.5.

Questions for sub-objective 2:

a. What is the appropriate structure of an input file for performing classification using the proposed network?

   Figure 4 shows the structure of the input image for FCN-SNet4.2 architecture by concatenating the images in spectral then temporal dimension into one file and feed this file into the network. Handling the temporal information in a different way as implemented in FCN-SubNet approach does not improve the classification result. FCN-SubNet treats the temporal information separately in the beginning by separating the stream for every date of acquisition of the image.

b. What are the optimal hyper-parameters values for the proposed network to be used for performing crops classification using MTSI?

   Section 5.5.3 presents the selected hyper-parameter value to implement. Those values are the optimal hyper-parameters values for Dataset 1 and Dataset 2. The differences of the selected values for the different datasets and different networks indicate the hyper-parameters values depend on the used dataset input. These selected values can be used as an approximation value to initialise crops classification for other datasets.

c. How significant are the contributions of the spatial, spectral and temporal information for the classification result?

   Section 6.5 provides the experiment results to answer this question. Table 48 presents a summary of the experiments. Based on the classification results, temporal information has proven to have the most significant contribution to the classification accuracies. Our experiments also reveal that single-date images obtain less satisfactory results.

   By excluding the spatial information, we obtain less accurate classification results as well. We assume that the spatial information contributes less to the classification accuracy because crops identification is not highly dependent on very large neighbourhood pixels. In addition, incorporating multi-spectral information outperforms the classification obtained by using a single-based image as input.

d. What assessment and evaluation are relevant to measure the performance of the proposed network?

   Section 3.3 explains the assessment and evaluation used to measure the performance of the classification quantitatively (by calculating Overall Accuracy, User's Accuracy, Producer's Accuracy, F-Measure) and qualitatively (by inspecting the classification map visually).

Questions for sub-objective 3:

a. Which method performs better based on the performance assessment?

   According to Section 6.4, we deduce that the FCN-SNet4.2 performs better than other evaluated methods after applying the classification evaluation and accuracy assessment.

b. What aspect of the method that contributes to the classification result?

   Based on Section 6.3.2.7 and 6.3.3.7, FCN-SNet can be used for different type of input, i.e., NDVI and reflectance images. For reflectance images with multi-spectral information, it requires a specific structure of the input images (stacking) as shown in Figure 4 which is different from the FCN-SubNet input images. By concatenating the spectral and temporal information in the third dimension of the input images in combination with an adequate number of filters to extract the feature maps, we could obtain satisfactory results compare to the other methods. In the case of DTW, we could only provide one spectral reference pattern to apply the method. Spectral information is limited to the NDVI data. It also ignores the spatial information that is considered by the FCN approach. Similar to DTW, SVM only utilises the spectral and temporal information of the input image while ignoring the spatial information.

   Besides that, the flexibility of FCN in defining the hyper-parameter values is the other aspect that contributes to classification result. SVM has a less hyper-parameter value to set, and DTW has no hyper-

parameter value. Furthermore, our study reveals that providing the proper number of samples for training and test is important for all methods in the pre-processing stage.

Additional

a. How to optimise reference dataset as input for the network?

There are several ways to optimise the dataset:

- Inspect the temporal pattern by using NDVI value
- Checking for the spatial distribution of the samples
- Adequately divide reference data into training and test samples
- Increasing the number of samples by adding more sample polygons and/ or training patches from the available reference data

7.3. Recommendation

We summarise some suggestions for future research:

- FCN-SubNet, regarding the complexity, will be more useful if we want to generate a classification map for every acquisition image or do change detection for it and produce the annual classification map.
- Using different setup for the sub-networks, such as use the spectral information as sub-network in FCN-SubNet may useful to utilise different temporal information provision.
- Inspect the effect of the spatial resolution of the input images to the patch size.
- To improve the classification accuracies, add the number of polygons samples by using ancillary data, such as additional ground measurements, available base map, aerial images with higher resolution. These additional data are used to increase the number of training samples without reducing the test samples and vice versa. In this condition, we need to pay attention to the quality of the ancillary data to prevent adding more data that is not correct.

LIST OF REFERENCES

Baumann, M., Ozdogan, M., Richardson, A. D., & Radeloff, V. C. (2017). Phenology from Landsat when Data is Scarce: Using MODIS and Dynamic Time-Warping to Combine Multi-Year Landsat Imagery to Derive Annual Phenology Curves. *International Journal of Applied Earth Observations and Geoinformation*, *54*, 72–83. https://doi.org/10.1016/j.jag.2016.09.005

Belgiu, M., & Csillik, O. (2018). Sentinel-2 Cropland Mapping Using Pixel-Based and Object-Based Time-Weighted Dynamic Time Warping Analysis. *Remote Sensing of Environment*, *204*, 509–523. https://doi.org/10.1016/j.rse.2017.10.005

Bergado, J. R., Persello, C., & Stein, A. (2018). Recurrent Multiresolution Convolutional Networks for VHR Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(11), 6361–6374. https://doi.org/10.1109/TGRS.2018.2837357

Bittner, K., Cui, S., & Reinartz, P. (2017). Building Extraction from Remote Sensing Data Using Fully Convolutional Networks. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* (Vol. 42, pp. 481–486). https://doi.org/10.5194/isprs-archives-XLII-1-W1-481-2017

Bruzzone, L., & Persello, C. (2009). Approaches Based on Support Vector Machine to Classification of Remote Sensing Data. In *Handbook of Pattern Recognition and Computer Vision* (pp. 329–352). World Scientific. https://doi.org/10.1142/9789814273398_0014

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 1–27. Retrieved from http://www.csie.ntu.edu.tw/~cjlin/libsvm

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, *37*(1), 35–46. https://doi.org/10.1016/0034-4257(91)90048-B

Congalton, R. G., & Green, K. (2010). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. *The Photogrammetric Record* (Second, Vol. 25). CRC Press. https://doi.org/10.1111/j.1477-9730.2010.00574_2.x

Earth Observation Portal. (2014). Copernicus: Sentinel-2 — The Optical Imaging Mission for Land Services. Retrieved November 18, 2018, from https://earth.esa.int/web/eoportal/satellite-missions/c-missions/copernicus-sentinel-2

Encyclopædia Britannica. (2018). Romania. Retrieved November 11, 2018, from https://www.britannica.com/place/Romania/Agriculture-forestry-and-fishing

EUMeTrain. (2010). Monitoring Vegetation from Space. Retrieved June 5, 2018, from http://www.eumetrain.org/data/3/36/print.htm#page_3.2.0

European Space Agency. (2018a). Sen2Cor | STEP. Retrieved November 11, 2018, from http://step.esa.int/main/third-party-plugins-2/sen2cor/

European Space Agency. (2018b). User Guides - Sentinel-2 MSI - Definitions - Sentinel Online. Retrieved June 5, 2018, from https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial

European Union. (2018). Utilised Agricultural Area by Categories. Retrieved November 11, 2018, from https://ec.europa.eu/eurostat/tgm/table.do?tab=table&plugin=1&language=en&pcode=tag00025

Foody, G. M. (2002). Status of Land Cover Classification Accuracy Assessment. *Remote Sensing of Environment*, *80*(1), 185–201. https://doi.org/10.1016/S0034-4257(01)00295-4

Fu, G., Liu, C., Zhou, R., Sun, T., & Zhang, Q. (2017). Classification for High Resolution Remote Sensing Imagery Using A Fully Convolutional Network. *Remote Sensing*, *9*(5), 1–21. https://doi.org/10.3390/rs9050498

Gao, D.-L., Zhang, R., & Xue, D.-X. (2017). Improved Fully Convolutional Network for the Detection of Built-up Areas in High Resolution SAR Images. https://doi.org/10.1007/978-3-319-71598-8_54

Gevaert, C. M., Persello, C., Nex, F., & Vosselman, G. (2018). A deep learning approach to DTM extraction from imagery using rule-based training labels. *ISPRS Journal of Photogrammetry and Remote Sensing*, *142*, 106–123. https://doi.org/10.1016/j.isprsjprs.2018.06.001

Gómez, C., White, J. C., & Wulder, M. A. (2016). Optical Remotely Sensed Time Series Data for Land Cover Classification: A Review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *116*, 55–72. https://doi.org/10.1016/j.isprsjprs.2016.03.008

Guan, X., Huang, C., Liu, G., Meng, X., & Liu, Q. (2016). Mapping rice cropping systems in Vietnam using an NDVI-based time-series similarity measurement based on DTW distance. *Remote Sensing*, *8*(1). https://doi.org/10.3390/rs8010019

Guo, R., Liu, J., Li, N., Liu, S., Chen, F., Cheng, B., … Ma, C. (2018). Pixel-Wise Classification Method for High Resolution Remote Sensing Imagery Using Deep Neural Networks. *ISPRS International Journal of Geo-Information*, *7*(3), 110. https://doi.org/10.3390/ijgi7030110

Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2003). A Practical Guide to Support Vector Classification (Vol. 101, pp. 1396–1400). Retrieved from https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *ICML*. Retrieved from https://arxiv.org/pdf/1502.03167.pdf

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). Support Vector Machine. In *An Introduction to Statistical Learning: with Applications in R* (Vol. 103, pp. 337–372). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-7138-7

Jensen, J. R. (2015). *Introductory Digital Image Processing A Remote Sensing Perspective* (4th ed.). Glenview: Pearson Education, Inc.

Ji, S., Zhang, C., Xu, A., Shi, Y., & Duan, Y. (2018). 3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images. *Remote Sensing*, *10*(1). https://doi.org/10.3390/rs10010075

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep Learning in Agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, 70–90. https://doi.org/10.1016/j.compag.2018.02.016

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems* (Vol. 1, pp. 1097–1105). Lake Tahoe, Nevada: Curran Associates Inc. Retrieved from http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Landgrebe, D. A. (1978). Useful Information from Multispectral Image Data: Another Look. In P. H. Swain & S. M. Davis (Eds.), *Remote Sensing, The Quantitative Approach* (pp. 336–374). United States of America: McGraw-Hill, Inc.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Li, Y., Chen, Y., Liu, G., Jiao, L., Li, Y., Chen, Y., … Jiao, L. (2018). A Novel Deep Fully Convolutional Network for PolSAR Image Classification. *Remote Sensing*, *10*(12), 1984. https://doi.org/10.3390/rs10121984

Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2016). Fully Convolutional Neural Networks for Remote Sensing Image Classification. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 5071–5074). IEEE. https://doi.org/10.1109/IGARSS.2016.7730322

Maus, V., CĂmara, G., Cartaxo, R., Sanchez, A., Ramos, F. M., & De Queiroz, G. R. (2016). A Time-Weighted Dynamic Time Warping Method for Land-Use and Land-Cover Mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(8), 3729–3739. https://doi.org/10.1109/JSTARS.2016.2517118

Mou, L., Bruzzone, L., & Zhu, X. X. (2018). Learning Spectral-Spatial-Temporal Features via a Recurrent Convolutional Neural Network for Change Detection in Multispectral Imagery, 1–12. Retrieved from http://arxiv.org/abs/1803.02642

Müller-Wilm, U. (2018). *S2 MPC Level 2A Input Output Data Definition*. Retrieved from http://step.esa.int/thirdparties/sen2cor/2.5.5/docs/S2-PDGS-MPC-L2A-IODD-V2.5.5.pdf

Ndikumana, E., Minh, D. H. T., Baghdadi, N., Courault, D., & Hossard, L. (2018). Deep recurrent neural network for agricultural classification using multitemporal SAR Sentinel-1 for Camargue, France. *Remote Sensing*, *10*(8), 1–16. https://doi.org/10.3390/rs10081217

Persello, C., & Stein, A. (2017). Deep Fully Convolutional Networks for the Detection of Informal Settlements in VHR Images. *IEEE Geoscience and Remote Sensing Letters*, *14*(12), 2325–2329. https://doi.org/10.1109/LGRS.2017.2763738

Petitjean, F., Inglada, J., & Gançarski, P. (2012). Satellite image time series analysis under time warping.

IEEE Transactions on Geoscience and Remote Sensing, *50*(8), 3081–3095. https://doi.org/10.1109/TGRS.2011.2179050

Rizaldy, A., Persello, C., Gevaert, C. M., & Oude Elberink, S. J. (2018). Fully Convolutional Networks for Ground Classification from LIDAR Point Clouds. https://doi.org/10.5194/isprs-annals-IV-2-231-2018

Rußwurm, M., & Körner, M. (2017). Multi-Temporal Land Cover Classification With Long Short-Term Memory Neural Network. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-1/W1*(1W1), 551–558. https://doi.org/10.5194/isprs-archives-XLII-1-W1-551-2017

Sakoe, H., & Chiba, S. (1978). *Dynamic Programming Algorithm Optimization for Spoken Word Recognition*. *IEEE Transactions on Acoustics, Speech, and Signal Processing* (Vol. 26). https://doi.org/10.1109/TASSP.1978.1163055

Sharma, A., Liu, X., & Yang, X. (2018). Land cover classification from multi-temporal, multi-spectral remotely sensed imagery using patch-based recurrent neural networks. *Neural Networks*, *105*, 346–355. https://doi.org/10.1016/j.neunet.2018.05.019

Shelhamer, E., Long, J., & Darrell, T. (2017). Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(4), 640–651. https://doi.org/10.1109/TPAMI.2016.2572683

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, *15*, 1929–1958. Retrieved from http://www.cs.toronto.edu/~rsalakhu/papers/srivastava14a.pdf

Stanford University. (2018). CS231n Convolutional Neural Networks for Visual Recognition. Retrieved January 30, 2019, from http://cs231n.github.io/convolutional-networks/

United Nations. (2015). Hunger and Food Security - United Nations Sustainable Development. Retrieved August 12, 2018, from https://www.un.org/sustainabledevelopment/hunger/

Vedaldi, A., & Lenc, K. (2015). MatConvNet - Convolutional Neural Networks for MATLAB. In *Proceeding of the {ACM} Int. Conf. on Multimedia*. https://doi.org/10.1145/2733373.2807412

Wang, X., & Zhong, Y. (2003). Statistical learning theory and state of the art in SVM. *Proceedings - 2nd IEEE International Conference on Cognitive Informatics, ICCI 2003*, (2), 55–59. https://doi.org/10.1109/COGINF.2003.1225953

World Bank. (2018). Agricultural land (sq. km). Retrieved November 18, 2018, from https://data.worldbank.org/indicator/AG.LND.AGRI.K2?locations=US

Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors*, *2017*. https://doi.org/10.1155/2017/1353691

Yang, J., Jiang, Y., Fang, H., Jiang, Z., Zhang, H., & Hao, S. (2018). Semantic segmentation of aerial image using fully convolutional network. In *Communications in Computer and Information Science* (Vol. 875, pp. 546–555). https://doi.org/10.1007/978-981-13-1702-6_54

Yu, F., & Koltun, V. (2016). Multi-Scale Context Aggregation by Dilated Convolutions. *CoRR*. Retrieved from https://arxiv.org/pdf/1511.07122.pdf

Zhai, Y., Qu, Z., & Hao, L. (2018). Land cover classification using integrated spectral, temporal, and spatial features derived from remotely sensed images. *Remote Sensing*, *10*(3). https://doi.org/10.3390/rs10030383

Zhang, T.-X., Su, J.-Y., Liu, C.-J., & Chen, W.-H. (2019). Potential Bands of Sentinel-2A Satellite for Classification Problems in Precision Agriculture, *16*(February), 16–26. https://doi.org/10.1007/s11633-018-1143-x

Zulkifli, H. (2018). Understanding Learning Rates and How It Improves Performance in Deep Learning. Retrieved February 11, 2019, from https://towardsdatascience.com/understanding-learning-rates-and-how-it-improves-performance-in-deep-learning-d0d4059c1c10