CRANFIELD UNIVERSITY

Alex Matthew Hamer

Mapping agricultural land in support of opium monitoring in Afghanistan with Convolutional Neural Networks (CNNs)

School of Water, Energy and Environment
PhD in Environment and Agrifood

PhD
Academic Year: 2017 - 2021

Supervisor: Dr Toby W. Waine
Associate Supervisor: Dr Daniel M. Simms
December 2021

CRANFIELD UNIVERSITY

School of Water, Energy and Environment
PhD in Environment and Agrifood

PhD

Academic Year 2017 - 2021

Alex Matthew Hamer

Mapping agricultural land in support of opium monitoring in Afghanistan with Convolutional Neural Networks (CNNs)

Supervisor: Dr Toby W. Waine
Associate Supervisor: Dr Daniel M. Simms
December 2021

Abstract

This work investigates the use of advanced image classification techniques for improving the accuracy and efficiency in determining agricultural areas from satellite images. The United Nations Office on Drugs and Crime (UNODC) need to accurately delineate the potential area under opium cultivation as part of their opium monitoring programme in Afghanistan. They currently use unsupervised image classification, but this is unable to separate some areas of agriculture from natural vegetation and requires time-consuming manual editing. This is a significant task as each image must be classified and interpreted separately. The aim of this research is to derive information about annual changes in land-use related to opium cultivation using convolutional neural networks with Earth observation data.

Supervised machine learning techniques were investigated for agricultural land classification using training data from existing manual interpretations. Although pixel-based machine learning techniques achieved high overall classification accuracy (89%) they had difficulty separating between agriculture and natural vegetation at some locations.

Convolutional Neural Networks (CNNs) have achieved ground-breaking performance in computer vision applications. They use localised image features and offer transfer learning to overcome the limitations of pixel-based methods. There are challenges related to training CNNs for land cover classification because of underlying radiometric and temporal variations in satellite image datasets. Optimisation of CNNs with a targeted sampling strategy focused on areas of known confusion (agricultural boundaries and natural vegetation). The results showed an improved overall classification accuracy of +6%. Localised differences in agricultural mapping were identified using a new tool called ‘localised intersection over union’. This provides greater insight than commonly used assessment techniques (overall accuracy and kappa statistic), that are not suitable for comparing smaller differences in mapping accuracy.

A generalised fully convolutional model (FCN) was developed and evaluated using six years of data and transfer learning. Image datasets were standardised across image dates and different sensors (DMC, Landsat, and Sentinel-2), achieving high classification accuracy (up to 95%) with no additional training. Further fine-tuning with minimal training data and a targeted training strategy further increased model performance between years (up to +5%).

The annual changes in agricultural area from 2010 to 2019 were mapped using the generalised FCN model in Helmand Province, Afghanistan. This provided new insight into the expansion of agriculture into marginal areas in response to counter-narcotic and alternative livelihoods policy. New areas of cultivation were found to contribute to the expansion of opium cultivation in Helmand Province.

The approach demonstrates the use of FCNs for fully automated land cover classification. They are fast and efficient, can be used to classify satellite imagery from different sensors and can be continually refined using transfer learning. The proposed method overcomes the manual effort associated with mapping agricultural areas within the opium survey while improving accuracy. These findings have wider implications for improving land cover classification using legacy data on scalable cloud-based platforms.

Keywords: Fully convolutional networks, convolutional neural networks, transfer learning, remote sensing, image classification

Acknowledgements

I would firstly like to thank Dr Toby Waine and Dr Daniel Simms for their continued support, patience, and guidance over the course of this PhD research. The skills I have developed during my time at Cranfield University are a credit to the fantastic supervision I have received. I'd also like to extend my appreciation to the Natural Environment Research Council (NERC) for providing the funding for this project [NE/M009009/1] and to the Data, Risk and Environmental Analytical Methods Centre for Doctoral Training (DREAM CDT). I would also like to thank the United Nations Office on Drugs and Crime (UNODC) for providing data used in this project.

I would like to thank Kirsty for the immeasurable amount of motivation and support she has given me throughout my studies. I would also like to thank my parents for their continued support and encouragement to always push myself and achieve my goals. I know my father would be proud of my achievements. Also, to my brother, Marcus, who never fails to put a smile on my face and battling through our degrees together.

Finally, I would also like to acknowledge all my friends who have kept me going during my studies. Ben, Ewelina, Karolina, Laura, Kayu, Juan, Eduarda, Ali, Martina, Tom, Rhodri, Abi and Mary – thank you!

Table of contents

List of figures

List of tables

List of abbreviations

ANN	Artificial Neural Network
CNN	Convolutional Neural Network
DMC	Disaster Monitoring Constellation
EE	End-to-End
FAO	Food and Agriculture Organisation
FCN	Fully Convolutional Network
FN	False Negative
FP	False Positive
fwIoU	Frequency-weighted Intersection over Union
IoU	Intersection over Union
IR-MAD	Iteratively Reweighted Multivariate Alteration Detection
ISODATA	Iterative Self-Organising Data Analysis Technique
mIoU	Mean Intersection over Union
NDVI	Normalised Difference Vegetation Index
NSIA	National Statistic and Information Authority
OA	Overall Accuracy
ResNet50	Residual Network 50
SRTM	Shuttle Radar Topography Mission
SVM	Support Vector Machine
TL	Transfer Learning
TN	True Negative
TP	True Positive
UNODC	United Nations Office on Drugs and Crime
VGG	Visual Geometry Group

Chapter 1. Introduction

1.1 Approaches for image classification in remote sensing

Image classification is widely used in remote sensing to arrange data into groups based on spectral, textual and spatial properties to extract knowledge for the region of interest with applications ranging from flood detection mapping to land cover/land use mapping (Bellón *et al.*, 2017; Clement, Kilsby and Moore, 2017). Change detection approaches often use image classification for identifying differences across the landscape using legacy data (Hansen *et al.*, 2014). The two main forms of image classification are traditional statistical-based classifiers and machine learning classifiers. Maximum likelihood uses a statistical algorithm to evaluate the likelihood of a pixel belonging to each land cover class. The log likelihood function is commonly used, therefore the data requires a gaussian distribution (Otukei and Blaschke, 2010). Traditional statistical-based algorithms used for classification have been cited as inappropriate for using Earth observation data as they are designed to identify the relationships between variables first to inform the prediction. There has been a transition to machine learning techniques to efficiently utilise these large datasets and has been instigated from greater accessibility to the required computational resources to facilitate accurate prediction (DeFries & Chan, 2000; Gislason *et al.*, 2006).

Machine learning is an empirical approach for unsupervised and supervised classification of non-linear systems. Machine learning algorithms learn underlying patterns and trends from a set of training data and adapt to new data without human interaction. Machine learning has increased in popularity as these techniques can alter data into an appropriate feature space and have been found to be insensitive to noise within datasets unlike traditional statistical models (Su *et al.*, 2017). The *k*-means algorithm is a popular technique used for unsupervised classification (Han, Champeaux and Roujean, 2004), where an arbitrary number of classes are selected for classification based on spectral, textural and spatial differences. The technique is more computationally efficient than supervised algorithms by negating the requirement for separation of training data. However,

the producer has no control over the classes selected by the algorithm and smaller land cover classes may be omitted (Keuchel *et al.*, 2003).

Random Forest is a supervised learning algorithm that builds multiple decision trees to identify the underlying trends from input data (Pal & Mather, 2003). Support Vector Machines (SVMs) fit a hyperplane into a 3D-feature space of the training data to separate land cover classes. SVMs outperform other machine learning algorithms with large training datasets (Huang et al., 2002). An Artificial Neural Network (ANN) is a machine learning algorithm that simulates the way a brain processes data and consists of hidden layers to identify patterns from input data. ANNs are often used because of their ability to generalise inputs whilst remaining robust to noisy incomplete datasets (Zulhaidi *et al.*, 2007). Optimal performance has been found using between 5% and 10% of the training data (Zhuang *et al.*, 1994). ANNs were found to train faster using less training data, although classification accuracy decreased with less training data (Verbeke, Vancoillie and De Wulf, 2004). ANNs have been found to be less accurate than other machine-learning classification algorithms, such as SVM, particularly with small training datasets by allowing superior generalisation (Shao & Lunetta, 2012). Although, neural networks have previous outperformed tree-based classification machine-learning classifiers for land cover change detection (Rogan *et al.*, 2008).

Deep learning is a type of machine learning based on ANNs, where multiple layers are used to process input data to extract higher level features (e.g. objects formed from low level features, such as blobs and edges). These algorithms require large datasets to train the networks. The increase in the availability of satellite imagery with labelled datasets has provided the opportunity to use state-of-the-art supervised classification techniques at various spatial scales. Convolutional Neural Networks (CNNs) are the most common form of deep learning and are becoming increasingly common among image processing, predominately due to their support and application for *N*-dimensions and high-level feature extraction (Ball, Anderson and Chan, 2017). The CNN classifier utilises an image filter to extract common features using convolutions across an

image while preserving features with greater importance. State-of-the-art classification accuracies have been achieved with this deep-learning approach (Maggiori *et al.*, 2017). However, they requires large volumes of training data for image classification (Nogueira, Penatti and dos Santos, 2017). The main constraint found with other machine learning approaches are their inability to generalise across image dates due to differences in the radiometry (Ball, Anderson and Chan, 2017). CNNs not only use individual pixels or objects for training, but image features in context from input data which are found across imagery regardless of image timing.

The spatial scale for classification varies in remote sensing classification, with object-based and pixel-based classification. Object-based segmentation splits the image scene into groups of similar adjacent pixels based on edge identification, image gradients and spectral similarity (Laliberte *et al.*, 2004). Object-based classifications have previously been found to be superior to pixel-based approaches based on per-scene overall classification accuracy in medium resolution datasets (+2% overall accuracy for Landsat-5 TM (30 m)) (Geneletti and Gorte, 2003). The technique is useful for storing data into GIS databases unlike pixel-based approaches, although studies have found poor overall accuracy with minority classes and land cover types with small objects compared to pixel-based techniques, such as the maximum likelihood classifier (Dingle Robertson and King, 2011). Objects use surrounding values to provide context to the classification, but have been found to generalise the landscape at object boundaries (Chen *et al.*, 2012).

1.2 Overview of land cover classification accuracy assessment

1.2.1 Importance for accuracy assessment in image classification

Land cover classifications are important for land management, habitat conservation, spatial planning, and monitoring ecosystem services. The quality of these classifications is vital to provide a good representation of the area for use as decision making tools (UNODC, 2018b). Two common types of error in mapping the spatial distribution of land cover are positional uncertainty and attribute uncertainty and estimating these errors is an important step to assess

the quality of the information derived from them and ensures they are fit for purpose (Foody, 2002). Within remote sensing there are other common methods used to determine the accuracy of image classifications (e.g. overall accuracy, kappa statistic) (Tateishi *et al.*, 2011; Fu *et al.*, 2017; Pouliot *et al.*, 2019). These methods are often selected based on the type of classification approach selected and data-specific attributes (Olofsson *et al.*, 2014). The size of the area used to train and evaluate classifications varies based on the classification technique selected (Pal, 2005; Lucas *et al.*, 2011; Demir and Başayiğit, 2019), commonly these are either individual pixels or objects (Stehman and Wickham, 2011). Objects are often not uniform, unlike individual pixels which are of equal size, which make them unsuitable for comparison between classifications (Ye, Pontius Jr and Rakshit, 2018). Segmenting satellite imagery into homogenous groups also causes increased pixel mixing (Geneletti and Gorte, 2003) and can generalises the landscape and remove the subtle variation in land cover changes.

Image classifications are evaluated using ground reference data and the evaluation is commonly conducted using the confusion matrix (Foody, 2002). Ground reference data are considered to be a true representation of the landscape, but there is potential uncertainty from their creation and have the potential to skew classification accuracy metrics. Quality control by trained interpreters is a common approach to mitigating uncertainty for reference data (Foody, 2009). The most common metrics include overall accuracy, user accuracy, producer accuracy and kappa statistic (Morales-Barquero *et al.*, 2019). This increasing complexity in remote sensing image classification has given rise for the need to transition away from traditional pixel accuracy assessments to accommodate for these new workflows. Recently more studies are using other accuracy assessment metrics derived from machine learning applications, such as Receiver Operating Characteristic - Area Under the Curve (ROC-AUC) (Maggiori *et al.*, 2017), F1-score (Pouliot *et al.*, 2019) and Intersection over Union (IoU) (Shelhamer, Long and Darrell, 2016), to provide further evaluation of their image classifications.

1.2.2 Validation sample selection

Common practice across both machine learning and remote sensing applications is to use an independent set of hold-out validation ground data to perform accuracy assessments. The sampling strategy for validation sample selection is a highly influential factor for assessing the performance of land cover and poorly devised strategies may lead to misleading accuracy metrics (Foody, 2002). The most common approach used for land cover classification is stratified random sampling to provide examples across minority and majority classes (Olofsson *et al.*, 2014). Stratification provides class accuracy across all land cover types and ensures there is sufficient representation across minority classes. The number of samples per strata is often dependant on the amount of reference data available for each stratum, therefore should be appropriate for the intended use (Stehman, 2009). The number of unique samples is also dependant on the size of the area selected with individual pixels having more independent samples for selection than objects (although sometimes the centroid pixel or sub-samples of the object are used (Pouliot *et al.*, 2019)).

The main objective of stratification is to group homogenous areas of the landscape to identify distinctive regions within the image to aid classification. The data used for stratification has a substantial role in the ability to separate these distinctive areas. Data-driven stratification is common within remote sensing, where the grouping of similar spectral bands from an image is widely adopted. However, other forms of spatial data can be used to stratify an area, such as vegetation indices, elevation data and other forms on ancillary data (e.g. distance to water). Map-to-map comparisons are common in land cover classification tasks, especially when developing new classification approaches for comparison between classification outputs and reference data (Foody, 2002). The importance of accurate statistics for assessing change between areas is imperative, therefore time-series classifications require the appropriate accuracy assessments (Gómez, White and Wulder, 2016). Stratification based on map-derived reference data (e.g. from another image classification task) may introduce error based on the certainty the labels represent the truth, therefore confidence measures have

been suggested to minimise error introduced by the reference data (Foody, 2002).

The balancing of datasets for equal thematic class distribution is common practice to avoid under-representation of minority classes in the accuracy assessment (Olofsson *et al.*, 2014; Colditz, 2015; Maxwell, Warner and Fang, 2018). Balanced datasets for individual pixels are created by under-sampling the majority class to match the number of minority class samples (Douzas *et al.*, 2019). For samples with areas larger than a single pixel, augmentation of samples from the minority class can be carried out to increase the number of samples during training and are an advantage of object-based samples (Zeiler and Fergus, 2013). These augmented samples should not be used during validation to ensure no bias is introduced to the accuracy calculations. Studies that use highly unbalanced datasets are still able to carry out accuracy assessments but require weighted accuracy metrics to limit bias from the majority class. These weighted accuracy assessments are useful for proportional class distribution, which use a stratified sampling approach to collect samples based on their presence across the classification. The limitation of using this strategy is the need for dense labels of the study site prior to classification to select the correct proportions across the land cover types. This approach would also suppress the performance of minority classes, whilst favouring the majority classes for accuracy assessment. Stehman (2009) notes the importance of spatially balanced datasets, where samples are selected across the study extent or limited to the areas of interest to limit bias during accuracy assessment from the same sampling locations.

1.2.3 Accuracy metrics from the confusion matrix

The confusion matrix is the most widely used approach for accuracy assessment in remote sensing studies (Foody, 2002). The matrix visualises classification performance using an n x n matrix, where n is the number of classes assessed. Each column is used for each reference class and each row is used for each prediction class. An example of the confusion matrix is shown in Table 1-1 with a total sample size of 100 with 50 pixels of agriculture detected and 50 pixels of

non-agriculture detected. The diagonal numbers across the matrix are the number of cases that match in both the reference and prediction. These are known as true positives (TP) and true negatives (TN), where a correctly classified agriculture pixel is a true positive and a correctly classified non-agriculture pixel is a true negative (Table 1-1). A false positive (FP), also known as a type I error, is where the prediction has classified the presence of agriculture when there is not and a false negative (FN), also known as a type II error, is where the prediction has stated there is no presence of agriculture when there is.

Table 1-1. Example of a confusion matrix for the presence of agriculture across 100 validation samples (50 agriculture and 50 non-agriculture). TP = True Positive, FP = False Positive, FN = False Negative, TN = True Negative and OA = Overall accuracy.

		Reference		**User accuracy**
		Agriculture	Non-agriculture	
	Agriculture	**34 (TP)**	24 (FP)	34/(34+24) = **59%**
Prediction	Non-agriculture	16 (FN)	**26 (TN)**	26/(16+26) = **65%**
Producer accuracy		34/(34+16) **= 68%**	26/(24+26) = **52%**	**(34+26)/100 = 60% OA**

Many of the metrics used in remote sensing and machine learning literature are derived from this confusion matrix (Foody, 2002). The overall accuracy, also known as pixel accuracy, provides the number of correct predictions made by the model. Overall accuracy is the most widely used metric in remote sensing and machine learning literature, but all metrics derived from the confusion matrix require balanced validation samples between classes for a representative accuracy assessment (Stehman, 2009). A commonly cited target overall accuracy is 85% with no class less than 70% (Thomlinson, Bolstad and Cohen, 1999), but acceptable target accuracy is often subjective and dependant on the purpose of the land cover classification. Many studies argue overall accuracy alone is not enough for evaluating the results of image classification (Foody, 2002; Myint *et al.*, 2008; Olofsson *et al.*, 2014). Precision measures the proportion of predicted positives to the number of reference positives, also known as commissions, where high precision means low false positive rate. Precision is also used to derive user accuracy by subtracting the commission error from 100.

Recall measures the proportion of the number of reference positives and the number of true positives, also known as omissions, where high recall means low false negative rate. Producer's accuracy is derived using recall by subtracting the omission error from 100. Precision and recall are important to identify confusion between classes based on common omissions and commissions and should be reported together for transparency of the classification's error (Olofsson *et al.*, 2014).

The F1-score calculates the weighted mean between precision and recall. The false positives and false negatives are crucial to this accuracy metric and unlike overall accuracy the F1-score can be used with class imbalance (Johnson and Khoshgoftaar, 2019). The main caveat to this metric is the limited use within the remote sensing community, therefore makes comparison between classifications difficult unlike overall accuracy (Morales-Barquero *et al.*, 2019). The specificity and sensitivity, also referred to as the false positive and true positive rate, are used to the Receiver Operating Characteristics curve (ROC), which is a visual representation of these two metrics at different decision thresholds irrespective of class balance (Maggiori *et al.*, 2017). Specificity measures the proportion of the number of reference negatives and the number of true negatives and sensitivity measures the proportion of the number of reference positives and the number of true positives. Specificity is often cited as the true positive and $1 - sensitivity$ is referred to as the false positive rate. The ROC curve is also used to derive the Area Under the Curve (AUC) metric, which measures the 2-dimensional area under the ROC curve. However, the AUC does not identify individual class performance unlike the metrics derived from the confusion matrix.

The kappa coefficient (Cohen, 1960) is often cited in remote sensing literature and is used to evaluate how well the classification is performing compared to random values based on pixel agreement (Sexton *et al.*, 2013; Dronova *et al.*, 2015; Y Zhai, Qu and Hao, 2018). Although it has been recommended kappa should no longer be used for accuracy assessment as the calculation is potentially based on incorrect assumptions and provides redundant information (Pontius and Millones, 2011; Delgado and Tibau, 2019). Pontius and Millones

(2011) suggested the use of quantity and allocation disagreement, which measures the overall level of disagreement from the confusion matrix, as a replacement of the kappa statistic. Quantity disagreement is the difference between the reference and prediction samples based on the differences in sample proportions, therefore can be used with unbalanced datasets. Allocation disagreement is the difference between the reference and prediction samples based on the spatial allocation, which is similar recall by identifying areas of commission using false negatives and false positives. By only focusing on the areas of disagreement they suggest it is a more effective metric than kappa which compares the total agreement to an arbitrary random baseline. This metric has been used by the remote sensing community (Moreno Navas, Telfer and Ross, 2012; Pandey, Joshi and Seto, 2013; Aldwaik, Onsted and Pontius, 2015), although the uptake of these metrics is limited and the kappa statistic remains widely cited. The main metrics used to summarise the performance of image classification in remote sensing are overall accuracy and the kappa statistic.

The spatial distribution of error is not considered when using global accuracy assessment metrics derived from the confusion matrix, which is often bound by the size of the area used for accuracy assessment (Stehman and Wickham, 2011). The advantage with using areas greater than a single pixel is the ability to assess the performance of pixels in context. Using objects and blocks of pixels provides the more spatial context to features during accuracy assessments. A common metric used in deep learning for semantic segmentation is known as Intersection over Union (IoU), shown in Figure 1-1 (Long, Shelhamer and Darrell, 2015), which measures the agreement of a block of pixels between the reference and prediction based on the proportion of overlap between these areas. Mean and frequency-weighted are the two common forms of IoU. Mean IoU calculates IoU for each class and provides the mean average across the number of classes. Frequency-weighted IoU calculates IoU for each class and provides the weighted average based on classes present. Each class has its own IoU to assess the performance of individual classes with the common threshold for a ‘good’ IoU score being >0.5 (Li *et al.*, 2018) (Table 1-2). The main disadvantage with using IoU in remote sensing is that very few studies have utilised this metric and it does

not give explicit information on the false positives or false negatives of the classification. Accuracy assessments are traditionally carried out at the pixel-level, rather than at the level of an object or block of pixels, therefore is not applicable. This method is similar to the calculation for overall accuracy, where it looks at the total number of pixels in agreement compared to the overall number of pixels. The advantage of IoU, in comparison to overall accuracy, is the ability to add weighting to the metric to remove bias of the majority class. The increase in methods that require objects or image chips as sources of input and validation would provide more opportunity to use this metric in the field of remote sensing.

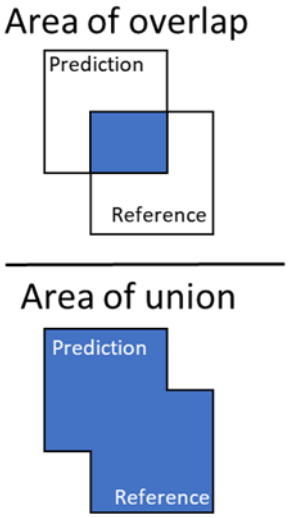

Figure 1-1. Visual description of intersection over union using the bounding boxes of the reference and prediction

The spatial representation of error and uncertainty is an important aspect that is being developed further due to the changes in classification approaches (Stehman and Wickham, 2011). Long, Shelhamer and Darrell (2015) explored the effect of spatial error caused through classification reconstruction of a Fully Convolutional Network (FCN) by altering the resampling rate by increasing the spacing between predictions of commercial photography (stride length). Generalisation of the prediction was found to increase with greater stride lengths during the resampling process (-7% pixel accuracy). They highlight the need for

error assessments over larger areas during the reconstruction of classifications and not only for isolated samples. The transition to larger areas for validation allows for further understanding of spatial error and provides visual interpretation of accuracy in context to the landscape.

Most of the metrics currently used in remote sensing studies for image classification use those derived from the confusion matrix. The transition to feature-based classification techniques and size of the validation sample area has allowed other metrics outside of individual pixel accuracy assessments to be carried out. The importance of accuracy in a spatial context is advantageous for understanding the spatial distribution of error and provides the ability to visualise these errors spatially across the landscape. No single accuracy metric is applicable for every study; therefore, values and metrics should be interpreted based on the context of the study. A summary of all accuracy assessment metrics reviewed are presented in Table 1-2.

Table 1-2. Summary of common accuracy metrics used across remote sensing disciplines. AUC-ROC = Area Under Curve of Receiver Characteristic Operator and IoU = Intersection over Union

Metric	Advantages	Disadvantages	References
Overall accuracy	Most common metric cited in remote sensing	Only considers the rate of true positives	(Dronova *et al.*, 2015; Parente and Ferreira, 2018; Yadav and Congalton, 2018)
	Identifies correctly classified areas	Sensitive to class imbalance	
Precision / user accuracy	Identifies commission errors both overall and per-class	Sensitive to class imbalance	(Alatorre *et al.*, 2011; Fu *et al.*, 2017; Liu *et al.*, 2018)
Recall / producer accuracy	Identifies omission errors both overall and per-class	Sensitive to class imbalance	(Alatorre *et al.*, 2011; Fu *et al.*, 2017; Liu *et al.*, 2018)
F1 score	Works for uneven class distributions	Not widely cited in remote sensing	(Inglada *et al.*, 2015; Pouliot *et al.*, 2019)
AUC-ROC	Able to handle class imbalance	Does not provide classes-level statistics	(Alatorre *et al.*, 2011; Khatami, Mountrakis and Stehman, 2017; Maggiori *et al.*, 2017)
Kappa statistic	Normalised metric widely cited in remote sensing literature	Potential flaws in its calculation	(Dronova *et al.*, 2015; Fu *et al.*, 2017; Y Zhai, Qu and Hao, 2018)
Allocation disagreement	Useful for only assessing disagreement	Not widely cited in remote sensing literature	(Moreno Navas, Telfer and Ross, 2012; Pandey, Joshi and Seto, 2013; Aldwaik, Onsted and Pontius, 2015)
	Similar to recall by relying on false negatives and false positives	Does not consider agreement of the samples	
Quantity disagreement	Useful for highlighting disagreement based on proportions	Does not consider agreement of the samples	(Moreno Navas, Telfer and Ross, 2012; Pandey, Joshi and Seto, 2013; Aldwaik, Onsted and Pontius, 2015)
		Able to use with unbalanced datasets	
Mean IoU	Able to use it across areas greater than a single pixel	Requires dense reference labels	(Long, Shelhamer and Darrell, 2015; Xu *et al.*, 2018; Weinstein *et al.*, 2020)
	Useful if background class is to be taken into consideration	Sensitive to unbalanced samples	
		Unable to consider false negatives/positives	
Frequency weighted IoU	Able to use it across areas greater than a single pixel	Requires dense reference labels	(Long, Shelhamer and Darrell, 2015)
	Balanced metric and can be used with high sample variance	Can suppress problems with majority class	
		Unable to consider false negatives/positives	

1.2.4 Remote sensing and uncertainty in area estimation

Remote sensing data is a useful tool for extracting agricultural statistics across large areas or in remote and inaccessible locations. The most common approach is by quantifying the total area by measuring the absolute area for each crop type

and could provide an absolute value for under- or over-classification if reference data are available, also referred to as pixel counting. These are more common in regional and national studies (Lucas *et al.*, 2011), but require large amounts of labelled data across all stratum to conduct. Traditional pixel-based accuracy assessments (e.g. overall accuracy) do not require large continuous areas of labelled data to be calculated. Comparison in absolute area would require large areas of labelled data to cross-reference with the predicted area but can be limited to smaller areas where continuous data are available. Pixel counting is also high dependant on the spatial resolution of the dataset with lower spatial resolutions having a greater impact on the estimate. The scale is also an important consideration if multiple spatial resolutions are used for area estimation where there will be more spectral mixing.

Remote sensing data are often used in a regression estimator to estimate the presence of a crop within an area (Carfagna and Gallego, 2005). The primary reason for is the use of remote sensing data reduces the amount of ground data needed for estimation. Land cover classifications are commonly used within linear regression estimators by estimating the mean of an area of crop (y) with a known area (reference data) by using the number of pixels classified as the crop from the land cover classification (x) as the two populations should be correlated. The coefficient of the regression line (b) can have a fixed value for the regression estimator to avoid bias between y and x during the classification otherwise it is the least-squares estimate of the regression slope between y and x (Gallego, 2004).

Confidence intervals are often used with crop level area estimates to quantify the uncertainty of the estimate with 95% as the most common unit (Olofsson *et al.*, 2014). The range between the confidence intervals can often be explained from the error matrix. For example, a higher estimation than the crop area would refer to high omission error in the classification and lower estimation than the crop area refers to high commission error in the classification. Using these area estimation approaches avoids pixel counting bias and decreases the standard error within the estimated crop area.

1.3 Agricultural practices in Afghanistan

Agriculture accounts for ~ 25% of Afghanistan's GDP with crop agriculture producing the greatest income, where 12% of Afghanistan's land area is arable (8 million ha) (The World Bank, 2014). Water shortages are the primary constraint for arable agricultural production with irrigated agriculture output up to three times greater than rainfed, where most cultivated land receives less than 400 mm of rain annually (Kawasaki *et al.*, 2012; The World Bank, 2014). Irrigated arable land accounts for 25% of arable land and the remaining land is rainfed or fallow.

The region has limited diversity in their growing practices with more than 65% of cultivated area dedicated to food crops and predominately concentrate on wheat production and accounts for 2.7 – 3 million ha of cultivated land (The World Bank, 2014). Farm sizes in Afghanistan are small with 60% smaller than 1 ha (CSO, 2018). The Ministry of Agriculture, Irrigation and Livestock (MAIL) undertake annual wheat cultivation surveys using visual interpretation of satellite imagery and ground sample data to estimate wheat production and manage food security in the country (Tiwari *et al.*, 2020). Other major cereal crops include rice, maize and barley which account for 15% of the total cereal area. The crop calendar for these vary across Afghanistan with wheat and barley sown from October until December and harvested from April until July. Water shortages have resulted in high fluctuations in cereal production and has led to more importation of these crops to manage food security (The World Bank, 2018).

The United Nations Office on Drugs and Crime (UNODC) (2020) estimates Afghanistan produced 84% of the world's illicit opium over the last five years, despite the decline in cultivation over recent years. Opiates are psychoactive substances derived from the dried milky sap of unripe poppy seed capsules. Illicit drug use poses serious health risks to the global community and burdens healthcare systems for the prevention and treatment of drug use (UNODC, 2020). Opium cultivation occurs throughout Afghanistan with annual variation in the total area under cultivation and its spatial distribution, even though cultivation is illegal in the country. Opium poppy is typically sown between October and November and harvested between April and May (Tiwari *et al.*, 2020). The main cultivating

provinces in 2019 were Helmand, Kandahar, Uruzgan, Badghis and Farah (Figure 1-2). Opium poppy is primarily grown as it has been more profitable in comparison with cereals such as wheat, especially for small farms with limited arable land and is more drought-resistant than most other crops (UNODC, 2021a).

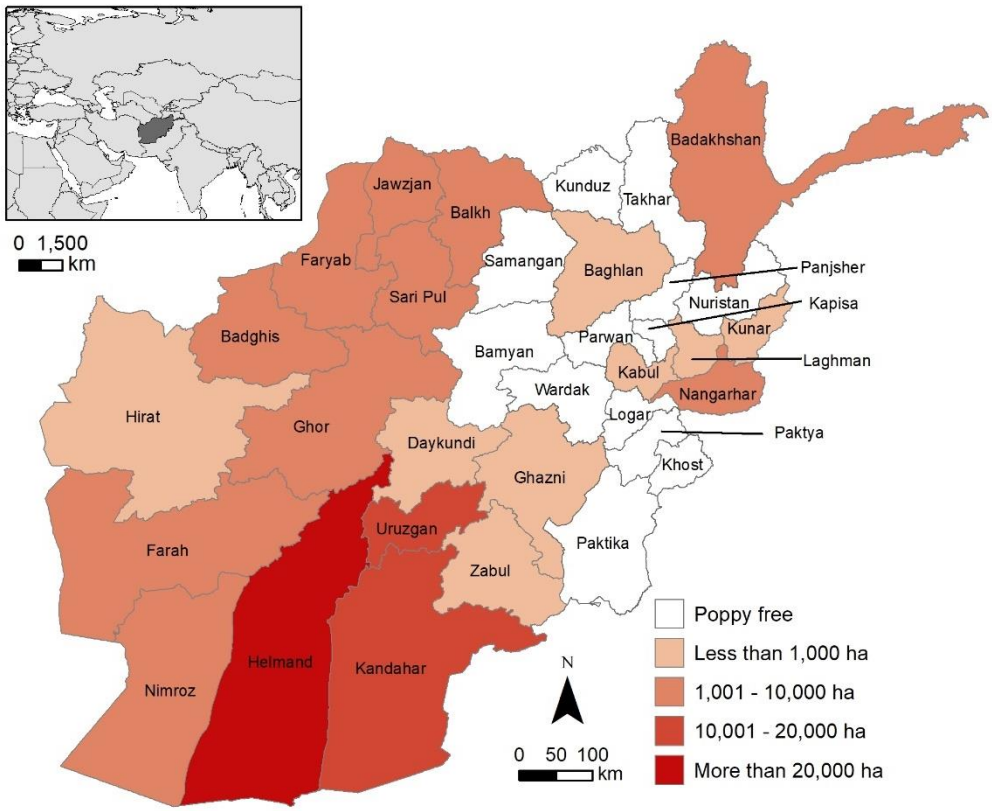

Figure 1-2. Distribution of opium poppy cultivation across Afghanistan from the 2019 annual survey (UNODC, 2021)

1.4 Global opium monitoring using remote sensing

The UNODC and Afghanistan's National Statistic and Information Authority (NSIA) operate an annual survey to monitor opium cultivation in Afghanistan to aid counter-narcotic efforts. Remote sensing has been used to conduct the survey since 2002 because of difficulty in accessing areas under cultivation. Afghanistan is known for its poor infrastructure and variable terrain (Chabot and Dorosh, 2007), therefore remote sensing satellites provide a pragmatic approach for agricultural land and opium poppy discrimination. The main area estimation

method used for provinces with high opium cultivation (>1,000 ha) is a ratio-based sampling approach using the agricultural mask and the proportion of poppy within the selected samples (Figure 1-2) (UNODC, 2018b). The agricultural mask refers to the agricultural area under cultivation, often referred to as the risk area. There are two types of agricultural mask for different purposes: 1) the active mask and 2) the potential mask. The active agricultural mask delineates only agriculture currently under cultivation and is useful for monitoring annual changes to the agricultural area. The potential mask delineates all agriculture areas previously under cultivation with new active agricultural areas added from Landsat-8 and Disaster Monitoring Constellation (DMC) imagery on an annual basis after each survey (UNODC, 2015).

The agricultural area is important in the opium survey for targeting sample locations to delineate opium poppy cultivation. The potential mask is used by the UNODC to ensure no areas of agriculture are omitted in their calculation of opium cultivation estimates. Opium monitoring conducted by the UNODC favours over classification of the risk area to ensure no areas of opium poppy are missed from the survey. The potential mask is created by using an unsupervised ISODATA classification using medium resolution imagery to identify new areas of agriculture that were not in the previous year's mask. These new areas undergo quality control by visual human interpretation as the security risks for surveyors to carry out ground truthing are too greater to conduct a systematic accuracy assessment (UNODC, 2018b). Human interpreters use true colour, false near-infrared and short-wave infrared composites of satellite imagery to distinguish new areas of agriculture. Manual interpretation is needed as the unsupervised classification is unable to separate some areas of agriculture from natural vegetation, such as steppe and semi-desert vegetation which has been used for grazing livestock on a seasonal basis (FAO, 2019). Natural vegetation can be distinguished from arable agriculture in satellite imagery from differences in the association with other land covers (e.g. mountainous terrain), colour and shape.

UNODC use area sampling frames to estimate opium poppy cultivation and are commonly used for conducting crop production estimates and to gather other

agricultural information (UNODC, 2015; Boryan *et al.*, 2017). Each province is split into a 5 × 5 km grid to create individual samples and form the sampling frame. These sample locations are intersected with the potential agricultural area and samples with less than 0.25 km^2 of potential agricultural area are removed. The number of samples selected for each province is determined based on the size of the agricultural area and the variability of poppy cultivation (UNODC, 2015). These 5 × 5 km samples are intersected with high resolution satellite imagery timed to coincide with peak opium biomass for trained human interpreters to delineate the area under poppy cultivation within the agricultural area for each sample. The mean ratio of opium poppy and agricultural land for each province is subsequently extrapolated across the potential mask outside the frame samples to quantify the opium estimate using (Equation 1-1):

$$Y_k = \mathrm{X}\frac{\sum_{i=1}^{n_k} y_i}{\sum_{i=1}^{n_k} x_i} \qquad 1\text{-}1$$

where, Y_k is the area of opium poppy cultivation, $\mathrm{n_k}$ is the number of 5 × 5 km image locations selected from the frame sample across the province k, y_i is the area of poppy cultivation in image i, x_i is the potential agricultural land available for poppy cultivation (potential agricultural mask) in image i and X is the total potential land available for poppy cultivation in province k (UNODC, 2021a).

District level estimations use only samples located within the district, and where there are no samples the agricultural area and poppy cultivation is accounted for in the neighbouring district (UNODC, 2018b). The other two types of estimation are a targeted approach and a village survey. The targeted approach is used for provinces with low levels of opium cultivation (<1,000 ha) (Figure 1-2) and estimates are derived from cultivation areas intersected with satellite imagery (UNODC, 2018b). Village surveys are only conducted in poppy-free status provinces.

Outside of Afghanistan the two other main areas for illicit opium production are Myanmar and Mexico. The UNODC use similar monitoring methodologies for these areas. Myanmar is estimated to have grown 29,500 ha of opium poppy in

2020 with 84% grown in Shan State (UNODC, 2021c). The potential area for opium cultivation in Myanmar is identified annually from a land cover image classification of 22 m resolution DMC imagery from February 2011 as there is limited change in the agricultural area. Prior knowledge of opium free areas based on ground information are factored into defining the potential area for the frame sample selection. Opium cultivation in Myanmar is only grown at altitudes higher than 600 m and the area is limited to these higher altitudes, especially in Shan State where opium poppy is grown to achieve favourable diurnal temperatures (Tian *et al.*, 2011). Opium poppy fields are manually delineated using very high resolution imagery (Pleiades at 50 cm spatial resolution). Opium poppy fields in Myanmar are small and larger areas of agriculture are removed from the potential area as they are assumed to be poppy-free. The same ratio estimate for Afghanistan is used to estimate total opium cultivation by extrapolating the ratio of opium poppy to agricultural land to the risk area outside of the frame sample.

Mexico is estimated to have grown 21,500 ha of opium poppy in 2019 with cultivation primarily in the states of Oaxaca, Guerrero and Nayarit (UNODC, 2021b). Frame sampling is used by the UNODC for Mexico opium cultivation statistics by using the ratio estimate between agricultural area and opium poppy. The area under cultivation does not alter drastically between years therefore the same 300 frame samples have been used in previous years (UNODC, 2021b). Manual interpretation of opium poppy fields within the potential agricultural mask is carried out by human interpreters using several band combinations on medium and high-resolution imagery (SPOT-6 and 7, Geo Eye-1 and Worldview 2 and 3) to distinguish the difference between opium poppy and other arable agriculture.

The main difference between opium monitoring in Afghanistan and other regions is the constantly changing arable agricultural area and requires annual delineation for estimating opium poppy cultivation. The constraint of the method used by the UNODC is this area is very large and would be time-consuming to visually inspect the active agricultural mask year-on-year, therefore they use the potential mask to only identify new agricultural areas. Agricultural areas out of rotation are not removed every year from the potential mask which leads to a

larger agricultural mask. The ratio-based estimation used for opium production estimates by the UNODC uses the potential agricultural mask as a multiplier to the proportion of opium poppy found within each frame sample, therefore opium production estimates are likely to be inflated as it does not include only active areas of agriculture.

1.5 Research context

The work the UNODC and NSIA do to accurately estimate opium poppy is essential for informing counter-narcotic policy. Satellite imagery is used to identify potential areas under agricultural cultivation and quantify the proportion related to illicit opium production. The current approach uses an automated unsupervised image classification with manual editing by human interpreters to determine the potential area of agriculture (UNODC, 2021a). The main challenge of this methodology is to produce an accurate agricultural area estimate. This is important as the area is used as a multiplier in the ratio-based poppy cultivation estimate (Equation 1-1). The agricultural area is also used for targeting the sample locations for manual interpretation of opium poppy fields to determine the ratio between opium poppy and available agricultural land for cultivation. These areas can be used to inform counter-narcotic eradication programmes to reduce opium cultivation. The automated classification is not able to separate between similar land cover types (e.g. agriculture and natural vegetation) and requires time consuming manual interpretation to remove these areas. The active agricultural mask can be difficult and time-consuming to manually edit because of the dynamic nature of crops, visual confusion with natural vegetation and inter-seasonal variation. The current methodology implemented by the UNODC does not consider the spatial distribution of opium poppy and only provides an estimate of the area. Further understanding of the distribution of annual change within the active agricultural mask is important for supporting counter-narcotic policy by informing national and international stakeholders. New knowledge on land cover changes related to opium cultivation across Afghanistan can be achieved by utilising historical data to aid agricultural land classification. Outside of opium poppy estimates, crop production estimates are often used for agricultural

decision making (FAO, 2016); therefore, it is essential to provide both a practical and effective methodology to accurately quantify agricultural areas.

The volume and variety of these Earth observation datasets provides opportunities for agricultural patterns to be extracted from legacy datasets across Afghanistan (Casu *et al.*, 2017). Agricultural productivity has intra-seasonal variation (Simms *et al.*, 2014), therefore agricultural monitoring requires timely data processing and dissemination. The lack of data-driven processing is limiting the full extent of data analytics within remote sensing to facilitate further understanding of vegetation dynamics (Guo and Mennis, 2009). Previously, remote sensing studies have primarily focused on single date image classifications to extract agricultural areas with multi-temporal analysis becoming increasingly common for image classification (Petitjean, Ketterlin and Gançarski, 2011; Verbesselt, Zeileis and Herold, 2012; Yongguang Zhai, Qu and Hao, 2018).

The main barrier for using these supervised techniques for land cover classification has been access to large labelled datasets of the landscape (e.g. agriculture and non-agriculture land cover classes) (Ball, Anderson and Chan, 2017). These are scarce in the remote sensing community because their creation is labour intensive (Vali, Comai and Matteucci, 2020). Labelled datasets are required to train and validate data-driven classifiers and limited access to these emphasises the importance of developing methodologies to take full advantage of all available datasets. The restricted availability of the UNODC's labelled datasets of agricultural area for this thesis provide the opportunity to utilise data intensive machine learning techniques, such as Convolutional Neural Network (CNN) and Fully Convolutional Network (FCN), for image classification. The limited use of CNNs and FCNs for pixelwise classification within remote sensing provides an exciting opportunity to utilise, optimise and explore the use of this technique (Ball, Anderson and Chan, 2017). Traditionally labelled individual pixels or objects with associated satellite imagery are used to train machine learning classifiers, whereas image chips can only be used for implementation into the CNN and FCN models and is a unique approach to image classification.

This thesis critically evaluates cutting-edge technologies with historical data to classify and monitor active agricultural land to understand annual changes in cultivation practices and remove the need for human interpretation. The aim of this research is to derive information about annual changes in land-use related to opium poppy cultivation using convolutional neural networks with Earth observation data.

The objectives of this study are to:

1. Critically evaluate the use of CNNs for delineating Afghanistan's active agricultural mask compared to human interpretation
2. Develop generalised CNN models for classifying the agricultural mask and evaluate the uncertainty on measuring the active agricultural area
3. Understand the annual changes in agricultural land use in relation to opium poppy in Helmand Province, Afghanistan between 2010 and 2019

1.6 Thesis structure

This thesis is made up of 3 separate analytical chapters followed by a discussion and summary of the findings for each thesis objective. This introduction provides a literature review of the current uses of remote sensing data with an emphasis on the use of image classification techniques and accuracy metrics for land cover mapping. Afghanistan is presented as a case study to utilise these image classification techniques with a background for monitoring opium poppy using remote sensing and the annual opium survey methodology currently used by the UNODC and the context of this research.

Chapter 2 presents a strategy for training and evaluating CNNs for active agricultural mask classification in comparison to human interpretation (thesis objectives 1). Multiple images were critically evaluated between 2007 and 2009 and highlighted the potential for transferring knowledge between different years.

Chapter 3 provides a methodology for training generalised CNNs for automated agricultural land classification across image datasets (thesis objective 2). Understanding of how CNNs learn agricultural patterns using spectral, spatial, and textural data is presented. Further understanding on how transferable CNNs

are between different image datasets (DMC, Landsat, and Sentinel-2) using different standardisation procedures is presented. A localised accuracy assessment called ‘localised intersection over union’ is presented for assess local differences in classification accuracy to better understand the strengths and uncertainties of automated classification.

Chapter 4 presents a case study for monitoring agricultural land in Helmand Province across different image datasets with generalised CNN models from Chapter 3 between 2010 and 2019 (thesis objective 3).

Chapter 5 presents a discussion of the main findings for each thesis objective and the contributions to knowledge and wider impact of the research for opium monitoring and image classification are presented.

Chapter 6 summarises the main conclusions from each thesis objective and recommendations are provided for using data-driven techniques to further understand landscape change related to opium cultivation in Afghanistan.

Chapter 2. Replacing human interpretation of agricultural land in Afghanistan with a deep convolutional neural network

This chapter investigates the development of a Convolutional Neural Network (CNN) to automate the classification of agriculture from medium resolution satellite imagery as an alternative to manual interpretation agricultural delineation across Helmand and Kandahar, Afghanistan (thesis objective 1). The effect of input image chip size, training sampling strategy, elevation data and multi-seasonal imagery are investigated to accurately delineate active agricultural land.

Underlying data for this chapter is available at:
https://doi.org/10.17862/cranfield.rd.13228634

Published as: Hamer, A.M., Simms, D.M. and Waine, T.W. (2021) Replacing human interpretation of agricultural land in Afghanistan with a deep convolutional neural network. *International Journal of Remote Sensing*, 42 (8), pp. 3017-3038. doi: https://doi.org/10.1080/01431161.2020.1864059

2.1 Introduction

The United Nations Office on Drugs and Crime (UNODC) and Government of Afghanistan conduct an annual survey to estimate the production of opium in Afghanistan, a country responsible for 82% of global production (UNODC, 2019b). The opium trade fuels poverty, political instability and insurgency; hampering development efforts. The survey plays an important role in monitoring the extent and evolution of illicit opium production for the development of counter-narcotic policy. Within the survey, the accurate mapping of agricultural land, known as the agricultural mask, is essential for robust statistical sample design and production estimates. The mask is reviewed each year because of the large variation in the annual distribution of agricultural land (UNODC, 2018b). The current method uses unsupervised classification of medium resolution satellite imagery, such as Land-Remote Sensing Satellite System (Landsat) 8 (30 m). This approach has difficulty separating natural vegetation from agriculture, so requires post-classification manual refinement. This is time-consuming and uses trained interpreters with knowledge of local agronomic practice in order to accurately map agricultural land.

Machine learning techniques have been shown to increase the accuracy of image classification (Lecun, Bengio and Hinton, 2015; Belgiu and Drăgu, 2016; Yamashita *et al.*, 2018; Pouliot *et al.*, 2019). Of particular importance are Convolutional Neural Networks (CNNs), which can outperform other machine learning classifiers such as Support Vector Machines (about 19%) at image labelling (Russakovsky *et al.*, 2015) and Random Forests (7%) on mapping from medium resolution imagery (Pouliot *et al.*, 2019), and can match human performance in certain image related tasks (Haenssle *et al.*, 2018). They are inspired by the connections between neurons in the cerebral cortex (Ball, Anderson and Chan, 2017) and are made up of convolutional layers that encode the spatial and spectral elements of image features from large training datasets. The rapid improvements in accuracy have been achieved through the development of new CNN architectures for image classification (Zeiler and Fergus, 2013; Simonyan and Zisserman, 2015). CNNs are able to capture complex contextual information in a similar way to manual image-interpretation,

where associated features and the spatial context of observations are used as interpretation keys (e.g. field patterns, irrigation canals and topography). This overcomes one of the limitations of pixel-based classification for mapping agricultural land in Afghanistan, where spectral separation alone is not able to discriminate between natural vegetation and agriculture.

The aim of this study is to determine whether CNNs can perform the role of a human interpreter in delineating agricultural land from medium-resolution imagery. Access to densely labelled agricultural masks from opium surveys in Afghanistan provide the necessary data to develop an optimal CNN training strategy for agricultural delineation and evaluate its performance across multiple years.

2.2 Convolutional neural networks

CNNs are widely used in image classification because of their high performance and ability to accept multi-dimensional pixel arrays (Lecun, Bengio and Hinton, 2015). These networks are designed to adaptively learn spatial features from a set of labelled examples through a backpropagation algorithm (Yamashita *et al.*, 2018). Each convolutional layer in the neural network runs a fixed-sized filter matrix across the image at a defined spacing, or stride, to generate a feature map, which forms the input to the next layer. A rectified linear unit (ReLU) activation function is used to introduce non-linearity and avoid saturation during learning (Nogueira, Penatti and dos Santos, 2017). Pooling layers are used to reduce the dimensionality of feature maps, using a maximum or average filter matrix, by downsampling the spatial resolution of the input layers. Fully connected layers join each node from the previous layer, flattening them out into one-dimensional feature maps. The final layer is a fully connected layer that calculates class probabilities for each instance using a classification activation function, usually a Softmax (Goodfellow, Bengio and Courville, 2016).

The network is trained using a gradient-based optimisation algorithm, most commonly Adam (Kingma and Ba, 2015), Stochastic Gradient Descent (SGD), Adaptive Gradient Algorithm (AdaGrad) (Duchi, Hazan and Singer, 2011) or Root Mean Square Propagation (RMSprop) (Tieleman and Hinton, 2012), and a loss

function that measures the agreement between the model predictions and the reference data labels. Normalisation is used as a data pre-processing step, usually *z*-score normalisation (Chollet, 2017), to scale input data to a common range. The optimisation algorithm minimises the loss by altering the layer weights for each batch of reference data fed into the CNN. Training stops once there is no longer any significant decrease in the loss, usually after many epochs, where an epoch represents one complete pass of the reference data through the network during training.

There are two approaches for CNN training, known as end-to-end and transfer learning. End-to-end learning uses the input data alone to identify the target object's features from randomly initialised filter weights. Transfer learning uses pre-trained filter weights from a previous application. ImageNet, a dataset of commercial photographs used for visual object recognition, is commonly used for image transfer learning, particularly where training data are limited (Shin *et al.*, 2016). Transfer learning has been beneficial for classification as similar features often transcend individual image recognition tasks and reduce the requirement for large labelled datasets (Yosinski *et al.*, 2014).

Common CNN architectures include Visual Geometry Group (VGG) 16 and VGG 19 that use small convolutional filters (3 × 3) across their 16 and 19 layer networks to achieve state-of-the-art classification accuracy on ImageNet (Simonyan and Zisserman, 2015). The residual network (ResNet50) architecture, a 50 layer network, found deeper networks were beneficial to classification accuracy and has outperformed VGG CNNs (He *et al.*, 2016).

Existing applications of CNNs for remote sensing data often use imagery benchmark datasets, including University of California (UC) Merced land use dataset (Yang and Newsam, 2010), Aerial Image Dataset (AID) (Xia *et al.*, 2017) and Brazilian coffee scenes (Penatti, Nogueira and Santos, 2015; Nogueira, Penatti and dos Santos, 2017; Deng *et al.*, 2018; Zhang, Tang and Zhao, 2019). These use a similar approach to photography-based object recognition and classify whole images, or image subsets known as chips, with a single label. Across these benchmark datasets there are differences in training sample sizes,

input image chip sizes and CNN model architectures. Image chip sizes vary greatly, with the UC Merced dataset (0.3 m) (Yang and Newsam, 2010) at 256 × 256 pixels, AID dataset (0.5 to 0.8 m) (Xia *et al.*, 2017) at 300 × 300 pixels and the Brazilian coffee scenes dataset (10 m) (Penatti, Nogueira and Santos, 2015) at 64 × 64 pixels. Sample sizes are constrained by the amount of labelled data available with samples ranging from 10s (Fu *et al.*, 2017) to 100s (Deng *et al.*, 2018) to 1000s per class (Cheng, Han and Lu, 2017). CNN image categorisation is used outside of benchmark datasets (Koga, Miyazaki and Shibasaki, 2018; Liu *et al.*, 2018; Feng *et al.*, 2019), but is again constrained by the requirement for large datasets. These applications use the normalised spectral bands exclusively for prediction, while for other machine learning algorithms, ancillary data, such as distance to water, elevation and economic indicators (Gislason, Benediktsson and Sveinsson, 2006; Lucas *et al.*, 2011) are used to improve classification performance.

Land cover classifications are validated using hold-out samples to calculate accuracy metrics. Overall accuracy (OA) is the number of correctly classified pixels in comparison to the reference data and widely adopted within remote sensing (Foody, 2002) (Equation 2-1),

$$\mathrm{OA} = \frac{\Sigma_i n_{i,i}}{\Sigma_i t_i} \qquad 2\text{-}1$$

where $n_{i,i}$ is number of pixels predicted as class i belonging to class i and t_i is total number of pixels belonging to class i in the reference data. The Kappa coefficient (K) is used to quantify the statistical significance in comparison to random performance (Cohen, 1960) (Equation 2-2),

$$K = \frac{p_o - p_e}{1 - p_e} \qquad 2\text{-}2$$

where p_o is the empirical probability of correctly labelled samples and p_e is the expected probability of correctly labelled samples by random chance.

The Intersection over Union (IoU), also known as Jaccard index, gives the similarity between the predicted region and reference region by identifying

overlapping regions (Long, Shelhamer and Darrell, 2015). Two variations of IoU commonly used for deep learning applications are the mean IoU (mIoU) (Russakovsky *et al.*, 2015) (Equation 2-3),

$$\mathrm{mIoU} = \frac{1}{k}\sum_{i}\frac{n_{i,i}}{t_i + \sum_j n_{j,i} - n_{i,i}} \qquad 2\text{-}3$$

and frequency weighted IoU (fwIoU) (Equation 2-4),

$$\mathrm{fwIoU} = \frac{1}{\sum_i t_i}\sum_{i}\frac{t_i n_{i,i}}{t_i + \sum_j n_{j,i} - n_{i,i}} \qquad 2\text{-}4$$

where $n_{j,i}$ is the number of pixels predicted as class j belonging to class i and k is the number of classes in the reference data.

2.3 Materials and methods

2.3.1 Study site

The study area is the provinces of Helmand and Kandahar in the south of Afghanistan, covering an area of 81,383 km^2 (Figure 2-1). These are the largest opium producing provinces in Afghanistan with an estimated 160,208 ha grown in 2018, accounting for 61% of national opium cultivation (UNODC, 2018b). They also contain the highest proportions of irrigated agricultural land in Afghanistan, 342,172 ha and 312,465 ha respectively (FAO, 2016). The main area of cultivation is in the Helmand valley with large areas of natural vegetation in northern Kandahar (Figure 2-1). Helmand and Kandahar contain a wide range of agricultural landscapes, including rain-fed agriculture in lowland and highland areas, fruit trees, vineyards, marginal agriculture and natural needle leaved forests (FAO, 2016). Agriculture in Afghanistan is predominately reliant on snowpack melt to supply sufficient groundwater for irrigation. Water availability is vital for agricultural production and is the main driver for changes in agricultural area (Shahriar Pervez, Budde and Rowland, 2014; UNODC, 2019b).

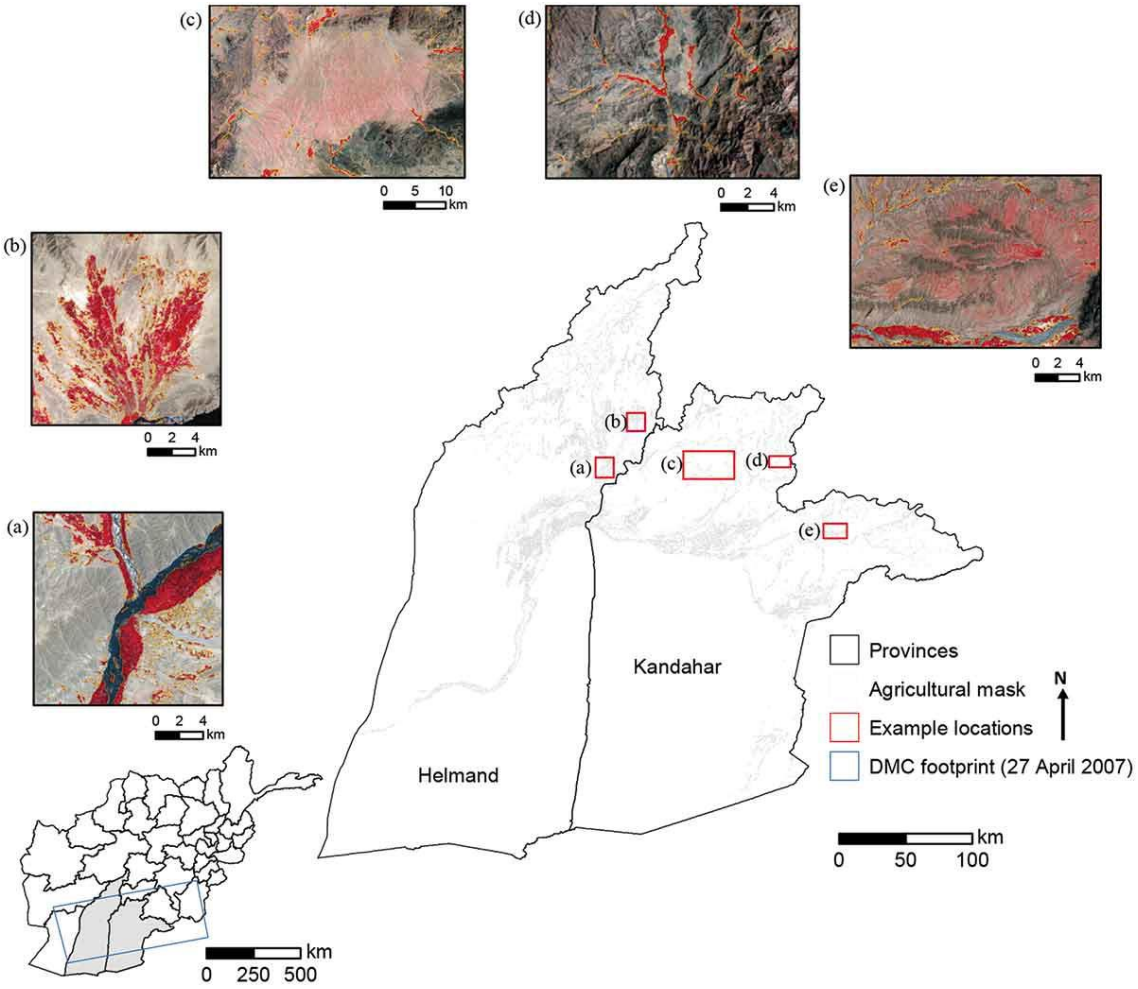

Figure 2-1. Helmand (centre 31.3636° N, 63.9586° E) and Kandahar (centre 31.6289° N, 65.7372° E) provinces, Afghanistan showing the agricultural area in 2007. Insets show locations used for detailed evaluation with 2007 agriculture delineated in yellow; (a) and (b) show areas of intensive agriculture, (c) and (e) show areas with natural vegetation and (d) shows agriculture in the highlands of Kandahar. Inset background is a false colour DMC image (NIR, R, G at 32 m) from 27 April 2007.

2.3.2 Image data and agricultural masks

The agricultural masks and associated Disaster Monitoring Constellation (DMC) imagery from the 2007 to 2009 opium cultivation surveys were used as labelled reference datasets of agricultural land. These image years were used for analysis as they are the latest complete agricultural masks available for the study. These densely labelled data were originally created from an unsupervised classification of orthorectified multispectral DMC imagery, with near-infrared (NIR) (0.76 to 0.90

μm), red (R) (0.63 to 0.69 μm), and green (G) (0.52 to 0.62 μm) bands at 32 m spatial resolution. DMC imagery was used because its high temporal frequency (up to daily) and wide area coverage were well suited to target the peak of opium poppy biomass. Images were collected on: 27 April 2007, 24 March, 7 April and 24 April 2008 and 25 March, 3 April, and 8 April 2009. The same area was used for analysis between 2007 and 2009 based on the provincial boundaries of Helmand and Kandahar and the DMC footprint from 2007 (Figure 2-1), which resulted in multiple images for 2008 and 2009. The unsupervised classification was performed using the Iterative Self-Organising Data Analysis Technique (ISODATA) by separating the imagery into 30 classes with each output cluster manually labelled as agriculture or non-agriculture. These classifications were then manually edited as some clusters contained pixels of both agriculture and natural vegetation. Editing was done by trained interpreters with access to ancillary information from high resolution commercial IKONOS imagery (Taylor *et al.*, 2010). Finally, the masks were quality checked and compared with data from other years to ensure consistency of interpretation.

2.3.3 Model selection

The best performing CNN model for chip classification was selected from three CNN models with previously high performance in other image classification tasks (Simonyan and Zisserman, 2015; He *et al.*, 2016): ResNet50, VGG16 and VGG19. Firstly, the input image from 27 April 2007 was split into pixel chips as CNN architectures require images rather than individual pixels to function. Pixel chips of size 33 × 33 pixels were created using a non-overlapping grid as this was the smallest input image size that can be used based on these model architectures. The class of the centre pixel was used as the label for each chip as the goal was to classify whole images pixel-by-pixel through reconstructing overlapping chips (Kampffmeyer, Salberg and Jenssen, 2016). All chips were *z*-score normalised and a 75% random sub-sample was selected for each class for training and the remaining 25% were used for independent validation. The agriculture samples in the training datasets were augmented at 90°, 180° and 270° angles (also known as horizontal and vertical flipping) to increase the

number of samples by a factor of 2. The training and validation datasets were balanced by under sampling the majority class (non-agriculture) to match the number of samples in the agriculture class. This resulted in a total of 11,664 training samples and 1,944 validation samples across the two classes using the reference agricultural mask data. Each model was then trained end-to-end and using an ImageNet transfer learning model with an Adam optimizer and a learning rate of 0.0001. Model performance was assessed on the validation samples using overall accuracy and the Kappa coefficient. All experiments were undertaken using Keras (Chollet, 2015) with a TensorFlow (Abadi *et al.*, 2015) backend on a workstation with a Intel Xeon E5-2687W v3 CPU, NVIDIA Quadro K2200 GPU and 64 GB of RAM. As a benchmark the CNNs were compared to a Random Forest, a pixel-based machine learning classifier, to provide comparison between a pixel-based and chip-based classifier. The number of trees (100), tree-depth (2) and maximum features used to split each internal node (10) were determined as the optimal hyper-parameters by grid search using a stratified 3 fold cross validation on the training data. More information on this experiment can be found in Appendix A.

All CNN models were able to classify agricultural chips better than the Random Forest classifier with up to a 9% improvement. The ResNet50 architecture achieved the highest overall accuracy and Kappa coefficient using transfer learning, 99.02% and 0.98 respectively (Table 2-1) and was used for all further experiments.

Table 2-1. Summary of ResNet50, VGG16 and VGG19 CNN model performance for end-to-end (EE) and transfer learning (TL) training parameters for 10 epochs using 33 × 33 pixel image chips from DMC (NIR, R, G) imagery for 27 April 2007 across Helmand and Kandahar. The best performing model and metrics are highlighted in bold.

CNN model	Training approach	Model performance (10 epochs) (%)	
		Overall accuracy	*K*
ResNet50	EE	96.81	0.94
	TL	**99.02**	**0.98**
VGG16	EE	98.82	**0.98**
	TL	97.48	0.95
VGG19	EE	98.30	0.97
	TL	96.40	0.93
Random Forest	N/A	89.22	0.88

2.3.4 Experiment 1: Image chip size and CNN training strategy

Image chipping is an important pre-processing step for CNNs with fully connected layers. Three sets of fixed non-overlapping grids at small, medium and large scale with a centre pixel (33 × 33 pixels, 65 × 65 pixels, and 129 × 129 pixels) were created to provide model input data at different spatial scales to investigate the effect of chip size. The chips were *z*-score normalised and the reference data label for each chip was assigned based on the centre pixel.

The agriculture class is heavily under-represented in the data and accounts for only 5% of samples for all image chip sizes (Table 2-2). Data augmentation was used for all experiments to increase the number of agriculture samples by a factor of 2. Datasets for all experiments were balanced in number between non-agriculture and augmented agriculture samples using a stratified random sample from the non-agriculture group. Sub-stratification was carried out within the non-agriculture group to select chips at the boundary of agricultural land and areas of natural vegetation, which are known confusion areas for agricultural mapping (Simms *et al.*, 2016). Non-agriculture chips with natural vegetation were selected using an Otsu threshold (Otsu, 1979) on a Normalised Difference Vegetation Index (NDVI) image of the study extent with the reference agricultural mask applied. Masking agricultural land forces the NDVI threshold to focus on selecting samples located in areas of natural vegetation.

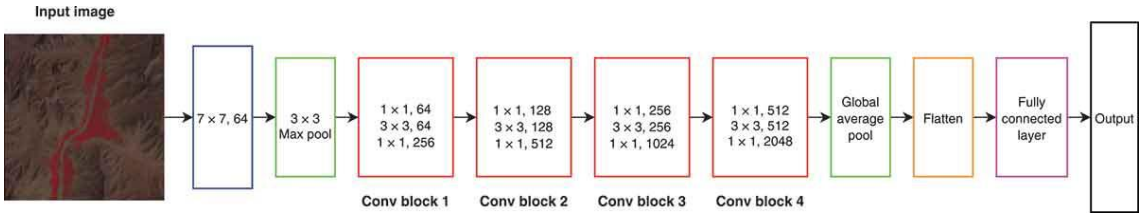

Figure 2-2. The ResNet50 model architecture used for agricultural mask production

Three separate strategies were used to identify how best to train the ResNet50 CNN (Figure 2-2). These were: (1) random sampling of agriculture and non-agriculture; (2) sampling from boundary cases; and (3) targeted sampling from non-agriculture chips containing natural vegetation and boundary cases. Boundary cases are defined as chips with a non-agriculture label that contain agriculture within the chip. These samples have been introduced to provide more difficult interpretation cases to train and validate the model.

The number of selected training and validation samples (from the total number in Table 2-2) for each chip size remained consistent, regardless of training strategy. Chip size 33 × 33 used 11,664 training and 1,944 validation samples, 2,868 training and 478 validation samples were used for chip size 65 × 65 and 724 training and 120 validation samples were used for chip size 129 × 129.

Table 2-2. Total number of chips (*n*) in the study area for agriculture, non-agriculture, and boundary samples (non-agriculture chip label, but with agriculture present) for each size of image chip for 2007 data with the percentage of total area

Input image chips size (pixels)	Agriculture		Boundary		Non-agriculture	
	n	Percentage area (%)	*n*	Percentage area (%)	*n*	Percentage area (%)
33 × 33	3,899	5.4	10,699	14.8	57,556	79.8
65 × 65	1,012	5.5	4,414	24.0	12,961	70.5
129 × 129	250	5.5	1,599	35.1	2,711	59.5

A separate ResNet50 CNN model was also trained to include Shuttle Radar Topography Mission (SRTM) elevation data (resampled to 32 m and min-max normalised) to investigate the effect of ancillary data. The raw SRTM elevation values are provided rather than other representations (e.g. slope or aspect). Aspect is not considered in these experiments as augmentation of training

samples has removed directionality from model learning. Slope is not provided as the CNN uses kernel operations to identify underlying patterns, which slope is a kernel function therefore does not need to be explicitly encoded. Most CNN architectures are developed for photographs, with input channels restricted to three. The green band was substituted for elevation as the NIR and R spectral bands were considered to be of higher importance for monitoring of vegetation (Panda, Ames and Panigrahi, 2010).

The CNN outputs a single prediction for each image chip, so reconstruction is required to classify a whole image pixel-by-pixel. The reconstruction process used in this study applies the trained CNN model to each pixel in the image using an overlapping sliding window the same dimensions as the image chip used during training. This achieves a pixelwise classification by labelling the centre pixel of each sliding window with the model prediction for the chip (Figure 2-3).

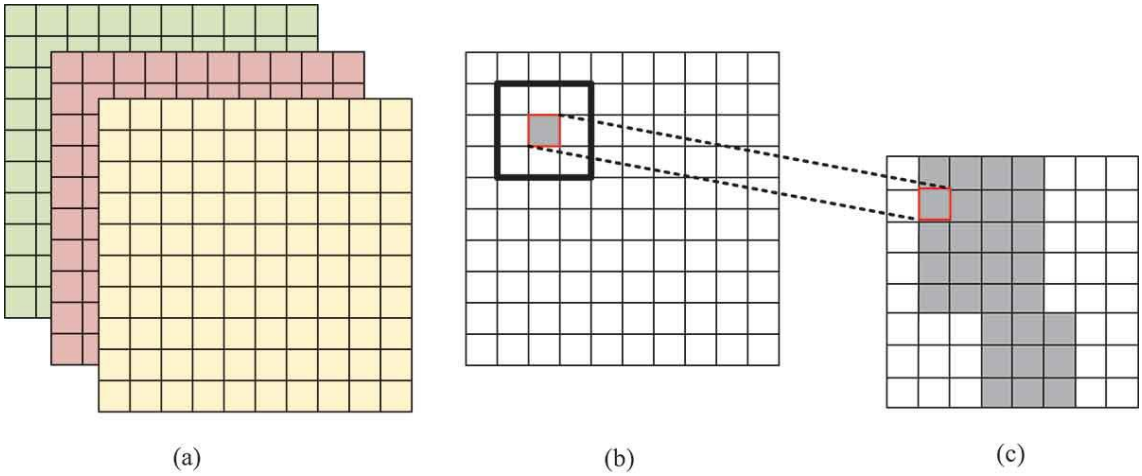

Figure 2-3. The process for pixelwise agricultural mask prediction using a sliding window with a trained CNN model. (a) 3 band satellite image chip, (b) sliding window applying CNN model e.g. 3 × 3 and (c) pixelwise agricultural mask production

2.3.5 Experiment 2: Transfer learning across multiple seasons

The ability to retrain CNN models with new data is a desirable attribute for image classification. The transferability of agricultural features for continual refinement of a multi-seasonal classifier was explored using 2007, 2008 and 2009 data. Transfer learning was used to update the model previously trained on 2007 data

with data from 2008, the combined model was then updated with 2009 data. The best performing training strategy from experiment 1 was used to create the 2008 and 2009 input data, which were balanced using the same image augmentation as the 2007 dataset. The total number of training samples used for 2008 and 2009 were 11,960 and 12,032 with 1,994 and 2,006 validation samples, respectively. The proportion of training data was varied (25%, 50%, 75% and 100%) to identify the number of samples required to update each year's model to a similar level of accuracy.

2.4 Results

The CNN model outputs for agricultural delineation were found to consistently achieve higher accuracy with ImageNet transfer learning across all chip sizes than end-to-end learning (Figure 2-4). The major difference between transfer and end-to-end learning is shown during the initial 5 epochs with higher initial training and overall accuracy, where the training accuracy is the overall accuracy of the training data. Training accuracies for all image chip sizes were found to achieve a similar accuracy after 50 epochs, unlike validation accuracies. The 129 × 129 chips were found to plateau faster than the other chip sizes across all training strategies and 65 × 65 chips took the longest to train in both transfer and end-to-

end learning. End-to-end CNN models across all training strategies were also found to require more epochs of training than transfer learning.

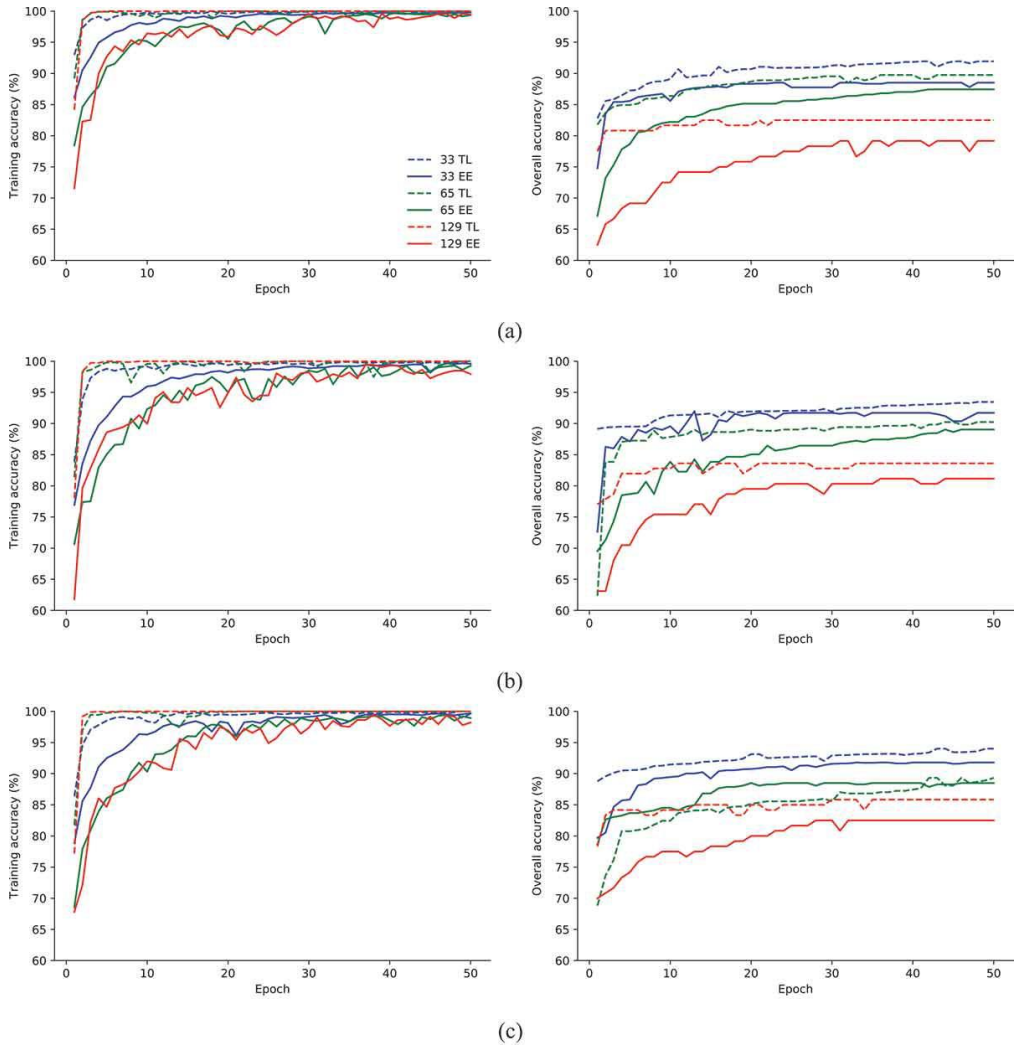

Figure 2-4. Evaluation of training and validation overall accuracy for three different training strategies for different image chips sizes (33 × 33, 65 × 65 and 129 × 129 pixels) using transfer (TL) and end-to-end learning (EE): (a) strategy 1: random agriculture and non-agriculture classes; (b) strategy 2: random agriculture and boundary classes; and (c) strategy 3: random agriculture class, boundary cases and NDVI targeted non-agriculture class

Smaller chips were found to be less generalised than larger chips across various agricultural landscapes after image reconstruction (Figure 2-5). The 129 × 129 chip classification delineates the overall agricultural boundary extent but performs poorly on smaller non-agricultural areas and edge cases. This can be seen as a buffering effect along agricultural boundaries. Boundaries between the

agriculture and non-agriculture class were found to be well-defined for the 33 × 33 chip size. High visual agreement with the reference agricultural mask was achieved, particularly for strong edge cases such as the river valley.

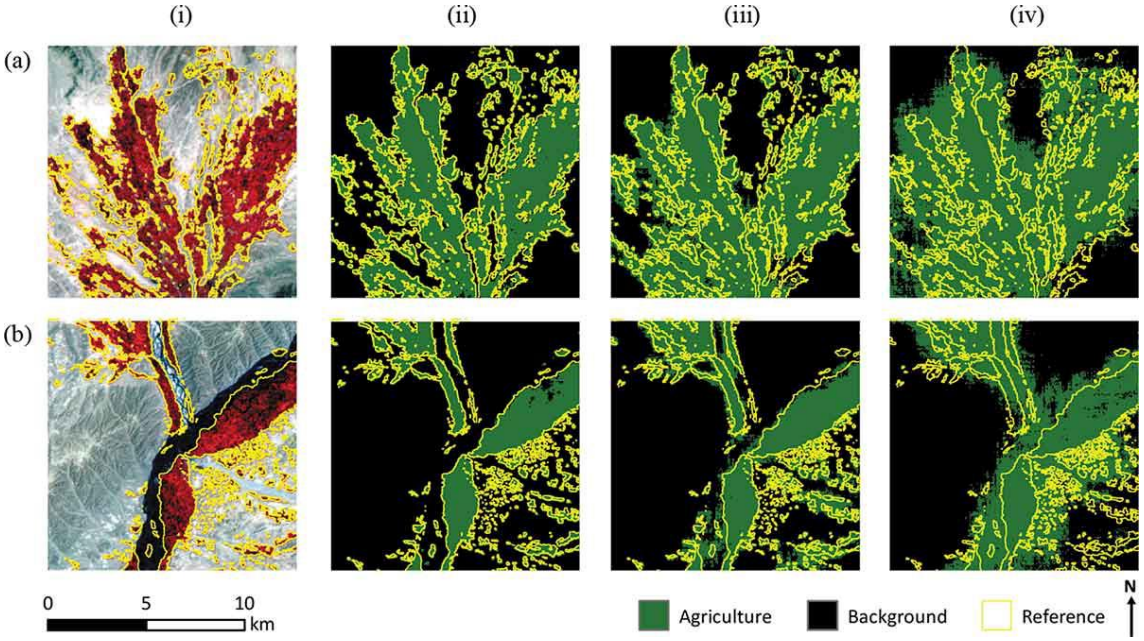

Figure 2-5. False colour DMC imagery (NIR, R, G at 32 m) from 27 April 2007 (i) for an (a) agriculture dominated area and (b) non-agriculture dominated area with corresponding agriculture delineation for three image chip sizes ((ii) 33 × 33, (iii) 65 × 65 and (iv) 129 × 129) using the best-performing ResNet50 CNN (strategy 3, random agriculture, boundary cases and NDVI targeted non-agriculture class with transfer learning)

The ResNet50 CNN model performance summarised in Table 2-3 shows the overall accuracy, Kappa coefficient, mean IoU and frequency weighted IoU, for the three different chip sizes, two model training methods (end-to-end and transfer learning) and the three different training strategies. Overall accuracy is highest for 33 × 33 chip size, with transfer learning outperforming all end-to-end training strategies (Figure 2-4). The best performing model was transfer-trained using data from strategy 3 and 33 × 33 chips, with an overall accuracy of 94.01%, Kappa coefficient of 88.02% and mean and weighted IoU of 50.33 and 67.61 respectively (Table 2-3). Strategy 3 produced the best performing models with the exception of the 65 × 65 size chips, which achieved marginally higher overall

accuracy (+0.5% improvement using transfer learning), Kappa coefficient, mIoU and fwIoU for both end-to-end and transfer learning with strategy 2.

Table 2-3. Evaluation of input chip sizes (129 × 129, 65 × 65 and 33 × 33 pixels) and strategies for CNN models trained using both end-to-end (EE) and transfer learning (TL) across Helmand and Kandahar provinces on DMC (NIR, R, G) imagery in April 2007. Strategy 1: random sampling of agriculture and non-agriculture classes; strategy 2: random sampling of agriculture and boundary classes; strategy 3: random agriculture, boundary cases and NDVI targeted non-agriculture class. The best performing validation metrics for each training strategy are highlighted in bold

Training strategy	Validation metric (50 epochs) (%)	Training approach					
		EE (pixels)			TL (pixels)		
		129	65	33	129	65	33
Strategy 1							
	OA	79.17	87.45	88.53	82.50	89.75	**91.94**
	K	58.33	74.90	77.07	65.00	79.50	**83.88**
	mIoU	30.31	34.80	47.75	30.04	40.32	**49.03**
	fwIoU	34.56	41.73	63.70	33.86	52.04	**65.35**
Strategy 2							
	OA	81.15	89.04	91.72	83.61	90.24	**93.47**
	K	62.30	78.09	83.44	67.21	80.48	**86.93**
	mIoU	32.05	37.36	48.56	28.31	41.66	**49.18**
	fwIoU	36.96	45.40	63.79	31.89	52.95	**65.42**
Strategy 3							
	OA	82.50	88.49	91.79	85.83	89.33	**94.01**
	K	65.00	76.10	83.57	71.67	78.66	**88.02**
	mIoU	28.13	36.33	48.57	29.63	38.17	**50.33**
	fwIoU	31.78	44.43	64.63	33.29	47.58	**67.61**

Table 2-4. Best performing CNN training strategies based on overall accuracy for each image chip size with prediction times using DMC (NIR, R, G at 32 m) imagery samples across Helmand and Kandahar provinces in April 2007. Strategy 2: random sampling agriculture and boundary classes and strategy 3: random agriculture, boundary cases and NDVI targeted non-agriculture class

Input image chip size (pixels)	Best performing training strategy	Prediction time (s) (*n* = 250,000)
33 × 33	Strategy 3: transfer learning	384
65 × 65	Strategy 2: transfer learning	677
129 × 129	Strategy 3: transfer learning	1,677

The fastest prediction times were found using smaller image chips for the same area extent (800 ha). The shortest time was achieved using 33 × 33 chips (384 seconds for 250,000 samples), as opposed to 1677 seconds for 129 × 129 chips (Table 2-4).

Visual evaluation of the Helmand and Kandahar agricultural mask for April 2007 found most complex agricultural areas and highland vegetation were well delineated (Figure 2-6). The CNN identified the distinct difference between features of the background class and agriculture. Large extents of natural vegetation were found to be correctly classified, with small areas of over-classification of agriculture (also referred to as a commission error) (Figure 2-6 (a)). There are also regions of high commission in low lying areas surrounding highland regions in Kandahar (Figure 2-6 (c)).

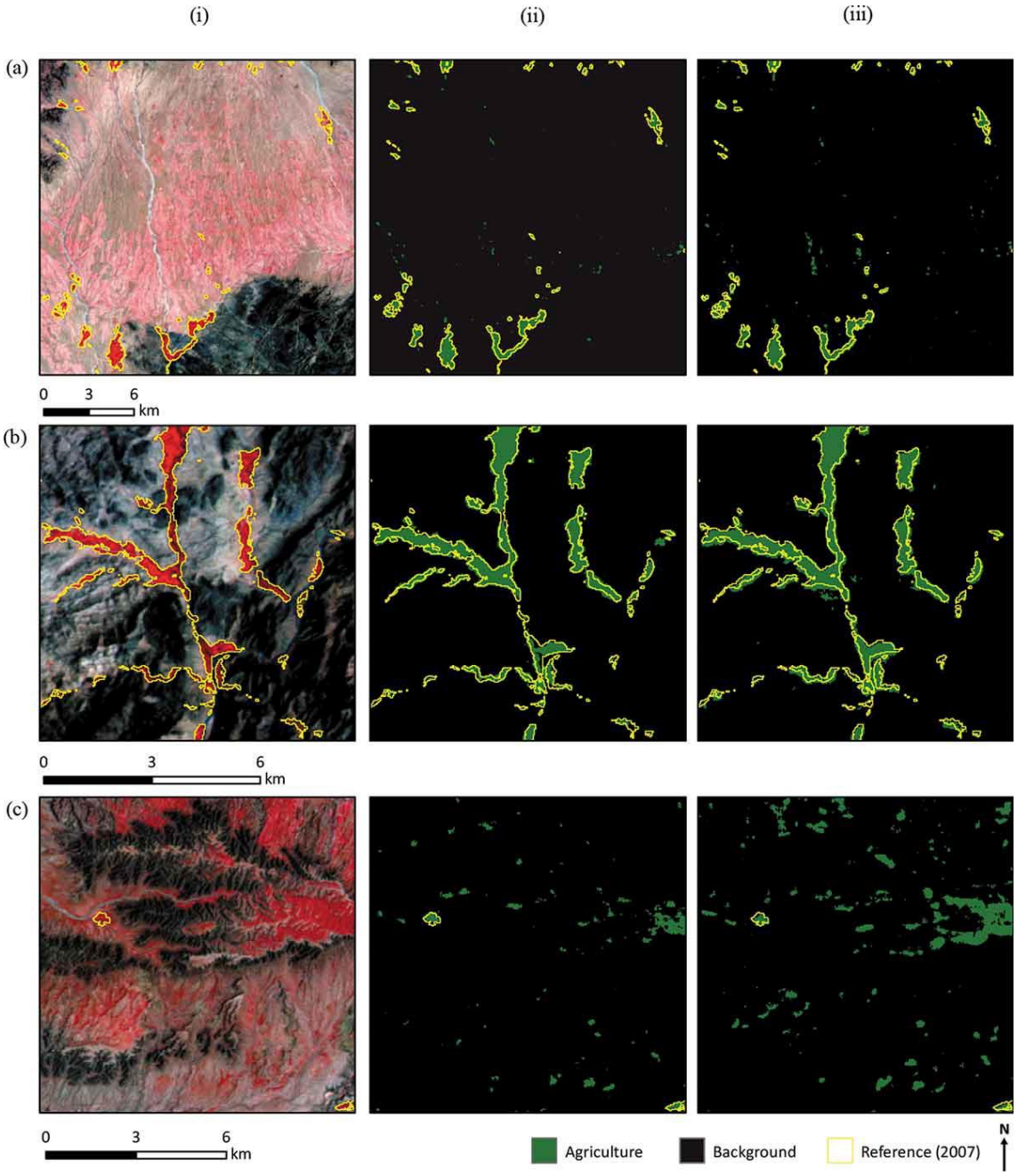

Figure 2-6. Visual interpretation of (i) DMC imagery for the best performing CNN classification model (training strategy 3: random agriculture, boundary cases and NDVI targeted non-agriculture class with transfer learning) with input image chip size 33 × 33 pixels for (ii) spectral and (iii) SRTM elevation data (resampled to 32 m). Image extents have been selected based on prior knowledge of confusion areas for interpretation. (a) Large extent of natural vegetation, (b) well-delineated agriculture in highland areas and (c) commission of agriculture surrounding highland areas in Kandahar, Afghanistan. False colour DMC imagery (NIR, R, G at 32 m) for 27 April 2007.

Substituting elevation data (SRTM) with the green spectral band (Table 2-5) and training with the best performing strategy resulted in marginally lower performance than using only spectral data (-0.63% overall accuracy). The visual comparison of spectral and elevation CNNs for 2007 (Figure 2-6) show little difference between natural vegetation in highland areas and an increase in the commission error of agriculture in the mountains of Kandahar province (Figure 2-6 (c)).

Table 2-5. Evaluation of using elevation data (SRTM) and DMC imagery (NIR, R, G at 32 m) across Helmand and Kandahar provinces in April 2007 for agricultural delineation using transfer learning, targeted background sampling (training strategy 3) and image chip size 33 × 33 pixels

Training data	Model performance (50 epochs) (%)			
	Overall accuracy	*K*	mIoU	fwIoU
NIR, R, G	94.01	88.02	50.33	67.61
NIR, R, SRTM	93.38	86.78	49.43	65.47

2.4.1 Multi-seasonal CNN application

A multi-season model was trained starting with the best performing 2007 CNN (Table 2-5). The black dotted line in Figure 2-7 (a) shows the overall accuracy of the 2007 model on classifying the 2008 validation data (80.16%). The 2007 model trained faster over the first 10 epochs when updated with 75% (977 ha) of the available 2008 data than the other proportions (25%, 50%, and 100%) and achieved a similar overall accuracy to a single-season 2008 model (92.83% and 91.78% green line and red dotted line respectively in Figure 2-7 (a)). Adding data from 2008 increased the overall accuracy of the 2007 model by +12.67% to 92.83% and the Kappa coefficient, mean and weighted IoU also increases by 24.04%, 7.29 and 11.46, respectively (Table 2-6).

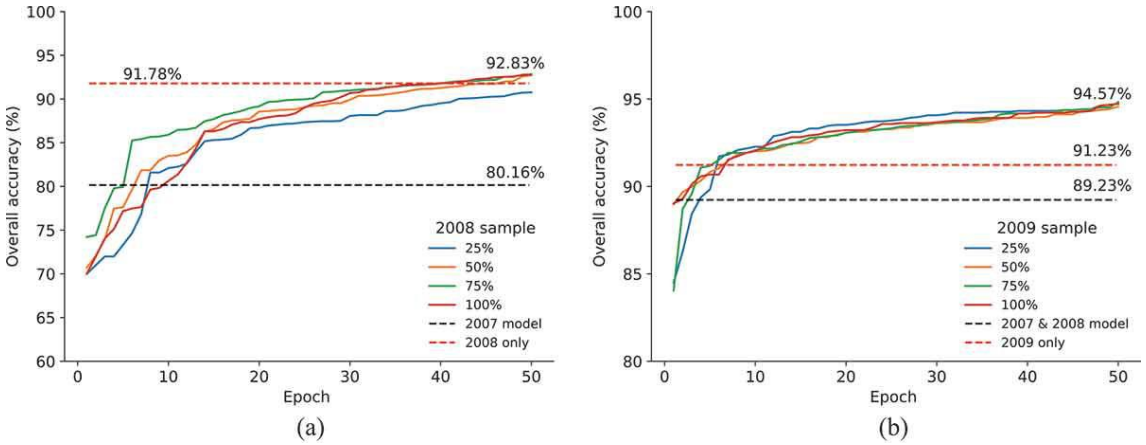

Figure 2-7. Evaluation of updating the best performing 2007 model using multiple training sample proportions (25%, 50%, 75% and 100% of total training data available) from 2008 and 2009 training data. (a) Transfer learning of 2008 training data using the 2007 model and (b) transfer learning of 2009 training data using the 2007 model updated with 75% of 2008 training data. Previous year's model with no additional training is shown by the black dashed line. Target year's model trained from the ImageNet dataset using 100% of available training data is shown by the red dashed line

The analysis was repeated for the 2009 agriculture mask classification (Figure 2-7 (b)). The CNN was trained using only 2009 data to provide a single-year model with and overall accuracy of 91.23% (the red dotted line in Figure 2-7 (b)). The previous years' combined model (trained on 2007 and 75% of 2008 data) with no training from 2009 achieved an overall accuracy of 89.23% (the black dotted line in Figure 2-7 (b)). This was an increase of +9.07% compared with using the 2007 model on 2008 imagery, showing a year-on-year improvement with additional data across seasons. The differences in the agricultural extent between image dates may affect year-on-year accuracy, but using the same validation locations aims to reduce this influence. Updating the previous years' model with 25% (317 ha) of available data trained faster than other sample proportions, improving the overall accuracy by 5.34% with the Kappa coefficient, mean and weighted IoU increasing by 10.68, 9.86 and 16.18, respectively (Figure 2-7 (b) and Table 2-6).

Table 2-6. Evaluation of transfer learning (TL) using a CNN trained on DMC imagery (NIR, R, G at 32 m) from 2007, 2008 and 2009 for Helmand and Kandahar provinces. *Retrain* uses the optimal percentage of available training samples with CNN TL from the previous year. The retrained 2008 model uses the 2007 model as a starting point and the retrained 2009 model uses the retrained 2007 model with 2008 data (75%)

Training year & model	Model performance (50 epochs) (%)			
	Overall accuracy	*K*	mIoU	fwIoU
2008				
2007 TL only	80.16	61.62	39.39	48.63
Retrain (75%)	92.83	85.66	46.68	60.09
2009				
2008 TL only	89.23	78.44	41.05	52.23
Retrain (25%)	94.57	89.12	50.91	68.41

Visual inspection of the multi-seasonal classifications in areas of known confusion between agriculture and natural vegetation show differences between years (see Figure 2-6, Figure 2-8 and Figure 2-9). In 2008 there are fewer agriculture commission errors than in 2009, despite the higher overall accuracy of the 2009 model. There are noticeable seasonal differences in natural vegetation between all three years. Larger areas of natural vegetation are found in 2007 (Figure 2-6 (a) and (c)) and 2009 (Figure 2-9 (a) and (c)), but little natural vegetation is found across the same area for 2008 (Figure 2-8 (a) and (c)).

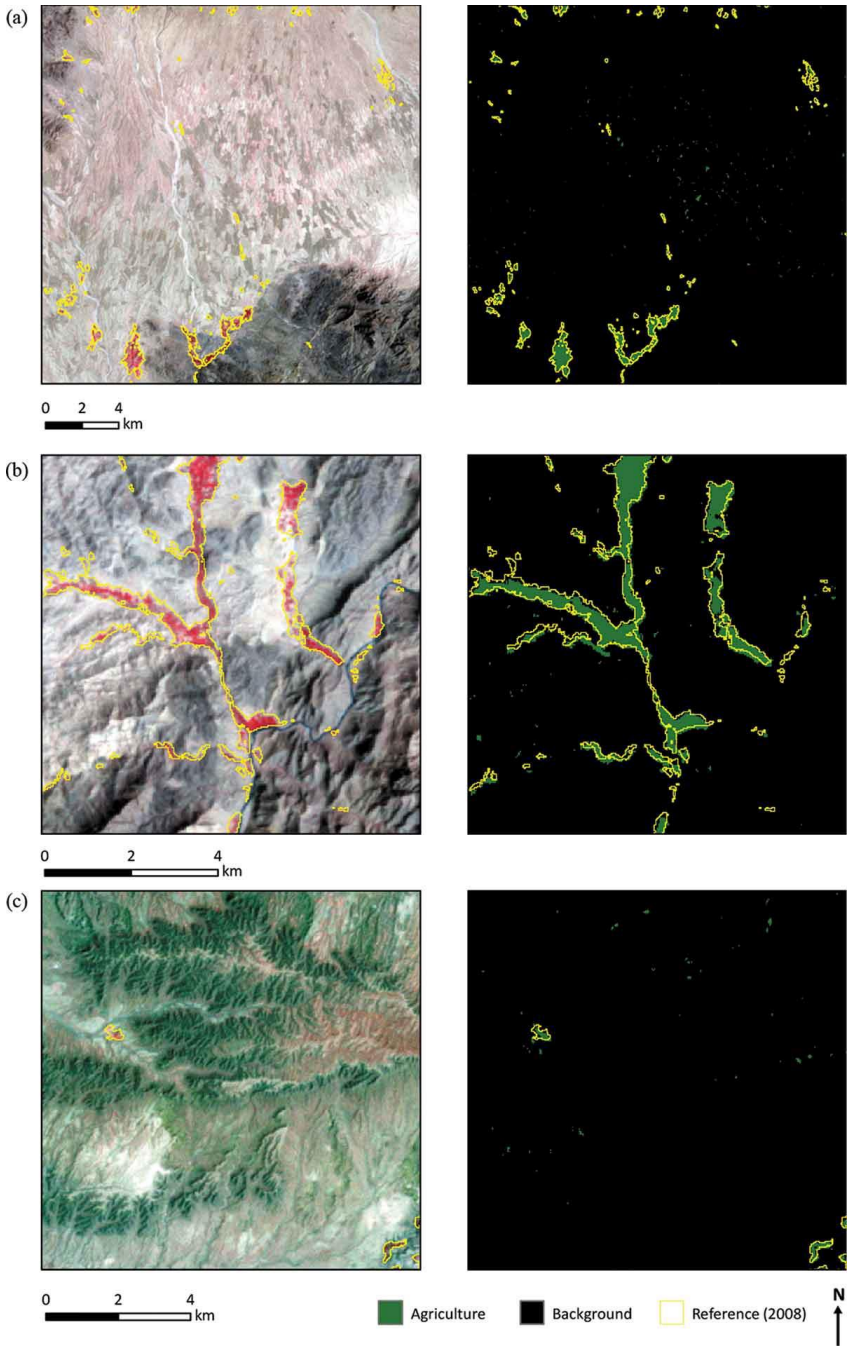

Figure 2-8. Visual interpretation of the 2008 CNN model with 75% of available training data using training strategy 3 (random agriculture, boundary cases and NDVI targeted non-agriculture classes with transfer learning) with input image chip size 33 × 33 pixels. Image extents have been selected based on prior knowledge of confusion areas for interpretation. (a) Large extent of natural vegetation, (b) well-delineated agriculture in highland areas and (c) commission of agriculture in highland areas in Kandahar, Afghanistan. False colour DMC imagery (32 m) for April 2008

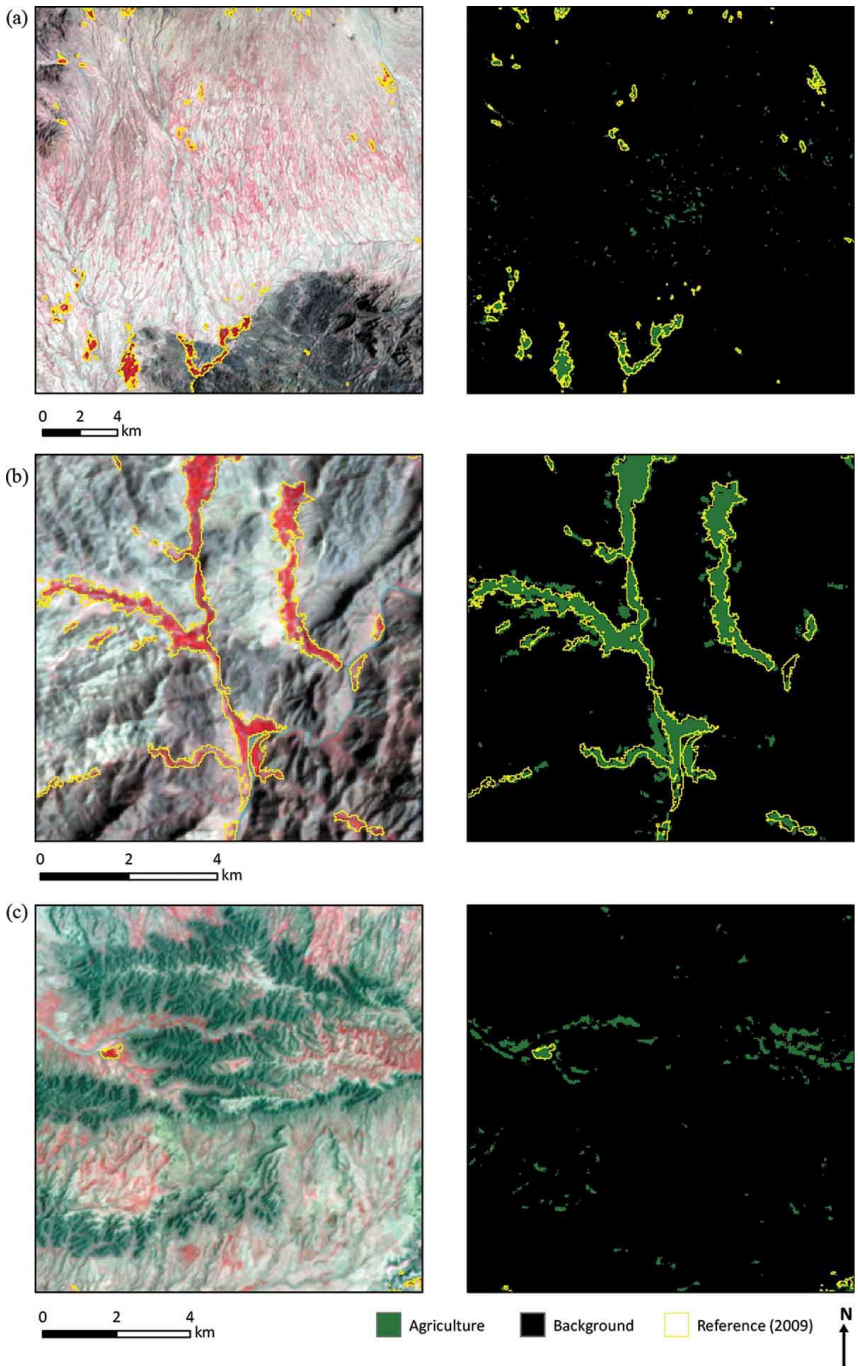

Figure 2-9. Visual interpretation of the 2009 CNN model with 25% of available training data using training strategy 3 (random agriculture, boundary cases and NDVI targeted non-agriculture classes with transfer learning) with input image chip size 33 × 33 pixels. Image extents have been selected based on prior knowledge of confusion areas for interpretation. (a) Large extent of natural vegetation, (b) well-delineated agriculture in highland areas and (c) commission of agriculture in highland areas in Kandahar, Afghanistan. False colour DMC imagery (32 m) for April 2009

2.5 Discussion

2.5.1 Importance of contextual information

In the operational opium survey, image interpreters manually refine the agricultural area defined by an unsupervised classification to adjust boundaries and remove areas of confusion, using contextual information to support their decisions (UNODC, 2018b). Contextual information includes field boundaries contrasting with desert and rock, buildings, and river channels. The CNN overcomes the limitations of pixel-based unsupervised classification by encoding the surrounding landscape features within each chip, with the scale of features dependent on the chip size. For example, in Figure 2-10 there are local field parcels and texture visible in the 33 × 33 chip (Figure 2-10 (a)), many more fields and the surrounding desert in the 65 × 65 chip (Figure 2-10 (b)) and a much greater proportion of desert and lower proportion of local information (relative to the centre pixel) within the 129 × 129 chip (Figure 2-10 (c)). Smaller chip sizes result in a set of more localised features in the CNN and larger chip sizes have greater generalisation, which is analogous to how a human interpreter will use local context (using a larger mapping scale) to refine a boundary.

In the CNN output each prediction is based on a single chip, with pixel-by-pixel classification achieved by sliding a chip-sized window across the input image, assigning the prediction to the centre pixel. A one pixel shift of a small chip results in a more substantial change to the surrounding contextual information than larger chips, explaining why smaller chips are more sensitive to localised change and result in increased classification accuracy (Kroupi *et al.*, 2019). Also, classification for the whole image is much more efficient with smaller chips as prediction for each sliding window is faster, despite the longer training times. A limitation of the fixed sampling grid is that a different number of samples are created for each chip size. An equal number of samples for each chip scale could be produced by introducing an overlapping sampling grid with the same centre pixel but would result in non-independent samples. Non-independent samples would introduce bias during the accuracy assessment process by validating on areas the classifier has already seen.

Making the image chip smaller might further improve the accuracy of the agricultural mask but there is likely to be a trade-off as the amount of contextual information decreases. The ResNet50 architecture limits the smallest chip size to 32 × 32 (33 × 33 was used here to accommodate a centre pixel), but to investigate smaller chips would require changing the network architecture. Alternatively, Fully Convolutional Networks (FCNs) could be used to extract agricultural features from arbitrary chip sizes as the model architecture allows for dynamic image chip scales for training and inference unlike the ResNet50 architecture which requires the same size. FCNs also provide pixel-by-pixel predictions and overcome the limitation of the centre pixel and improve the speed of image classification (Paisitkriangkrai *et al.*, 2015).

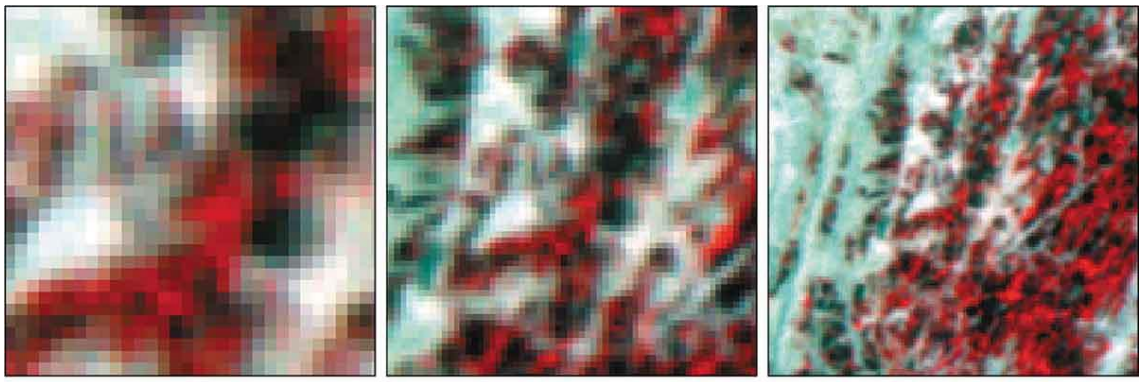

Figure 2-10. Examples of the sliding window sizes (a) 33 × 33, (b) 65 × 65 and (c) 129 × 129 pixels using the same centre pixel for agriculture. False colour DMC imagery (NIR, R, G at 32 m) for 27 April 2007

The influence of the centre pixel is an important consideration for pixel-by-pixel classification (Zhang and Lu, 2019) as it is used to label training and validation samples into agriculture and non-agriculture classes. As an experiment into pixel bias, the centre pixels for agriculture validation samples for each chip size were altered to the mean average of non-agriculture chips for each input channel (NIR, R, G). There was a negligible difference in overall pixel accuracy (-0.36%) for the smallest chip (33 × 33), and no differences were found with the other chip sizes. This shows the centre pixel plays no individual role in prediction and why pixel-level classification using image chips generalises as chip size increases.

Human interpreters have other sources of information at their disposal to help in the delineation of agricultural land. In our study area, using elevation data with the CNN model made little difference to classification accuracy (-0.63%). Highland and lowland areas are visually different, and it is likely that the spatial and spectral information in the chip is not improved further by adding explicit height information. In future work, other image sensors (e.g. hyperspectral (Dell'Acqua *et al.*, 2006) and Synthetic Aperture Radar (SAR) (Liu, He and Li, 2017), more ancillary data, and new deep learning architectures with an increased number of input channels could be investigated to improve agricultural land classification.

2.5.2 Year-on-year transfer learning for agricultural mask production

Transfer learning has similarities in the way a trained interpreter gains experience as they are both able to build upon existing knowledge. Whereas end-to-end learning is similar to an inexperienced interpreter with no prior knowledge of image classification. Transfer learning was consistently faster than end-to-end learning demonstrating some similarity between the underlying features across years. Transfer learning is unique to CNNs and other machine learning approaches, such as Random Forest, are unable to take advantage of faster training. The accuracy for CNN models trained by transfer learning were also generally higher than their end-to-end counterparts, which is consistent with previous studies of transfer learning for remote sensing data (UC Merced land use, RS19 and Brazilian coffee scenes) (Nogueira, Penatti and dos Santos, 2017).

Transfer learning increases the total number of samples used to train the CNN. However, sampling of the inter-annual changes between 2007 and 2009 was still required to refine the model. Even with very little (25%), or no training data from the target year, the model's performance increases. Fewer samples are required each year for training as multiple years' worth of different landscapes and examples of agricultural features adequately extract and predict common features. This alleviates the burden for complete labelled datasets for CNN classification, which remains a challenge in remote sensing. Transfer learning

from remote sensing data provides the opportunity to provide timely initial predictions without the need for additional labelled datasets. Updating the model for subsequent years may only require 25% or less of the total sample fraction.

Targeted sampling using boundary samples and an NDVI threshold (strategy 3) to identify samples containing natural vegetation was the optimal training strategy. By directing the training to areas of known confusion in the background class the CNN was better able to separate edge cases, which supports other studies reporting a decrease in CNN performance with random sampling compared to a selective strategy (Van Grinsven *et al.*, 2016). A human interpreter would also improve their delineations with more experience of difficult interpretation cases. Future CNN applications could include additional samples using post-classification refinement to further improve classification performance.

Using historical information to inform predictions of agricultural land has the potential to substantially decrease the manual effort associated with current agricultural mask production by the UNODC. Change detection techniques could be used by interpreters to focus manual editing on those areas identified as having changed. Misclassified areas could then be used as training samples to improve the model for the next year's agricultural mask. The outlined rationale could be used to develop a data-driven classification based solely on historical knowledge of agricultural land within Afghanistan. Utilising existing knowledge to derive upcoming agricultural masks without the need for additional data is an exciting prospect for timely image classification.

2.6 Conclusions

The overall accuracy for the ResNet50 CNN was >94% for agricultural land classification in all years (2007 to 2009). The best results were achieved using a chip size of 33 × 33 pixels and a NDVI-based sampling strategy, which targeted the main source of confusion between natural and agricultural vegetation. Transfer learning using a pre-trained model (ImageNet) was found to achieve higher overall accuracy than end-to-end learning (+2.2%). Substitution of the green spectral band for elevation data achieved marginally lower performance (-0.6%).

When considering transfer learning of the CNN model year-on-year, the classification of 2008 imagery using the 2007 ResNet50 model, with no additional training, resulted in an accuracy of 80.2%, improving to 89.3% for 2009 imagery using a combined 2007 and 2008 model. Using only 25% of the 2009 data to update the combined model further improved classification accuracy to 94.6%. High classification performance coupled with continual model refinement from additional data shows the potential for CNNs to replace human interpreters for the UNODC's agricultural mask production. Reducing the manual effort in the production of the mask to a small proportion of the total area would improve the speed and efficiency of the survey, reducing the overall cost. Deep transfer learning across multiple years presents an exciting opportunity for timely and efficient land cover classification.

Chapter 3. Understanding and developing generalised FCNs for agricultural land mapping in Afghanistan

This chapter presents a method for developing a generalised Fully Convolutional Network (FCN) for agricultural land classification across medium-resolution imagery datasets in Helmand, Afghanistan (thesis objectives 1 and 2). The influence of different image features (spatial, textural, and spectral) on an FCN's ability to learn representations of agricultural land is shown to understand how these networks learn. Standardisation of spectral image features is presented to transfer knowledge across image datasets between 2007 and 2017. A new method called localised intersection over union is presented for assessing localised error in land cover classification.

Underlying data for this chapter is available at:
https://doi.org/10.17862/cranfield.rd.14447400

3.1 Introduction

In Chapter 2, Convolutional Neural Networks (CNNs) were used to classify agricultural land and outperformed other pixel-based machine learning approaches (Random Forest). They demonstrated the potential for replacing manual approaches. The ResNet50 CNN used achieved high classification performance across three training years (>94% overall accuracy). The CNN outperformed other supervised classifiers as it is able to encode the surrounding contextual information into fixed sized image chips as inputs. However, classification of whole images using a sliding window is time-consuming and caused generalisation at difficult interpretation cases (e.g. boundaries between natural vegetation and agriculture). The CNN models with fully connected layers are not well suited for pixel-level classification because of the difficulties in allocating a pixel-level label from a chip-level prediction.

Fully Convolutional Networks (FCNs) are a type of CNN which are able to produce pixel-level predictions by upsampling outputs from consecutive convolutional layers back to the native resolution of the input image. High performance among image classification tasks suggests these models may be more appropriate for pixel-level land cover classification (Fu *et al.*, 2017; Simms, 2020). A limitation of CNNs and FCNs is the requirement for large amounts of labelled data to train accurate models (Vali, Comai and Matteucci, 2020) with up to 20% improvement using larger datasets with ResNet (Brigato and Iocchi, 2021). In Chapter 2, agricultural land classification improved with more training data and with transfer learning from multiple years. This amount of training data is uncommon, or only available as legacy data from historical studies. Building general models for these data are advantageous for training CNN and FCN models, especially as access to these data are limited.

Developing general models for agricultural land classification that can be trained and used across a range of image sensors is complicated because of the differences in their radiometric, spectral, spatial, and temporal resolutions. Imagery from different sensors generally require different analysis and models,

especially for supervised classification (Talukdar *et al.*, 2020). Standardised imagery datasets would provide consistent data for training and classification of agricultural land between sensors and through time. In Chapter 2, *Z*-score normalisation was previously used to successfully standardise imagery from the same sensor (DMC) between 2007 and 2009, which transforms input data into a gaussian distribution for classification. This approach, from the field of computer vision, may not be appropriate for use of image datasets from different sources because of their radiometric, temporal and spectral differences.

Understanding the image features that are important for CNNs to identify agriculture would help in the development of a general model. This would identify which features require standardisation to train general models across image datasets. However, the underlying patterns in data extracted by CNNs are often difficult for humans to interpret. These types of techniques are commonly referred to as black-box approaches. Previously, visualising heat maps has been used (Park *et al.*, 2020), but intermediary layers are often uninterpretable due to their high number of output channels. Separating the spatial information from input data is difficult for binary image processing tasks, as multiple classes cannot be combined to create a background class (e.g. the agricultural mask). A binary mask was previously used with outlines of multi-class target objects of commercial photography at the pixel-level (Shelhamer, Long and Darrell, 2016). The best results were found by only masking the foreground whilst maintaining the background spectral values to suggest the network can learn spatial patterns. Another approach isolates the global shape of their commercial photography data using the outline of the target object and filling it with the spectral values of a different photograph (Baker *et al.*, 2018). The study found the texture was more important than the global shape for chip-level CNN classification as a CNN is able to encode the local shape features from texture unlike the global shape. Further investigation of these image features could help optimise CNN performance.

Assessing the performance of CNNs is essential for establishing whether the model is optimally trained to classify agricultural land. Global accuracy metrics, such as overall accuracy (Foody, 2009), are commonly used to quantify how well

a model is performing. In Chapter 2, there were only marginal improvements in overall accuracy whilst the mapped agricultural area in difficult interpretation areas (e.g. boundary locations) had improved delineation. Global metrics do not provide enough information on the distribution of error during classification. Intersection over Union (IoU) (Russakovsky *et al.*, 2015) was found to be a more sensitive global accuracy metric for assessing global performance of CNN models in Chapter 2. A method for mapping the distribution of local error across different agricultural areas would be useful for identifying strengths and weaknesses of the general models.

The aim of this chapter is to develop a generalised model for classifying agricultural areas from features across different datasets. The research questions are: (1) What are the respective spatial, textural, temporal, and spectral features used by the CNN for agricultural land prediction?, (2) Can these features be standardised across image sensors for transfer learning from one imagery source to another? and (3) Can the local differences in agricultural land classification performance be assessed?

This chapter begins by evaluating an FCN-8 and ResNet50 architecture for agricultural land classification using the same training and validation datasets to select the best-performing CNN architecture. A new method for evaluating local classification performance is presented as localised IoU to map the spatial distribution of IoU. The subsequent sections conduct a series of experiments to identify the influences of different image features on CNN training. Spectral features are standardised for training and evaluating a generalised model for agricultural land classification across image datasets.

3.2 Fully convolutional network

The FCN-8 architecture (Long, Shelhamer and Darrell, 2015) is used in this chapter because of its high performance on pixel-based image processing tasks. The architecture is modelled on VGG-16 (Simonyan and Zisserman, 2015) and built up of convolutional layers to extract distinctive features from progressive layers based on sets of filter weights of shape $m \times n \times c$ (Figure 3-1). These filters act as kernel operations for individual images, where m and n are the vertical and

horizontal dimensions of the image and *c* is the number of input features or spectral bands. Kernels operate on local spatial information at the defined stride length, also known as the receptive field, to perform convolution and pooling operations. Pooling uses downsampling to learn more complex spatial patterns by reducing the number of parameters. Upsampling of these layers rebuilds a dense spatial prediction. The FCN is able to produce a pixel-by-pixel output of classes (*k*) by replacing the fully connected layer of the VGG-16 architecture with a bilinear upsampling convolutional layer, also known as a deconvolutional layer. Lower layers are summed with the upsampled layers using skip connections to encode finer spatial structures and improve pixel-level predictions.

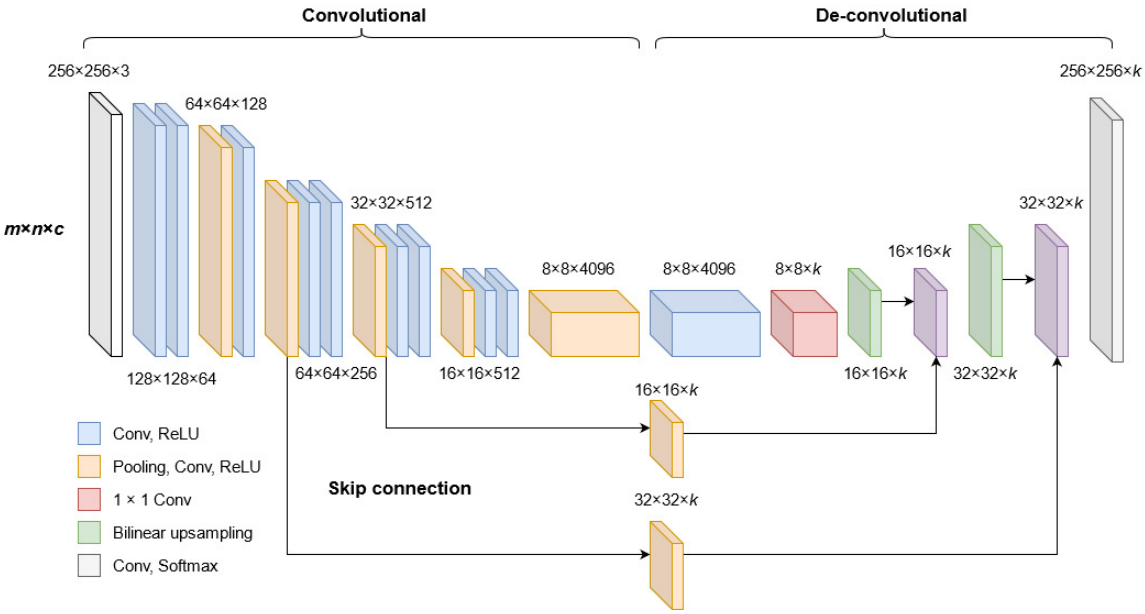

Figure 3-1 The FCN-8 model architecture used for agricultural land classification using 256 × 256 pixel image chips. *m* × *n* are the vertical and horizontal dimensions, *c* is the number of input features (three for DMC imagery (NIR, R, G)) and *k* is the number of classes (Adapted from Piramanayagam *et al.* (2018))

The upsampling layers in the FCN-8 architecture overcomes the constraint for reconstruing the agricultural mask in Chapter 2 by creating a dense labelled prediction at the native resolution of the input image, rather than a single prediction for each image chip. The FCN-8 model configuration implemented in Simms (2020) is used in this chapter and was trained end-to-end using 0.0001 learning rate with an Adam optimiser (Kingma and Ba, 2015) across 50 epochs.

All experiments were conducted on a NVIDIA Quadro K2200 GPU and the TensorFlow library (Abadi *et al.*, 2015).

The FCN-8 and ResNet50 architectures were compared for agricultural land classification to evaluate their performance using overall accuracy, kappa statistic (*K*) and frequency-weighted intersection over union (fwIoU) (§2.2). The same training dataset from Level-1A DMC data (NIR, R, G at 32 m) between 2007 and 2009 across Helmand and Kandahar Province, Afghanistan were used to train both models. The same validation dataset from Level-1A DMC data from 2009 was used to validate both models. The targeted training strategy outlined in Chapter 2 using 33 × 33 pixel chips was implemented for sample selection. The FCN-8 outperformed the ResNet50 for agricultural land classification (+3.34% overall accuracy) with higher frequency weighted IoU (+13.36%) (Table 3-1). The boundaries between agriculture and non-agriculture were the main areas of difference between the ResNet50 and FCN-8 classification, where the FCN had less generalisation (Figure 3-2). The smaller areas of agriculture were also better delineated in the FCN-8 classification and smaller variations within larger agricultural fields have improved to reduce agriculture over-classification. The FCN-8 is used in all subsequent experiments based on its superior classification of agricultural land.

Table 3-1. Comparison between using the ResNet50 and FCN-8 architectures for classification of agricultural land using a DMC-trained model on 2007, 2008 and 2009 data and validated on DMC imagery from 2009 (NIR, R, G at 32m)

Model training data (*n* = 11,624)	Validation metrics (%) (*n* = 1,936)		
	Overall accuracy	*K*	fwIoU
ResNet50	91.31	82.68	67.61
FCN-8	94.65	88.51	80.97

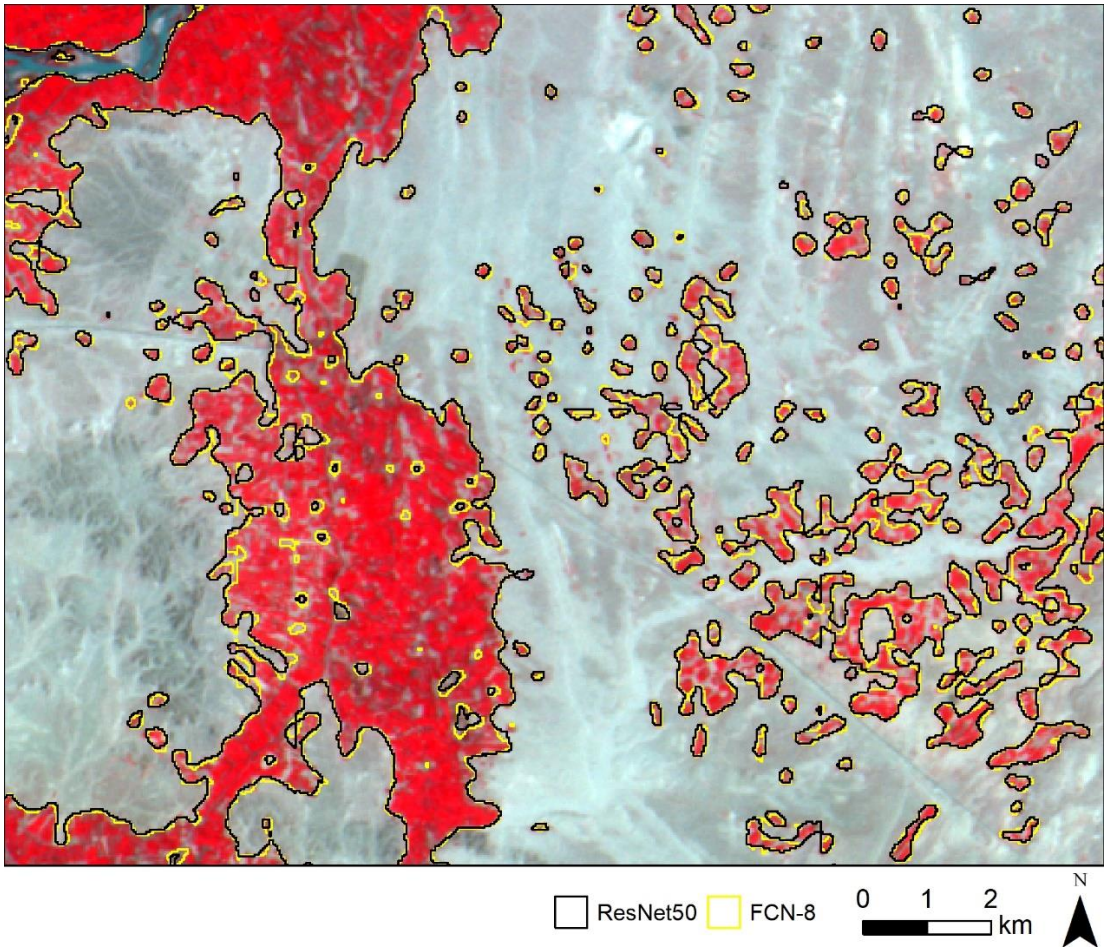

Figure 3-2. Visual evaluation of agricultural delineation in Helmand between CNN and FCN classification in 2009 using models trained on the same image data from 2007 and 2009. Image (31.4643° N, 64.4414° E): False colour (NIR, R, G) DMC imagery from 2009

3.3 Materials and methods

3.3.1 Image data and agricultural masks

The study area was Helmand and Kandahar Province, Afghanistan which accounts for 61% of national opium cultivation with an estimated 160,208 ha grown in 2018 (UNODC, 2018b) (Figure 3-3). Cloud-free scenes from satellite imagery were selected to target timing for peak opium poppy biomass to coincide with the maximum active agricultural area (Table 3-2) and identified from MODIS Normalised Difference Vegetation Index (NDVI) temporal profiles (Simms *et al.*, 2014). The North-South cutlines were also used to select cloud-free imagery for

active agricultural mask production based on known differences in peak opium biomass (approximately 1-2 weeks) (Taylor *et al.*, 2010). Four different image sensors are used to evaluate the interoperability of FCN models for agricultural land classification (DMC, Landsat-5, Landsat-8, and Sentinel-2).

Table 3-2 Image specifications of near-infrared (NIR), red (R) and green (G) spectral bands for DMC, Landsat-5, Landsat-8 and Sentinel-2 imagery. North (N) or South (S) imagery are defined by the cutline in Figure 3-3. Some footprints overlap in North and South areas and the same cloud-free images were used for the analysis.

Imagery	Year	Peak image dates	Spatial resolution (m)	Central wavelength (µm)	Revisit time
DMC	2007	27 Apr (N & S)	32	0.83 (NIR) 0.66 (R) 0.57 (G)	Up to daily
	2008	24 Mar (S) 7 Apr (S) 24 Apr (N)			
	2009	25 Mar (S) 3 Apr (N) 8 Apr (S)			
Landsat-5	2009	28 Mar (N & S) 30 Mar (N) 5 Apr (N & S)	30	0.83 (NIR) 0.66 (R) 0.56 (G)	16 days
Landsat-8	2015	7 Apr (N & S) 16 Apr (N & S)	30	0.87 (NIR) 0.66 (R) 0.56 (G)	16 days
	2016	24 Mar (N & S) 18 Apr (N & S)			
	2017	27 Mar (N & S) 5 Apr (N & S)			
Sentinel-2	2017	8 Apr (N & S) 15 Apr (N & S)	10 (resampled to 30 m for comparison)	0.84 (NIR) 0.67 (R) 0.56 (G)	10 days

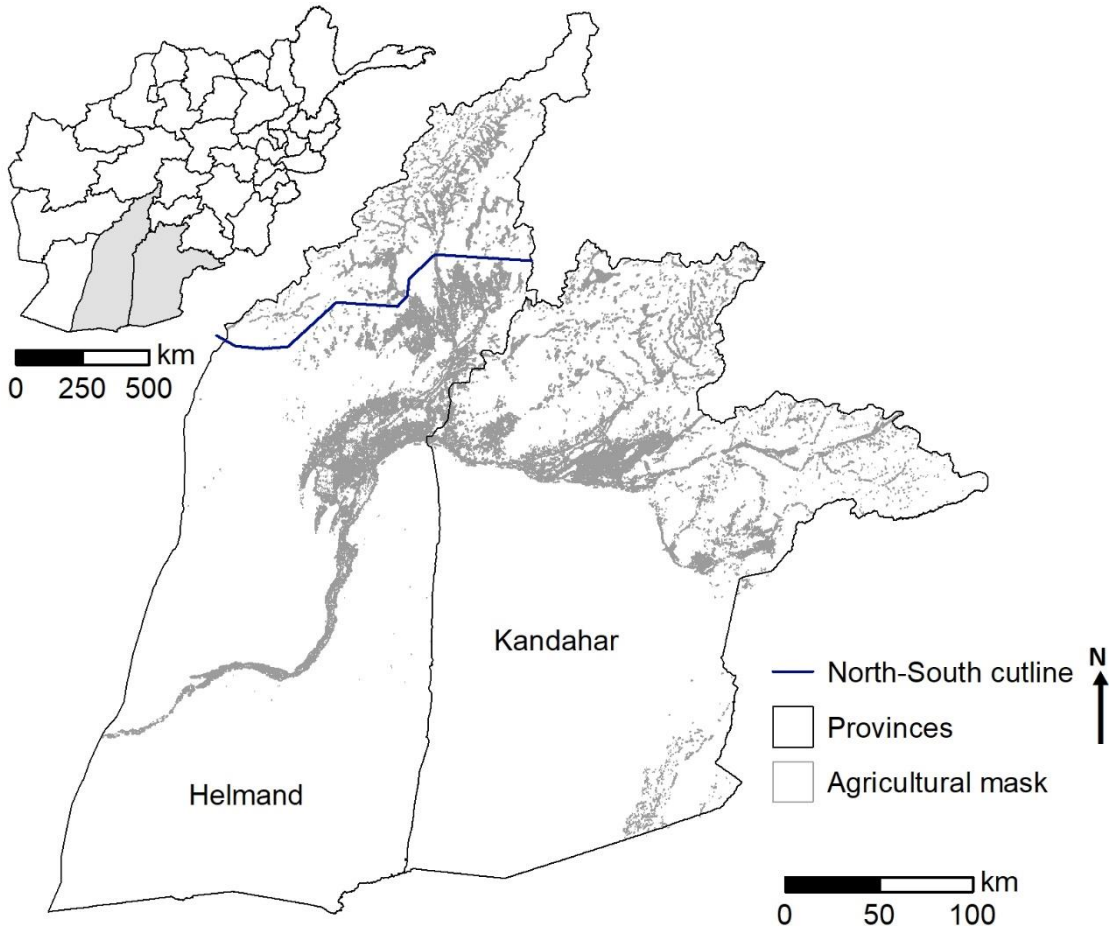

Figure 3-3. Helmand (centre 31.3636° N, 63.9586° E) and Kandahar (centre 31.6289° N, 65.7372° E) Province, Afghanistan showing the active agricultural area from 2009 DMC imagery with the image cutline for selecting optimally timed imagery

Annual active agricultural masks (§1.4) from Helmand and Kandahar were used for training and validating FCN models between 2007 and 2009, for further information on their production see §2.3.2. Sample locations for 2015, 2016, and 2017 were selected from previous UNODC surveys in Helmand Province and manually interpreted using medium resolution imagery coinciding with peak vegetation to be consistent with the 2007-2009 datasets.

3.3.2 FCN sampling strategy for agricultural land

The targeted training strategy outlined in Chapter 2 was used to train the general models. Input imagery from each year was split into a non-overlapping 256 x 256 pixel grid to create input image chips to be fed to the model. This pixel grid size was selected as it was the largest image size that could be stored in the system's graphic card memory during training (NVIDIA Quadro K2200 4GB), but larger image chips could be used. The samples were subsequently stratified into three groups using the reference agricultural mask for each target training year: (1) chips containing agriculture, (2) chips containing only non-agriculture (the background class) and (3) chips containing only non-agriculture with high NDVI, as defined by the Otsu threshold (Otsu, 1979), to introduce samples with natural vegetation. In Chapter 2, optimised CNN models were found to distinguish the difference between agriculture and non-agriculture using only samples with agriculture present because of the abundance of the background non-agriculture class across Helmand. Samples containing agriculture were first ordered based on their respective proportion of agriculture. Then training and validation datasets were created by using a 75%/25% split of these samples and selecting the first three samples for training and the fourth sample for validation to ensure representative sets of labelled data.

3.3.3 Experiment 1: Image features for agricultural land classification

FCNs can encode localised information into their convolutional filters to provide high image classification performance. Individual satellite imagery scenes have three main features for these techniques to learn from: (1) spatial, (2) textural and (3) spectral. These individual features were extracted from level-1A DMC data in 2009 across Helmand and Kandahar Province to understand how important each feature is for training an FCN to delineate agricultural land. Time-series of satellite imagery would also provide the temporal differences, which have been successful in land cover classification (Rußwurm and Körner, 2018), but this experiment focuses on extracting features from individual scenes using pre-trained FCN models which do not allow for temporal signatures. A total of 415

training samples and 137 validation samples from 2009 imagery (Figure 3-4) were used for this experiment.

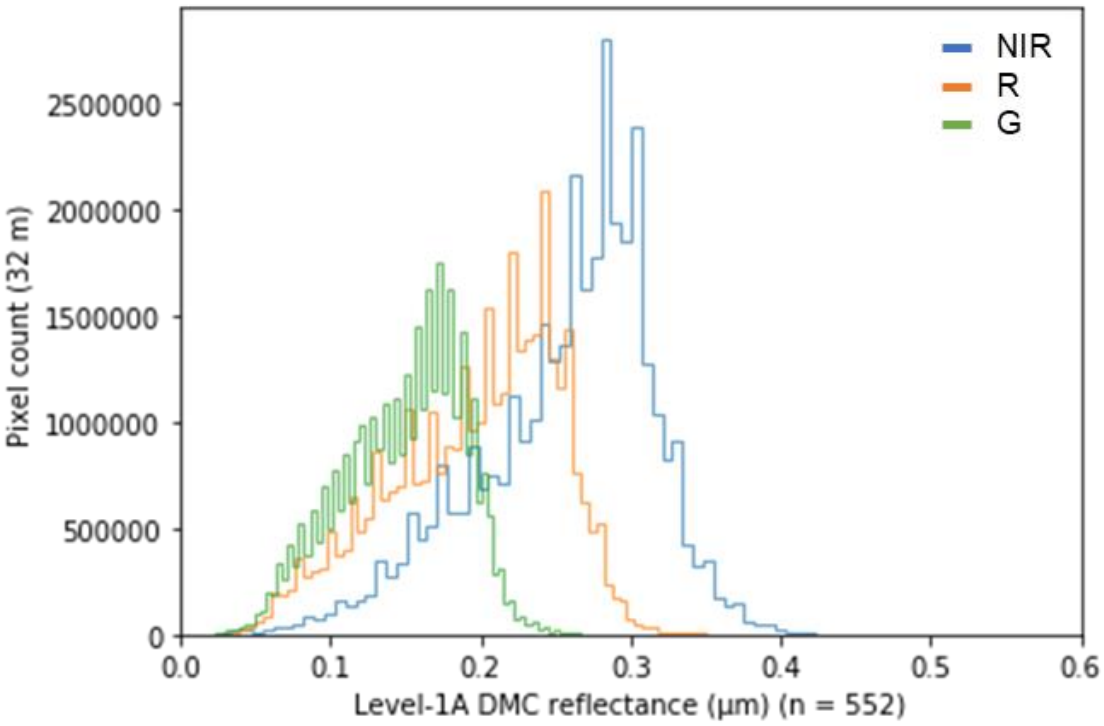

Figure 3-4 Histograms for the NIR (0.76 - 0.90 µm), R (0.63 – 0.69 µm) and G (0.52 – 0.62 µm) spectral bands of the 2009 training and validation image chips (*n* = 552)

The approach used to separate the spatial component for the agricultural mask uses a method where the spectral values are filled with the opposite class (Baker *et al.*, 2018). The assumption is by replacing the spectral values with the opposite class the FCN is relying solely on the encoded spatial information to learn features of agricultural land. In this experiment, the training data were ordered based on the proportion of agriculture. Then, every other sample had their agricultural area filled with spectral values from samples with 100% non-agriculture spectral values. The sample's non-agricultural area was filled with spectral values from samples with 100% agriculture proportion spectral values (Figure 3-5). A combination of synthetic and original Level-1A data were used to train the FCN-8 model and validated on the original Level-1A validation samples.

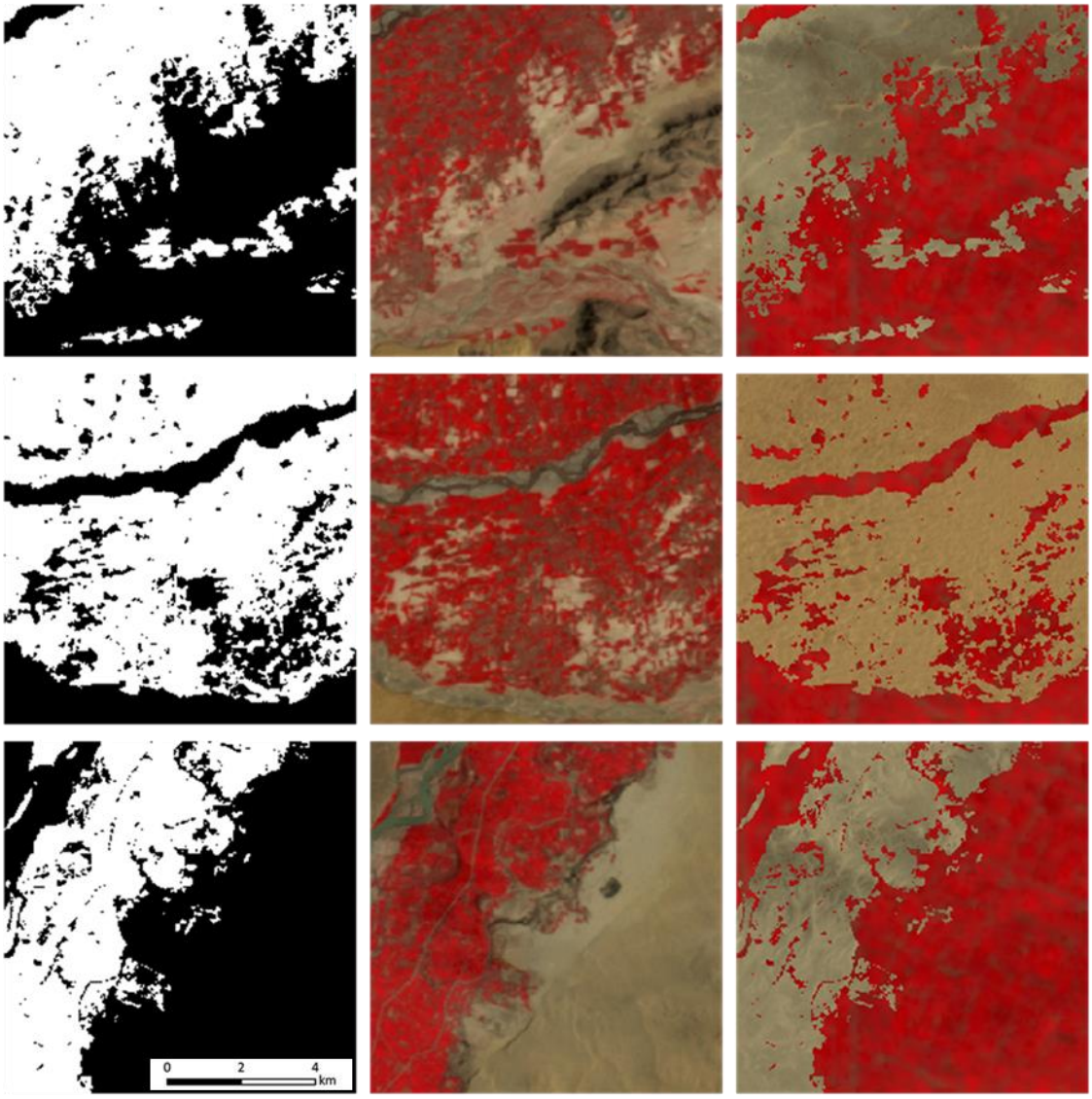

Figure 3-5 Example 256 x 256 training chips after spatial pre-processing (i) labels, (ii) Level-1A image, and (iii) synthetic image, both (ii) and (iii) are used as training samples. Agriculture is denoted in white and non-agriculture in black in the labels image.

Textural components were extracted by applying a grey-level occurrence matrix (GLCM) on DMC imagery from 2009. The grey-level co-occurrence matrix (GLCM) extracts textural features based on the spatial and spectral relationship of each pixel (Haralick, Shanmugan and Dinstein, 1973). The three textural metrics with the greatest variance were used as inputs to the FCN model to investigate texture, which are homogeneity (Equation 3-1), entropy (Equation

3-2), and correlation (Equation 3-3) (Figure 3-6), where $P_{i,j}$ is the probability of column (i) and row (j) labels occurring in adjacent pixels using the fixed kernel window of the GLCM, μ is the mean and σ is the standard deviation (Hall-Beyer, 2017).

$$\text{Homogeneity} = \sum_{i,j=0}^{N-1} \left(\frac{P_{i,j}}{1 + (i - j)^2} \right) \quad \text{3-1}$$

$$\text{Entropy} = \sum_{i,j=0}^{N-1} P_{i,j} \left(-\ln P_{i,j} \right) \quad \text{3-2}$$

$$\text{Correlation} = \sum_{i,j=0}^{N-1} P_{i,j} \left[\frac{(i - \mu_i)(j - \mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right] \quad \text{3-3}$$

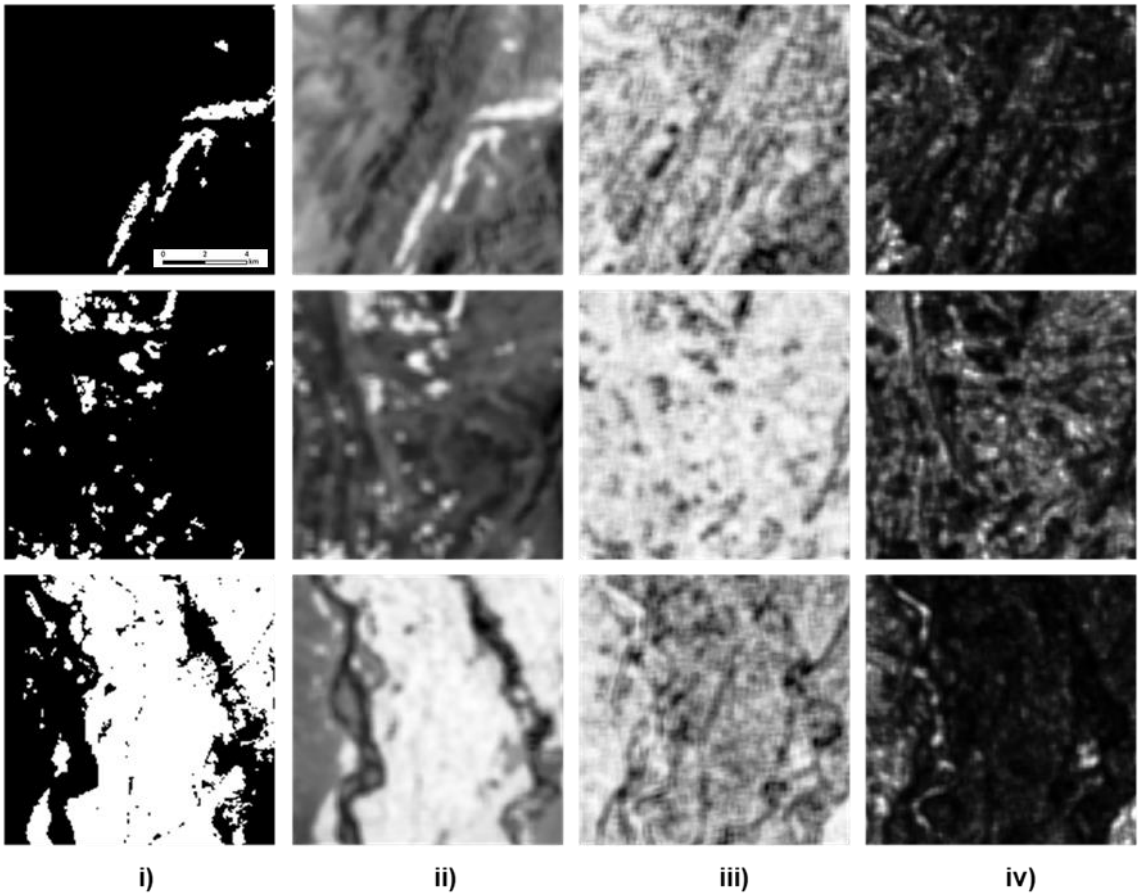

Figure 3-6 Example 256 x 256 training chips after textural pre-processing. i) labels, ii) homogeneity, iii) entropy, and iv) correlation. Agriculture is denoted in white and non-agriculture in black in the labels image.

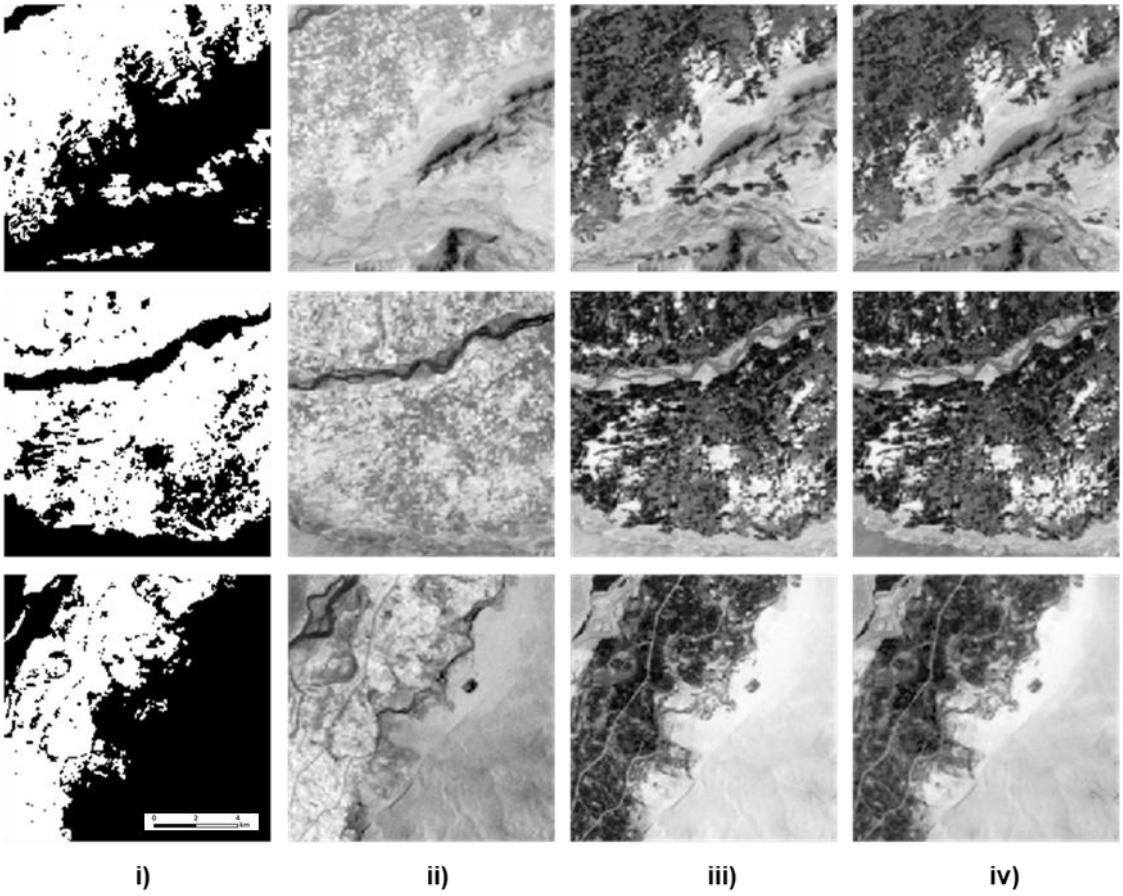

Figure 3-7 Example 256 x 256 training chips after spectral pre-processing. i) labels, ii) near-infrared, iii) red, and iv) green. Agriculture is denoted in white and non-agriculture in black in the labels image.

The spectral components were separated using the individual spectral bands from the DMC imagery (NIR, R, G) (Figure 3-7). Texture is encoded within these components based on the land cover therefore, it is not possible to exclude the influence from the texture in these spectral components. When referring to the spectral component for FCN model training in this study it also encompasses the textural component.

3.3.4 Experiment 2: Standardisation of image data

Standardisation is required for image data to reduce the influence of radiometric differences caused by the atmosphere, illumination, or sensor specifications to ensure differences between images are actual changes on the Earth's surface. This experiment aims to understand how standardisation of spectral data affects

agricultural land delineation between image sensors using FCNs. A total of 820 training samples were used from DMC data across two training years (2007 and 2008) and the model was validated using 137 validation samples from DMC data in 2009. Landsat-5 imagery (NIR,R,G at 30m) was used to validate the between sensor performance of the trained FCN by selecting overlapping cloud-free imagery at similar dates to DMC imagery (Table 3-2). Landsat-5 underwent nearest neighbour resampling to match the spatial grid of the validation data (32 m) to allow validation using the labels derived from the DMC data. The same DMC validation sample locations from 2009 data were selected to compare the performance of the classifier between the two satellite sensors.

Four different types of standardisation are used to assess their interoperability with the spectral data between satellite sensors for FCN training: (1) Top of Atmosphere (TOA) reflectance calibration (Level-1A), (2) Iteratively Reweighted Multivariate Alteration Detection (IR-MAD) of Level-1A images, (3) pixel intensity and (4) Normalised Difference Vegetation Index (NDVI). Level-1A imagery, also known as TOA reflectance, is a common radiometric calibrated product which removes the radiation scattered and emitted by the atmosphere to estimate surface reflectance. In order to convert to TOA reflectance, Equation 3-4 is used where, L_λ is at sensor radiance, d^d is Earth-Sun distance in astronomical units, $ESUN_\lambda$ is the solar exo-atmospheric irradiance and θ is the solar zenith angle.

$$\rho = \frac{\pi \, L_\lambda \, d^d}{ESUN_\lambda \, \mathrm{Cos}(\theta)} \qquad 3\text{-}4$$

IR-MAD is a radiometric normalisation technique for bi-temporal analysis (Canty and Nielsen, 2008), which maximises the differences related to land cover changes. IR-MAD uses invariant features between images to match the radiometry using orthogonal regression. Invariant features from the scalar difference image (Equation 3-5) are determined using canonical correlation analysis to subtract the canonical components from each pixel position between the target image (T) and reference image (R) across each spectral band (N) (Equation 3-6). The probability of change is calculated using the chi-square

distribution for each pixel from the sum of the squares of the MAD variates (M_i) (Equation 3-7). The observations are iteratively reweighted using the probability of no-change until the threshold of change falls below the defined threshold (>0.95).

$$U = a^T R \ - \ b^T T \tag{3-5}$$

$$M_i = U_{N-i+1} - V_{N-i+1}, i = 1, \dots, N \tag{3-6}$$

$$Z = \sum_{i-1}^{N} \left(\frac{M_i}{\sigma_{M_i}} \right) \tag{3-7}$$

The experiment used Level-1A DMC and Landsat-5 imagery from 2007 to 2009 across Helmand and Kandahar Province to run the IR-MAD analysis from code available from Canty (2014). The reference image used during IR-MAD radiometric calibration is a single cloud free DMC image across Helmand and Kandahar Province acquired on 27 April 2007, as this is the oldest image in the analysis. All target imagery (DMC, Landsat, and Sentinel-2) can be radiometrically matched to this single image across the same image extent (Figure 3-8).

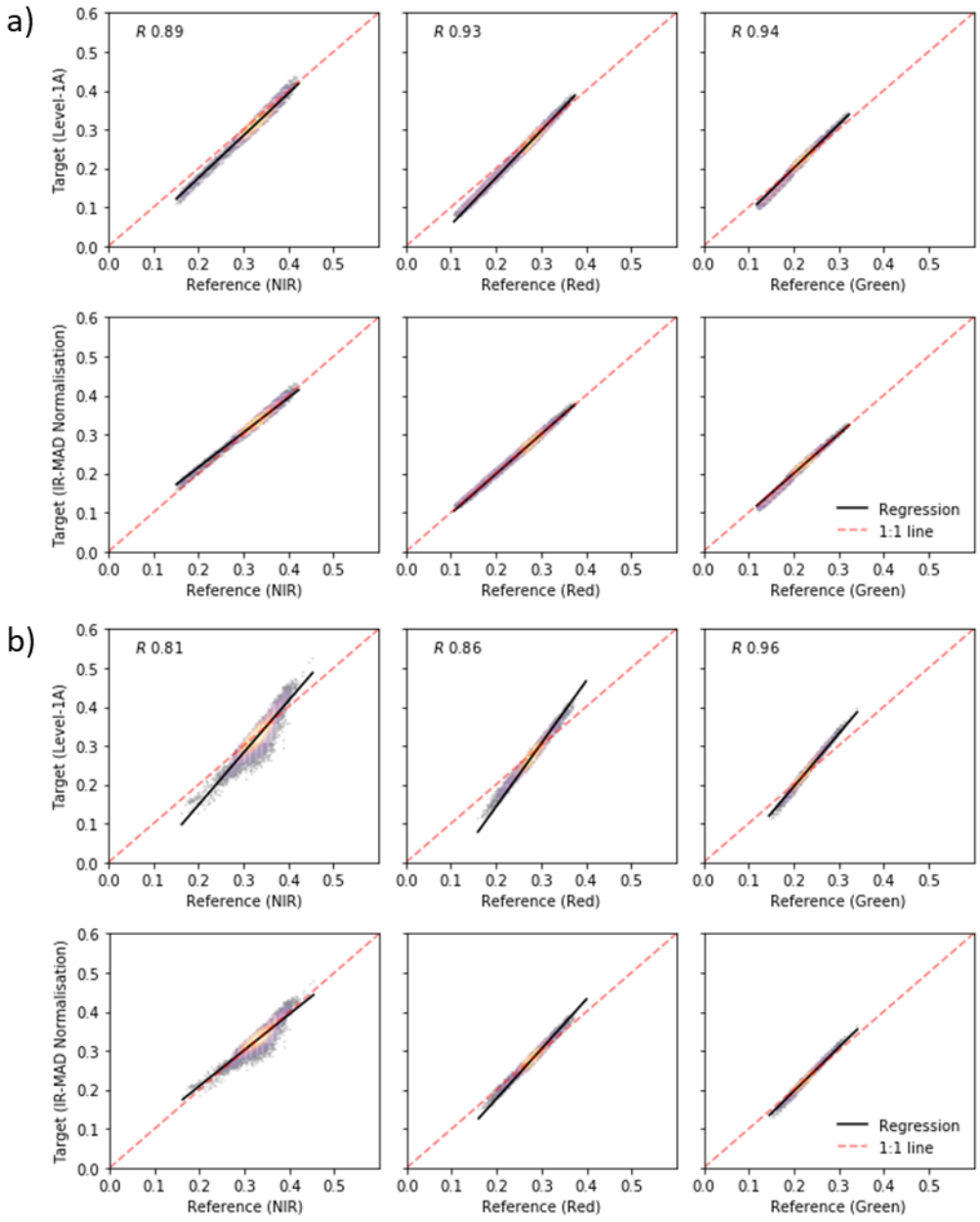

Figure 3-8 Orthogonal regressions for spectral bands NIR, red and green on Level-1A a) DMC imagery on 25 March 2009 and b) Landsat-5 imagery on 5 April 2009 before and after IR-MAD normalisation. Only unchanged pixels are plotted.

Pixel intensity is used to assess whether a single grey-level band derived from the NIR, R and G bands can separate agriculture and derived by calculating the weighted sum of the spectral bands. The NDVI (Equation 3-8) is a commonly used vegetation index to separate agriculture using the near-infrared (NIR) and

red (R) spectral bands and is used in the operational methodology by the UNODC to perform their unsupervised agricultural mask classifications.

$$\frac{(NIR - R)}{(NIR + R)} \qquad 3\text{-}8$$

3.3.5 Generalised FCN model training with few labels

The workflow presented in Figure 3-9 applies the proposed methodology for updating the FCN with only new areas of agriculture in Chapter 2. These new agricultural areas are areas of expansion between the target year and the previous year's agricultural mask. Updating the model with a smaller subset of labelled data creates a sparse dataset, therefore the FCN-8 model needs to be altered to reflect this change. The FCN-8 network masks existing areas of agriculture between the two target years by providing a weighting layer during each epoch to only take into account predictions of new agricultural areas to fine-tune the model, in a similar approach used to ensure class balance (Long, Shelhamer and Darrell, 2015).

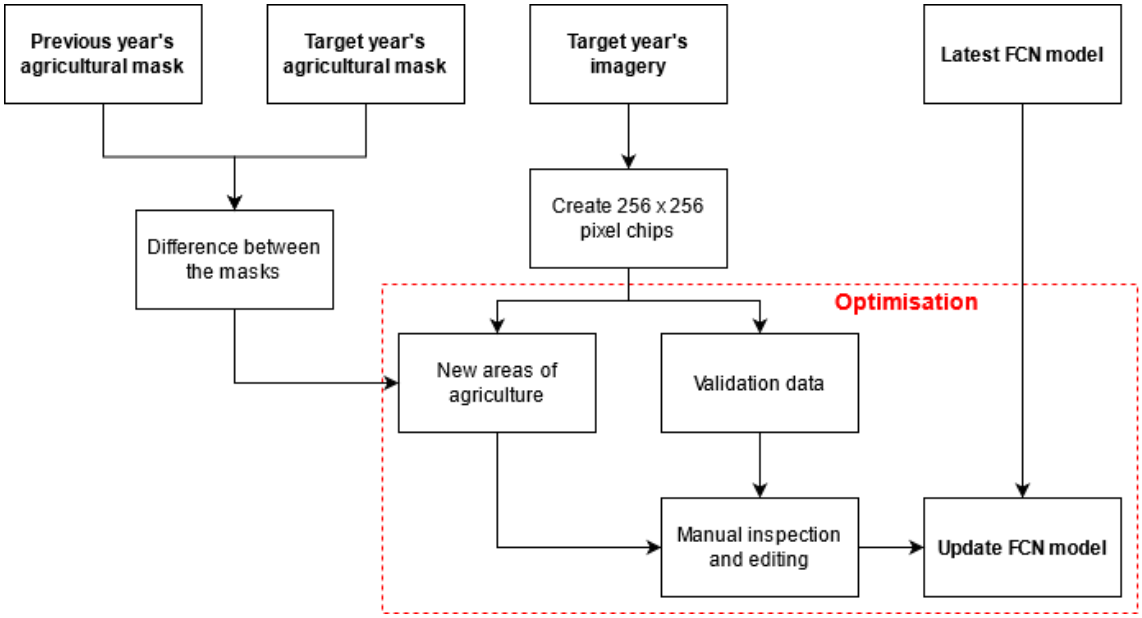

Figure 3-9. Workflow for active agricultural mask production using sparse data by fine-tuning an existing FCN model with only new areas of agriculture between agricultural masks

The best-performing FCN model trained with data from 2007 to 2009 for between-sensor performance used IR-MAD normalisation and is used to fine-tune with new data between 2015 and 2017. All imagery were standardised using IR-MAD radiometric normalisation of Level-1A satellite imagery (Figure 3-10). The reference image used for all normalisation is the same DMC image from 27 April 2007 across Helmand Province. Standardised imagery (NIR, R, G) and active masks are split into a non-overlapping grid of 256 x 256 pixels, where these samples are subsequently stratified using the same approach outlined in §3.3.2. The 2015 dataset had a total of 273 available training samples and 91 validation samples. From Chapter 2, it was found not all training data was required to fine-tune the model between years, therefore only 25% of the training dataset is used to fine-tune 2015 (69 samples).

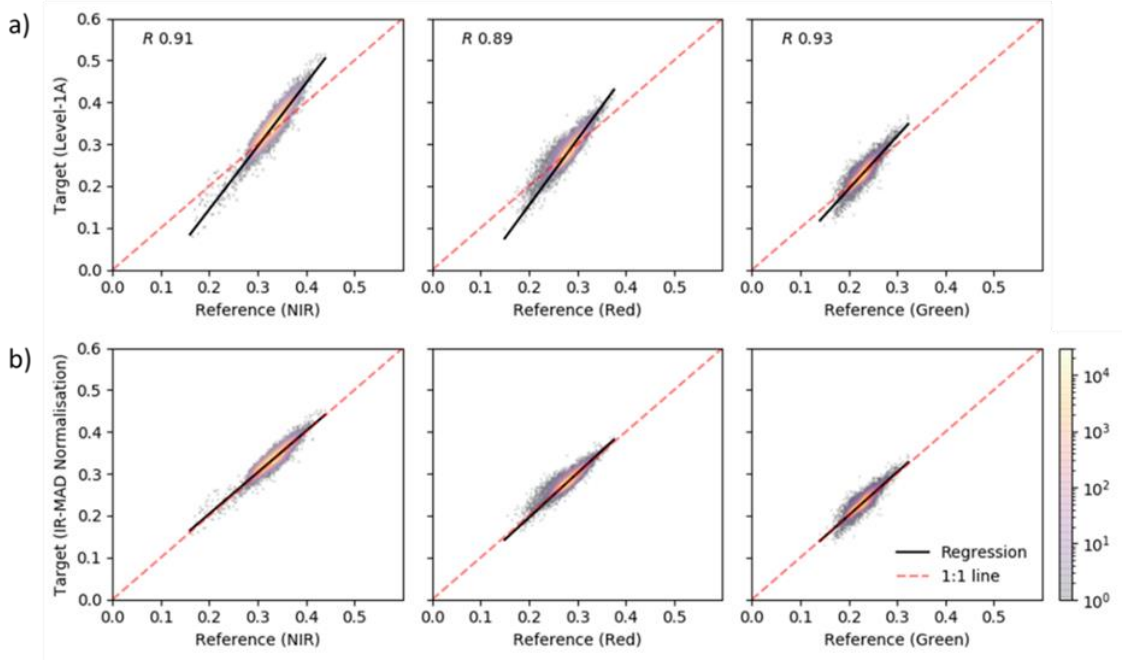

Figure 3-10. Orthogonal regressions for spectral bands NIR, red and green on Level-1A calibrated Landsat-8 imagery on 18 April 2016 a) before IR-MAD normalisation and b) after IR-MAD normalisation. Only unchanged pixels are plotted

The training sampling approach used for updating the model only includes samples where new areas of agriculture are above 1% of its proportion. The 2016 dataset has 54 samples for updating the 2015 model and the 2017 dataset has 34 samples for updating the 2016 model. The same validation sample locations are used for 2015, 2016 and 2017 to assess the performance of the generalised model. Uncertainty of the generalised model is expressed using the 95% prediction intervals using the total reference and prediction areas of the 2009, 2015, 2016 and 2017 validation samples (Equation 3-9) where, x_k is the predicted value, $\hat{y}_h$ is the response value, $t_{\left(\frac{\alpha}{2},\ n-2\right)}$ is the critical t-value (1.96) with multiplier and $\sqrt{MSE \times \left(1 + \frac{1}{n} + \frac{(x_k - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}\right)}$ is the standard error of prediction.

$$\hat{y}_h \pm t_{\left(\frac{\alpha}{2},\ n-2\right)} \times \sqrt{MSE \times \left(1 + \frac{1}{n} + \frac{(x_k - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}\right)} \qquad 3\text{-}9$$

Accurate agricultural area estimates are essential for the annual opium cultivation estimates as it is used in the ratio-based sampling approach (Equation 1-1). The use of different imagery resolutions is an important consideration for comparison between area estimates as the influence of generalisation in agricultural mask production is detrimental to cultivation estimates (Simms, Waine and Taylor, 2017). The influence of spatial resolution on FCN-8 performance are also investigated using localised IoU with Sentinel-2 imagery from 2017 at 10 m and 30 m resolution across Helmand Province.

3.3.6 Experiment 3: Influence of image timing for active agriculture classification

Understanding how early in the crop season agricultural land can be delineated is important for opium survey planning, particularly for new areas of agriculture and understanding growing intentions. The fine-tuned 2015 FCN-8 model from §3.3.5 was run across all available Landsat-8 imagery between 18 February and 25 May 2015 to understand the differences in image timing and its impact on imagery selection across Helmand Province. All images have undergone IR-MAD normalisation and the same validation samples are used for comparison between the predictions and the maximum area of active agriculture for 2015 (n=91). Cloud cover has restricted the number of validation samples used in these assessments with the limited subset for each image date presented in Table 3-3.

Table 3-3. Suitable validation samples for each image date between 18 February and 25 May 2015 in Helmand, Afghanistan

Image date	Available validation samples	Scene cloud cover (%)
18 February	11	50.32
27 February	45	28.01
6 March	32	3.79
15 March	44	29.67
22 March	17	27.85
31 March	27	22.59
7 April (Peak - South)	32	0.23
16 April (Peak - North)	79	1.41
23 April	28	0.02
2 May	82	0.26
9 May	10	63.99
18 May	58	4.54
25 May	35	0.34

3.3.7 Localised intersection over union

A new method called localised Intersection over Union (IoU) is presented to further understand the distribution of spatial error in agricultural land classification. IoU is often determined at the global image level, as there is often only a single object of interest. For the application of agricultural mask production, the end user is interested in the spatial variation of accuracy across the classification at varying spatial scales. This method labels connected regions of pixels with the same class into separate individual objects (e.g. agriculture), where an object is defined as a connected area greater than 1 pixel (Figure 3-11). The objects are subset from the reference agricultural mask (Figure 3-11 (a)) and the predicted agricultural mask (Figure 3-11 (c)).

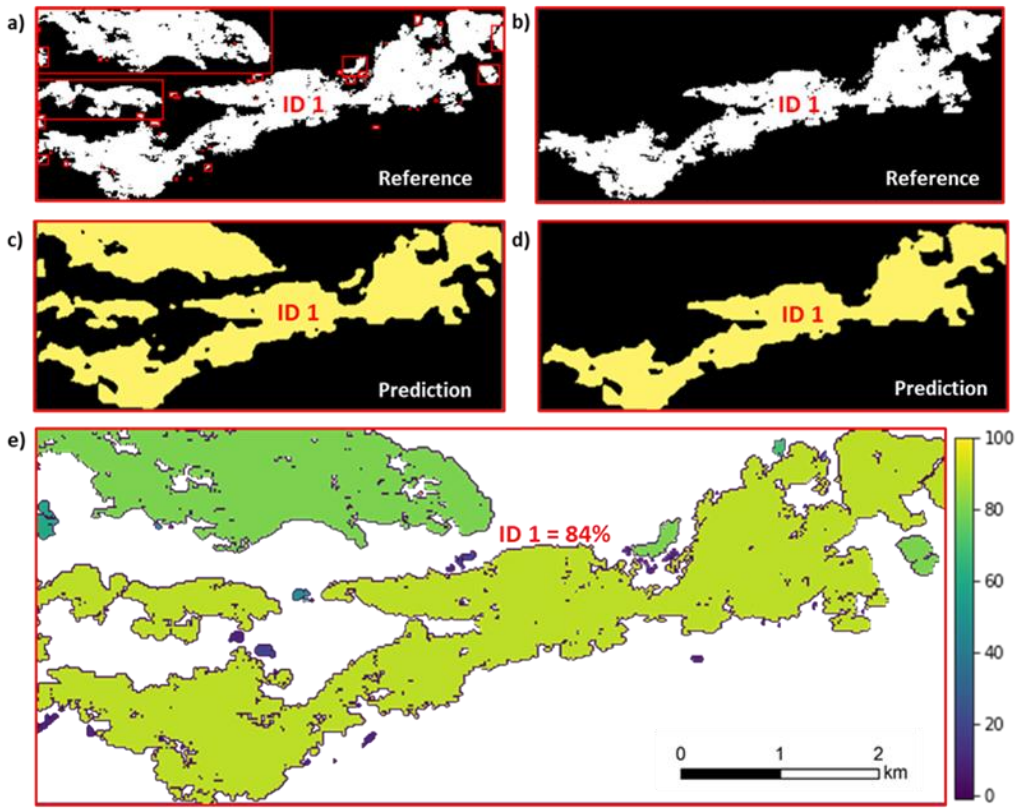

Figure 3-11. Overview of calculating localised intersection over union (IoU) between the reference and predicted agricultural masks for each area of agriculture. a) reference agricultural mask for 2009, b) isolated object (ID 1) in the reference agricultural mask, c) FCN-8 prediction, d) isolated object (ID 1) in the predicted agricultural mask and e) visualisation of all IoU values for each object in (a).

Larger blocks of agriculture intersect with other smaller blocks of agriculture. The same technique is used on these objects to remove other objects in the same area. This is carried out to avoid introducing multiple objects into the IoU calculation. If more than a single object is identified, the object with the highest pixel count is counted as the primary object and all other objects are removed from the image subset in the reference agricultural mask (Figure 3-11 (b)).

The reference object is subsequently used to isolate the same objects in the prediction (Figure 3-11 (d)). If multiple objects are present in the prediction, only those that intersect the reference object are selected to ensure only objects related to the reference object are counted. These two sets of objects can be used to calculate frequency-weighted IoU (Equation 3-10) at different spatial scales of agriculture, where:

$$\mathrm{fwIoU} = \frac{1}{\sum_i t_i} \sum_i \frac{t_i n_{i,i}}{t_i + \sum_j n_{j,i} - n_{i,i}} \qquad 3\text{-}10$$

$n_{j,i}$ is the number of pixels predicted as class j belonging to class i, $n_{i,i}$ is number of pixels predicted as class i belonging to class i and t_i is total number of pixels belonging to class i in the reference data. This method provides local variations in IoU when mapped back to each agricultural area (Figure 3-11 (e)).

3.4 Results

3.4.1 Experiment 1: Image features for FCN model training

The underlying spatial, textural, and spectral properties were explored using 2009 DMC imagery to investigate the influence of each image feature for FCN learning. Using modified training samples, the shape of agricultural land was unable to separate agriculture with poor segmentation accuracy with an IoU of 12.84 and 53% overall accuracy (Table 3-4). The textural features alone achieved an accuracy of 73%. The red and green spectral information increased texture-only accuracy by up to 14% and both performed similarly with only 1% difference in their IoU and accuracy. The addition of spectral information of the NIR band provided negligible additional contribution to the textural component by

increasing both accuracy and IoU by less than 1%. The combination of all spectral and textural features achieved an accuracy of 94% and 66% IoU.

Table 3-4. Summary of experiments used to isolate the spectral, textural, and spatial features of DMC imagery (NIR, R, G at 32m) for Fully Convolutional Network (FCN) classification of agricultural land using 2009 data

Model training data (*n* = 415)	Validation metrics (%) (*n* = 137)	
	Overall accuracy	fwIoU
Level-1A (benchmark)	**93.74**	**65.51**
Spatial		
Shape of agriculture area	53.20	12.84
Textural		
Homogeneity, entropy, and correlation	72.87	55.27
Spectral & textural		
NIR	73.65	56.25
R	87.77	62.78
G	86.77	61.84

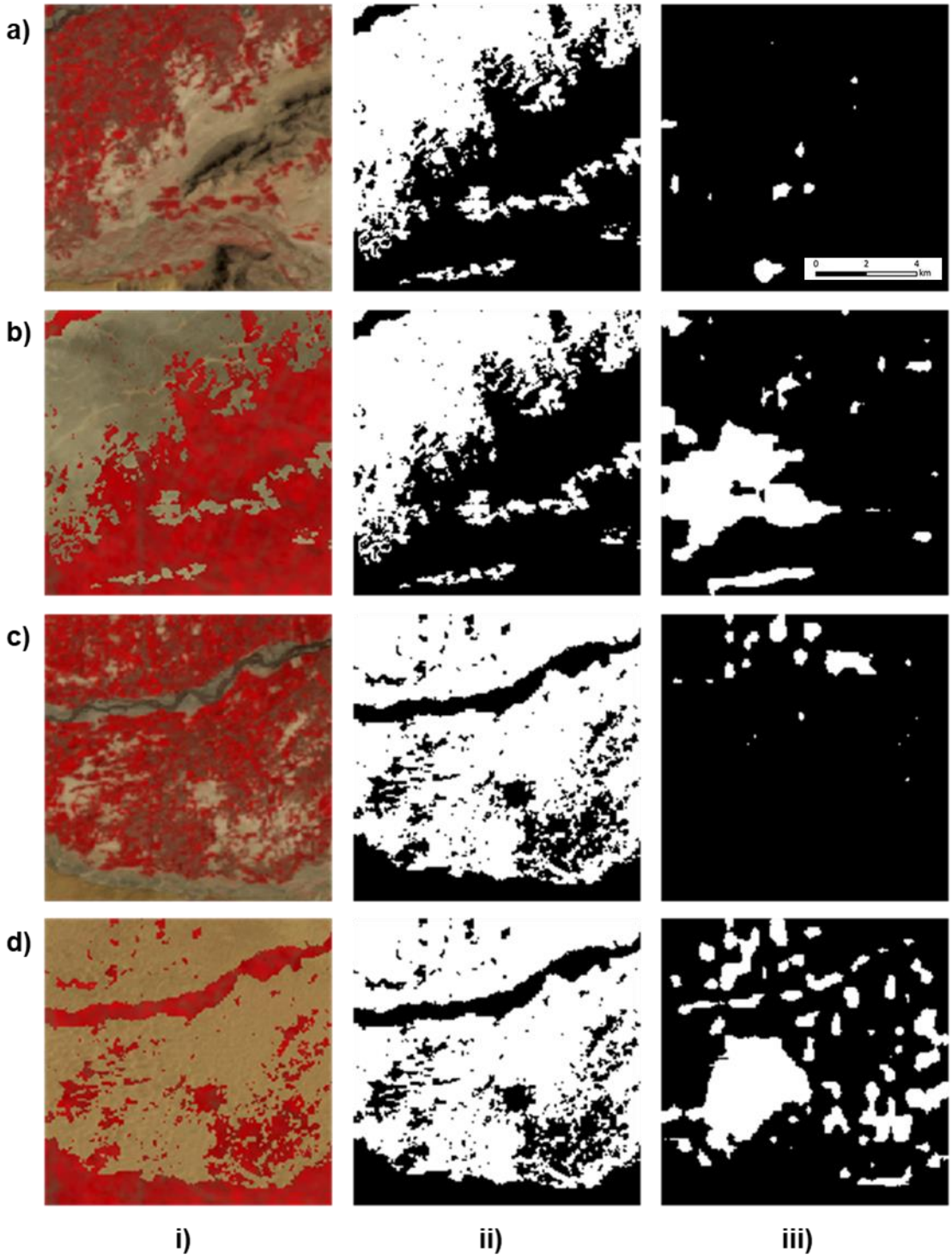

Figure 3-12. Validation of an example 256 x 256 pixel chip with the global shape FCN model trained using 2009 data. i) Level-1A image chip (original (a and c) and synthetic (b and d)), ii) image labels and iii) prediction with Level-1A chip and inverted Level-1A chip. Imagery: 2009 DMC (NIR,R,G) at 32m. Agriculture is denoted in white and non-agriculture in black in the labels and prediction image.

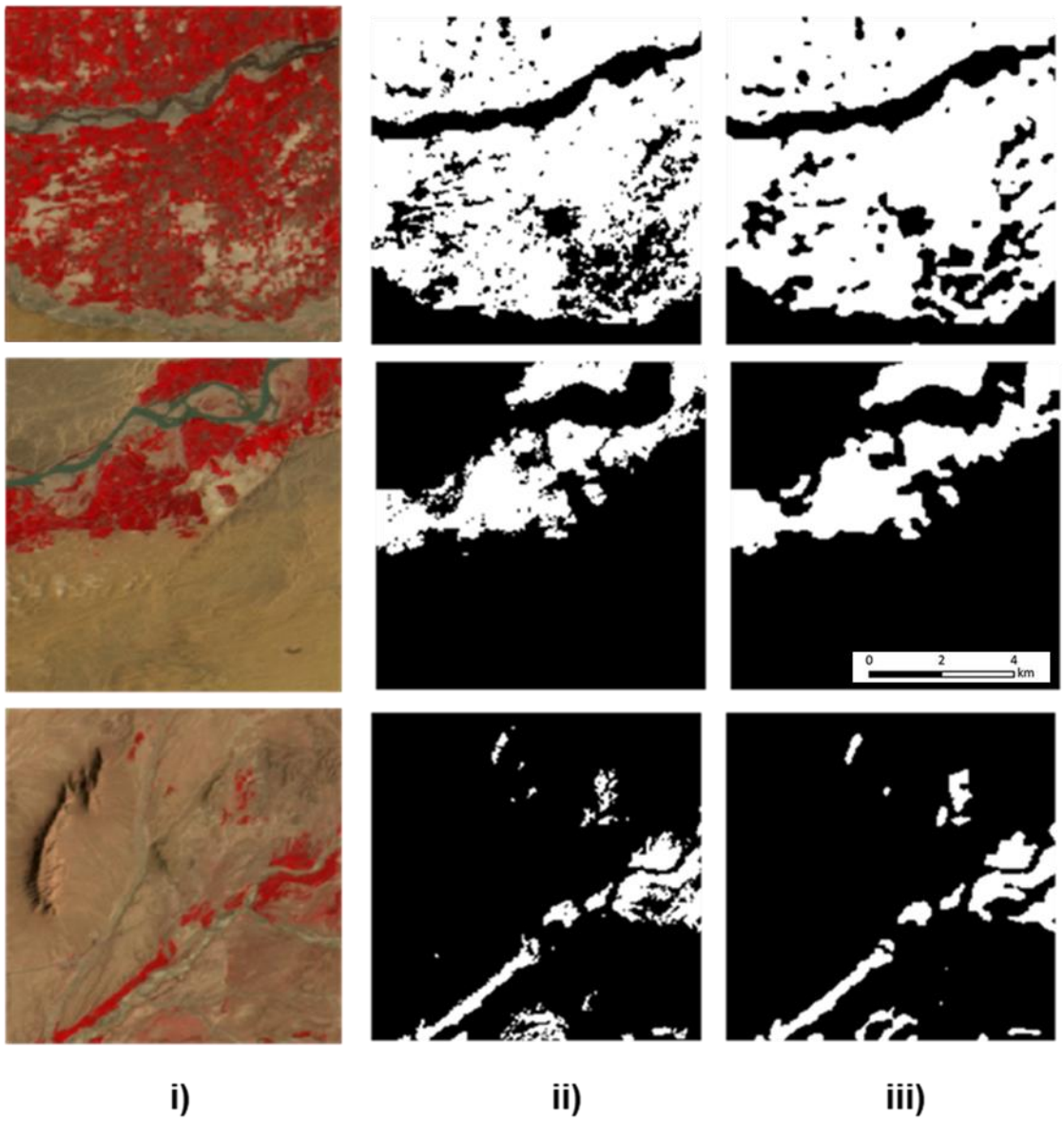

Figure 3-13. Validation of example chips with the texture FCN model trained using 2009 data of homogeneity, entropy, and correlation. i) Image data, ii) labels and iii) prediction. Imagery: 2009 DMC (NIR, R, G) at 32 m. Agriculture is denoted in white and non-agriculture in black in the labels and prediction image.

The visual delineation for classification using the spatial component was found to perform very poorly with no consistent delineation patterns using the unaltered validation sample (Figure 3-12). An inverted version of the chip was also found to perform poorly and only isolated a thin block of agriculture in the south which is generalised at the boundaries. The textural model successfully separated the distinct boundaries across agricultural areas of different sizes and shapes (Figure

3-13). The noticeable difference between the textural classification and the Level-1A benchmark classification is the increased generalisation at the boundaries, especially in larger blocks of agriculture.

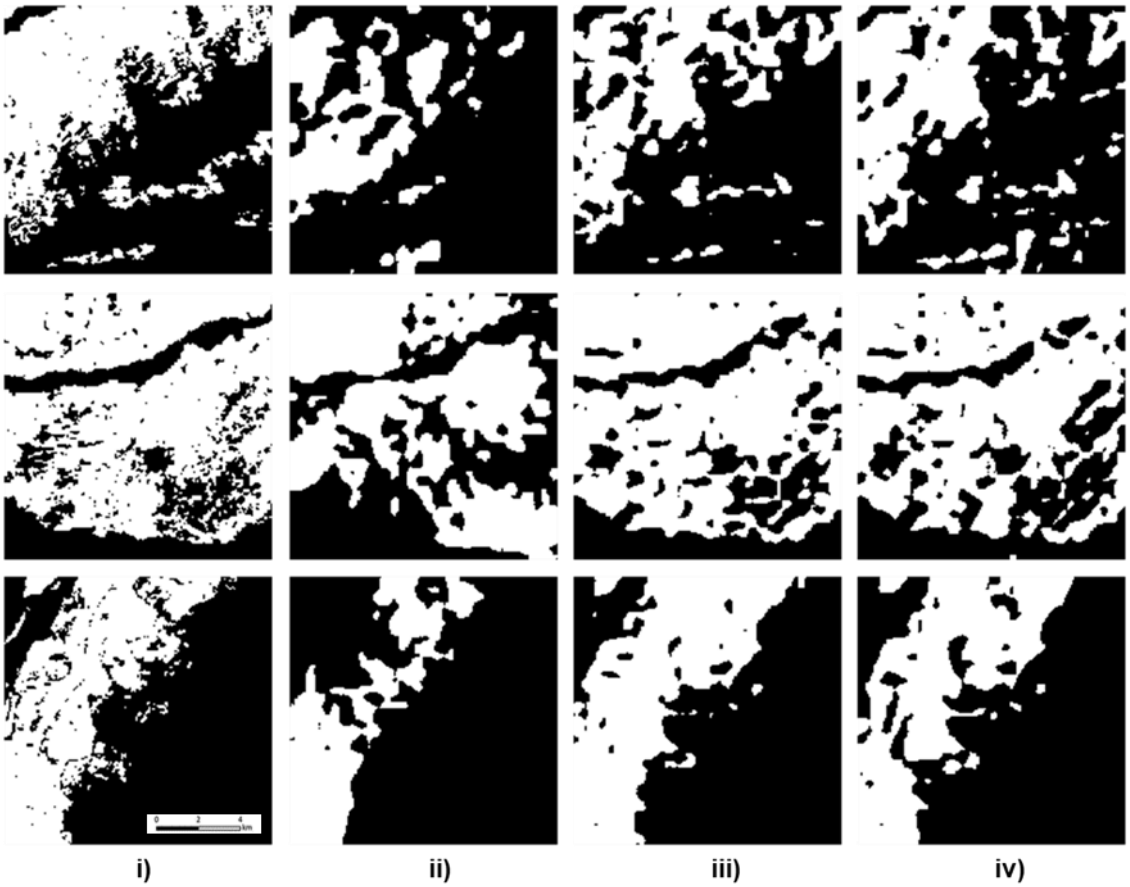

Figure 3-14 Validation of example chips with the spectral FCN models trained using 2009 data of the near-infrared (NIR), red (R), green (G) spectral bands. i) labels, ii) NIR spectral band prediction, iii) R spectral band prediction and iv) G spectral band prediction. Imagery: 2009 DMC (NIR, R, G) at 32 m. Agriculture is denoted in white and non-agriculture in black in the labels and prediction images.

The spectral features of satellite imagery are more important for FCN learning from this experiment (87.77% overall accuracy for the red spectral band). The NIR band is able to classify large continuous areas of agriculture (Figure 3-14), but has difficulty separating edge cases and those with greater spectral confusion (Figure 3-7). The red and green bands have similar classification results, but the green band omits areas of agriculture in boundary areas. The red band has tighter classification, but still has generalisation in these areas. The importance of the

spectral information for FCN learning emphasises the need for standardisation to evaluate transfer learning across image datasets.

3.4.2 Experiment 2: Image standardisation and model transferability between image sensors (2007-2009)

The agricultural features learnt from DMC images were transferable to Landsat-5 classification and achieved >89% overall accuracy across all standardisation techniques (Table 3-4). The best-performing model for DMC imagery used IR-MAD image matching to achieve 94.49% overall accuracy. The IR-MAD marginally outperformed Level-1A by 0.10% for DMC imagery and outperformed Level-1A for Landsat-5 imagery by 1.97%. Level-1A corrected imagery marginally outperformed NDVI for agricultural mask delineation of both DMC and Landsat-5, +0.95% and +0.34% respectively. Pixel intensity performed well by achieving 90% accuracy for agriculture delineation using Landsat-5. The intersection over union was also greater for IR-MAD normalisation with the highest performance for DMC imagery and Landsat-5 imagery. The IoU segmentation results of the two sensors were more comparable between intensity with -4.66 difference between the two sensors. Level-1A and NDVI had greater intersection over union, but have greater difference between DMC and Landsat, -9.57 and -10.77 respectively.

Table 3-5. Summary of experiments using a Fully Convolutional Network (FCN) classification of agricultural land to transfer knowledge between DMC and Landsat-5 by using a DMC-trained model on 2007 and 2008 data and validated on DMC and Landsat-5 2009 data (NIR, R, G at 32 m)

Model training data (2007 and 2008: *n* = 830)	2009 imagery	Validation metrics (%) (2009: *n*=137)	
		Overall accuracy	fwIoU
IR-MAD	DMC	94.49	67.96
	Landsat-5	93.98	62.84
Level-1A	DMC	94.39	67.21
	Landsat-5	92.01	57.64
NDVI	DMC	93.44	65.89
	Landsat-5	91.67	55.12
Intensity	DMC	91.87	57.81
	Landsat-5	89.96	53.15

The delineations for DMC imagery are visually similar in (Figure 3-15), although the IR-MAD is able to separate the difficult boundaries between larger agricultural

fields better than Level-1A. The IR-MAD has a similar delineation along the distinct agricultural boundaries with desert to Level-1A, but had difficulty in separating the spectrally mixed responses within the agricultural area (e.g. missing part of the agricultural area in the south-east of Figure 3-15). Both classifications have omitted smaller areas of agriculture in the north-west of the image extent, whilst delineating the larger blocks of agriculture surrounding these areas.

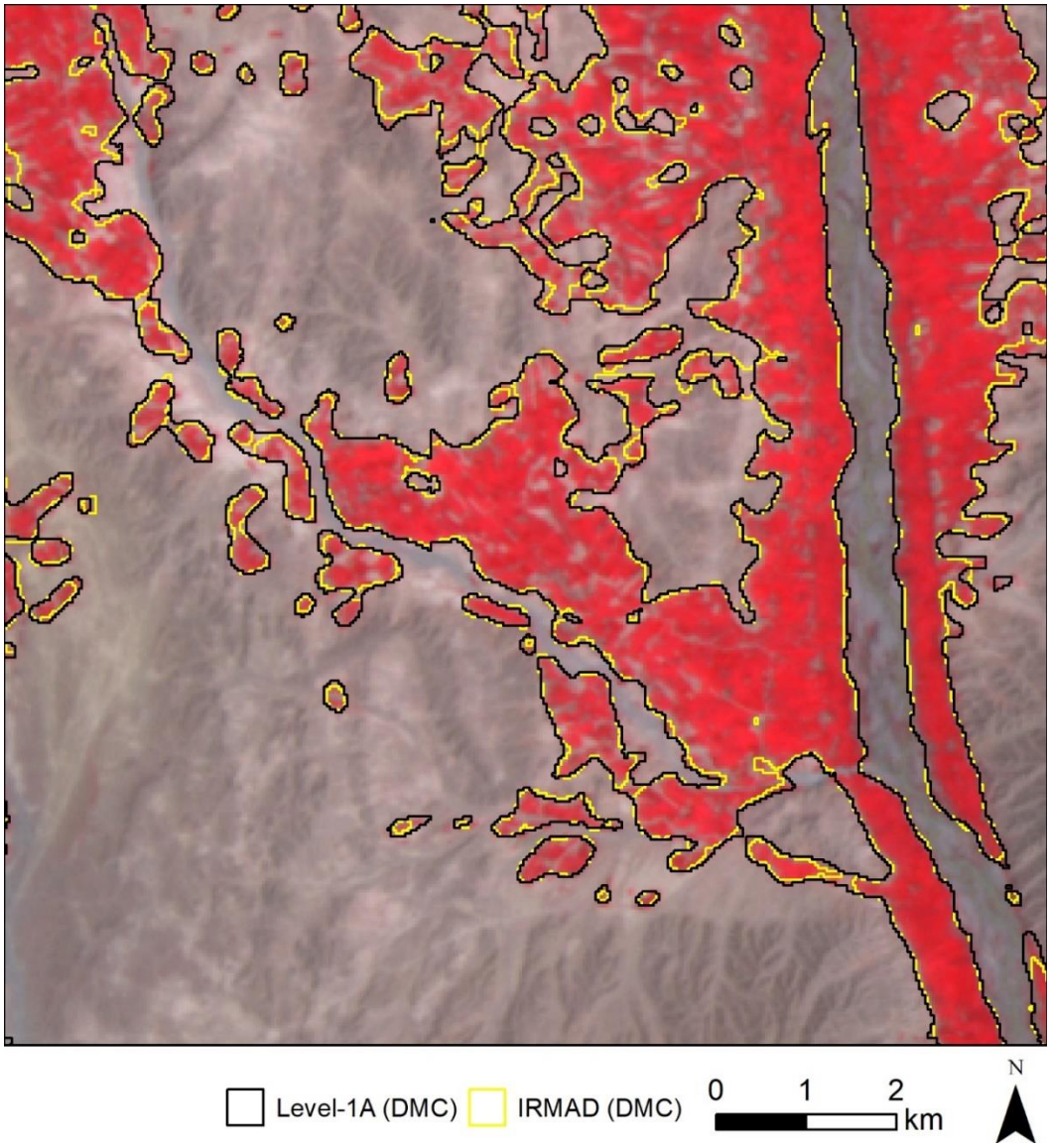

Figure 3-15. Visual evaluation of agricultural delineation in the main Helmand valley using Level-1A and IR-MAD radiometric calibration techniques for DMC imagery. Image (32.0904° N, 64.4758° E): False colour (NIR, R, G) DMC imagery (25 March 2009)

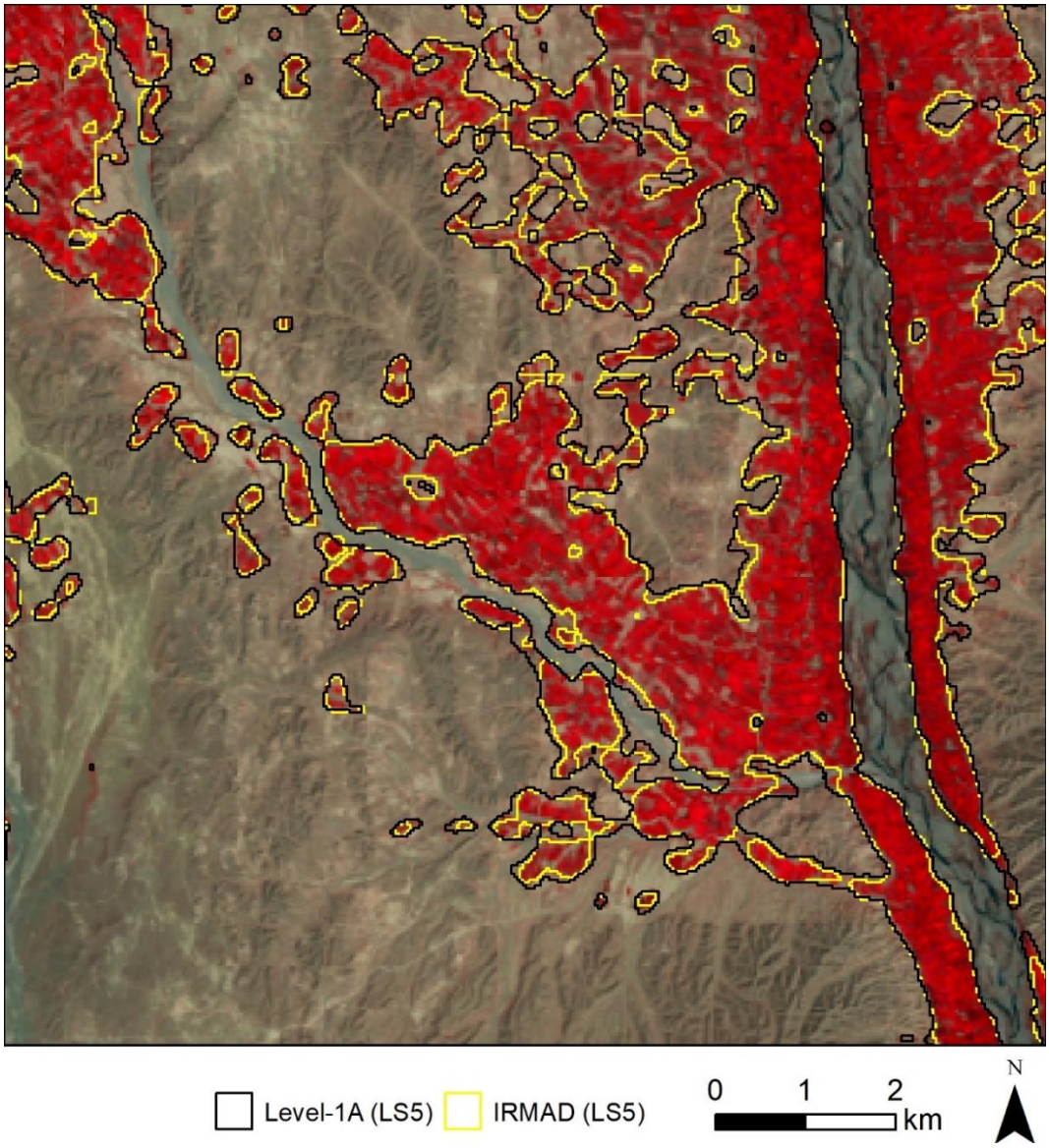

Figure 3-16. Visual evaluation of agricultural delineation in the main Helmand valley using Level-1A and IR-MAD radiometric calibration techniques for Landsat-5 imagery. Image (32.0904° N, 64.4758° E): False colour (NIR, R, G) Landsat-5 imagery (5 April 2009)

The delineation using IR-MAD normalisation for Landsat-5 imagery was found to be similar to DMC imagery with good delineation at distinctive boundaries between agriculture and desert (Figure 3-16). The IR-MAD's performance is greater between blocks of agriculture with less generalisation, which is found in the Level-1A imagery and has difficulty in delineating complicated areas with spectral confusion between non-agriculture and agriculture. In the DMC imagery

there were more areas of agriculture commission in these complicated areas, but in the DMC imagery the IR-MAD is more conservative in its delineation and resulted in more areas of omission on the boundaries of agricultural areas.

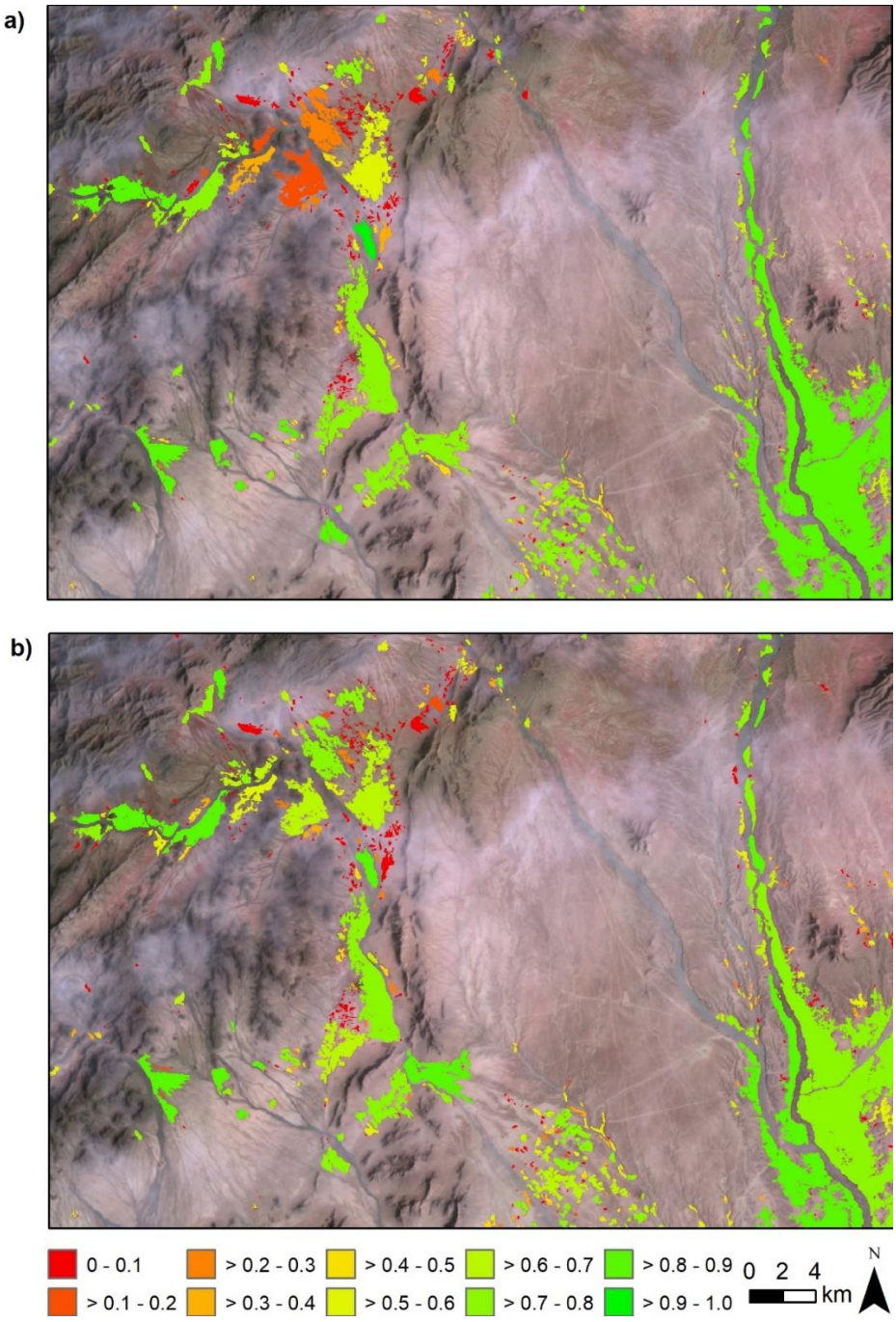

Figure 3-17. Comparison of the FCN-8 predictions using Level-1A (a) and IR-MAD normalisation with the reference agricultural mask for 2009. Image (32.2228° N, 64.3931° E): False colour composite of DMC imagery for 3 April 2009 (NIR, R, G at 32m).

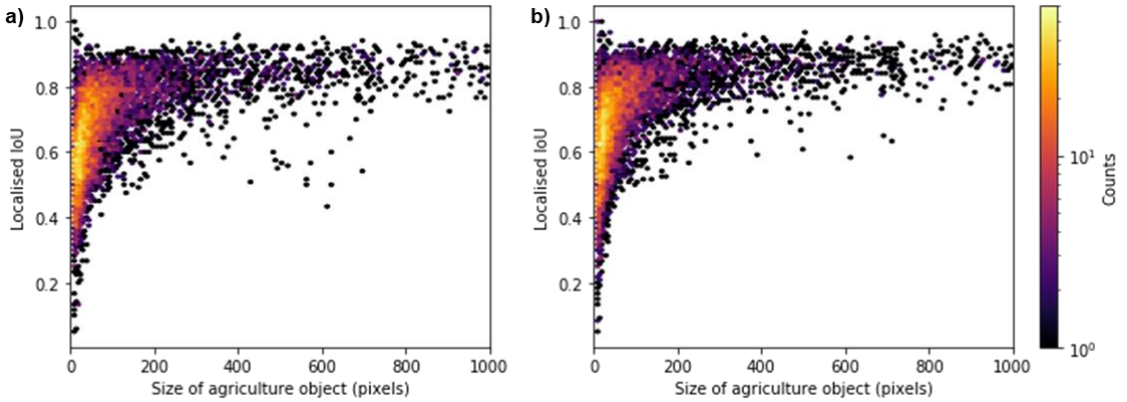

Figure 3-18 Comparison between the size of agricultural area and localised intersection over union using a) Level-1A and b) IR-MAD normalisation for 2009 DMC imagery (NIR, R, G at 32 m)

Using the new localised IoU method, the variation in IoU between the reference agricultural mask and Level-1A (Figure 3-17 (a)) and IR-MAD (Figure 3-17 (b)) FCN-8 predictions for DMC imagery can be visualised. Larger areas of agriculture have tight classification at their boundaries in Level-1A and IR-MAD resulting in high IoU. Smaller agricultural areas are generally better classified in the IR-MAD. These areas are shown in the north-west corner of the image subsets, where the tighter classification from IR-MAD has resulted in higher localised IoU across these areas of agriculture. Using localised IoU provides further understanding of the best-performing classification (IR-MAD) at different spatial scales of agriculture (see Appendix B). Smaller agricultural areas were found to be more affected by generalisation because the pixel population count for agriculture is smaller (Figure 3-18). Smaller differences over larger areas are reduced in the localised IoU calculation, even when using the frequency-weighted IoU because there is a higher population count. Level-1A (Figure 3-18 (a)) and IR-MAD (Figure 3-18 (b)) had similarities in their localised IoU distribution with greater variation for samples less than 200 pixels. The IR-MAD model was able to improve IoU for both large and small objects.

3.4.3 Generalised model evaluation using transfer learning between image sensors (2007-2017)

The 2009 FCN-8 model trained on DMC imagery was fine-tuned using 2015 data from Landsat-8 to achieve an overall accuracy of 93.01% on the validation samples (Table 3-6). Without any additional training the 2009 FCN-8 model achieved 90.99%. The 2015 model was further updated with only new areas of agriculture from differences between the 2015 and 2016 agricultural mask to achieve an overall accuracy of 93.01%. The 2016 model was further updated with new areas of agriculture from 2017 to achieve an overall accuracy of 95.98% on Landsat-8 imagery. Sentinel-2 imagery was available over the study area for 2017 and achieved 95.58% overall accuracy (-0.40%) from the same validation samples using resampled 30 m resolution imagery.

Table 3-6. Summary of classification performance using IR-MAD radiometric normalisation for agriculture area delineation in Helmand, Afghanistan for the 2009 FCN model, 2015 model, 2016 model and 2017 model on Landsat-8 data (2015, 2016 and 2017) and Sentinel-2 data (2017). Image dates for these datasets can be found in Table 3-2. UA is user accuracy and PA is producer accuracy.

Classified image (30 m)	FCN-8 model name	Model transfer learning	Additional training	Number of training years	Accuracy metrics (%) (*n*=91)		
					Overall accuracy	Agriculture UA	Agriculture PA
2015 Landsat-8	2009	2007, 2008, 2009 (DMC)	None	3	90.99	89.32	83.81
2015 Landsat-8	2015	2009 FCN-8	2015 (Landsat-8)	4	93.01	91.76	87.54
2016 Landsat-8	2015	2015 FCN-8	None	4	95.12	91.34	89.11
2016 Landsat-8	2016	2015 FCN-8	2016 (Landsat-8)	5	96.11	92.01	89.55
2017 Landsat-8	2016	2016 FCN-8	None	5	95.01	91.26	85.71
2017 Landsat-8	2017	2016 FCN-8	2017 (Landsat-8)	6	95.98	91.91	89.03
2017 Sentinel-2	2017	2017 FCN-8	None	6	95.58	91.23	88.81

The generalised model performance for each FCN-8 model was evaluated using 95% prediction intervals on the total agricultural area across validation samples for 2009, 2015, 2016 and 2017 (Figure 3-19). Evaluating the performance of the model on total agricultural area shows the uncertainty of area estimates using this approach. Orthogonal regression between each validation sample's

agricultural area across all image years had high correlation (0.99 r^2) with fewer samples in large agricultural areas. The total agricultural area is overestimated by +3.2% in comparison with the reference data using the generalised FCN-8 model (Table 3-7). The overestimation of agricultural area is prominent in larger areas of sparse agriculture which the model has difficulty separating between (Figure 3-20 (a)). Strong boundary edges do not experience the same generalisation and have tight classification (Figure 3-20 (b)).

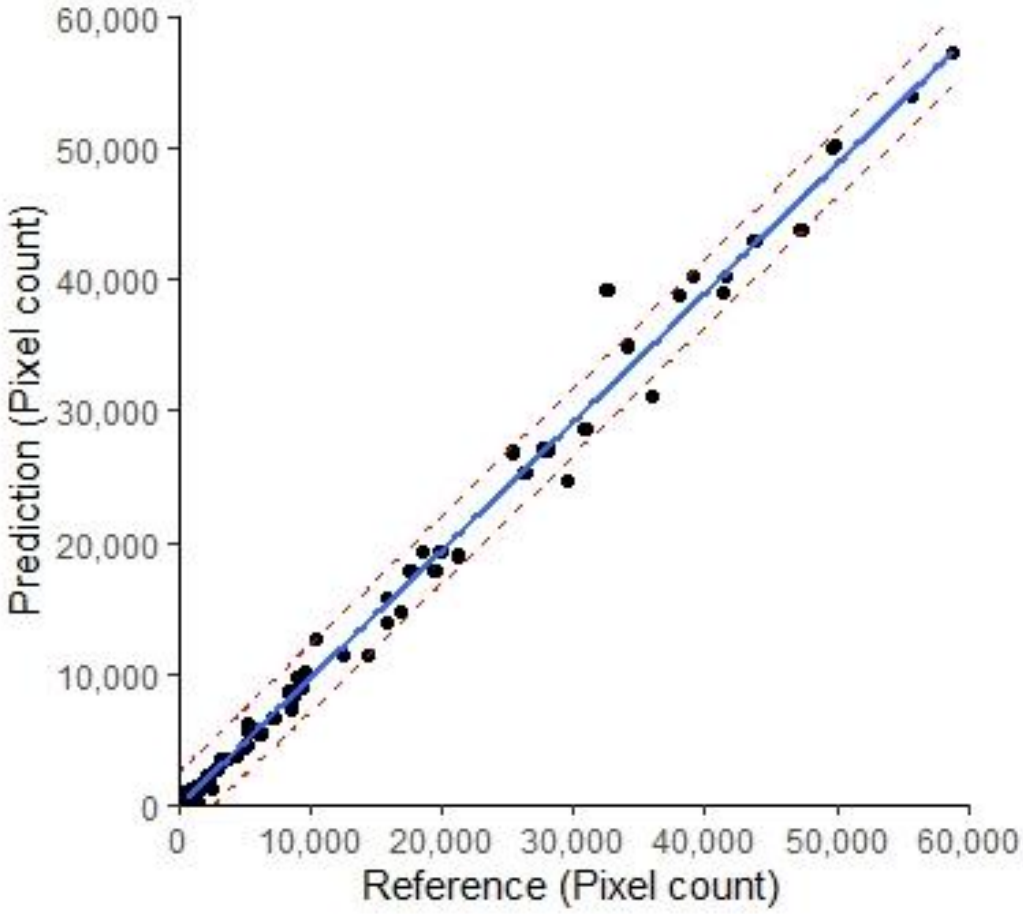

Figure 3-19 95% prediction intervals for the generalised FCN-8 model using the reference and predicted agricultural area in 2009, 2015, 2016 and 2017 validation samples

Table 3-7 Total agriculture area of the reference data and FCN-8 prediction across validation datasets (2009, 2015, 2016 and 2017) with 95% prediction intervals.

Validation data	Total agriculture area (ha) (*n* = 364)		
	Reference	Prediction	95% prediction interval range
2009, 2015, 2016 and 2017	12,057	12,455	12,343 – 12,668

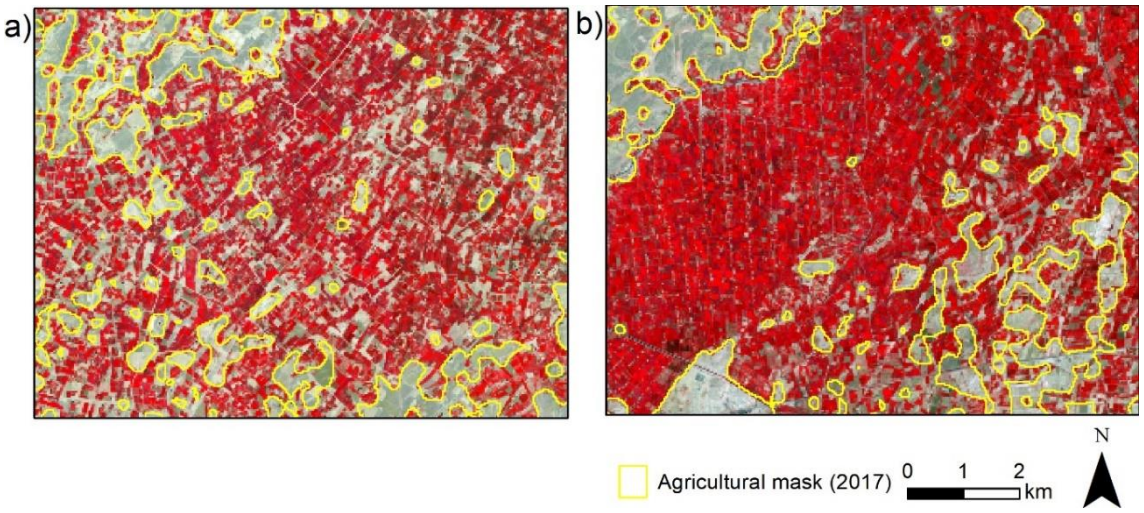

Figure 3-20 Comparison between a (a) sparse agriculture dominated area and (b) dense agriculture dominated area. Image (31.3333° N, 64.901° E): False colour Sentinel-2 imagery (NIR, R, G at 30 m) from 8 April 2017

3.4.4 Impact of spatial resolution on area estimates

Variations in spatial resolution between image sensors may present differences in the total agricultural area classified by generalised FCN-8 models. The differences in the classified agricultural area for central Helmand is shown in Figure 3-21. The main difference between the two resolutions is the tighter agricultural delineation at field boundaries in the 10 m resolution imagery (Figure 3-21 (a and d)). The generalisation in the 30 m resolution caused by the complexity in the agricultural area results in a lower IoU as there is more agricultural area detected in comparison to the 10 m resolution.

Higher IoU is found in dense areas of agriculture with limited complexity between field boundaries, especially along the Helmand river (Figure 3-21 (b and c). The higher spatial resolution also detects smaller areas of agriculture which are missing from the 30 m classification, as found in Figure 3-21 (c) with a large area

of dispersed agriculture with low IoU. Spectral confusion between non-agriculture and agriculture is reduced with higher spatial resolution and forms a more representative classification of the agricultural land (Figure 3-21 (a)). Figure 3-21 (d) shows poor quality crops have been classified in the middle of the large block of agriculture in the 10 m imagery. The 30 m has omitted these areas which is likely due to the spectral confusion from the low spectral response.

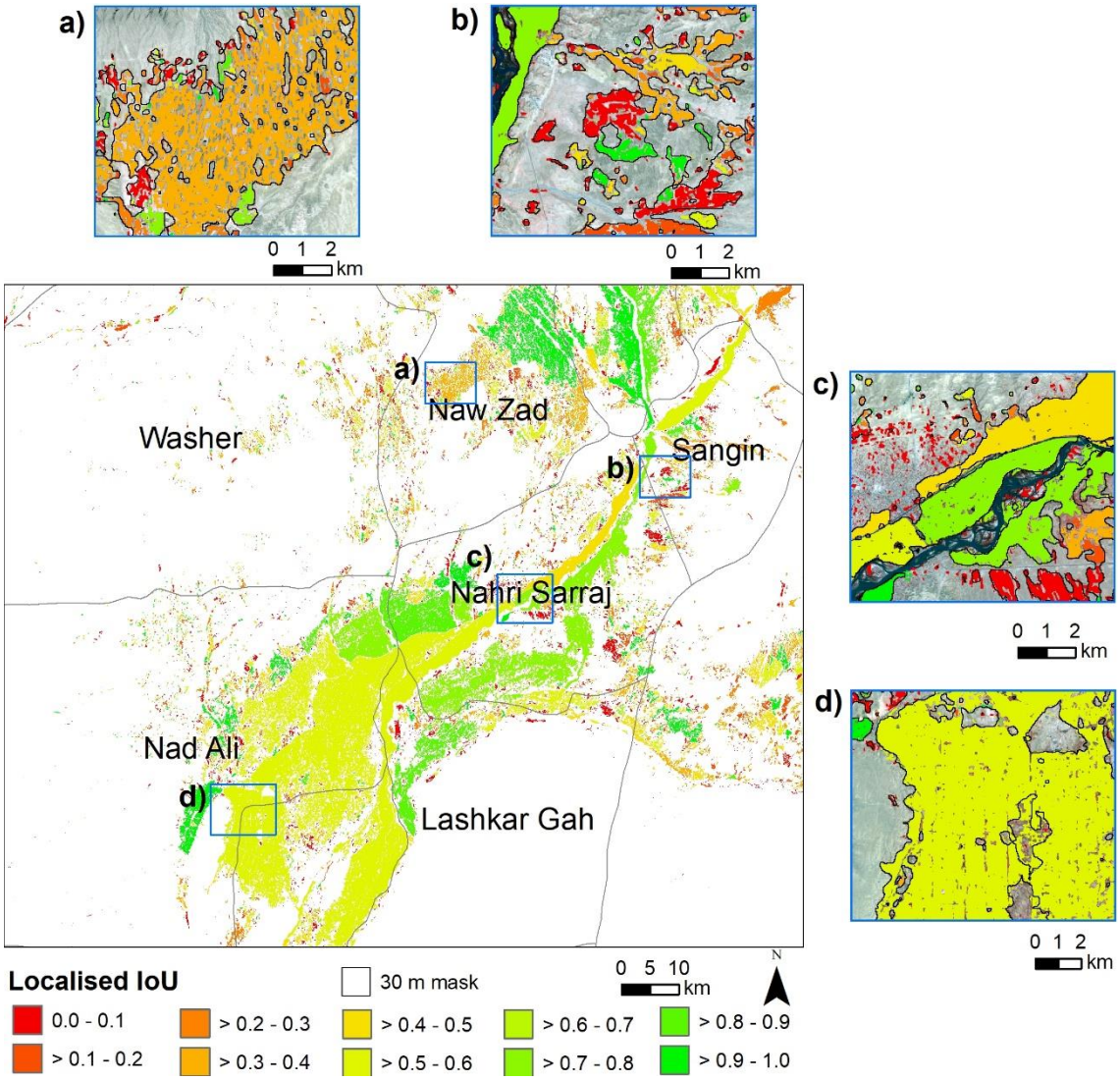

Figure 3-21. Comparison between 10 m and 30 m agricultural land classification from the 2017 generalised FCN-8 model using localised IoU. Sentinel-2 at 10 m is displayed as the coloured base map and resampled Sentinel-2 at 30 m is displayed as the black outline in the insets across central Helmand, Afghanistan. Inset images: False colour (NIR, R, G) Sentinel-2 on 8 April 2017

Tighter agricultural delineations within larger areas of sparse agriculture in the high resolution and overestimation in the lower resolution are the differences between classifications. Variations in agricultural area caused by differences in the spatial resolution would also occur with human interpretation of agricultural land where the trained interpreter would produce tighter agriculture delineations with higher resolution imagery. Differences in agricultural area would alter the opium cultivation estimates using the sample mean ratio methodology (UNODC, 2017b), where an overestimation within the opium frame samples would inflate the production estimates. In Nad Ali, the district with the largest agricultural area and opium cultivation, the area for 10 m resolution classification in 2017 was 64,564 ha and is 14,263 ha lower in comparison with the 30 m resolution. Using higher resolution imagery would reduce the overestimation caused at weak edge cases within areas of sparse agriculture.

3.4.5 Experiment 3: Performance of generalised models through time (2015)

Standardisation provides comparable values not only between image sensors, but also through time. Deploying generalised models on multiple dates from the same year provides information on agricultural practices before agricultural areas achieved their peak extent. The variation in NDVI derived from Landsat-8 shows the differences in the average NDVI profiles and its relationship with user accuracy of validation samples between the north and south of Helmand (Figure 3-22). A subset of the validation samples with the same sample locations from the 2015 image timing analysis have been separated into north and south Helmand to investigate the differences in agricultural land delineation. The south of Helmand reaches around ¾ biomass in early March based on the NDVI profile. The north of Helmand shows a later peak on 15 April, therefore the variation in peak NDVI timing in the north may require a later image date to identify these agriculture areas (after 6 March).

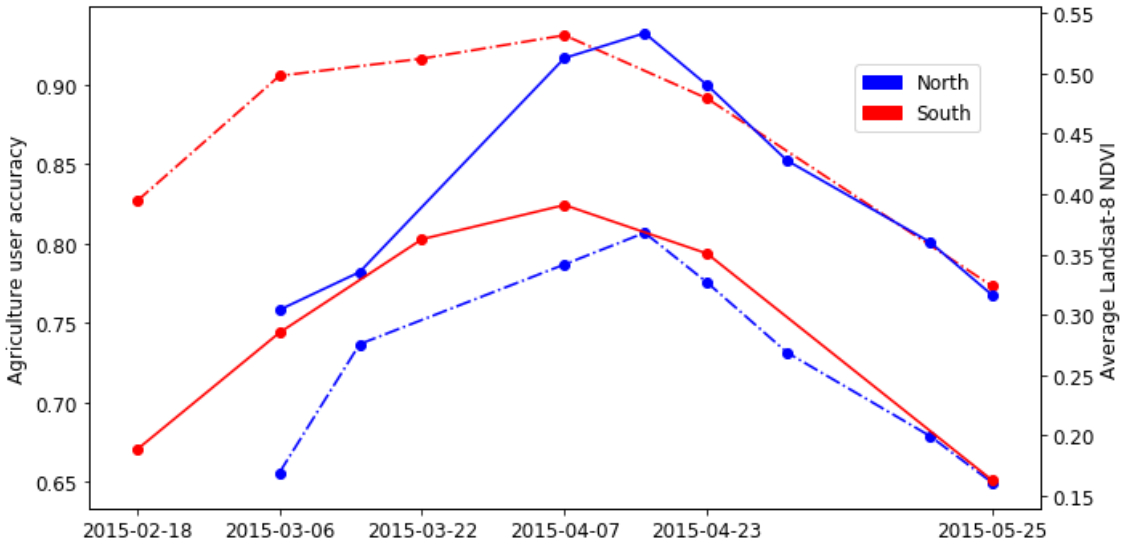

Figure 3-22. Comparison between average agriculture user accuracy and Landsat-8 NDVI between first crop cycle image dates across the whole of Helmand in 2015 (North: *n*=5 and South: *n*=8*). Dotted line shows agriculture user accuracy and solid line shows the average NDVI for the samples. * Fewer samples have been used to ensure all validation samples selected have the same image date range for accuracy assessment

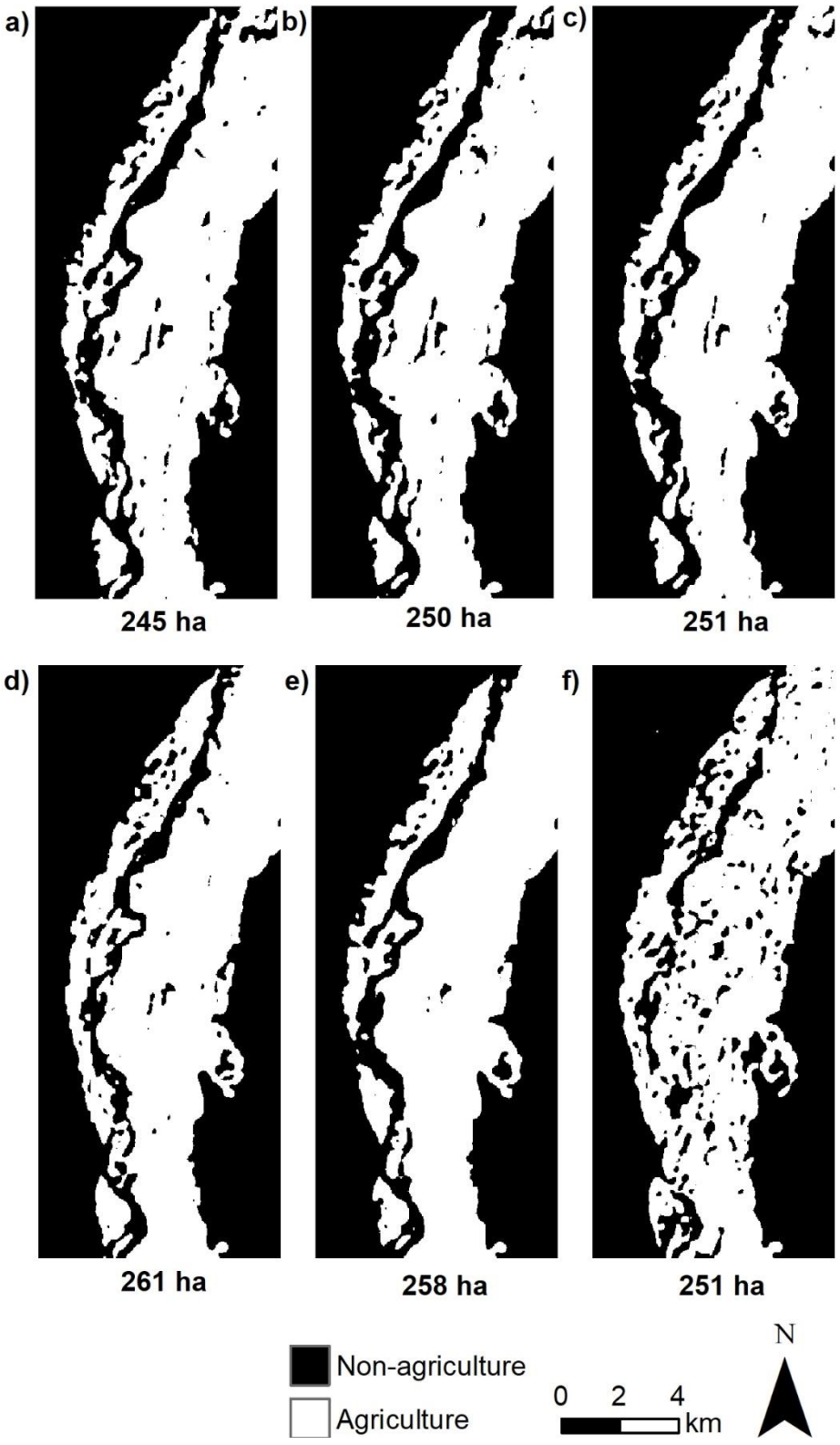

Figure 3-23 Classification of Landsat-8 images between 18 February and 25 May 2015 across Reg district, central Helmand (30.5333° N, 64.776° E). a) 18 February, b) 6 March, c) 22 March, d) 7 April (peak), e) 23 April and f) 25 May

The effect of the variation in the timing for agricultural land delineation is an important consideration when selecting the optimal date for agriculture land classification or inferring early agricultural land estimates (Taylor *et al.*, 2010). The early Landsat-8 can identify the overall agricultural area in Reg district from mid-February (Figure 3-23 (a)), but as the image dates approach the peak date in the South (7 April 2015) the agricultural land reaches its maximum extent (Figure 3-23 (d)). Between these two dates the classification has an increasing ability to distinguish agriculture as the spectral response increases. After the peak the agricultural land reduces as it is harvested and areas within the main body of the agricultural area are removed and the spectral response reduces (Figure 3-23 (f)).

3.5 Discussion

3.5.1 Standardisation of multispectral images for agricultural land classification with FCNs

Understanding the relative importance of image features aids the training strategy used for agricultural land classification between image sensors with FCNs. The shape of agriculture was defined by the landscape and management practices of arable land (e.g. the field parcels and their size) and was found to have negligible influence on the FCN's ability to train the network for agricultural mask delineation (53% overall accuracy). Using the synthetically altered chips to train the model did not perform well with the validation data (12.84% IoU). The edges of the training data are very different to those found in the validation dataset, where there is a sudden transition between non-agriculture and agriculture without the spectral mixing between classes. The original validation dataset would have the natural transition between the two classes, which the FCN is able to identify from prior experiments using the original samples as training data and having seen these boundary cases before. The shape was also found not to be a main contributing factor for Baker *et al.* (2018), where texture was the main contributing factor and not the global shape using their CNN model at the image level. Shelhamer, Long and Darrell (2016) found shape was a minor contributing factor in their multi-class FCN model with a mean IoU of 29.1% at the pixel-level, in comparison with 84.8% using all imagery features for training. The binary

application of this experiment may be attributed to the cause of little contribution from the global shape of agriculture. The visual human interpreted differences between agriculture and non-agriculture are distinct (e.g. desert in comparison to agricultural land), but there is not a consistent shape for agriculture (e.g. ribbon valleys, dense agriculture, and sparse agriculture). Whereas in multi-class classification with commercial image datasets (e.g. ImageNet), there are consistencies between shapes (e.g. the outlines of a cat or dog).

The textural component was influential in agricultural mask delineation (72% overall accuracy). The spectral components are dependent on similar values in specific parts of the electromagnetic spectrum to identify the differences between agriculture and non-agriculture. Surprisingly, the NIR band had the lowest performance (73%) in comparison with the other spectral bands across this landscape (e.g. 87% and 86% for red and green bands, respectively) (Table 3-4). Although NIR wavelengths are synonymous with agriculture detection, due to their high reflectance of this portion of the electromagnetic spectrum, it was found not to be as important as the red or green bands. Further investigation found the NIR band had lower standard deviation than the other spectral bands for 2009 DMC imagery in areas with high proportions of desert (Figure 3-24). Desert areas have high reflectance in the NIR spectrum, similar to agricultural land, and is difficult to separate because of the limited spectral contrast (Figure 3-7). The greatest standard deviation was found in the red band, which provides greater spectral contrast, and is also the best-performing of the spectral bands. Texture relies on spectral contrast or backscatter differences; therefore, this component can be derived in both optical and passive imagery. Further studies could investigate the transfer of knowledge using textural features between optical and passive imagery, which could provide further synergy between these two types of imagery for classification tasks.

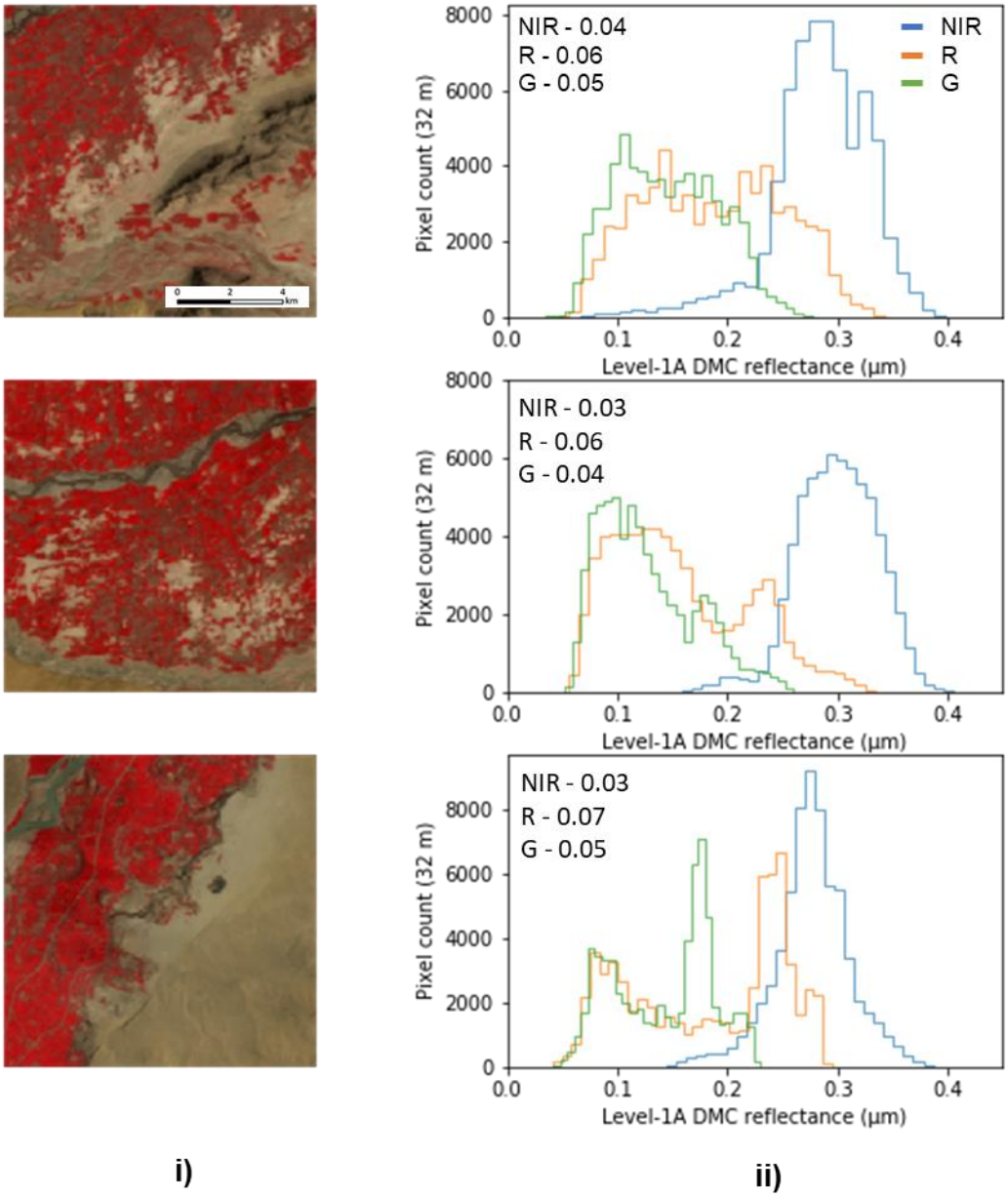

Figure 3-24 Image histograms of example Level-1A DMC validation chips with standard deviations for each spectral band (NIR (0.76 – 0.90 µm), R (0.63 – 0.69 µm) and G (0.52 – 0.62 µm) at 32 m) from Figure 3-24

The importance of the spectral data emphasises the importance of standardising imagery data for interoperable classification techniques, where consistent values are required for classification. IR-MAD normalisation outperformed Level-1A data primarily at boundary locations where consistency amongst the spectral values is essential for correct classification. The use of localised intersection over union aided classification evaluation by providing a localised evaluation of the

classification on individual blocks of agriculture. This localised evaluation method was useful in determining the strengths and weaknesses of each model. Global accuracy metrics are useful as an initial benchmark for classification performance, but visualising differences locally is advantageous for identifying well- and poorly-delineated areas. The localised IoU provided information not available in the global accuracy metrics used. Larger areas of agriculture without irrigation canals or gaps in crop cultivation had high IoU. Larger areas with breaks in crop distribution resulted in lower IoU because of the boundary generalisation.

The FCN was proven to be interoperable between imagery datasets by radiometrically normalising Level-1A imagery using the IR-MAD technique. High classification performance was maintained across DMC and Landsat-5 at the same resolution with tighter agricultural land delineations. This overcomes the main disadvantage of traditional techniques, which is their inability to be transferred between datasets (Vali, Comai and Matteucci, 2020). The localised IoU found the greatest differences between the IR-MAD and Level-1A classifications (Figure 3-17) were at edge cases where there is greater spectral confusion. Level-1A calibration provides a similar visual result on larger areas of dense agriculture and global accuracy metrics to IR-MAD and is a common radiometrically calibrated product from imagery providers (Dwyer *et al.*, 2018). Level-1A may be more appropriate for initial deployment over long time-series imagery where precise classification is not required. Where end users require accurate delineations of agricultural land the additional IR-MAD pre-processing is appropriate.

3.5.2 Continuous improvement of a generalised model using transfer learning

The FCN-8 model outperformed ResNet50 in agricultural land classification across Helmand Province with tighter delineations at boundary locations (Figure 3-2). The reconstruction of pixelwise land cover classifications was found to be more efficient for FCNs as the removal of the fully connected layer in a traditional CNN provided predictions at the native resolution of the image sensor and did not require a sliding window to achieve the same result (Chapter 2). The trained

FCN model can also be fed images of different sizes to those used to train the model to increase efficiency further. CNNs use a fully connected layer requiring inputs to be the same size during training and inference. Computer memory was found to be the only limiting factor when scaling the FCN over larger areas but can be overcome by decreasing the input image size during inference. The main constraint of training an FCN from scratch is the requirement for dense labelled datasets, which for Afghanistan involved time-consuming manual interpretation of image classifications.

Classification performance improved year-on-year when more data was used to train the FCN-8 model. The generalised model was able to be updated using sparsely labelled data of observations in new areas of agriculture and led to incremental improvements (+1%), in comparison to using the previous year's model. Updating with only new areas of agriculture from subsequent years' agricultural delineations further reduces the manual effort required to maintain the model. Over the five years of transfer learning, the accuracy of the FCN models improved when introducing more training samples from additional years. Further work is required to understand how many years of data are required before the model no longer improves. Updating the model may only be required if substantial changes occur (e.g. introduction of new crops). The generalised model was found to overestimate agricultural land at boundary edges between complex field parcels in medium resolution imagery (Figure 3-20 (a)). The generalised FCN model can be deployed on any size image across different spatial resolutions provided standardisation is used with higher resolutions resulting in tighter agricultural land classification. The overestimation in difficult areas can be mitigated by using high resolution imagery, where 10 m Sentinel-2 separated field boundaries in sparse agricultural areas which were not present in the resampled 30 m dataset. Peak opium biomass was different between northern and southern Helmand, but using generalised models allowed for classification of agricultural land at any time in the season. Generalised models are under-utilised in remote sensing and would allow for a wider range of legacy datasets, regardless of image timing, resolution or sensor, to be used to derive new knowledge on monitoring land cover change.

3.6 Conclusions and recommendations

An FCN-8 model achieved 94.65% overall accuracy to outperform the ResNet50 CNN for agriculture classification across Helmand and Kandahar by +3% using DMC imagery between 2007 and 2009. Upsampling during FCN classification is more efficient than a chip-level sliding window required for pixel-level CNN classification. FCN architectures are recommended for pixel-level land cover classification based on their superior performance and ease of prediction.

Localised IoU was developed as a new method to quantify and visualise the spatial distribution of classification error across agricultural land at different scales. Use of local accuracy methods provides further insight into the strengths and weaknesses of the classification at the object level. The FCN-8 performs well in dense agricultural areas with strong edges between the desert, but has difficulty classifying weak edges in sparse and small areas of agriculture. Classification of weak edge cases improves with increased spatial resolution, where the spectral confusion is reduced. Future studies are encouraged to use localised IoU to provide an evaluation of the spatial distribution of classification error beyond global accuracy metrics.

Textural and spectral properties are important for agricultural land classification with FCN-8 (72% and 87%, respectively). Whereas, the shape of agriculture was not an important cue for FCN learning (53%). The importance of the spectral image features emphasised the requirement for standardisation between image datasets for generalised FCN-8 training. Matching image datasets using IR-MAD marginally improved classification accuracy (94.49%) in comparison to Level-1A (+0.10%). The additional pre-processing required for IR-MAD is appropriate for accurate agricultural area estimates, especially for use between image sensors, with tighter classification for higher resolution imagery. Standardisation of time-series data allows for classification of agricultural land at any time in the opium growth season and provides early indication of total agricultural area.

Without any additional training data, the generalised FCN-8 model trained on six years of IR-MAD data achieved high classification performance (>95%) for agricultural land discrimination on Sentinel-2 imagery. Further fine-tuning with

either dense or sparsely labelled datasets targeting new agricultural areas improved classification accuracy between image years (2016-2017) by up to 3%. This demonstrates the potential for reducing the effort required for updating these models and would increase the efficiency of agricultural mask production. Transferring knowledge between image sensors and across years opens up an exciting opportunity to automate monitoring agricultural change from new and long-term Earth observation programmes.

Chapter 4. Exploiting generalised FCNs to understand agricultural land change related to opium cultivation in Helmand, Afghanistan

This chapter presents a case study for using generalised FCNs trained on historical data from Chapter 3 to monitor changes in agricultural land in Helmand, Afghanistan between 2010 and 2019 across image datasets (Landsat and Sentinel-2). This work demonstrates the potential of using these models to understand the relationship between expansion in agricultural land and opium cultivation (thesis objective 3).

Underlying data for this chapter is available at:
https://doi.org/10.17862/cranfield.rd.14447400

4.1 Introduction

Opium cultivation in Afghanistan increased by 33% between 2010 and 2019 leading to increased instability, insurgency, environmental damage from overexploitation of land and funding to terrorist organisations (UNODC, 2021a). The main drivers of recent opium cultivation include the introduction of new technologies for agricultural practice, high opium yields, lack of government control and poor security (Mansfield, 2019; UNODC, 2021a). Farmers illegally cultivate opium poppy because it is a lucrative cash crop, in comparison to other alternative crop production (e.g. wheat, onion and vegetables) (UNODC, 2019a). The higher farm-gate price of opium poppy provides farmers with the ability to purchase new technology to increase agricultural expansion (e.g. solar-powered pumps and fertiliser) (Mansfield, 2019). Technical support has been given to farmers through the World Bank's National Horticulture and Livestock Productivity Project (2013-2020) (The World Bank, 2018) and has inadvertently promoted adoption for use in opium cultivation.

Helmand is the largest opium producing province in Afghanistan and was responsible for 73% of opium cultivation in 2019 (UNODC, 2021a). Counter-narcotic programmes were introduced into the province to reduce cultivation, such as the Helmand Food Zone (HFZ) (2008-2012) which highlighted the existing ban on opium cultivation and encouraged farmers to grow alternative crops (Mansfield, 2018). Despite these efforts, cultivation in Helmand continued to increase, reaching a record high of 144,019 ha in 2017. Since then cultivation has decreased to 136,798 ha in 2018 and 90,727 ha in 2019 (UNODC, 2021a).

The relationship between the area of agricultural land and opium cultivation are important for understanding trends in agricultural expansion and reduction across Helmand Province. There are disagreements in the drivers for opium cultivation expansion in relation to the total agricultural land. UNODC (2018b) suggested the increase in opium cultivation was attributed to farmers shifting arable land to poppy, as agricultural area remained consistent whilst opium cultivation increased. Mansfield (2019) has suggested new areas of agricultural land was

the driving cause of cultivation increase in southern Afghanistan. Primarily, low opium yields are driving farmers to relocate from canal-irrigated areas to former desert areas for higher opium yields (Mansfield, 2017), especially surrounding the Boghra canal (Mansfield, 2019). Monitoring these shifts in active agricultural land with generalised FCN-8 classifications would provide further understanding of opium cultivation and its drivers for expansion.

The aim of this chapter is to investigate the change in agricultural land using generalised models (described in Chapter 3) to understand the relationship between new areas of agriculture outside the main Helmand valley and opium cultivation. The main objectives of this chapter are to: (1) demonstrate the use of trained FCN-8 models for automating agricultural mask production across time-series datasets, (2) identify trends in agricultural area change between 2010 and 2019 in Helmand and (3) evaluate how land use change is driven by opium poppy cultivation throughout Helmand Province.

4.2 Materials and methods

The generalised FCN-8 models from Chapter 3 were used to classify agricultural land in Helmand between 2010 and 2019 to assess changes in the agricultural extent at the province and district level (Figure 4-1). These models have been trained using transfer learning across image datasets between 2009 and 2017 (Disaster Monitoring Constellation (DMC) and Landsat-8) (Table 4-1). The same methodology found in §3.3.5 was followed to identify and standardise cloud-free imagery across Helmand province. No suitable imagery was available over Helmand for 2012 because of the limited acquisitions during the decommission period of Landsat-5, therefore is omitted from agricultural land estimation. Sentinel-2 imagery was resampled to the same resolution as Landsat imagery (30 m) for consistency in reporting total agricultural area and comparison over the time-series.

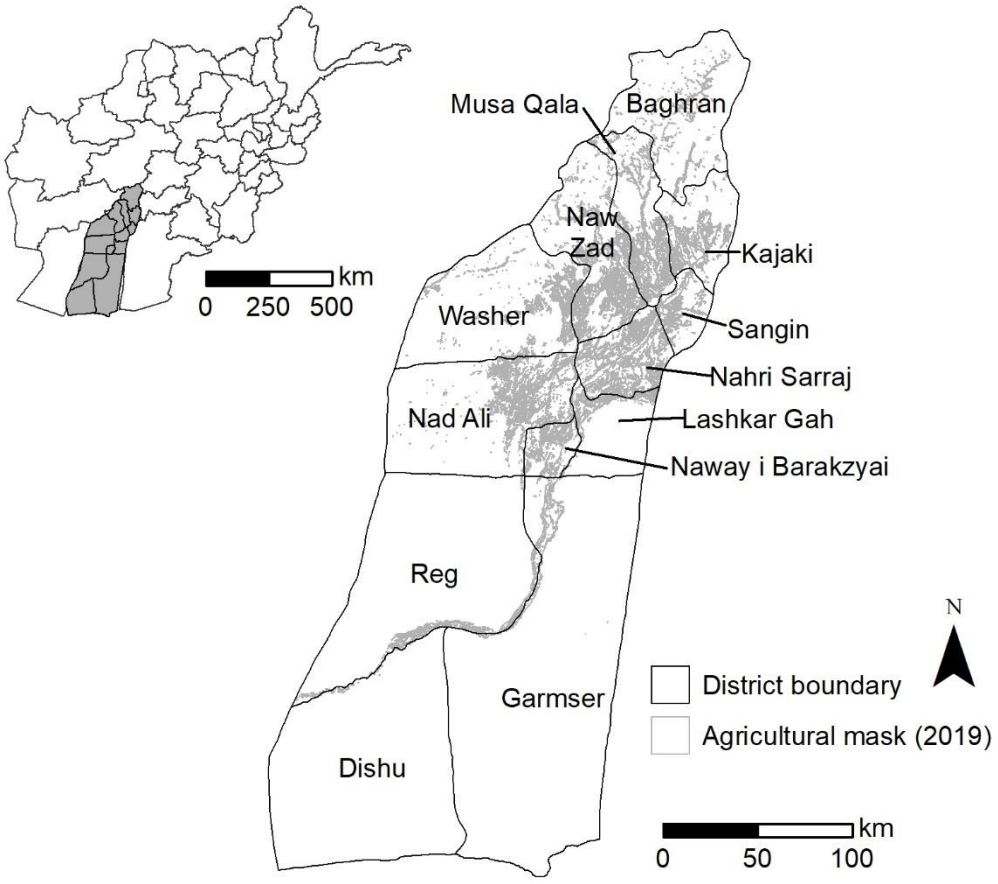

Figure 4-1 Location of districts in Helmand Province, Afghanistan with the active agricultural mask for 2019

Table 4-1. Image collections for assessing land cover change across Helmand between 2010 and 2019 with training data used for transfer learning FCN-8 models from Chapter 4 for active agricultural mask classification

Year	Image source	Image dates	FCN-8 model	Training data
2010	Landsat-5	8 Mar & 2 Apr	2009	2007-2009 DMC
2011	Landsat-5	20 Mar & 12 Apr	2009	2007-2009 DMC
2013	Landsat-8	10 Apr & 26 Apr	2009	2007-2009 DMC
2014	Landsat-8	3 Mar & 28 Mar	2009	2007-2009 DMC
2015	Landsat-8	7 Apr & 16 Apr	2015	2007-2009 DMC, 2015 Landsat-8
2016	Landsat-8	24 Mar & 18 Apr	2016	2007-2009 DMC, 2015-2016 Landsat-8
2017	Sentinel-2	8 Apr & 15 Apr	2017	2007-2009 DMC, 2015-2017 Landsat-8
2018	Sentinel-2	31 Mar & 3 Apr	2017	2007-2009 DMC, 2015-2017 Landsat-8
2019	Sentinel-2	29 Mar & 5 Apr	2017	2007-2009 DMC, 2015-2017 Landsat-8

4.3 Results

4.3.1 Changes in agricultural land across Helmand Province between 2010 and 2019

Using generalised FCN-8 models for automated agricultural land classification, Helmand Province has seen an increase in agricultural area between 2010 and 2019 with the largest agricultural area in 2019 (460,735 ha) (Figure 4-2). The districts with the highest agricultural area from 2019 are Nad Ali (85,928 ha), Nahri Sarraj (73,396 ha) and Naw Zad (65,643 ha) (Table 4-2). Total agricultural area reduced in 2011 and can be attributed to both the enforcement of the HFZ, which also lead to a reduction in opium cultivated (UNODC, 2011), and low yields from the main cash crop wheat (Mansfield, 2011).

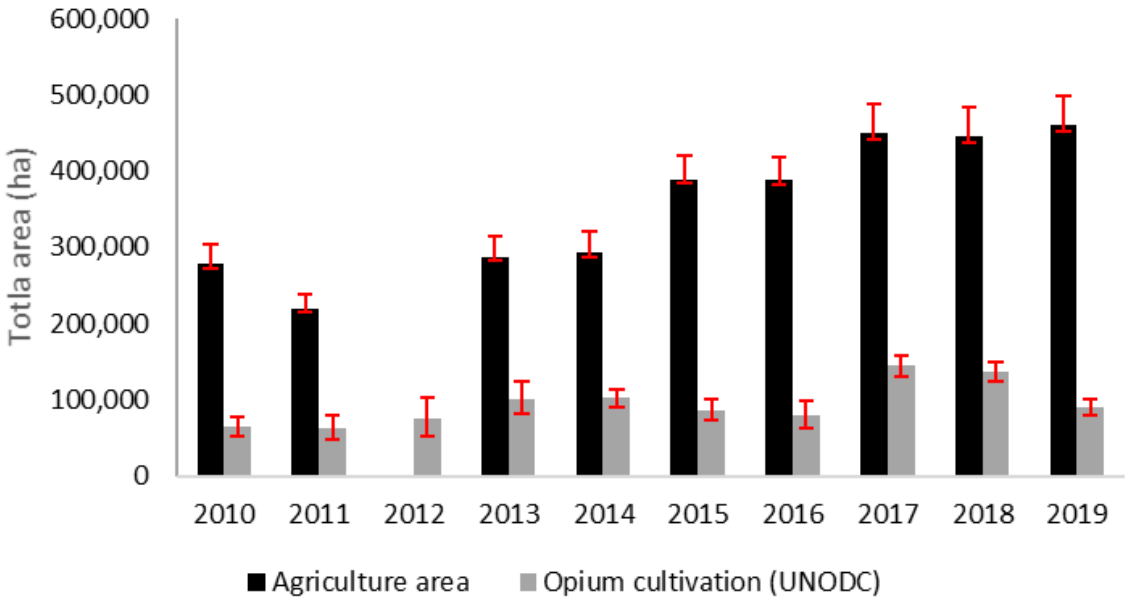

Figure 4-2. Annual active agricultural area derived from FCN classifications and UNODC opium cultivation estimates from annual UNODC opium cultivation reports for Helmand province, Afghanistan between 2010 and 2019. The vertical lines represent the upper and lower bounds of the 95% prediction intervals

Table 4-2. Annual agricultural area for Helmand province, Afghanistan between 2010 and 2019.

District	Agricultural area (ha)								
	2010	2011	2013	2014	2015	2016	2017	2018	2019
Baghran	10,890	7,649	10,180	10,837	12,082	11,821	9,691	8,694	8,773
Dishu	1,322	2,758	5,840	5,201	5,855	5,977	7,166	5,824	5,624
Garmser	22,921	18,509	25,412	20,223	27,633	27,203	28,121	26,148	24,214
Kajaki	16,666	12,686	13,430	29,314	29,870	29,971	35,204	34,286	37,245
Lashkar Gah	14,468	10,668	15,122	13,061	20,172	16,123	19,853	18,427	19,671
Musa Qala	22,378	15,317	24,029	27,453	31,089	33,450	39,015	41,107	46,550
Nad Ali	49,522	42,599	48,809	43,391	69,421	68,783	78,827	77,891	85,928
Nahri Sarraj	44,631	33,872	45,395	50,748	63,416	59,978	69,751	68,867	73,396
Nawzad	25,133	14,939	28,169	28,421	34,209	43,454	56,074	59,856	65,643
Naway-i-Barakzai	36,148	30,703	32,531	31,423	43,170	39,431	42,992	39,597	36,390
Reg	20,523	18,204	24,678	15,960	32,039	30,201	33,875	31,817	18,631
Sangin	8,677	7,097	7,883	10,780	11,024	11,804	13,972	13,503	14,463
Washer	4,373	3,792	6,133	6,474	8,843	10,053	15,846	20,462	24,207
Helmand	**277,652**	**218,793**	**287,611**	**293,286**	**388,823**	**388,249**	**450,387**	**446,479**	**460,735**

The overall opium cultivation reported by UNODC between the same time period saw a steady increase overall, but a small reduction occurred between 2014 and 2015 (-16%) (UNODC, 2021a). There was an increase in opium poppy cultivation between 2016 and 2017 (+79%), where there is also an increase in the amount of agricultural land across the province. UNODC (2018a) attributed the rise to an increase in the overall area under cultivation and high opium yield. Opium cultivation has begun to decline from after 2017 to similar levels found in 2016, whilst the agricultural area has continued to rise.

4.3.2 Changes in available agricultural land for districts across Helmand Province

The individual districts within Helmand province are presented as groups based on their geographic location. The total agricultural area derived from the classification and opium cultivation from the UNODC statistics decrease between 2010 and 2019 in Baghran (Figure 4-3) with limited expansion of agriculture into new areas from 2010 (Figure 4-4). The agricultural area and opium cultivation has increased in Kajaki and Musa Qala (Figure 4-3), predominately surrounding the Kajaki reservoir and Helmand river (Figure 4-4 (b)) and into desert areas after 2017 in south-west Musa Qala (Figure 4-4 (a)).

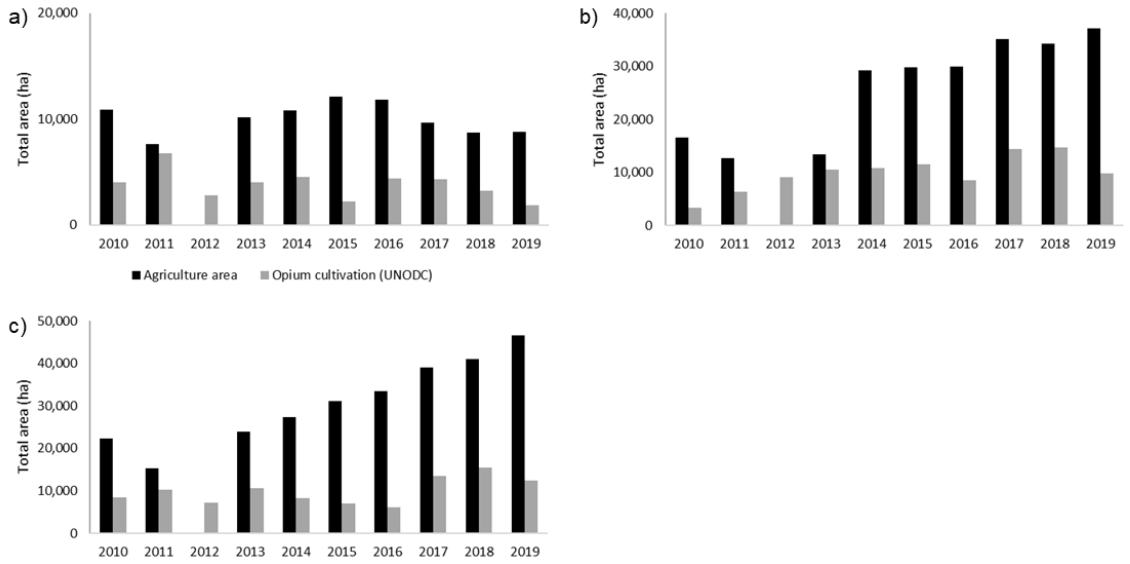

Figure 4-3 Annual agricultural area and UNODC opium cultivation estimates for northern districts a) Baghran, b) Kajaki and c) Musa Qala. Agricultural area estimates are derived from FCN-8 classifications of active agricultural masks across Helmand, Afghanistan

The total agricultural area has increased substantially in Naw Zad and Washer, particularly from 2017 (Figure 4-5), whilst opium cultivation has increased over the same period the increase in available land is greater. The areas of expansion are predominately expansion from existing areas of agriculture in Naw Zad with additional areas in desert locations (Figure 4-6 (a)). The expansion is also in existing areas of agriculture in Washer, but larger areas of new agriculture have been identified in desert areas in the south-east (Figure 4-6 (b)).

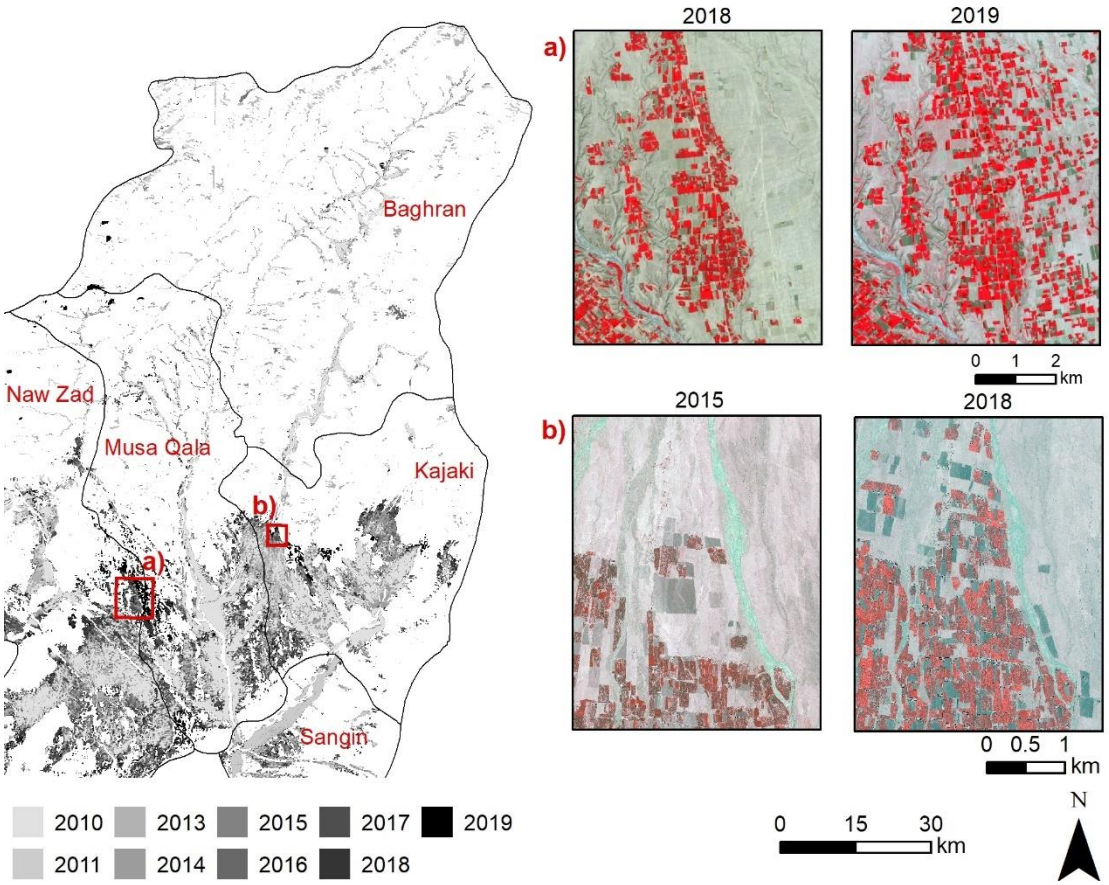

Figure 4-4. Annual areas of active agriculture in Baghran, Kajaki and Musa Qala between 2010 and 2019. Newer areas are denoted as darker colours. False colour imagery (NIR, R, G): a) 2018 and 2019 Sentinel-2 imagery (10 m) and b) 2015 and 2018 pan-sharpened Pleiades imagery (0.5 m)

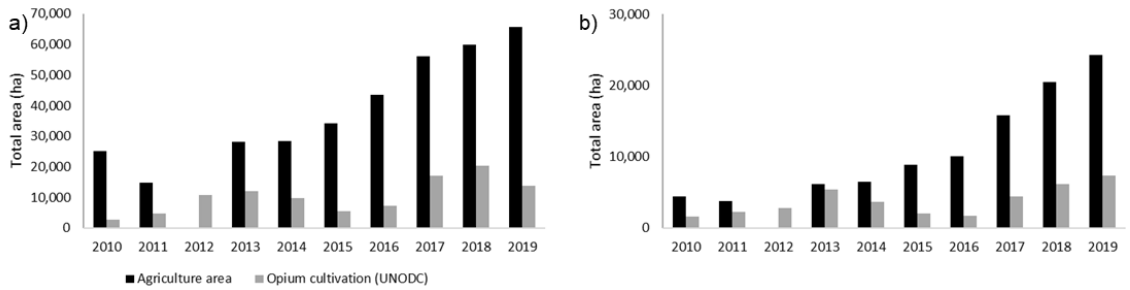

Figure 4-5 Annual agricultural area and UNODC opium cultivation estimates for north-western districts a) Naw Zad and b) Washer. Agricultural area estimates are derived from FCN-8 classifications of active agricultural masks across Helmand, Afghanistan

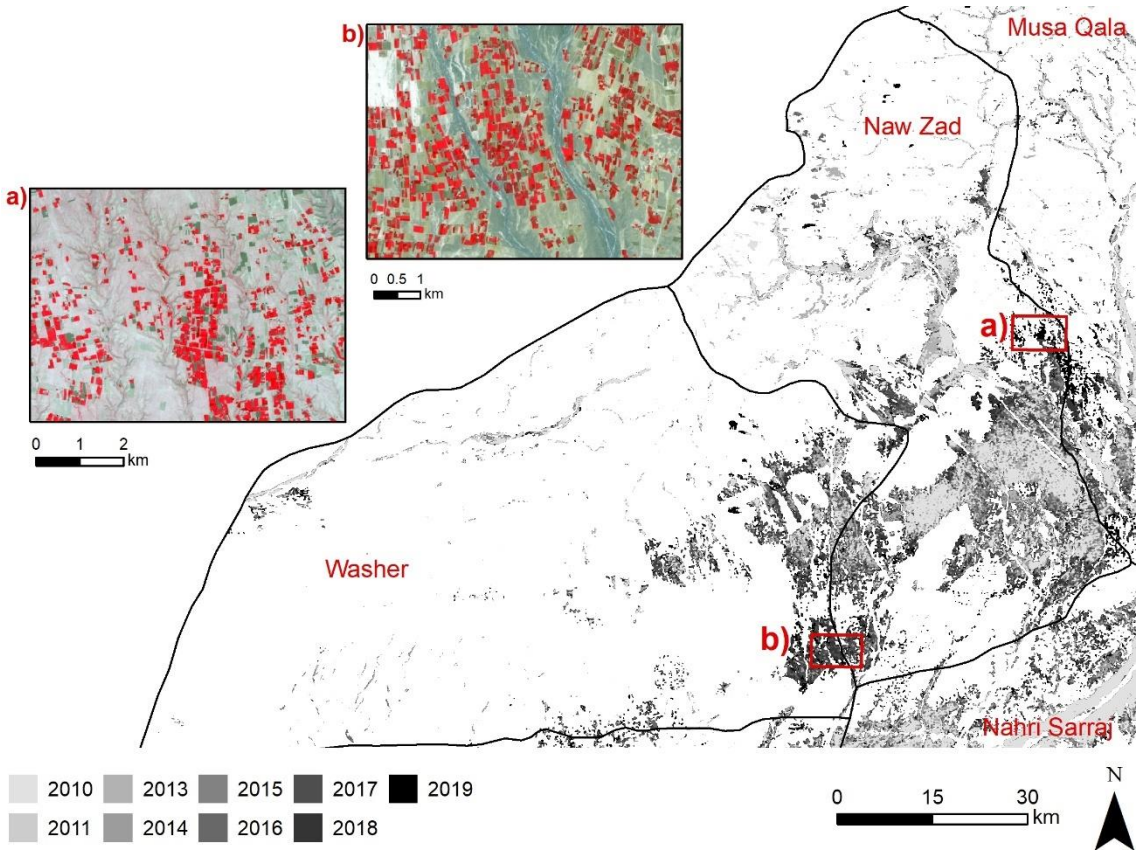

Figure 4-6. Annual areas of active agriculture in Naw Zad and Washer between 2010 and 2019. Newer areas are denoted as darker colours. False colour imagery (NIR, R, G at 10 m): a) 2019 Sentinel-2 and b) 2018 Sentinel-2

The total agricultural area in Nad Ali has increased over the study period, particularly from 2015 onwards (Figure 4-7 (a)), whilst opium cultivation also steadily increased. The main areas of expansion are from existing areas of agriculture (Figure 4-8 (b)) and in desert areas surrounding the Boghra canal and away from the Helmand river (Figure 4-8 (a) and (c)). Naway-i-Barakzai has low opium cultivation compared to other districts in Helmand with little expansion over the study period (Figure 4-7 (b)). There has been little expansion in the agricultural area identified for Naway-i-Barakzai with these areas of expansion extending off the main area of agriculture surrounding the Helmand river (Figure 4-8).

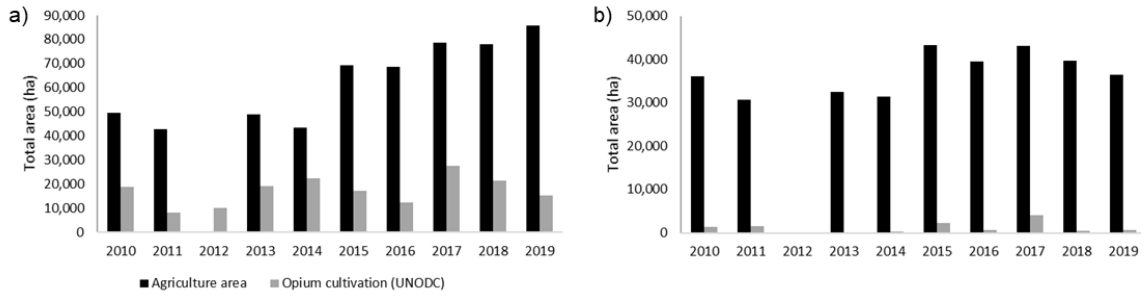

Figure 4-7 Annual agricultural area and UNODC opium cultivation estimates for central districts a) Nad Ali and b) Naway-i-Barakzai. Agricultural area estimates are derived from FCN-8 classifications of active agricultural masks across Helmand, Afghanistan

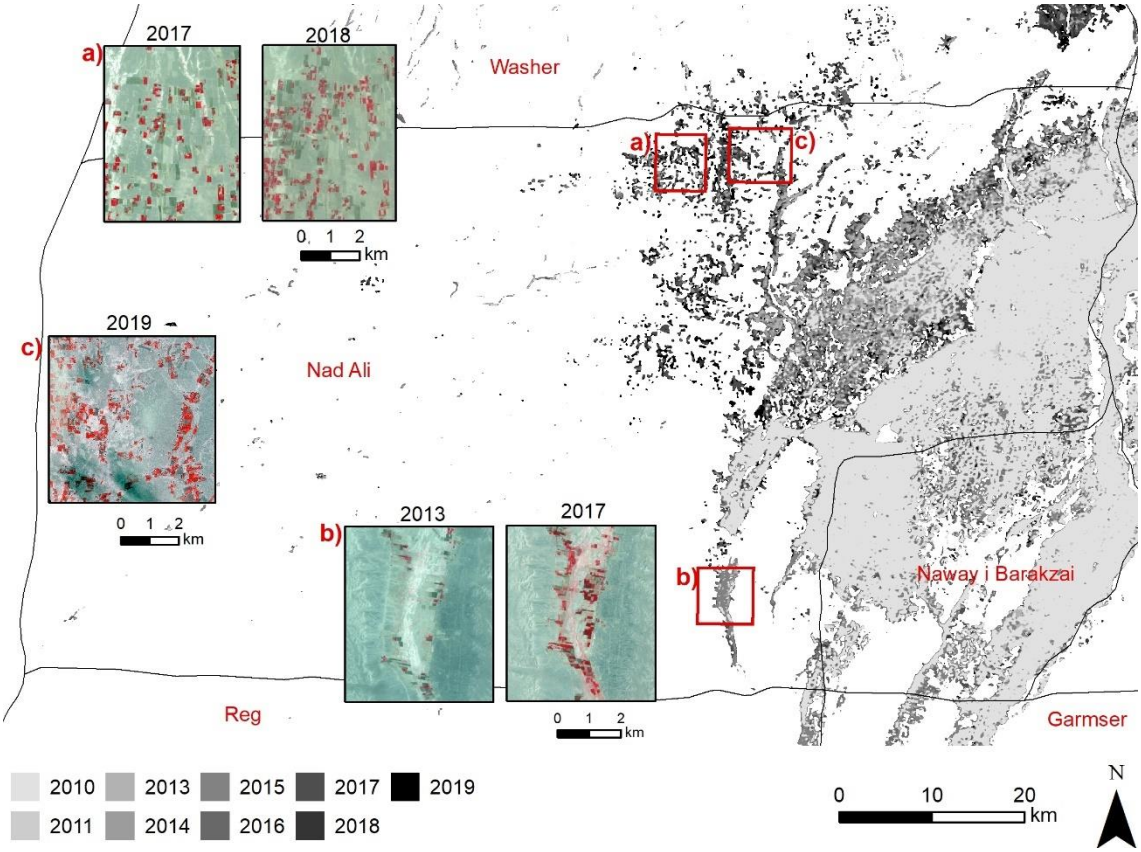

Figure 4-8. Annual areas of active agriculture in Nad Ali and Naway-i-Barakzai between 2010 and 2019. Newer areas are denoted as darker colours. False colour imagery (NIR, R, G): a) 2017 and 2018 Sentinel-2 (10 m), b) 2013 Landsat-8 (30 m) and 2017 Sentinel-2 (10 m) and c) 2019 Pleaides imagery (0.5 m)

The agricultural areas in Lashkar Gah, Nahri Sarraj and Sangin have steadily increased between 2010 and 2019 with opium cultivation comprising a low

proportion of the available agricultural land (Figure 4-9). The primary areas of expansion are off the main agricultural block surrounding the Helmand river (Figure 4-10 (a) and (c)) and the majority of new areas are cultivated on the southern boundary with Naw Zad (Figure 4-10 (b)).

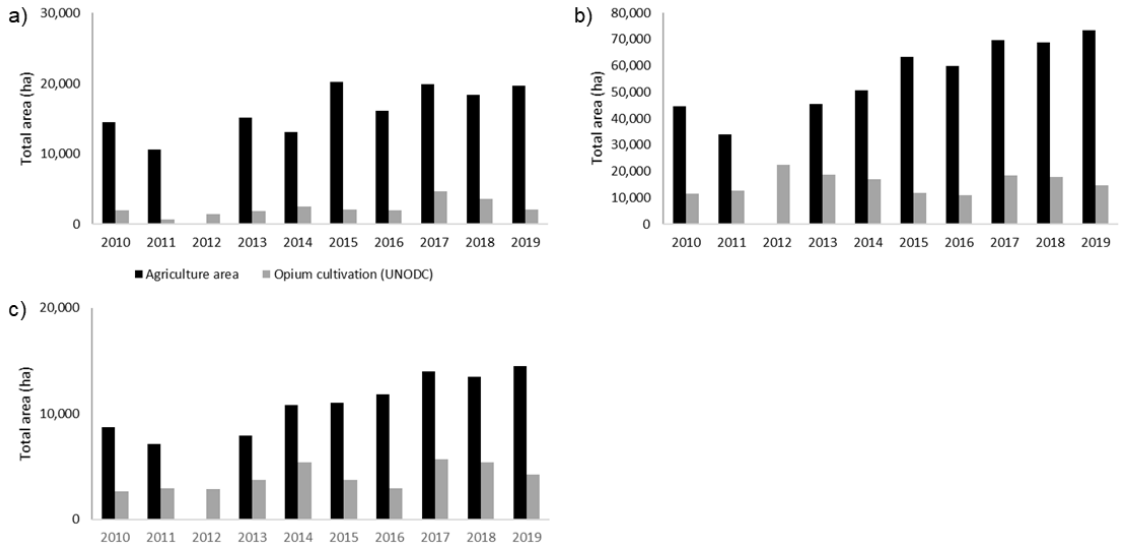

Figure 4-9 Annual agricultural area and UNODC opium cultivation estimates for eastern districts a) Lashkar Gah, b) Nahri Sarraj and c) Sangin. Agricultural area estimates are derived from FCN-8 classifications of active agricultural masks across Helmand, Afghanistan

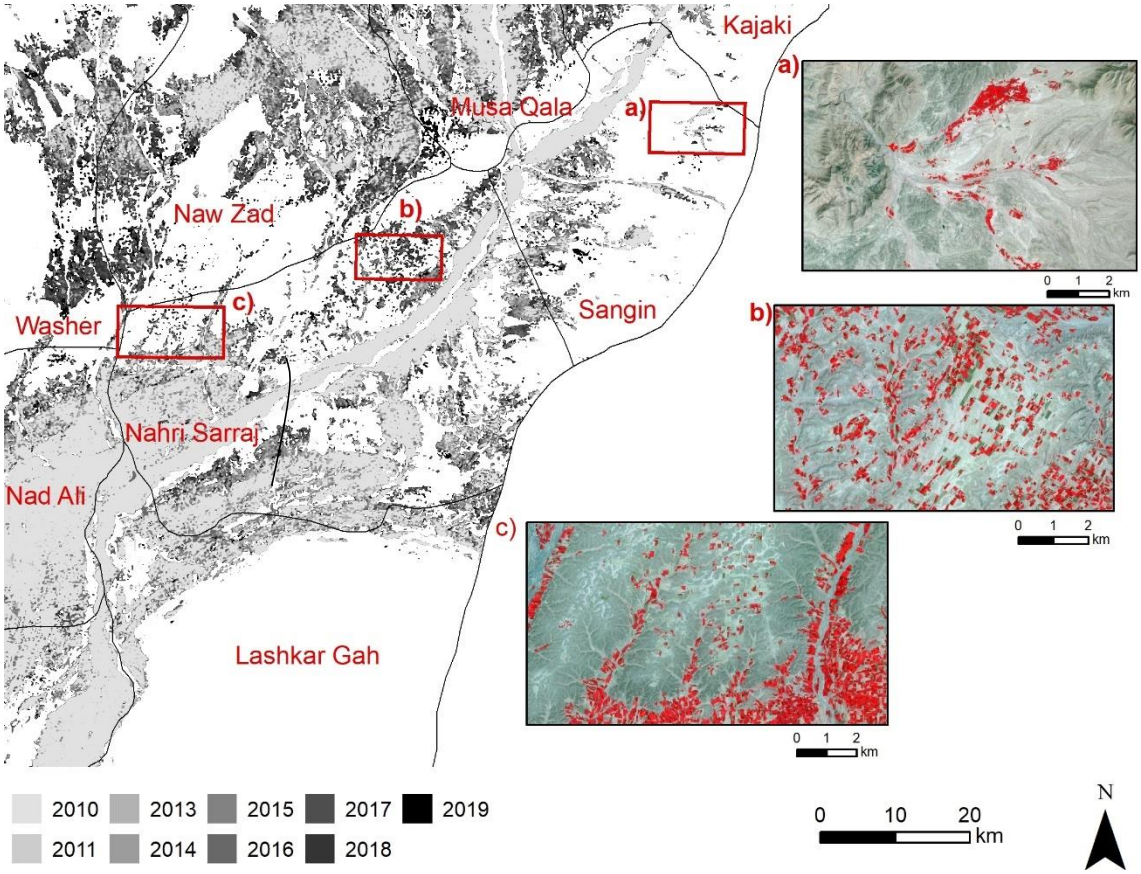

Figure 4-10. Annual areas of active agriculture in Lashkar Gah, Nahri Sarraj and Sangin between 2010 and 2019. Newer areas are denoted as darker colours. False colour imagery (NIR, R, G at 10 m): a) 2017 Sentinel-2 and b) 2019 Sentinel-2 and c) 2019 Sentinel-2

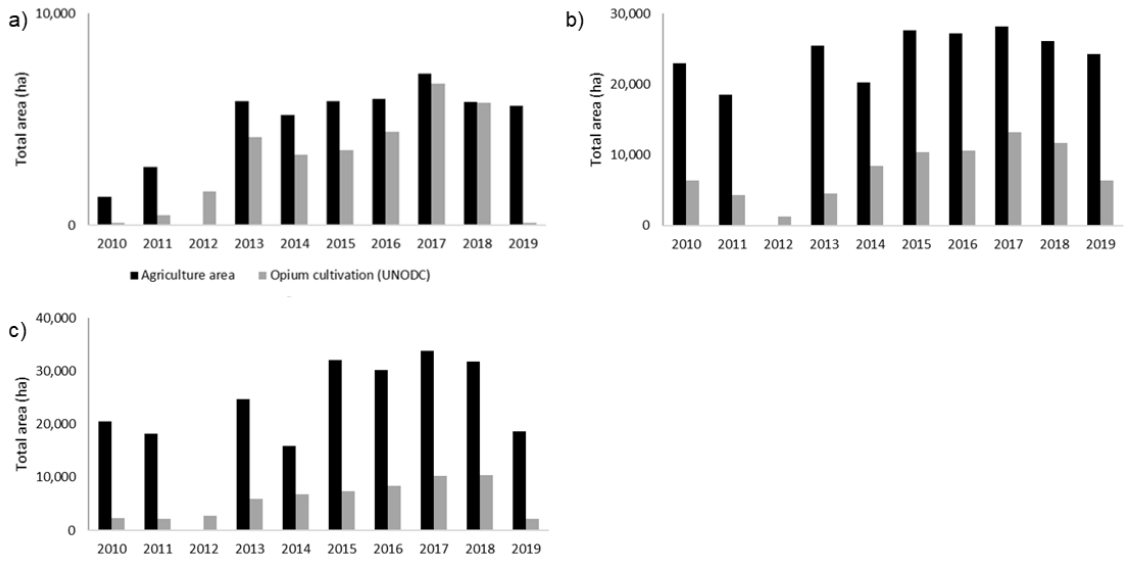

Figure 4-11 Annual agricultural area and UNODC opium cultivation estimates for southern districts a) Dishu, b) Garmser and c) Reg. Agricultural area estimates are derived from FCN-8 classifications of active agricultural masks across Helmand, Afghanistan

Between 2010 and 2019 there has been an increase in the total agricultural area and opium cultivation in Dishu, Garmser and Reg (Figure 4-11). The proportion of opium cultivation to available agricultural land identified is high in Dishu in comparison to the other southern districts. The main areas of expansion are from the existing continuous area of agriculture surrounding the Helmand river (Figure 4-12 (a)), predominately in between Reg and Dishu (Figure 4-12 (b)), with smaller areas detected in marginal desert areas (Figure 4-12 (c)).

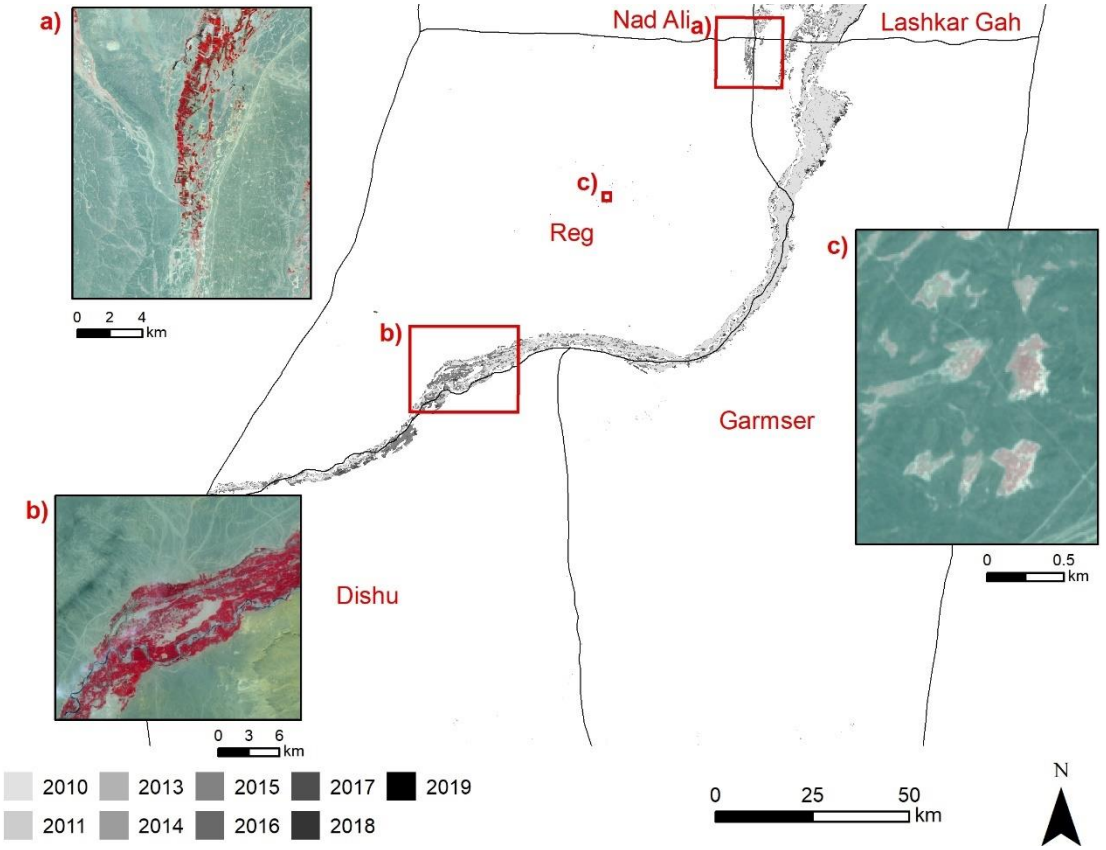

Figure 4-12. Annual areas of active agriculture in Dishu, Garmser and Reg between 2010 and 2019. Newer areas are denoted as darker colours. False colour imagery (NIR, R, G at 10 m): a) 2017 Sentinel-2 and b) 2018 Sentinel-2 and c) 2017 Sentinel-2

4.4 Discussion

4.4.1 Ability to monitor agricultural land changes with an active mask

In 2016, 21% of agricultural land was used for opium cultivation in Helmand and increases further to 32% in 2017, which are similarly reported by UNODC, 20% and 33% respectively (UNODC, 2018a). However, the total agricultural land in Helmand remains unchanged across Helmand in UNODC estimates between 2015 and 2018 (401,300 ha) (UNODC, 2015, 2016, 2017b, 2018b). This analysis has identified year-on-year expansion of arable land during the same time period (Figure 4-2). The limitation of UNODC using the potential mask for their survey is only new areas of agriculture will be added to their area under potential cultivation

and omit annual rotations of agricultural areas between years. Using the potential mask will overestimate agricultural land as it is not identifying only active agriculture from each target year, but ensures no areas are omitted from the sampling frame in the opium survey. The ability to train generalised models for application on any imagery provides further insight into the expansion of agricultural land, which the UN estimates currently are unable to do.

4.4.2 Security and enforcement driving agricultural land change in Helmand

Political instability and enforcement across opium producing villages are cited as a primary driver for opium cultivation expansion (UNODC, 2021a). The main driver for low opium production over 2011 was the successful counter-narcotic efforts of the HFZ, which was cited as the main reason for farmers not cultivating opium (UNODC, 2011), and resulted in a decrease in agricultural area across Helmand. The HFZ programme stopped at the end of the 2012 season (Mansfield, 2018), which led to an increase in opium cultivation and total agricultural area in 2013. These increases are reflected in a rise of the total agricultural area in Helmand, particularly in southern districts such as Dishu (Figure 4-11) where the proportion used for opium cultivation is high from 2013 suggesting expansion is driven by opium cultivation after the enforcement of the HFZ.

The increase in agricultural area has previously been attributed to farmers relocating away from canal areas to former desert areas as counter-narcotic activities had reduced in these area (Mansfield, 2018). The main drivers for this decision are these areas are more difficult to access and control and are attractive to opium farmers. FCN-8 classifications have provided further insight into expansion above the Boghra canal in Nad Ali (see Appendix C), which saw a dramatic increase in agricultural area after the enforcement of the HFZ had ceased.

4.4.3 Water availability and increasing access to technology driving agricultural land expansion in Helmand

UNODC (2017) stated opium cultivation increases across Helmand have largely been driven from converting existing agricultural area to opium poppy and increased yield, particularly the increase between 2016-2017 (Figure 4-2). Approximately one third of all available agricultural land had been used for opium cultivation in 2017 (UNODC, 2018a). The FCN-8 classifications found increases in agricultural land is primarily found away from larger blocks of existing agriculture after 2013 and 2014, which would be unsuitable without access to additional means of irrigation. This is providing additional insight into agricultural change in Helmand Province and supports findings in Mansfield (2019) for using new technologies to convert desert areas into arable land through increasing access to irrigation.

The main year of agricultural area increase for most districts was in 2015 (+32%), which occurred during the widespread adoption of solar powered deep wells for crop irrigation because of falling water tables and the increased use of fertilisers and pesticides (Mansfield, 2019). The World Bank Afghanistan Horticulture and Livestock Productivity Product supported farmers to improve their uptake of these new technologies to increase agricultural productivity (The World Bank, 2018). Opium cultivation did not increase at the same rate as the introduction of new areas of agriculture for all districts and suggests these new areas are used for other agricultural purposes. The increase in agricultural land across Helmand in 2017 and 2018 has also been attributed to the adoption of solar power for irrigation and use of fertilisers to convert desert areas to arable land (UNODC, 2017b, 2018b), in addition to the high yields of opium which increased farm-gate prices.

Rapid expansion in opium cultivation in Nad Ali has been driven by better irrigation practices by using water from the canal system and utilising pumped water from deep wells (UNODC, 2017a). The success of the wider adoption of utilising new technologies for increasing water security was attributed to alleviating the impact of 2018’s widespread drought in Afghanistan to Helmand

Province in comparison to other southern provinces (e.g. Nimroz) (UNODC, 2019a). However, the increasing adoption of irrigation from deep wells has increased concerns over the future impact to agricultural practices caused by further lowering of the water table (Mansfield, 2019). Districts surrounded by the Helmand river or in areas unsuitable for agricultural expansion have no need for access to these technologies and have resulted in consistent agricultural areas and opium cultivation estimates. All districts in Helmand Province had an increase in total agricultural area over the study period, apart from Baghran, which can be attributed to its limited access to water and areas for new agricultural areas because of its mountainous terrain. Naway-i-Barakzai is surrounded by the Helmand river and is another example of not needing access to these new technologies because of the limited available areas for agricultural expansion.

4.4.4 Socio-economic drivers for agricultural land change in Helmand

Opium cultivation fell in 2019 across many districts in Helmand whilst the agricultural area increased (Table 4-2), particularly in Dishu (Figure 4-11), suggesting a transition to other crops. Opium production had been high between 2017-2019 which saturated the market for opium and led to falling prices and impacted farmers cropping decisions with some transitioning to licit crops such as wheat (UNODC, 2021a). An increase in wheat production led to a surplus in Helmand Province because of increases in wheat's farm-gate pricing, even despite low rainfall (NSIA, 2019). Farmers were found to not cultivate every year and the decrease in farm-gate prices for opium would have discouraged impoverished farmers from cultivating opium. Non-poppy farmers cultivating cannabis earned four times more than poppy farmers (UNODC, 2021a) and rapid increases in these price changes since 2018 resulted in an increase in cannabis cultivation in 2019 (Bradford and Mansfield, 2019).

4.5 Conclusions and recommendations

Automated agricultural land classification from generalised FCN-8 models, trained on legacy data, demonstrates the potential for utilising these models for monitoring land cover changes across image datasets. The main advantage in

comparison to other techniques is the ability to classify image datasets that were different to those used during training. The agricultural area has increased year-on-year whilst the proportion of opium cultivation has remained static between 2010 and 2016. The increase in agricultural land from 2017 and 2018 increased opium cultivation, but cultivation decreased to prior levels in 2019 suggesting a shift to alternative agricultural crops.

Trends in the changes to agricultural land alongside existing literature provided an insight into the main drivers of agricultural land expansion and its relationship with opium cultivation. Agricultural expansion in Helmand Province has been largely driven by shifting agricultural practices from canal-irrigated areas and areas away from the main Helmand valley to marginally desert areas to increase the potential land available for opium cultivation (e.g. Nad Ali, Naw Zad and Washer). The adoption of new technologies for increasing water security was one of the main drivers for agricultural expansion into otherwise inarable land, including solar-powered deep wells, fertilisers, and pesticides. Districts along the main Helmand valley have witnessed less agricultural expansion because of limited areas for expansion and better access and control by counter-narcotic efforts. The case study presented exemplifies the ease of utilising FCNs across large volumes of image datasets to derive new information about past and on-going changes to agricultural practices to support national and international stakeholders.

Chapter 5. Discussion

This chapter contains the discussion of the research on convolutional neural networks for automated agricultural land classification in the preceding chapters. The main findings from this thesis are discussed and consider the contributions to knowledge and wider impact of the research.

5.1 Introduction

The increasing volume, velocity, veracity, and variety of satellite imagery requires new data-driven methodologies to derive new knowledge. This thesis has developed generalised models for agricultural land classification using Fully Convolutional Networks (FCNs) trained with historical datasets. The FCN models were used across images from different years and different sensors to map agricultural land change across Helmand Province, Afghanistan.

5.2 Utilising generalised CNNs to advance land cover classification

Convolutional Neural Networks (CNNs) outperformed other machine learning techniques (Artificial Neural Networks and Random Forest) in separating agricultural land by encoding the surrounding landscape from each image during model training. Similarities can be drawn to the human interpreter who manually edits the agricultural area from unsupervised classification using contextual information (UNODC, 2018b). In pixel-based approaches elevation data is recommended to improve overall accuracy (Khatami, Mountrakis and Stehman, 2016). The importance of contextual spectral information was demonstrated when adding height information had negligible difference to the overall accuracy (-0.63%). The highland and lowland areas are visually different, and the CNN is able to understand this from the spatial and spectral information encoded in each chip, whereas pixel observations require explicit height information. The limitation of using CNN architectures with fully-connected layers (e.g. ResNet50) for pixelwise classification is the need for a sliding window, which caused generalisation at edge cases caused by too much global contextual information. The FCN-8 architecture is recommended for future implementation in remote sensing studies, which is able to retain the native resolution of inputs during classification by upsampling convolutional layers.

The main limitation with using CNNs are without labelled datasets from manually edited classifications of agricultural land training these models would not have

been possible. Real-world labelled datasets are scarce in the remote sensing discipline and CNN models are commonly evaluated on benchmark datasets (e.g. UC Merced (Yang and Newsam, 2010) and AID (Xia *et al.*, 2017)). Limited access to real world datasets stresses the need for appropriate training strategies to best utilise all available data. The optimal training strategy for agricultural land classification with CNNs used samples containing edge cases and natural vegetation. The targeted strategy outperformed other random sampling strategies because it forces the CNN to learn from the most difficult cases rather than continuous areas of agriculture or desert areas. The limiting factor is legacy labelled datasets may not have the necessary data to sample from difficult interpretation locations if it is outside of their sampling frame (e.g. natural vegetation). Models performed better when trained with samples in areas of known confusion caused by natural vegetation and boundary areas. These locations would also be difficult for untrained human interpreters, but their interpretations would improve with more exposure to these areas.

Transfer learning alleviates the need for large amounts of labelled data by using previous knowledge of image features from trained CNNs during model initialisation. Transfer learning, which is unique to deep learning, is essential for continual improvement of CNN models and speeds up learning compared to random model initialisation (Nogueira, Penatti and dos Santos, 2017). Transfer training models on ImageNet, an unrelated image dataset, achieved higher performance than training from scratch with remote sensing data (Figure 2-4). Future studies are recommended to use previous model initialisations to improve classification performance, especially with limited access to training data and enhance the generalisation of models. CNNs are already noted for their ability to generalise well, even for applications they were not originally trained for (Penatti, Nogueira and Santos, 2015). Improved year-on-year performance was also achieved using transfer learning from previous years' models on sparse datasets targeting new areas of agriculture (Table 3-6). Over time less data are needed (e.g. 25% of the total dataset) to update CNN models, which demonstrates image features are transferable between images with fewer training samples needed for fine-tuning. FCNs have demonstrated with transfer learning they can train on

sparsely labelled datasets and further improve classification performance. This is important in the remote sensing discipline, where limited labelled datasets have previously been a barrier for using deep learning techniques (Ball, Anderson and Chan, 2017).

The lack of labelled data for training supervised classification models emphasises the importance of training models across sensors to use a wider range of legacy datasets. Training across image datasets is difficult because of the temporal, atmospheric, radiometric, and spatial differences and are a barrier for supervised learning (Song *et al.*, 2019). Ensemble models have previously been used to overcome the differences between image sensors (Lei *et al.*, 2020). An FCN model has demonstrated high classification performance across image years by encoding localised features during training without the need for complex ensemble models.

FCNs encode spatial, spectral, and textural features during training and understanding the importance of these features aids the sampling design for training between image sensors. Isolating image features is difficult for the FCN model as they use convolutional filters rather than isolated observations. Artificial samples were used to invert the spectral information between agriculture and non-agriculture classes to force the FCN to rely on the shape of agricultural area. The spectral information cannot be separated from texture because of the encoded localised patterns from convolutional filters. Investigation into the isolated image features found that textural and spectral features were important for agricultural land classification, 55% and 62% IoU respectively. The shape of blocks of agricultural land was not useful in the classification of agriculture (12% IoU). This is different to findings that found shape to be an important predictor for pixel-level classification using CNNs (Shelhamer, Long and Darrell, 2016). However, these studies use commercial photographs where the classes of objects are consistent (e.g. a cat or dog) unlike patterns in agricultural land, which do not have a recognisable shape. Texture is an important feature for both agricultural land classification and multi-class chip-level classification of commercial photography using CNNs (Baker *et al.*, 2018). Textural and spectral

image features are used to separate agriculture from natural vegetation and desert within the FCN. Understanding the important features of agricultural land to a FCN provides further insight into how best to train these model across image sensors for greater generalisation.

5.3 Agricultural mapping across Helmand province

Automated unsupervised classification of satellite imagery is unable to separate agriculture and non-agriculture from satellite imagery across Afghanistan (Taylor *et al.*, 2010). Time-consuming manual editing is required in areas of confusion (e.g. natural vegetation) to accurately map agricultural land. Generalised FCN models, trained using standardised image datasets, were used to successfully map agricultural land in Helmand Province between 2010 and 2019 (see Appendix C). These models, trained on legacy datasets, were used to classify image data for years where there are no training data available. The latest model for each image year was used for a simulated real-world application. FCNs have proven their ability to classify agricultural land by outperforming machine learning techniques and showing potential for replacing human interpreters for the UNODC's agricultural mask production. Sources of error occur predominately at weak edge cases (e.g. smaller agricultural areas and in larger areas of sparse agriculture) and achieved only +3.2% increase in total agricultural area on validation samples across image years. The only requirement to use these models was IR-MAD normalisation of input imagery, regardless of image sensor or image timing. IR-MAD normalisation introduces an additional pre-processing step but improves classification at difficult locations to minimise sources of error (e.g. edge cases).

UNODC (2017) stated increases in opium cultivation across Helmand have largely been driven by farmers converting existing agricultural area to opium poppy. UNODC use the potential mask for their survey where only new areas of agriculture are added annually, therefore retain areas out of rotation between years. Investigating annual changes in active agricultural land is now possible by comparing the automated year-on-year classifications from generalised models. The total agricultural land in Helmand remained static in UNODC estimates

between 2015 and 2018 (UNODC, 2015, 2016, 2017b, 2018b), but the automated classifications identified year-on-year expansion. No change in the UNODC estimates in agricultural land between 2015 and 2018 whilst opium cultivation increases suggests either no new areas outside of previous delineations have been added or the density of cultivation has increased.

The variation in agricultural land between 2010 and 2017 found the decreases in agricultural area across Helmand during 2011 are related to the enforcement of the Helmand Food Zone (HFZ). Opium cultivation and agricultural area began to increase after the counter-narcotic programme ceased at the end of 2012 (Figure 4-2). The main areas of expansion were found outside the main Helmand valley. This finding confirms that enforcement of the HFZ led farmers to cultivate in marginal areas with less counter-narcotic activities (Mansfield, 2019). Widening access to new technologies have been cited as a driving factor for the agricultural expansion outside of the main Helmand valley (Mansfield, 2019). The main area of expansion was found above the Boghra canal in Nad Ali (Figure 4-8), where increases are caused by farmers relocating during the HFZ enforcement to use canal irrigation and utilise solar powered deep wells and fertiliser (Mansfield, 2018).

Opium cultivation increases over the study period, but not at the same rate as the introduction of new areas of agriculture identified from classifications and suggests a proportion of these new areas are used for other agricultural purposes. In 2019, opium cultivation reduced across districts in Helmand whilst the agricultural area continued to increase suggesting more farmers transitioning to alternative crop production. Falling farm gate prices for opium caused by a saturated market from previous years of high yield are suggested as another reason for farmers switching to other crops (UNODC, 2021a).

Generalised models developed in this thesis demonstrate their potential scalability for real-world applications across remote sensing disciplines. They can classify up-to-date imagery in near-real time to understand changes in land cover without any manual editing and minimal fine-tuning. Without generalised models, classification between image sensors, whilst maintaining high accuracy (>95%),

would not be possible and all available legacy and future image datasets cannot be utilised.

The agricultural area increases during the season until peak opium biomass is reached, where agricultural land is at its maximum extent. Generalised FCN models were trained using imagery coinciding with the maximum agricultural extent, but standardisation allowed for classification at any point in the season (Figure 3-23). Differences between unstandardised image datasets have been cited as a major limitation for applying models across image datasets, which often only use few images to demonstrate their application (Cheng, Han and Lu, 2017; Song *et al.*, 2019). Northern Helmand had a later maximum agricultural extent than southern Helmand (Figure 3-22), as previously noted in Taylor *et al.* (2010). This thesis developed models that can classify at any point in the season to monitoring changes irrespective of image timing and provides an early indication of growing intentions before the maximum agricultural extent is reached. Training generalised models across image datasets is also more efficient than training multiple models for each image sensor which require large volumes of labelled data. Future studies are recommended to focus on image standardisation when combining labelled datasets across image sensors to improve classification performance.

5.4 Uncertainty from classification to area estimation

Global accuracy metrics are used across all remote sensing image classification because they provide common metrics to assess classification performance. The main limitation is the uncertainty of the labelled data as they are generated from another classification. However, these data have been manually inspected by trained interpreters to ensure they are representative to minimise error from incorrect label allocation. The evaluation of agricultural land classifications is difficult because of high global accuracy caused by the imbalance of labelled data (Douzas *et al.*, 2019; Feng *et al.*, 2019). For example, in Helmand and Kandahar only 5% of the total area is agriculture. Targeted sampling in areas of confusion (e.g. edge cases and natural vegetation) still had high overall accuracy as larger areas of agriculture are well-classified. The difficult areas of classification are

supressed in global metrics (e.g. edge cases and areas of confusion between agriculture and natural vegetation) because they are only a small proportion of the total area.

Localised IoU was developed to map the spatial distribution of error in agricultural land classification and provided further insight into model performance beyond global accuracy statistics. Separating the landscape based on each area of agriculture gives more information than other techniques to visualise local differences e.g. image differencing, which provides areas of omission and commission. The scale of the agricultural area is omitted in this technique, but localised IoU provides the end user with a corresponding IoU value for each agricultural area. This information can be used to identify common classification trends and further understand the limitations and uncertainties of the FCN.

The localised IoU found large blocks of agriculture in areas of high contrast had tight classification. Smaller and large blocks of agriculture with poor contrast were shown to have poor localised IoU. The limitation of the localised IoU method is its inability to consider the population of agriculture pixels for each object during calculation and leads to greater differences in IoU for smaller areas. Smaller areas have greater sensitivity to subtle differences between the reference and prediction values than larger objects. More differences are found in smaller areas as these relate to the strong contrasting edge cases contributing to high IoU and weaker contrasts contributing to lower IoU. Localised IoU is scalable over larger mapping areas (e.g. see Helmand Province in Appendix B) and aids existing techniques for visualising local variation in classification performance (e.g. image differencing). These techniques will only be further exasperated as the community utilise remote sensing data to map over larger areas. Mapping the distribution of error with localised IoU is recommended with global metrics for providing further evaluation of classification accuracy across landscapes at varying scales.

The agricultural area estimation from classification was found to consistently overestimate the active agricultural area. The overestimation would lead to an overestimation of opium poppy using the UN's ratio-based estimation approach.

However, the current approach uses the potential agricultural area which would still be greater than the active mask therefore using the active mask would still provide a more representative estimate. The main reason the potential mask is used is to ensure no potential areas under cultivation are missed during the survey. The classification found more commission for agriculture and it is unlikely using the active mask will omit potential areas under opium cultivation.

5.5 Analysis ready data and automated agricultural land classification

The contribution of the spectral image features during FCN training emphasises the need for standardising spectral values between image datasets for analysis ready data. Standardisation is common in remote sensing to minimise the variation in measured surface reflectance caused by atmospheric, topographic and illumination differences (Rußwurm and Körner, 2018; Zhang *et al.*, 2018). Consistency across imagery is advantageous to train generalised models for agricultural land classification. Standardisation using IR-MAD normalisation provided consistency between image datasets (DMC, Landsat and Sentinel-2) compared to TOA reflectance. Matching imagery with IR-MAD is able to remove the differences in the atmosphere between imagery, but Level-1A is the reflectance without atmospheric correction. The subtle differences in atmosphere between the IR-MAD matched imagery and Level-1A are the cause for little difference in localised IoU generally (Figure 3-18). Improvements in classification accuracy were seen in the localised IoU in areas of known confusion (e.g. edge cases and smaller field parcels) (Figure 3-17), where subtle differences in spectral values have the most influence during classification. Further improvements were found for weak edge cases in areas of sparse agriculture by increasing the spatial resolution using the same generalised model (Figure 3-21). By implementing the IR-MAD standardisation it is possible to have analysis ready data for use with the generalised model, but requires intersecting imagery over the same area to calculate the invariant features. The standardisation also requires additional time to process the imagery to be analysis ready data, but is beneficial for consistent classification.

The ease of using FCNs with analysis ready data for classifying agricultural land is advantageous for operational use. Encoding localised image features outperforms other machine learning techniques for agricultural land classification. Generalised models can be used to classify any image without any manual effort, regardless of image timing or image sensor. These models can be fine-tuned faster with transfer learning and updating models with less data and sparse datasets alleviates the need for large labelled datasets for continual improvement.

Using these models at scale requires suitable infrastructure to collect, store, process and classify across image datasets. Automated classification is possible by utilising cloud services with FCNs, such as Amazon Web Services (AWS) or Google Earth Engine (GEE) (Patel *et al.*, 2015). Cloud services provide access to large volumes of image datasets without the need to store or process them locally. Automated classification of image datasets would be possible when the trained model is hosted on the platform. Researchers can take advantage of the increased processing power available on cloud services by automating classification. For example, an automated workflow could be setup for new image datasets where the latest generalised FCN model for agricultural land would automatically classify near-real time without the need for manual input.

UNODC currently use a potential agricultural mask for their annual opium survey, while the FCN models developed in this thesis classify active areas of agriculture. However, active agricultural masks classified using the generalised model could be used to create potential masks by combining multiple years of these active masks. The cloud platform could be used to update the generalised model with newly classified areas of agriculture using the methodology from §3.3.5 and would only require manual inspection of these new areas for quality control. Localised IoU can also be used to monitor land cover changes by comparing the latest and previous year's active mask, where greater change is expressed as a lower IoU score.

Automated classification is advantageous for UNODC because generalised models can classify any image date with >80% user accuracy in southern Helmand

almost two months before the peak agricultural extent. Monitoring agricultural land before peak opium biomass will provide an early indication of new areas of agriculture to aid survey design, sample selection for the opium survey and identify new areas under opium cultivation. A fully automated workflow for agricultural land classification using FCNs would increase efficiency in the opium survey and allow for continual model improvement through fine-tuning on sparsely labelled datasets.

Chapter 6. Conclusions

This chapter contains the conclusions of the research on automated agricultural land classification and monitoring of land cover change presented in the preceding chapters and recommendations for future work.

6.1 Key findings from objective 1: critically evaluate the use of CNNs for delineating Afghanistan's active agricultural mask compared to human interpretation

Convolutional Neural Networks (CNNs) were able to separate between natural vegetation and agriculture. The improved classification accuracy was obtained by encoding local information using convolutional filters to separate between these areas of confusion (Chapter 2).

A targeted strategy was used to focus training the CNN on difficult classification areas (e.g. natural vegetation and edge cases between agriculture and non-agriculture) and improved classification accuracy (>94%) (Chapter 2).

Transfer learning using a pre-trained model on an unrelated image dataset (ImageNet) achieved higher overall accuracy than training from scratch (+2.2%) and training was faster. Updating the model between image years using transfer learning required only 25% of available labelled data with similar classification performance. Without additional training, models were able to classify imagery across sensors and years to achieve high classification accuracy (>95%) (Chapter 2 and Chapter 3).

A Fully Convolutional Network (FCN) outperformed the CNN and improved classification further by replacing the fully connected layer with an upsampling convolutional layer (+3%) (Chapter 3). A method was developed to fine-tune FCN models with sparsely labelled data for new areas of agricultural land and improved classification accuracy (+1%) (Chapter 3). FCNs can replace human interpretation of agricultural land.

6.2 Key findings from objective 2: develop generalised CNN models for classifying the agricultural mask and evaluate the uncertainty on measuring the active agricultural area

Further understanding of the underlying importance of the contextual information resulting in superior performance in agricultural land classification enabled generalised models to be trained. The textural and spectral image features were

more important than shape for separating agricultural land (Chapter 3). Generalised FCN models were successfully trained on legacy datasets across image sensors by standardising spectral image features with image matching (IR-MAD) (Chapter 3). These models classified agricultural land across image datasets (DMC, Landsat and Sentinel-2), even for image sensors not used during training and maintained high classification accuracy (>95%).

Localised intersection over union was developed as a new tool to assess the localised differences in classification and provided greater insight for comparing small differences in mapping accuracy (Chapter 3). Improvements to agricultural land classification from standardisation were identified using localised IoU at edges in mixed areas of agriculture and desert, where small differences in the spectral values have the largest effect during classification. Increased spatial resolution improved classification in these areas without the need for fine-tuning the generalised models.

These models can classify image datasets at any point within the season to monitor agricultural changes whilst maintaining high accuracy almost two months before peak opium biomass (>80%) (Chapter 3). Generalised FCN models are the best approach for utilising legacy data to understand changes in land-use across image sensors.

6.3 Key findings from objective 3: understand the annual changes in agricultural land use in relation to opium poppy in Helmand Province, Afghanistan between 2010 and 2019

Generalised models, trained on legacy data, were used to classify agricultural land from different image datasets and sensors to understand agricultural land change in Helmand Province between 2010 and 2019 (Chapter 4). All districts in Helmand Province were found to have year-on-year expansion of agricultural land and opium cultivation between 2010 and 2017. The UNODC opium estimates found opium cultivation began to decline after the peak in 2017, whilst agricultural area classified by the FCN continued to increase, suggesting a shift in agricultural practice. Helmand Province has undergone rapid agricultural expansion away from canal irrigated areas in the main Helmand valley between

2013 and 2019 into marginal desert areas with less control and security, predominately in Nad Ali, Naw Zad, and Washer. Increases in new agricultural land available are attributed to the introduction of new agricultural technologies (e.g. solar-powered deep wells and fertilisers) across the province to increase access to adequate irrigation. Agricultural expansion into marginal desert areas are associated with increased opium cultivation in Helmand Province. These technologies would otherwise be too expensive if not for the higher farm-gate price for opium than other cash crops (e.g. wheat). Districts located in the main Helmand valley had less agricultural expansion and opium cultivation remained static over the study period (e.g. Lashkar Gah, Nahri Sarraj, and Sangin). Southern districts were found to have less agricultural expansion, but opium cultivation rose suggesting intensification or improved yields (e.g. Dishu, Garmser, and Reg).

The models used in the analysis required no additional training for years without labelled data and demonstrate the potential for applying these models across large volumes of image data to derive new knowledge from legacy datasets. Improved and timely classification of agricultural land from satellite imagery can be used to gain further insight into land-use change associated with opium cultivation to inform national and international counter-narcotic strategy.

6.4 Implications of research for opium monitoring in Afghanistan

High classification performance coupled with continual model refinement from additional standardised data shows the potential for FCNs to aid human interpreters for UNODC's agricultural mask production. Utilising data-driven models from previous agricultural land delineation will improve the speed and efficiency of the survey and reduce the overall cost. Deep transfer learning across multiple years presents an exciting opportunity for timely and efficient classification and retraining existing models with only new areas of agriculture (sparse datasets) will further reduce the effort required to improve agricultural land delineation year-on-year. The ability to utilise generalised models on all

optical imagery types regardless of spatial resolution and timing is advantageous for monitoring the potential area under opium cultivation. Although acceptable classification accuracy (>80%) in comparison to peak agricultural extent is achieved from ¾ opium biomass, the active agricultural area can be classified earlier. Identifying the frame sampling locations early would increase the efficiently in conducting the ratio-based estimates for opium production. Utilising cloud imagery platforms, such as Google Earth Engine (GEE) and Amazon Web Services (AWS), with the most recently updated FCN model provides the opportunity for an automated agricultural mask production workflow. This would provide near real-time monitoring of the active agricultural area across image datasets by classifying each new image uploaded onto the platform.

6.5 Recommendations for future research

The potential for generalised FCN models to classify agricultural land across image datasets is an exciting prospect for land cover classification. Future research should focus on further understanding the use of fine-tuning existing FCN models with sparsely labelled data to reduce the requirements for large training datasets. The experiments conducted in Chapter 2 and Chapter 3 found classification performance increased year-on-year with more data from subsequent years, but is there a point where the model can no longer improve? The best performing FCN models are commonly limited to only three input channels for use with conventional photography (Red, Green, Blue), but satellite imagery often has more than three spectral bands (e.g. Landsat-8 (9 bands) and Sentinel-2 (13 bands)). Remote sensing applications calls for more model architectures to be built with more than three channels to fully utilise these image datasets. The success of using transfer learning from existing architectures with unrelated datasets (e.g. ImageNet) to train models more efficiently also calls for researchers to share their trained models with the wider community, especially for architectures built with more than three channels.

The FCN models trained in this thesis were focused on the southern provinces of Helmand and Kandahar, the two main opium producing provinces in Afghanistan. In the context of illicit crop monitoring in Afghanistan, further research could

investigate how well these models generalise across all provinces to map agricultural land, especially in northern provinces with different agricultural patterns. Utilising cloud imagery platforms with generalised FCN models would be advantageous for further understanding of crop dynamics across both the first and second crop cycle and aid near real-time monitoring of land cover change.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y. and Zheng, X. (2015) *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*.

Alatorre, L. C., Sánchez-Andrés, R., Cirujano, S., Beguería, S. and Sánchez-Carrillo, S. (2011) 'Identification of mangrove areas by remote sensing: The ROC curve technique applied to the northwestern Mexico coastal zone using Landsat imagery', *Remote Sensing*, 3(8), pp. 1568–1583. doi: 10.3390/rs3081568.

Aldwaik, S. Z., Onsted, J. A. and Pontius, R. G. (2015) 'Behavior-based aggregation of land categories for temporal change analysis', *International Journal of Applied Earth Observation and Geoinformation*, 35, pp. 229–238. doi: 10.1016/j.jag.2014.09.007.

Baker, N., Lu, H., Erlikhman, G. and Kellman, P. J. (2018) 'Deep convolutional networks do not classify based on global object shape', *PLoS Computational Biology*, 14(12), p. e1006613. doi: 10.1371/journal.pcbi.1006613.

Ball, J. E., Anderson, D. T. and Chan, C. S. (2017) 'A Comprehensive Survey of Deep Learning in Remote Sensing: Theories, Tools and Challenges for the Community', *Journal of Applied Remote Sensing*, 11(4), p. 042609. doi: 10.1117/1.JRS.11.042609.

Belgiu, M. and Drăgu, L. (2016) 'Random forest in remote sensing: A review of applications and future directions', *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, pp. 24–31. doi: 10.1016/j.isprsjprs.2016.01.011.

Bellón, B., Bégué, A., Lo Seen, D., de Almeida, C. and Simões, M. (2017) 'A remote sensing approach for regional-scale mapping of agricultural land-use systems based on NDVI time series', *Remote Sensing*, 9(6), p. 600. doi: 10.3390/rs9060600.

Boryan, C. G., Yang, Z., Willis, P. and Di, L. (2017) ‘Developing crop specific area frame stratifications based on geospatial crop frequency and cultivation data layers’, *Journal of Integrative Agriculture*, 16(2), pp. 312–323. doi: 10.1016/S2095-3119(16)61396-5.

Bradford, J. and Mansfield, D. (2019) ‘Known unknowns and unknown knows: what we know about the cannabis and Hashish trade in Afghanistan’, *EchoGéo*, 48. doi: 10.4000/echogeo.17626.

Brigato, L. and Iocchi, L. (2021) ‘A Close Look at Deep Learning with Small Data’, in *25th International Conference on Pattern Recognition*, pp. 2490–2497. doi: 10.1109/ICPR48806.2021.9412492.

Canty, M. J. (2014) *Image analysis, classification and change detection in remote sensing: with algorithms for ENVI/IDL and Python*. Third Revi. Taylor and Francis CRC Press.

Canty, M. J. and Nielsen, A. A. (2008) ‘Automatic radiometric normalization of multitemporal satellite imagery with the iteratively re-weighted MAD transformation’, *Remote Sensing of Environment*, 112(3), pp. 1025–1036. doi: 10.1016/j.rse.2007.07.013.

Carfagna, E. and Gallego, F. J. (2005) ‘Using Remote Sensing for Agricultural Statistics’, *International Statistical Review*, 73(3), pp. 389–404. doi: 10.1111/j.1751-5823.2005.tb00155.x.

Casu, F., Manunta, M., Agram, P. S. and Crippen, R. E. (2017) ‘Big remotely sensed data: tools, applications and experiences’, *Remote Sensing of Environment*, 202, pp. 1–2. doi: 10.1016/j.rse.2017.09.013.

Chen, G., Hay, G. J., Carvalho, L. M. T. and Wulder, M. A. (2012) ‘Object-based change detection’, *International Journal of Remote Sensing*, 33(14), pp. 4434–4457. doi: 10.1080/01431161.2011.648285.

Cheng, G., Han, J. and Lu, X. (2017) ‘Remote sensing image scene classification: benchmark and state of the art’, *Proceedings of the IEEE*, 105, pp. 1865–1883. doi: 10.1109/JPROC.2017.2675998.

Chollet, F. (2015) *Keras*. Available at: https://github.com/keras-team/keras.

Chollet, F. (2017) *Deep learning with Python*. New York: Manning Publications.

Clement, M. A., Kilsby, C. G. and Moore, P. (2017) 'Multi-temporal synthetic aperture radar flood mapping using change detection', *Journal of Flood Risk Management*. doi: 10.1111/jfr3.12303.

Cohen, J. (1960) 'A coefficient of agreement for nominal scales', *Educational and Psychological Measurement*, 20(1), pp. 37–46. doi: 10.1177/001316446002000104.

Colditz, R. R. (2015) 'An evaluation of different training sample allocation schemes for discrete and continuous land cover classification using decision tree-based algorithms', *Remote Sensing*, 7(8), pp. 9655–9681. doi: 10.3390/rs70809655.

CSO (2018) *Afghanistan Living Conditions Survey 2016-2017*. Kabul.

DeFries, R. S. and Chan, J. C. W. (2000) 'Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data', *Remote Sensing of Environment*, 74(3), pp. 503–515.

Delgado, R. and Tibau, X. A. (2019) 'Why Cohen's Kappa should be avoided as performance measure in classification', *PLoS ONE*, 14(9), p. e0222916. doi: 10.1371/journal.pone.0222916.

Dell'Acqua, F., Gamba, P., Casella, V., Zucca, F., Benediktsson, J. A., Wilkinson, G., Galli, A., Malinverni, E. S., Jones, G., Greenhill, D. and Ripke, L. (2006) 'HySenS data exploitation for urban land cover analysis', *Annals of Geophysics*, 49(1), pp. 311–318. doi: 10.4401/ag-3160.

Demir, S. and Başayiğit, L. (2019) 'Determination of Opium Poppy (Papaver Somniferum) Parcels Using High-Resolution Satellite Imagery', *Journal of the Indian Society of Remote Sensing*. doi: 10.1007/s12524-019-00955-1.

Deng, Z., Sun, H., Zhou, S., Zhao, J., Lei, L. and Zou, H. (2018) ‘Multi-scale object detection in remote sensing imagery with convolutional neural networks’, *IPRS Journal of Photogrammetry and Remote Sensing*, 145, pp. 3–22. doi: 10.1016/j.isprsjprs.2018.04.003.

Dingle Robertson, L. and King, D. J. (2011) ‘Comparison of pixel- and object-based classification in land cover change mapping’, *International Journal of Remote Sensing*, 32(6), pp. 1505–1529. doi: 10.1080/01431160903571791.

Douzas, G., Bacao, F., Fonseca, J. and Khudinyan, M. (2019) ‘Imbalanced learning in land cover classification: Improving minority classes’ prediction accuracy using the geometric SMOTE algorithm’, *Remote Sensing*, 11(24). doi: 10.3390/rs11243040.

Dronova, I., Gong, P., Wang, L. and Zhong, L. (2015) ‘Mapping dynamic cover types in a large seasonally flooded wetland using extended principal component analysis and object-based classification’, *Remote Sensing of Environment*, 158, pp. 193–206. doi: 10.1016/j.rse.2014.10.027.

Duchi, J., Hazan, E. and Singer, Y. (2011) ‘Adaptive Subgradient Methods for Online Learning and Stochastic Optimization’, *Journal of Machine Learning Research*, 12, pp. 2121–2159.

Dwyer, J., Roy, D., Sauer, B., Jenkerson, C., Zhang, H. and Lymburner, L. (2018) ‘Analysis Ready Data: Enabling Analysis of the Landsat Archive’, *Remote Sensing*, 10(9), pp. 1–24. doi: 10.20944/PREPRINTS201808.0029.V1.

FAO (2016) *The Islamic Republic of Afghanistan: Land cover atlas*.

FAO (2019) *The state of Afghanistan’s biodiversity for food and agriculture*. Available at: https://www.fao.org/3/CA3482EN/ca3482en.pdf.

Feng, Y., Diao, W., Sun, X., Yan, M. and Gao, X. (2019) ‘Towards automated ship detection and category recognition from high-resolution aerial images’, *Remote Sensing*, 11(1901). doi: 10.3390/rs11161901.

Foody, G. M. (2002) ‘Status of land cover classification accuracy assessment’, *Remote Sensing of Environment*, 80(1), pp. 185–201. doi: 10.1016/S0034-4257(01)00295-4.

Foody, G. M. (2009) ‘Sample size determination for image classification accuracy assessment and comparison’, *International Journal of Remote Sensing*, 30(20), pp. 5273–5291. doi: 10.1080/01431160903130937.

Fu, G., Liu, C., Zhou, R., Sun, T. and Zhang, Q. (2017) ‘Classification for high resolution remote sensing imagery using a fully convolutional network’, *Remote Sensing*, 9(5), pp. 1–21. doi: 10.3390/rs9050498.

Gallego, F. J. (2004) ‘Remote sensing and land cover area estimation’, *International Journal of Remote Sensing*, 25(15), pp. 3019–3047. doi: 10.1080/01431160310001619607.

Geneletti, D. and Gorte, B. G. H. (2003) ‘A method for object-oriented land cover classification combining Landsat TM data and aerial photographs’, *International Journal of Remote Sensing*, 24(6), pp. 1273–1286. doi: 10.1080/01431160210144499.

Gislason, P. O., Benediktsson, J. A. and Sveinsson, J. R. (2006) ‘Random forests for land cover classification’, *Pattern Recognition Letters*, 27(4), pp. 294–300. doi: 10.1016/j.patrec.2005.08.011.

Gómez, C., White, J. C. and Wulder, M. A. (2016) ‘Optical remotely sensed time series data for land cover classification: A review’, *ISPRS Journal of Photogrammetry and Remote Sensing*, 116, pp. 55–72. doi: 10.1016/j.isprsjprs.2016.03.008.

Goodfellow, I., Bengio, Y. and Courville, A. (2016) *Deep Learning*. Cambridge, MA, USA: MIT Press.

Van Grinsven, M. J. J. P., Van Ginneken, B., Hoyng, C. B., Theelen, T. and Sánchez, C. I. (2016) 'Fast Convolutional Neural Network Training Using Selective Data Sampling: Application to Hemorrhage Detection in Color Fundus Images', *IEEE Transactions on Medical Imaging*, 35(5), pp. 1273–1284. doi: 10.1109/TMI.2016.2526689.

Guo, D. and Mennis, J. (2009) 'Spatial data mining and geographic knowledge discovery - An introduction', *Computers, Environment and Urban Systems*, 33, pp. 403–408. doi: 10.1016/j.compenvurbsys.2009.11.001.

Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., Kalloo, A., Ben Hadj Hassen, A., Thomas, L., Enk, A. and Uhlmann, L. (2018) 'Man against Machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists', *Annals of Oncology*, 29(8), pp. 1836–1842. doi: 10.1093/annonc/mdy166.

Hall-Beyer, M. (2017) 'Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales', *International Journal of Remote Sensing*, 38(5), pp. 1312–1338. doi: 10.1080/01431161.2016.1278314.

Han, K. S., Champeaux, J. L. and Roujean, J. L. (2004) 'A land cover classification product over France at 1 km resolution using SPOT4/VEGETATION data', *Remote Sensing of Environment*, 92(1), pp. 52–66. doi: 10.1016/j.rse.2004.05.005.

Hansen, M. C., Egorov, A., Potapov, P. V, Stehman, S. V, Tyukavina, A., Turubanova, S. A., Roy, D. P., Goetz, S. J., Loveland, T. R., Ju, J., Kommareddy, A., Kovalskyy, V., Forsyth, C. and Bents, T. (2014) 'Monitoring conterminous United States (CONUS) land cover change with Web-Enabled Landsat Data (WELD)', *Remote Sensing of Environment*, 140, pp. 466–484. doi: 10.1016/J.RSE.2013.08.014.

Haralick, R. M., Shanmugan, K. and Dinstein, I. H. (1973) ‘Textural features for image classification’, *IEEE Transactions on Systems, Man, and Cybernetics*, 3, pp. 610–621.

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

Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P. and Koetz, B. (2015) ‘Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery’, *Remote Sensing*, 7(9), pp. 12356–12379. doi: 10.3390/rs70912356.

Johnson, J. M. and Khoshgoftaar, T. M. (2019) ‘Survey on deep learning with class imbalance’, *Journal of Big Data*, 6(1). doi: 10.1186/s40537-019-0192-5.

Kampffmeyer, M., Salberg, A. B. and Jenssen, R. (2016) ‘Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks’, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. Las Vegas, USA. doi: 10.1109/CVPRW.2016.90.

Kawasaki, S., Watanabe, F., Suzuki, S., Nishimaki, R. and Takahashi, S. (2012) ‘Current Situation and Issues on Agriculture of Afghanistan’, *ournal of arid land studies*, 22(1), pp. 345–348.

Keuchel, J., Naumann, S., Heiler, M. and Siegmund, A. (2003) ‘Automatic land cover analysis for Tenerife by supervised classification using remotely sensed data’, *Remote Sensing of Environment*, 86(4), pp. 530–541. doi: 10.1016/S0034-4257(03)00130-5.

Khatami, R., Mountrakis, G. and Stehman, S. V (2016) ‘A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research’, *Remote Sensing of Environment*, 177, pp. 89–100. doi: 10.1016/j.rse.2016.02.028.

Khatami, R., Mountrakis, G. and Stehman, S. V (2017) 'Mapping per-pixel predicted accuracy of classified remote sensing images', *Remote Sensing of Environment*, 191, pp. 156–167. doi: 10.1016/j.rse.2017.01.025.

Kingma, D. P. and Ba, J. L. (2015) 'Adam: A method for stochastic optimization', in *3rd International Conference on Learning Representations*.

Koga, Y., Miyazaki, H. and Shibasaki, R. (2018) 'A CNN-based method of vehicle detection from aerial images using hard example mining', *Remote Sensing*, 10(124). doi: 10.3390/rs10010124.

Kroupi, E., Kesa, M., Navarro-Sánchez, V. D., Saeed, S., Pelloquin, C., Alhaddad, B., Moreno, L., Soria-Frisch, A. and Ruffini, G. (2019) 'Deep convolutional neural networks for land-cover classification with Sentinel-2 images', *Journal of Applied Remote Sensing*, 13(2), p. 024525. doi: 10.1117/1.jrs.13.024525.

Laliberte, A. S., Rango, A., Havstad, K. M., Paris, J. F., Beck, R. F., McNeely, R. and Gonzalez, A. L. (2004) 'Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico', *Remote Sensing of Environment*, 93(1–2), pp. 198–210. doi: 10.1016/j.rse.2004.07.011.

Lecun, Y., Bengio, Y. and Hinton, G. (2015) 'Deep learning', *Nature*, 521, pp. 436–444. doi: 10.1038/nature14539.

Lei, T. C., Wan, S., Wu, S. C. and Wang, H. P. (2020) 'A new approach of ensemble learning technique to resolve the uncertainties of paddy area through image classification', *Remote Sensing*, 12, p. 3666. doi: 10.3390/rs12213666.

Li, Y., Huang, H., Xie, Q., Yao, L. and Chen, Q. (2018) 'Research on a surface defect detection algorithm based on MobileNet-SSD', *Applied Sciences*, 8(9). doi: 10.3390/app8091678.

Liu, H., He, L. and Li, J. (2017) 'Remote sensing image classification based on convolutional neural networks with two-fold sparse regularization', in *IEEE International Geoscience and Remote Sensing Symposium*. Texas, USA, pp. 992–995. doi: 10.1109/IGARSS.2017.8127121.

Liu, X., Tian, Y., Yuan, C., Zhang, F. and Yang, G. (2018) 'Opium Poppy Detection Using Deep Learning', *Remote Sensing*, 10(12), p. 1886. doi: 10.3390/rs10121886.

Lobell, D. B., Thau, D., Seifert, C., Engle, E. and Little, B. (2015) 'A scalable satellite-based crop yield mapper', *Remote Sensing of Environment*, 164, pp. 324–333. doi: 10.1016/j.rse.2015.04.021.

Long, J., Shelhamer, E. and Darrell, T. (2015) 'Fully convolutional networks for semantic segmentation', in *IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA, pp. 3431–3440. doi: 10.1109/CVPR.2015.7298965.

Lucas, R., Medcalf, K., Brown, A., Bunting, P., Breyer, J., Clewley, D., Keyworth, S. and Blackmore, P. (2011) 'Updating the Phase 1 habitat map of Wales, UK, using satellite sensor data', *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(1), pp. 81–102. doi: 10.1016/j.isprsjprs.2010.09.004.

Maggiori, E., Tarabalka, Y., Charpiat, G. and Alliez, P. (2017) 'Convolutional neural networks for large-scale remote-sensing image classification', *IEEE Transactions on Geoscience and Remote Sensing*, 55(2), pp. 645–657. doi: 10.1109/TGRS.2016.2612821.

Mansfield, D. (2011) *Managing concurrent and repeated risks: Explaining the reductions in opium production in Central Helmand between 2008 and 2011*. Kabul.

Mansfield, D. (2017) *Understanding control and influence: what opium poppy and tax reveal about the writ of the Afghan state*. Kabul.

Mansfield, D. (2018) 'Turning deserts into flowers: settlement and poppy cultivation in southwest Afghanistan', *Third World Quarterly*, 39(2), pp. 331–349. doi: 10.1080/01436597.2017.1396535.

Mansfield, D. (2019) 'On the Frontiers of Development : Illicit Poppy and the Transformation of the Deserts of Southwest Afghanistan', *Journal of Illicit Economies and Development*, 1(3), pp. 330–345. doi: 10.31389/jied.46.

Maxwell, A. E., Warner, T. A. and Fang, F. (2018) 'Implementation of machine-learning classification in remote sensing: an applied review', *International Journal of Remote Sensing*, 39(9), pp. 2784–2817. doi: 10.1080/01431161.2018.1433343.

Memarsadeghi, N., Mount, D. M., Netanyahu, N. S. and Moigne, J. L. (2007) 'A fast implementation of the ISODATA clustering algorithm', *International Journal of Computational Geometry and Applications*, 17(1), pp. 71–103.

Morales-Barquero, L., Lyons, M. B., Phinn, S. R. and Roelfsema, C. M. (2019) 'Trends in remote sensing accuracy assessment approaches in the context of natural resources', *Remote Sensing*, 11(19), pp. 1–16. doi: 10.3390/rs11192305.

Moreno Navas, J., Telfer, T. C. and Ross, L. G. (2012) 'Separability indexes and accuracy of neuro-fuzzy classification in Geographic Information Systems for assessment of coastal environmental vulnerability', *Ecological Informatics*, 12, pp. 43–49. doi: 10.1016/j.ecoinf.2012.06.006.

Mulla, D. J. (2013) 'Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps', *Biosystems Engineering*, 114, pp. 358–371. doi: 10.1016/j.biosystemseng.2012.08.009.

Myint, S. W., Yuan, M., Cerveny, R. S. and Giri, C. P. (2008) 'Comparison of remote sensing image processing techniques to identify tornado damage areas from Landsat TM data', *Sensors*, 8(2), pp. 1128–1156. doi: 10.3390/s8021128.

Nogueira, K., Penatti, O. A. B. and dos Santos, J. A. (2017) 'Towards better exploiting convolutional neural networks for remote sensing scene classification', *Pattern Recognition*, 61, pp. 539–556. doi: 10.1016/j.patcog.2016.07.001.

NSIA (2019) *Agricultural Prospective Report*. Available at: https://www.nsia.gov.af:8080/wp-content/uploads/2019/07/Agricultural-prospective-Report.pdf.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V, Woodcock, C. E. and Wulder, M. A. (2014) ‘Good practices for estimating area and assessing accuracy of land change’, *Remote Sensing of Environment*, 148, pp. 42–57. doi: 10.1016/j.rse.2014.02.015.

Otsu, N. (1979) ‘A threshold selection method from gray-level histograms’, *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), pp. 62–66. doi: 10.1109/TSMC.1979.4310076.

Otukei, J. R. and Blaschke, T. (2010) ‘Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms’, *International Journal of Applied Earth Observation and Geoinformation*, 12(1), pp. 27–31. doi: 10.1016/j.jag.2009.11.002.

Paisitkriangkrai, S., Sherrah, J., Janney, P. and Van-Den Hengel, A. (2015) ‘Effective semantic pixel labelling with convolutional networks and Conditional Random Fields’, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. Boston, USA. doi: 10.1109/CVPRW.2015.7301381.

Pal, M. (2005) ‘Random forest classifier for remote sensing classification’, *International Journal of Remote Sensing*, 26(1), pp. 217–222. doi: 10.1080/01431160412331269698.

Panda, S. S., Ames, D. P. and Panigrahi, S. (2010) ‘Application of vegetation indices for agricultural crop yield prediction using neural network techniques’, *Remote Sensing*, 2(3), pp. 673–696. doi: 10.3390/rs2030673.

Pandey, B., Joshi, P. K. and Seto, K. C. (2013) ‘Monitoring urbanization dynamics in india using DMSP/OLS night time lights and SPOT-VGT data’, *International Journal of Applied Earth Observation and Geoinformation*, 23, pp. 49–61. doi: 10.1016/j.jag.2012.11.005.

Parente, L. and Ferreira, L. (2018) ‘Assessing the spatial and occupation dynamics of the Brazilian pasturelands based on the automated classification of MODIS images from 2000 to 2016’, *Remote Sensing*, 10(4), p. 606. doi: 10.3390/rs10040606.

Park, J., Kim, D. I., Choi, B., Kang, W. and Kwon, H. W. (2020) ‘Morphological Analysis of Vector Mosquitoes using Deep Convolutional Neural Networks’, *Scientific Reports*, 10(1012). doi: 10.1038/s41598-020-57875-1.

Patel, N. N., Angiuli, E., Gamba, P., Gaughan, A., Lisini, G., Stevens, F. R., Tatem, A. J. and Trianni, G. (2015) ‘Multitemporal settlement and population mapping from Landsat using Google Earth Engine’, *International Journal of Applied Earth Observation and Geoinformation*, 35, pp. 199–208. doi: 10.1016/j.jag.2014.09.005.

Penatti, A. B., Nogueira, K. and Santos, J. A. (2015) ‘Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?’, in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 44–51. doi: 10.1109/CVPRW.2015.7301382.

Petitjean, F., Ketterlin, A. and Gançarski, P. (2011) ‘A global averaging method for dynamic time warping, with applications to clustering’, *Pattern Recognition*, 44(3), pp. 678–693. doi: 10.1016/j.patcog.2010.09.013.

Piramanayagam, S., Saber, E., Schwartzkopf, W. and Koehler, F. W. (2018) ‘Supervised classification of multisensor remotely sensed images using a deep learning framework’, *Remote Sensing*, 10(1429). doi: 10.3390/rs10091429.

Pontius, R. G. and Millones, M. (2011) ‘Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment’, *International Journal of Remote Sensing*, 32(15), pp. 4407–4429. doi: 10.1080/01431161.2011.552923.

Pouliot, D., Latifovic, R., Pasher, J. and Duffe, J. (2019) ‘Assessment of Convolution Neural Networks for Wetland Mapping with Landsat in the Central Canadian Boreal Forest Region’, *Remote Sensing*, 11(7), p. 772. doi: 10.3390/rs11070772.

Rogan, J., Franklin, J., Stow, D., Miller, J., Woodcock, C. and Roberts, D. (2008) ‘Mapping land-cover modifications over large areas: A comparison of machine learning algorithms’, *Remote Sensing of Environment*, 112(5), pp. 2272–2283. doi: 10.1016/j.rse.2007.10.004.

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

Rußwurm, M. and Körner, M. (2018) ‘Multi-temporal land cover classification with sequential recurrent encoders’, *ISPRS International Journal of Geo-Information*, 7(4), p. 129. doi: 10.3390/ijgi7040129.

Sexton, J. O., Urban, D. L., Donohue, M. J. and Song, C. (2013) ‘Long-term land cover dynamics by multi-temporal classification across the Landsat-5 record’, *Remote Sensing of Environment*, 128, pp. 246–258. doi: 10.1016/j.rse.2012.10.010.

Shahriar Pervez, M., Budde, M. and Rowland, J. (2014) ‘Mapping irrigated areas in Afghanistan over the past decade using MODIS NDVI’, *Remote Sensing of Environment*, 149, pp. 155–165. doi: 10.1016/j.rse.2014.04.008.

Shao, Y. and Lunetta, R. S. (2012) ‘Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points’, *ISPRS Journal of Photogrammetry and Remote Sensing*, 70(1), pp. 78–87. doi: 10.1016/j.isprsjprs.2012.04.001.

Shelhamer, E., Long, J. and Darrell, T. (2016) ‘Fully Convolutional Networks for Semantic Segmentation’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4). doi: 10.1109/TPAMI.2016.2572683.

Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D. and Summers, R. M. (2016) 'Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning.', *IEEE Transactions on Medical Imaging*, 35(5), pp. 1285–98. doi: 10.1109/TMI.2016.2528162.

Simms, D. M. (2020) 'Fully convolutional neural nets in-the-wild', *Remote Sensing Letters*, 11(12), pp. 1080–1089. doi: 10.1080/2150704X.2020.1821120.

Simms, D. M., Waine, T. W. and Taylor, J. C. (2017) 'Improved estimates of opium cultivation in Afghanistan using imagery-based stratification', *International Journal of Remote Sensing*, 38(13), pp. 3785–3799. doi: 10.1080/01431161.2017.1303219.

Simms, D. M., Waine, T. W., Taylor, J. C. and Brewer, T. R. (2016) 'Image segmentation for improved consistency in image-interpretation of opium poppy', *International Journal of Remote Sensing*, 37(6), pp. 1243–1256. doi: 10.1080/01431161.2016.1148290.

Simms, D. M., Waine, T. W., Taylor, J. C. and Juniper, G. R. (2014) 'The application of time-series MODIS NDVI profiles for the acquisition of crop information across Afghanistan', *International Journal of Remote Sensing*, 35(16), pp. 6234–6254. doi: 10.1080/01431161.2014.951099.

Simonyan, K. and Zisserman, A. (2015) 'Very Deep Convolutional Networks for Large-Scale Image Recognition', in *International Conference on Learning Representations 2015*.

Song, J., Gao, S., Zhu, Y. and Ma, C. (2019) 'A survey of remote sensing image classification based on CNNs', *Big Earth Data*, 3(3), pp. 232–254. doi: 10.1080/20964471.2019.1657720.

Stehman, S. V (2009) 'Sampling designs for accuracy assessment of land cover', *International Journal of Remote Sensing*, 30(20), pp. 5243–5272. doi: 10.1080/01431160903131000.

Stehman, S. V and Wickham, J. D. (2011) 'Pixels, blocks of pixels, and polygons: Choosing a spatial unit for thematic accuracy assessment', *Remote Sensing of Environment*, 115(12), pp. 3044–3055. doi: 10.1016/j.rse.2011.06.007.

Su, L., Gong, M., Zhang, P., Zhang, M., Liu, J. and Yang, H. (2017) 'Deep learning and mapping based ternary change detection for information unbalanced images', *Pattern Recognition*, 66, pp. 213–228. doi: 10.1016/j.patcog.2017.01.002.

Talukdar, S., Singha, P., Mahato, S., Pal, S., Liou, A. and Rahman, A. (2020) 'Land-use land-cover classification by machine learning classifiers for satellite observations-A review', *Remote Sensing*, 12(7), p. 1135. doi: 10.3390/rs12071135.

Tateishi, R., Uriyangqai, B., Al-Bilbisi, H., Ghar, M. A., Tsend-Ayush, J., Kobayashi, T., Kasimu, A., Hoan, N. T., Shalaby, A., Alsaaideh, B., Enkhzaya, T., Tana, G. and Sato, H. P. (2011) 'Production of global land cover data – GLCNMO', *International Journal of Digital Earth*, 4(1), pp. 22–49. doi: 10.1080/17538941003777521.

Taylor, J. C., Waine, T. W., Juniper, G. R., Simms, D. M. and Brewer, T. R. (2010) 'Survey and monitoring of opium poppy and wheat in Afghanistan: 2003-2009', *Remote Sensing Letters*, 1(3), pp. 179–185. doi: 10.1080/01431161003713028.

The World Bank (2014) *Islamic Republic of Afghanistan Agriculture Sector review*. Available at: https://documents1.worldbank.org/curated/en/245541467973233146/pdf/AUS9779-REVISED-WP-PUBLIC-Box391431B-Final-Afghanistan-ASR-web-October-31-2014.pdf.

The World Bank (2018) *The World Bank Afghanistan : National Horticulture and Livestock Productivity Project (P143841) Implementation Status and Key Decisions*. Available at: http://documents1.worldbank.org/curated/en/926371536341361243/pdf/Disclosable-Version-of-the-ISR-Afghanistan-National-Horticulture-and-Livestock-Productivity-Project-P143841-Sequence-No-11.pdf.

Thomlinson, J. R., Bolstad, P. V and Cohen, W. B. (1999) 'Coordinating Methodologies for Scaling Landcover Classifications from Site-Specific to Global: Steps toward Validating Global Map Products', *Remote Sensing of Environment*, 70(1), pp. 16–28. doi: 10.1016/S0034-4257(99)00055-3.

Tian, Y., Wu, B., Zhang, L., Li, Q., Jia, K. and Wen, M. (2011) 'Opium poppy monitoring with remote sensing in North Myanmar', *International Journal of Drug Policy*, 22(4), pp. 278–284. doi: 10.1016/j.drugpo.2011.02.001.

Tieleman, T. and Hinton, G. (2012) *Lecture 6.5 - RMSProp*.

Tiwari, V., Matin, M. A., Qamer, F. M., Ellenburg, W. L., Bajracharya, B., Vadrevu, K., Rushi, B. R. and Yusafi, W. (2020) 'Wheat Area Mapping in Afghanistan Based on Optical and SAR Time-Series Images in Google Earth Engine Cloud Environment', *Frontiers in Environmental Science*, 8(77). doi: 10.3389/fenvs.2020.00077.

Tucker, C. J. (1979) 'Red and photographic infrared linear combinations for monitoring vegetation', *Remote Sensing of Environment*, 8, pp. 127–150.

UNODC (2011) *Afghanistan Opium Survey 2011*. Available at: https://www.unodc.org/documents/crop-monitoring/Afghanistan/Afghanistan_opium_survey_2011_web.pdf.

UNODC (2015) *Afghanistan Opium Survey 2015 - Cultivation and Practice*. Available at: https://www.unodc.org/documents/crop-monitoring/Afghanistan/_Afghan_opium_survey_2015_web.pdf.

UNODC (2016) *Afghanistan Opium Survey 2016: Cultivation and Production*. Available at: https://www.unodc.org/documents/crop-monitoring/Afghanistan/Afghanistan_opium_survey_2016_cultivation_production.pdf (Accessed: 15 February 2017).

UNODC (2017a) *Afghanistan Opium Survey 2016: Sustainable development in an opium production environment*. Available at: https://www.unodc.org/documents/crop-monitoring/Afghanistan/Afghanistan_sustainable_development_for_web.pdf.

UNODC (2017b) *Afghanistan Opium survey 2017*. Available at: https://www.unodc.org/documents/crop-monitoring/Afghanistan/Afghan_opium_survey_2017_cult_prod_web.pdf.

UNODC (2018a) *Afghanistan opium survey 2017: Challenges to sustainable development, peace and security*. Available at: https://www.unodc.org/documents/crop-monitoring/Opium-survey-peace-security-web.pdf.

UNODC (2018b) *Afghanistan Opium Survey 2018*. Available at: https://www.unodc.org/documents/crop-monitoring/Afghanistan/Afghanistan_opium_survey_2018.pdf.

UNODC (2019a) *Afghanistan Opium Survey 2018: Challenges to sustainable development, peace and security*. Available at: https://www.unodc.org/documents/crop-monitoring/Afghanistan/Afghanistan_opium_survey_2018_socioeconomic_report.pdf.

UNODC (2019b) *World Drug Report 2019: Depressants*. Available at: https://wdr.unodc.org/wdr2019/.

UNODC (2020) *Word Drug Report 2020*. Available at: https://wdr.unodc.org/wdr2020/index.html.

UNODC (2021a) *Afghanistan Opium Survey 2019, socio-economic survey report: Drivers, causes and consequences of opium poppy cultivation*. Available at: https://www.unodc.org/documents/crop-monitoring/Afghanistan/20210217_report_with_cover_for_web_small.pdf.

UNODC (2021b) *México Monitoreo de Plantíos de Amapola 2018 - 2019*. Available at: https://www.unodc.org/documents/crop-monitoring/Mexico/Mexico_Monitoreo_Plantios_Amapola_2018-2019.pdf.

UNODC (2021c) *Myanmar Opium Survey 2020: Cultivation, Production, and Implications*. Available at: https://www.unodc.org/documents/crop-monitoring/Myanmar/Myanmar_Opium_survey_2020.pdf.

Vali, A., Comai, S. and Matteucci, M. (2020) ‘Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review’, *Remote Sensing*, 12(15), p. 2495. doi: 10.3390/RS12152495.

Verbeke, L. P. C., Vancoillie, F. M. B. and De Wulf, R. R. (2004) ‘Reusing back-propagation artificial neural networks for land cover classification in tropical savannahs’, *International Journal of Remote Sensing*, 25(14), pp. 2747–2771. doi: 10.1080/01431160310001652385.

Verbesselt, J., Zeileis, A. and Herold, M. (2012) ‘Near real-time disturbance detection using satellite image time series’, *Remote Sensing of Environment*, 123, pp. 98–108. doi: 10.1016/j.rse.2012.02.022.

Weinstein, B. G., Marconi, S., Bohlman, S. A., Zare, A. and White, E. P. (2020) ‘Cross-site learning in deep learning RGB tree crown detection’, *Ecological Informatics*, 56(101061). doi: 10.1016/j.ecoinf.2020.101061.

Xia, G., Hu, J., Hu, F., Shi, B., Bai, X. and Zhong, Y. (2017) ‘AID: a benchmark data set for performance evaluation of aerial scene classification’, *IEEE Transactions on Geoscience and Remote Sensing*, 55(7), pp. 3965–3981. doi: 10.1109/TGRS.2017.2685945.

Xu, Y., Zhu, M., Li, S., Feng, H., Ma, S. and Che, J. (2018) ‘End-to-end airport detection in remote sensing images combining cascade region proposal networks and multi-threshold detection networks’, *Remote Sensing*, 10(10), pp. 1–17. doi: 10.3390/rs10101516.

Yadav, K. and Congalton, R. G. (2018) ‘Accuracy assessment of Global Food Security-Support Analysis Data (GFSAD) cropland extent maps produced at three different spatial resolutions’, *Remote Sensing*, 10(11), p. 1800. doi: 10.3390/rs10111800.

Yamashita, R., Nishio, M., Kinh, R., Do, G. and Togashi, K. (2018) ‘Convolutional neural networks: an overview and application in radiology’, *Insights into Imaging*, 9, pp. 611–629. doi: 10.1007/2Fs13244-018-0639-9.

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

Ye, S., Pontius Jr, R. G. and Rakshit, R. (2018) ‘A review of accuracy assessment for object-based image analysis: From per-pixel to per-polygon approaches’, *ISPRS Journal of Photogrammetry and Remote Sensing*, 141, pp. 137–147. doi: 10.1016/j.isprsjprs.2018.04.002.

Yosinski, J., Clune, J., Bengio, Y. and Lipson, H. (2014) ‘How transferable are features in deep neural networks?’, *Advances in Neural Information Processing Systems*, 27, pp. 3320–3328.

Zeiler, M. D. and Fergus, R. (2013) *Stochastic Pooling for Regularization of Deep Convolutional Neural Networks*.

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

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

Zhang, H. K., Roy, D. P., Yan, L., Li, Z., Huang, H., Vermote, E., Skakun, S. and Roger, J. C. (2018) ‘Characterization of Sentinel-2A and Landsat-8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences’, *Remote Sensing of Environment*, 215, pp. 482–494. doi: 10.1016/j.rse.2018.04.031.

Zhang, W. and Lu, X. (2019) ‘The spectral-spatial joint learning for change detection in multispectral imagery’, *Remote Sensing*, 11(240), pp. 1–17. doi: 10.3390/rs11030240.

Zhang, W., Tang, P. and Zhao, L. (2019) ‘Remote Sensing Image Scene Classification Using CNN-CapsNet’, *Remote Sensing*, 11(5), p. 494. doi: 10.3390/rs11050494.

Zhuang, X., Engel, B. A., Lozano-Garcia, D. F., Fernández, R. N. and Johannsen, C. J. (1994) ‘Optimization of training data required for neuro-classification’, *International Journal of Remote Sensing*, 15(16), pp. 3271–3277. doi: 10.1080/01431169408954326.

Zulhaidi, H., Shafri, M., Suhaili, A. and Mansor, S. (2007) ‘The Performance of Maximum Likelihood, Spectral Angle Mapper, Neural Network and Decision Tree Classifiers in Hyperspectral Image Analysis’, *Journal of Computer Science*, 3(6), pp. 419–423.

Appendices

Appendix A Classification of the agricultural mask using machine learning techniques

A.1 Overview of the agricultural mask and machine learning classification

Crop production estimates are often used for agricultural decision making (FAO, 2016); therefore, it is essential to provide both a practical and effective methodology to accurately quantify agricultural areas. These provide vital information on the variation of cultivated area, crop stress and yield estimation (Mulla, 2013; Lobell *et al.*, 2015). In Afghanistan, UNODC extract arable agricultural area, also known as the agricultural mask, using an unsupervised image classification followed by manual post-refinement (UNODC, 2021a). This has resulted in over a decade's worth of accurate labelled agricultural image classifications. These classifications undergo quality control by visual human interpretation as the security risks for surveyors to carry out ground truthing are too greater to conduct a systematic accuracy assessment (UNODC, 2018b). There are two forms of agricultural mask: the active mask and the potential mask. The active mask delineates the area of agricultural land currently under arable cultivation for all crops in the target year (e.g. wheat, barley and opium poppy) (Taylor *et al.*, 2010). The potential mask is the delineation of agricultural land derived by adding new areas of agriculture into a single dataset from multiple years, hence the area has potential to grow crops in that season. The potential mask is used by UNODC to define the area-frame for sampling for their annual opium survey. Areas of long-term land-use change (e.g. agriculture to urban) are removed, and unused agricultural areas are removed after approximately five years (UNODC, personal communication). An annual active mask is useful for monitoring annual changes in agricultural land in relation to opium cultivation. This experiment aims to classify the active agricultural mask using data-driven machine learning approaches across Helmand and Kandahar Province, Afghanistan. The objective is to assess the performance of Random Forest and Artificial Neural Networks for agricultural land delineation in comparison to human interpretation.

A.2 Methods

The agricultural mask data used are derived from medium resolution Disaster Monitoring Constellation (DMC) satellite imagery (NIR,R,G at 32 m) acquired during peak opium biomass of the first cycle. The single date image used to derive the active agricultural mask is from 27 April 2007 (Figure A-). In the original Cranfield University study (Taylor et al., 2010), the delineation of agricultural areas was conducted using an unsupervised Iterative Self Organising Data Technique (ISODATA) (Memarsadeghi *et al.*, 2007) classification with Normalised Difference Vegetation Index (NDVI) imagery (Tucker, 1979) (Equation C-1).

$$\mathrm{NDVI} = \frac{NIR - Red}{NIR + Red} \qquad \text{C-1}$$

ISODATA was used to cluster a composite of NIR, R, G and NDVI values into a pre-defined number of classes. These clusters were manually labelled by trained interpreters into agriculture and non-agriculture classes. These were manually edited in areas of spectral confusion between classes, such as between agriculture and natural vegetation. Masks were quality checked using high resolution data (e.g. IKONOS) by comparing the resulting delineation with data from previous years to ensure consistent interpretation.

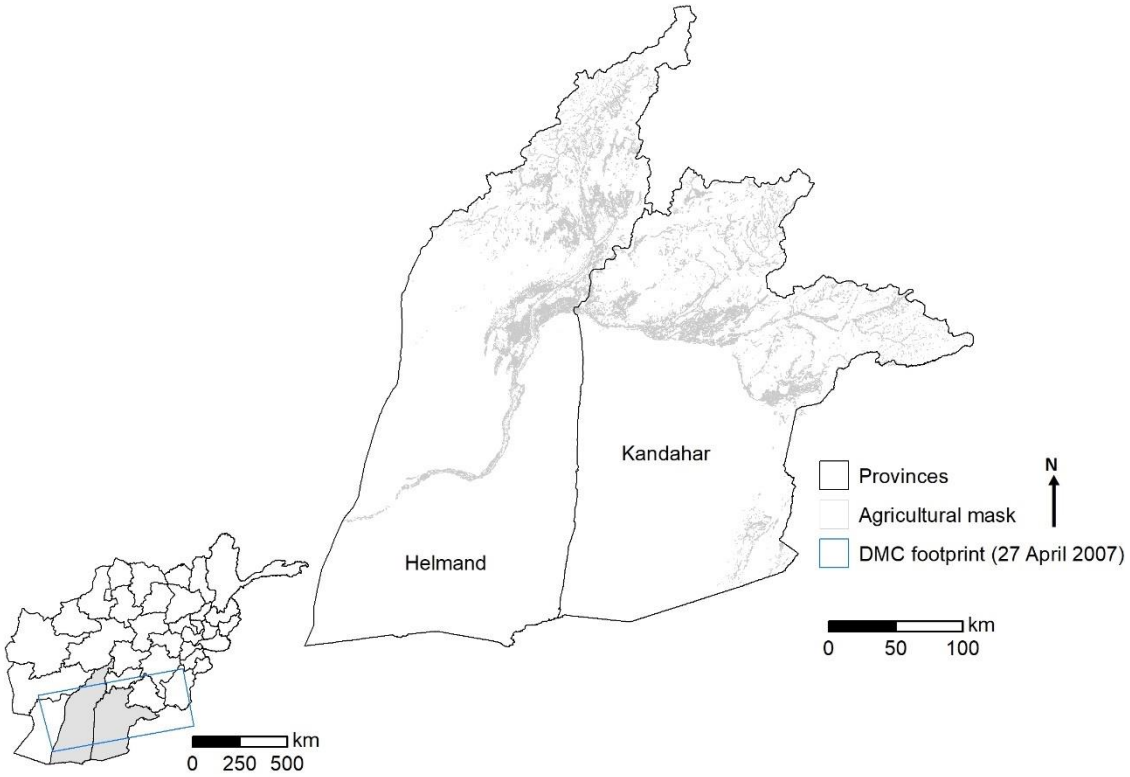

Figure A-1. Study extent based on the provincial boundaries of Helmand and Kandahar Province, Afghanistan and overlapping DMC imagery for 27 April 2007

The availability of large labelled datasets from previous agricultural masks allows supervised machine learning techniques to be used. The 2007 agricultural mask with associated DMC imagery (NIR, R, G) was used to assess the classification performance of active agricultural land using Random Forest and Artificial Neural Network. Stratified random sampling of agriculture and non-agriculture was carried out to create an equal number of training and validation samples of each class resulting in a total of 20,000 samples for training and 5,000 samples for validation. The hyper-parameters for both machine learning approaches were selected using a grid search approach. Random Forest used 100 trees with a maximum depth of 2 and Artificial Neural Network used 250 hidden layers with a learning rate of 0.01. The study site used for assessing the classification performance of the machine learning techniques is the image footprint of DMC imagery for 27 April 2007 across Helmand and Kandahar Province in Afghanistan (Figure A-1). These provinces are the highest opium producing provinces in

Afghanistan (UNODC, 2018b), therefore are ideal candidates to evaluate classification performance.

A.2.1 Results

The full comparison between the classifiers implemented on Helmand and Kandahar Province can be found in Table A-1. Random Forest has shown encouraging classification accuracy for Helmand and Kandahar with 96.73% overall accuracy and a kappa statistic of 91.32. Artificial Neural Network performed well with an overall accuracy of 92.88%, was outperformed by Random Forest (+3.85%). Additionally, Artificial Neural Network has achieved a moderate kappa statistic of 87.60 compared to Random Forest (91.32), which highlights the neural network classification is less robust when considering the probability of high overall accuracy occurring by chance. Random Forest outperforms Artificial Neural Network for agricultural land delineation for user and producer accuracy, +1.48% and +6.29% respectively.

Table A-1. Summary of Random Forest and Artificial Neural Network classification performance using DMC (NIR, R, G) imagery for 27 April 2007 across Helmand and Kandahar

Classification technique (n = 20,000 samples)	Accuracy metrics (%) (n = 5,000 samples)				
	Overall accuracy	User accuracy	Producer accuracy	Kappa	Difference in total area
Random Forest	96.73	97.12	95.01	91.32	+4.01
Artificial Neural Network	92.88	95.64	88.72	87.60	+10.09

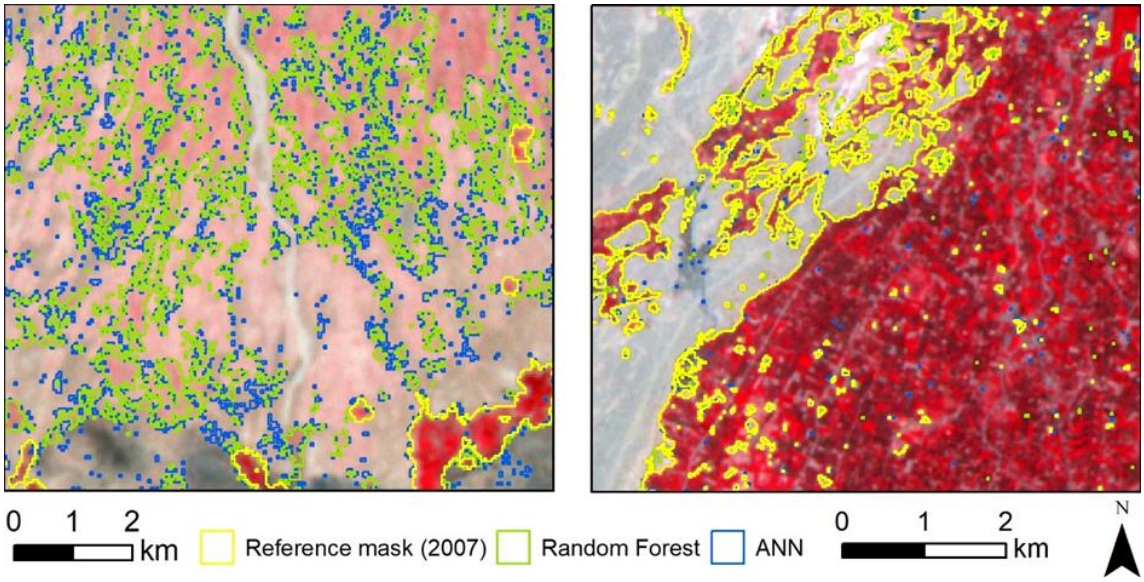

Figure A-2. False colour DMC imagery (NIR, R, G at 32 m) from 27 April 2007 for an (left) natural vegetation dominated area and (right) main block of agricultural area in the Helmand valley with corresponding agriculture delineation using Random Forest and Artificial Neural Network (ANN) approaches.

The accuracy has remained above 88% for all metrics used across both machine learning techniques, but the agricultural area estimate is up to 10% greater than the reference 2007 agricultural mask. The visual agricultural land delineation shown in Figure A-2 provides further understanding on where the classification approaches have worked well and performed poorly. Both classifications have performed well in well-defined agricultural areas (Figure A-2), where there is little difference between the reference data and predicted area. The main source of confusion to both classifications is the presence of natural vegetation, which is a known source of confusion for unsupervised classification approaches caused by the spectral similarity between agriculture and natural vegetation (Simms *et al.*, 2014). Human interpreters can easily separate between agriculture and natural vegetation because of visual differences in the structures of both land cover types.

A.3 Summary

The use of data-driven machine learning algorithms for agricultural land delineation have shown promising results with high classification accuracy for

Random Forest and Artificial Neural Network (>92%). Well-defined agricultural areas have tight delineations close to the reference data, but these approaches are unable to separate difficult interpretation cases (e.g. natural vegetation) from visual inspection. Human interpreters can separate these confusion locations by observing the areas in context, as natural vegetation displays different landscape characteristics than agriculture. Providing contextual information of these agricultural features to data-driven approaches may remove the confusion between these two land cover types. Convolutional Neural Networks use convolutional filters to extract encoded features from input imagery (e.g. field boundaries and structures) and have the potential to improve agricultural delineation. Validation metrics often only present global accuracy and ignore the spatial distribution of these errors across the classification. The size of the area used to validate image classifications needs to be further explored to further evaluate other data-driven image classification methodologies. Accurate agricultural land delineation is essential for monitoring opium cultivation and further investigation is required to overcome these sources of classification error from machine learning algorithms to perform as well as a human interpreter.

Appendix B Localised intersection over union for Helmand Province in 2009

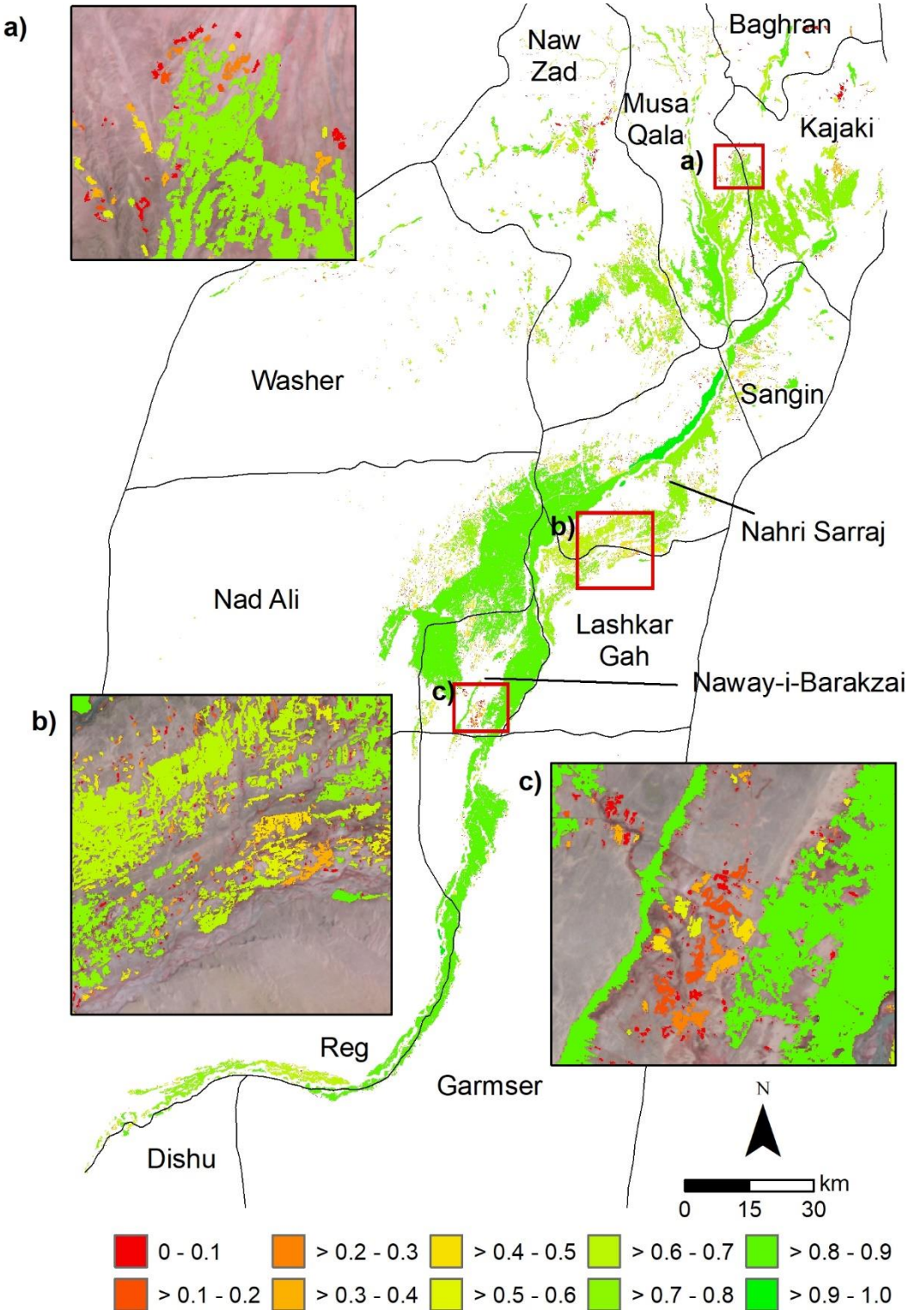

Figure B-1. Localised Intersection over Union for IR-MAD normalised DMC imagery in 2009 across the Helmand Province, Afghanistan extent used in Chapter 3

Appendix C Agricultural expansion in Helmand Province between 2010 and 2019

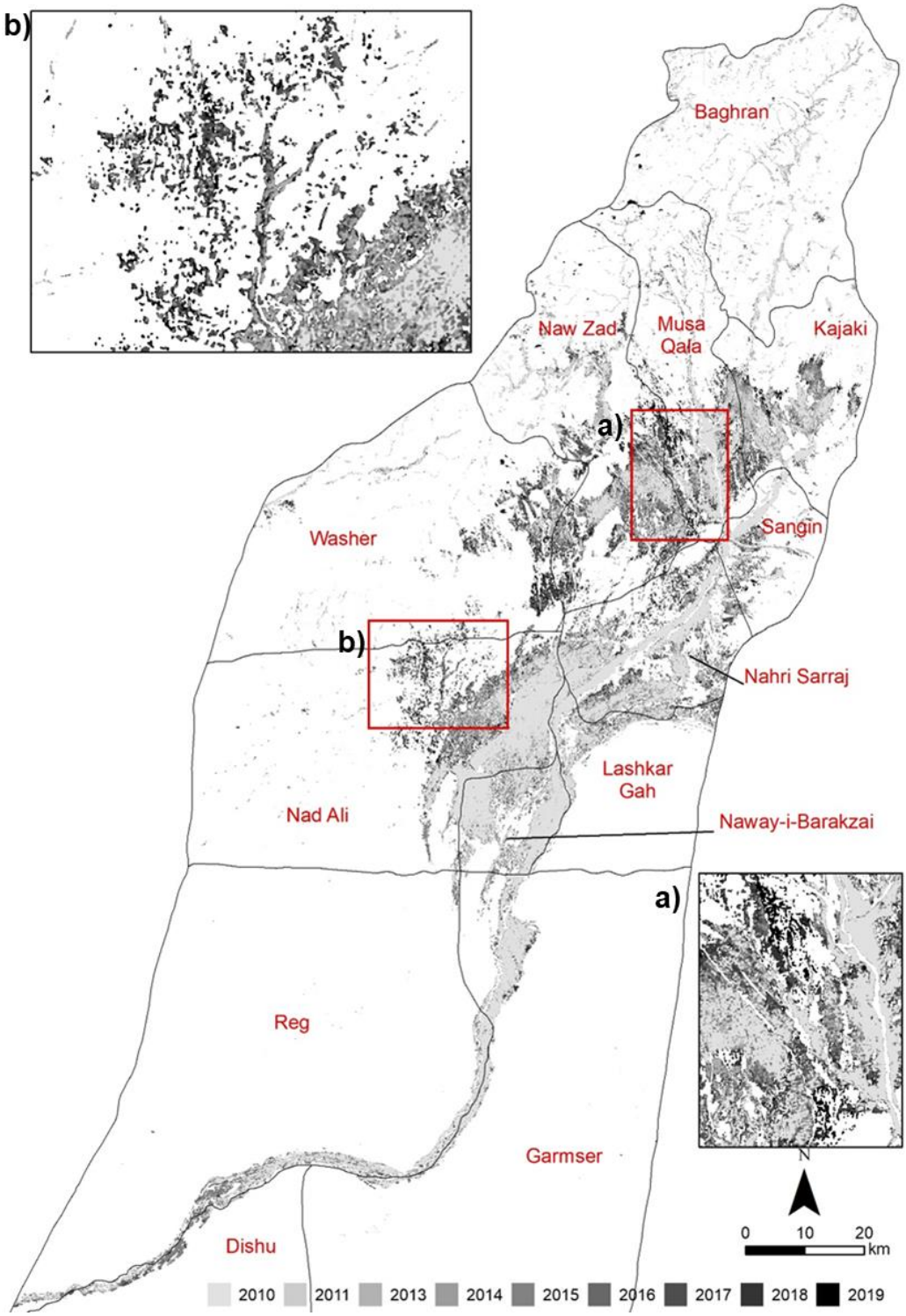

Figure C-1. Agricultural land expansion between 2010 and 2019 in Helmand Province, Afghanistan in Chapter 4