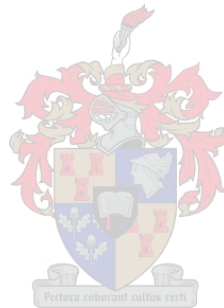


IMPROVING THE GENERALISABILITY OF A DEEP LEARNING MODEL  
FOR GLOBAL FOREST CLASSIFICATION THROUGH IMAGE  
NORMALISATION, ENHANCEMENT AND AUGMENTATION

By MICHAEL SWAINE



SUPERVISOR: DR. Z MÜNCH  
Department of Geography & Environmental Studies  
Stellenbosch University

March 2023

DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL STUDIES

## **DECLARATION**

By submitting this thesis/dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

March 2023

## SUMMARY

Effectively managing global forest resources, under threat from climate change, deforestation and fragmentation, requires the efficient extraction of a global tree cover dataset. The purpose of this research was to identify image enhancement and data augmentation methods that would improve the generalisability of a deep learning model for the classification of global tree cover.

In the first experiment we aimed to improve the accuracy of a deep learning model for global forest classification using Sentinel 2 optical data. We present several image enhancement methods widely used in natural image classification and biomedical imaging domains, including histogram equalisation (HE), contrast limited adaptive histogram equalisation (CLAHE) and global contrast normalisation (GCN), as pre-processing steps. The enhancement methods were compared with each other on a per biome basis, and both training and validation regions were selected to represent the heterogeneity within biomes. Selected images were captured within the local optimal foliage growing season and contained minimal or no clouds. A U-Net convolutional neural network model was trained for each enhancement per biome and used to perform inference on validation images for each of the corresponding biomes and enhancements. Random stratified samples were collated for all validation images per biome per enhancement for statistical analysis. Only GCN and CLAHE RGB returned higher means than the baseline dataset. The results showed that GCN most consistently improved classification results for tree cover across biomes, possibly due to the standardization of contrast levels of the training and validation images.

In the absence of accurately annotated training data for tree segmentation, training a robust, deep learning model for global tree cover classification remains a challenge. As its first objective, experiment 2 evaluated basic data augmentation methods and prediction frameworks that might lead to achieving an accurate, global tree cover classification. A training dataset was artificially inflated using common geometric and colour data augmentation methods borrowed from the computer vision domain. Their effectiveness in improving the generalisability of a U-Net model for tree classification was tested. Both geometric and colour augmentations, when applied individually, showed improvements in model accuracy. When applied together, the combined augmentations showed only marginal improvements over the individually applied augmentations. The second objective was to test two approaches towards achieving a global tree classification. The first was a model per biome approach, whereby a model was trained with data derived only from the respective biome. The second involved training a single globally representative model with training data from all biomes combined. This resulted in higher MCC scores than the multi-model approach. The diversity in training data appeared to increase model robustness. Thus, it was

found that training a single, globally representative model with a combination of colour and geometric augmentations led to an effective framework to infer a global tree classification.

#### KEY WORDS

Global tree cover, deep learning, U-Net, image enhancement, data augmentation, Sentinel-2

## OPSOMMING

Om woude regoor die wêreld, wat gebuk gaan onder die bedreiging van klimaatsverandering, ontbossing en fragmentasie, effektief te bestuur, vereis die doeltreffende skepping van 'n wêreldwye boombedekkingdatastel. Die doel van hierdie navorsing is om beeldverbetering en datavergrotingsmetodes te identifiseer wat die veralgemeenbaarheid van 'n diepleermodel vir die klassifikasie van wêreldwye boombedekking sal verbeter.

In die eerste eksperiment het ons daarop gemik om die akkuraatheid van 'n diepleermodel vir wêreldwye woudeklassifikasie te verbeter deur Sentinel-2 optiese data te gebruik. Ons bied verskeie beeldverbeteringsmetodes aan wat wyd gebruik word in natuurlike beeldklassifikasie en biomediese beeldingsdomeine, insluitend histogramgelykmaking (HE), kontrasbeperkte aanpasbare histogramgelykmaking (CLAHE) en globale kontrasnormalisering (GCN), as voorverwerkingstappe. Die verbeteringsmetodes is met mekaar vergelyk op 'n per-bioom basis, en beide opleiding- en validasiestreke is gekies om die heterogeniteit binne biome te verteenwoordig. Gekose beelde is binne die plaaslike groeiseisoen vasgelê en het geen of baie min wolke bevat. 'n U-Net konvolusionele neurale netwerkmodel is opgelei vir elke beeldverbetering per bioom en is gebruik om afleidings oor validasiebeelde vir elk van die ooreenstemmende biome en verbeterings uit te voer. Ewekansige gestratifiseerde steekproewe is vir alle validasiebeelde per bioom en per beeldverbetering vir statistiese analise ingesamel. Slegs GCN en CLAHE RGB het hoër gemiddeldes as die basislyndatastel opgelewer. Die resultate het getoon dat GCN die klassifikasie van boombedekking oor biome konsekwent verbeter het, moontlik as gevolg van die standaardisering van kontrasvlakke van die opleiding- en validasiebeelde.

In die afwesigheid van akkuraat geannoteerde opleidingsdata vir boomsegmentering, bly die opleiding van 'n robuuste, diepleermodel vir wêreldwye boombedekkingklassifikasie 'n uitdaging. As 'n eerste doelwit het die tweede eksperiment basiese datavergrotingsmetodes en voorspellingsraamwerke geëvalueer wat kan lei tot 'n akkurate, wêreldwye klassifikasie van boombedekking. 'n Opleidingsdatastel is kunsmatig vergroot deur gebruik te maak van algemene meetkundige en kleurdatavergrotingsmetodes wat van die rekenaarvisiedomein geleen is. Hulle doeltreffendheid in die verbetering van die veralgemeenbaarheid van 'n U-Net-model vir boomklassifikasie is getoets. Beide geometriese en kleurdatavergrotings het verbeterings in modelakkuraatheid getoon wanneer dit individueel toegepas is. Wanneer dit saam toegepas is, het die gekombineerde aanvullings slegs minimale verbeterings getoon teenoor die individueel toegepaste aanvullings. Die tweede doelwit was om twee benaderings tot die bereiking van 'n wêreldwye boomklassifikasie te toets. Die eerste benadering tot die probleem was een model per

bioom, waar die model opgelei is met data wat slegs van die onderskeie bioom verkry is. Die tweede benadering behels die opleiding van 'n enkele wêreldwye verteenwoordigende model met opleidingsdata van alle biome gekombineer. Dit het gelei tot hoër MCC-tellings as die multi-model benadering. Dit blyk dat die diversiteit in opleidingsdata die robuustheid van die model verhoog het. Dit is daarom bevind dat die opleiding van 'n enkele, wêreldwyd-vertteenwoordigende model met 'n kombinasie van kleur en geometriese datavergrotings gelei het tot 'n effektiewe raamwerk om 'n wêreldwye boomklassifikasie te skep.

## SLEUTELWOORDE

Wêreldwye boombedekking, diepleer, U-Net, beeldverbetering, datavergroting, Sentinel-2

## **ACKNOWLEDGEMENTS**

I sincerely thank:

- My supervisor, Dr. Zahn Münch for her continued support, guidance and encouragement;
- My family for always re-enforcing positivity, encouragement and specifically my wife, Ursula for giving birth to our son in the middle of this research thesis;
- Leigh Lotter for her unwavering support, advice and presentation coaching and for making the long after hours in the office more bearable;
- Robert Stevenson for his incredible editing efforts;
- LuxCarta for the opportunity and funding to undertake this research;
- Jano Barnard for his translation abilities;
- My fellow colleagues at LuxCarta for their encouragement.

## TABLE OF CONTENTS

<b>DECLARATION .....</b>	<b>i</b>
<b>SUMMARY .....</b>	<b>ii</b>
<b>OPSOMMING .....</b>	<b>iv</b>
<b>ACKNOWLEDGEMENTS.....</b>	<b>vi</b>
<b>TABLE OF CONTENTS.....</b>	<b>vii</b>
<b>LIST OF TABLES .....</b>	<b>x</b>
<b>LIST OF FIGURES .....</b>	<b>xi</b>
<b>ACRONYMS AND ABBREVIATIONS.....</b>	<b>xiii</b>
<b>1 CHAPTER 1: Introduction .....</b>	<b>1</b>
1.1 GLOBAL FOREST CLASSIFICATION .....	1
1.1 PROBLEM STATEMENT .....	3
1.2 AIMS AND OBJECTIVES .....	5
1.3 SIGNIFICANCE AND RATIONALE .....	5
1.4 RESEARCH METHODOLOGY AND DESIGN .....	6
<b>2 CHAPTER 2: Literature Review .....</b>	<b>9</b>
2.1 INTRODUCTION.....	9
2.2 REMOTE SENSING .....	9
2.2.1 ACTIVE AND PASSIVE REMOTE SENSING .....	9
2.2.2 SENSOR RESOLUTION .....	10
2.3 DATA AND GLOBAL LAND COVER.....	10
2.3.1 SENTINEL-2 OPTICAL SENSORS.....	10
2.3.2 GLOBAL LAND COVER.....	12
2.3.3 LARGE-SCALE LAND COVER CHALLENGES .....	13
2.3.4 BIOME APPROACH .....	13
2.4 IMAGE CLASSIFICATION .....	14
2.4.1 PIXEL-BASED CLASSIFICATION.....	14
2.4.2 OBJECT-BASED IMAGE ANALYSIS.....	14
2.4.3 CLASSICAL MACHINE LEARNING.....	14
2.4.4 DEEP LEARNING.....	15
2.5 IMAGE ENHANCEMENTS .....	19
2.6 DATA AUGMENTATION .....	21



2.7	SAMPLING METHODOLOGY .....	23
2.8	ACCURACY ASSESSMENT .....	24
2.9	SUMMARY OF LITERATURE .....	25
3	<b>CHAPTER 3: Image enhancement methods for improving the results yielded by a CNN based global forest classification model using Sentinel-2 optical data .....</b>	<b>27</b>
3.1	ABSTRACT .....	27
3.2	INTRODUCTION.....	27
3.2.1	GLOBAL FOREST MAPPING.....	28
3.2.2	MACHINE AND DEEP LEARNING .....	29
3.2.3	IMAGE ENHANCEMENT .....	30
3.3	DATA AND PROCESSING CHARACTERISTICS.....	32
3.3.1	SENTINEL-2 MULTI-SPECTRAL DATA.....	32
3.3.2	BASE IMAGE .....	32
3.3.3	TRAINING AND VALIDATION DATA SOURCING .....	32
3.3.4	COMPUTING RESOURCES .....	34
3.4	METHODS .....	35
3.4.1	IMAGE ENHANCEMENTS.....	35
3.4.2	DEEP LEARNING.....	40
3.4.3	ACCURACY AND STATISTICS.....	43
3.5	RESULTS & DISCUSSION.....	44
3.5.1	EXAMPLE OF ENHANCEMENTS .....	44
3.5.2	EFFECT OF ENHANCEMENT (MAD) .....	45
3.6	CONCLUSION.....	54
4	<b>CHAPTER 4: Data Augmentation methods for improving the result of a CNN based global forest classification model from Sentinel-2 optical data .....</b>	<b>56</b>
4.1	ABSTRACT .....	56
4.2	INTRODUCTION.....	56
4.2.1	CLASSIFICATION APPROACHES.....	57
4.2.2	DEEP LEARNING.....	57
4.2.3	BIOME BASED CLASSIFICATION .....	59
4.2.4	OVERVIEW .....	60
4.3	METHODS .....	61

4.3.1	DATA.....	61
4.3.2	EXPERIMENTAL FLOW .....	62
4.3.3	DATA AUGMENTATION.....	63
4.3.4	DEEP LEARNING MODEL.....	66
4.3.5	MULTI-MODEL VS. SINGULAR GLOBALLY TRAINED MODEL .....	67
4.3.6	ACCURACY AND STATISTICS.....	68
4.3.7	COMPUTING RESOURCES AND SOFTWARE LIBRARIES.....	70
4.4	RESULTS & DISCUSSION.....	70
4.4.1	EFFECTS OF AUGMENTATION .....	70
4.4.2	COMPUTATIONAL COST.....	76
4.4.3	MULTI-MODEL VS. SINGULAR GLOBAL MODEL .....	77
4.5	CONCLUSION.....	79
5	CHAPTER 5: DISCUSSION & CONCLUSION.....	81
5.1	REVISITING THE AIMS AND OBJECTIVES.....	81
5.2	SYNTHESIS: MAIN FINDINGS AND VALUE OF RESEARCH.....	82
5.3	STUDY LIMITATIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH.....	84
5.4	CONCLUSION.....	85
	REFERENCES .....	86

## LIST OF TABLES

Table 1: Spectral bands for the Sentinel-2 sensors (source: <a href="https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument">https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument</a> ).....	11
Table 2 Hyper parameters used by the UNet used in this study .....	41
Table 3 Example of enhancements applied to a false colour composite.....	44
Table 4 Overall MAD score and standard deviation across all training and test images per enhancement .....	46
Table 5 Mean MCC and OA per biome per enhancement, bold indicating cases outperforming no enhancement .....	49
Table 6 Example result of best performing enhancement for each biome .....	51

## LIST OF FIGURES

Figure 1 Graphical representation of the research design .....	7
Figure 2 Illustration of a multilayer perceptron .....	16
Figure 3 Visualising the architecture of a CNN (Sarker 2021) .....	17
Figure 4 Visual depiction of a U-Net architecture .....	18
Figure 5 Global distribution of training and validation locations .....	34
Figure 6 Experiment workflow .....	35
Figure 7 a) Histogram of image with no enhancement applied b) Histogram equalisation applied .....	36
Figure 8 a) Histogram of image with no enhancement applied b) Histogram of image with global contrast normalization applied. ....	38
Figure 9 a) Histogram of image with no enhancement applied b) Histogram of image with CLAHE applied on each channel in the RGB colour space .....	39
Figure 10 a) Histogram of image with no enhancement applied b) Histogram of image with CLAHE applied on V channel in HSV colour space .....	40
Figure 11 Graphical representation of the UNet architecture .....	41
Figure 12 Mean absolute residual difference (MAD) for each test and training image.....	46
Figure 13 Mean MCC score per enhancement per biome.....	47
Figure 14 Mean MCC across all biomes .....	53
Figure 15 Experiment flow, part one .....	63
Figure 16 Geometric transformations .....	64
Figure 17 Colour transformations .....	65
Figure 18 Shear transformation.....	66
Figure 19 Experiment flow, inference strategy .....	67
Figure 20 Global MCC by augmentation method.....	71
Figure 21 MCC of each test image per augmentation method.....	72
Figure 22 Visual effect of model applied with colour augmentations (c) vs. no augmentations (d) .....	73
Figure 23 Visualising the difference in model trained with colour augmentations (c) vs. geometric-with-shear augmentations (d).....	74
Figure 24 Visualising the difference between model trained with basic geometric augmentations (c) and shear with geometric augmentations (d) .....	74
Figure 25 Normalised difference in MCC between no augmentation and applied augmentations .....	75

Figure 26 Computation cost vs MCC.....	76
Figure 27 MCC of biome-specific model approach vs globally-trained model approach .....	77
Figure 28 Visual prediction of globally-trained model on Savanna biome vs Savanna-specific model prediction.....	78

## ACRONYMS AND ABBREVIATIONS

Artificial Neural Network	ANN
Central Processing Unit	CPU
Computer Vision	CV
Contrast Limited Adaptive Histogram Equalisation	CLAHE
Convolutional Neural Networks	CNN
Decision Trees	DT
Deep Learning	DL
Digital Number	DN
European Space Agency	ESA
False Negatives	FN
False Positives	FP
Fully Convolutional Networks	FCN
Generative Adversarial Network	GAN
Global Contrast Normalisation	GCN
Global Human Settlement	GHS
Graphical Processing Unit	GPU
Histogram Equalisation	HE
Hue, Saturation and Value	HSV
Intersection Over Union	IoU
Land Use Land Cover	LULC
Level-1-C	L1C

Level-2-A	L2A
Matthews Correlation Coefficient	MCC
Maximum Likelihood Classifier	MLC
Mean Absolute Difference	MAD
Multi-Layer Perceptron	MLP
Near Infrared	NIR
Neural Network	NN
Object Based Image Analysis	OBIA
Overall Accuracy	OA
Random Forest	RF
Rectified Linear Unit	ReLU
Remote Sensing	RS
Sentinel-2 Global Land Cover	S2GLC
Support Vector Machines	SVM
Synthetic Aperture Radar	SAR
True Negatives	TN
True Positives	TP
Universal Transverse Mercator	UTM
Very High Resolution	VHR

# **1 CHAPTER 1: INTRODUCTION**

## **1.1 GLOBAL FOREST CLASSIFICATION**

Remotely sensed data is required to fulfil a variety of decision-making needs. Deriving land cover data from satellite imagery is an application driven by natural resource management, land use policies, urban planning, and agricultural monitoring, among many others (DeFries, Foley & Asner 2004). Historically, extracting meaningful land cover data over large scales has been difficult due to the lack of available medium-resolution imagery.

Global land cover is a relatively new concept. While the first land observation satellite was only launched in 1972 (Gong et al. 2016), multiple global land cover products have been released at 1 km, 300 m and 30 m (Chen et al. 2015). However, the accuracy of these datasets is considered unsatisfactory for many applications (Chen et al. 2015; Gong et al. 2016). Recent technological improvements in satellite imaging sensors have resulted in a surge of freely available optical data at a medium spatial resolution (10 m).

The most notable contribution has been the introduction of the European Space Agency's (ESA) Copernicus programme, providing global coverage of higher resolution optically sensed data at a high frequency (3–5-day revisit period) via their constellation of Sentinel-2 satellites (Roy et al. 2017). Consequently, there has been increased interest in accurate algorithms to extract meaningful information from the vast amount of available satellite imagery (Buchhorn et al. 2020; Shendryk et al. 2019). The Copernicus Global Land Service (CGLS) Land Cover Collection 2 dataset (Buchhorn et al. 2020) consists of a 20-class, global discrete land cover map at 100m spatial resolution with overall accuracy around 80% (Buchhorn et al. 2020). Despite being useful for a wide range of applications such as forest and crop monitoring, biodiversity, conservation and climate modelling, this product is more suited to deducing general global trends, where fine-grained detail is less important. By exploiting the highest spatial resolution (10 m) available, the Sentinel 2 Global Land Cover (S2GLC) project (Lewinski 2017; 2019) delivers an automated pixelwise land cover classification using a multi-temporal test strategy at global scale, which has led to the release of the ESA World Cover 2020 (WC), annually updated (Zanaga et al. 2021) with an accuracy of 74.4% and 76.7% for WC 2020 v100 and WC 2021 v200 respectively. Both Google (Brown et al. 2022) and Esri (Karra et al. 2021) have also released global Sentinel-based 10 m landcover maps.

Increased computational power and the decreasing cost of hardware have promoted the use of machine learning in remote sensing. Machine learning is an effective empirical methodology for classification and regression problems (Lary et al. 2016). The most used types of supervised



machine learning algorithms are support vector machines (SVM), decision trees (DT) and random forests (RF) (Lary et al. 2016). RF has proven effective in land use and land cover classification (Buchhorn et al. 2020) as well as forest classification (Betts et al. 2017). For example, the RF algorithm was used to produce the global ESA WC product (Zanaga et al. 2021) from 141000 hand-labelled samples. RF produces higher accuracies than standard decision trees (Rodriguex-Galiano et al. 2012) due to the RF ensemble architecture where several classification trees are trained on a subgroup of training data (Breiman 2001). In contrast, SVM is a supervised machine learning algorithm which aims to find the optimal hyperplane to cluster training data into separate, predefined classes. A major advantage of SVM is the ability to classify data successfully with minimal input data samples, which makes SVM an attractive classifier in the remote sensing domain (Kulkarni & Lowe 2016).

Traditional classification approaches require expert knowledge to craft features as inputs to classifier models and have shown satisfactory results in local scale classification problems (Abdi 2020; Mirończuk & Hościło 2017; Noi & Kappas 2017). However, these methods are difficult to generalise on a global scale due to the intra-class variance of land cover classes, both temporally and spatially (McDermid, Franklin & LeDrew 2005; Woodcock et al. 2001).

Historically, land cover classification methods used the spectral value of each pixel (pixel-based classification), however assigning a class label to a pixel, based on spectral values only, can lead to noisy output classes due to intra-class variability (Tong et al. 2018). Texture metrics have been used in such pixel-based image analyses to gain additional spatial information (Khatami, Mountrakis & Stehman 2016). Object-based image analysis (OBIA) addresses the problem of noise by partitioning an image into segments based on pixel similarity. OBIA, therefore, allows richer spatial feature sets to be derived such as distance, topology and object shape (Blaschke 2010).

Within the last decade, deep learning (DL), a subset of machine learning, has achieved success in the field of Computer Vision (CV), particularly the medical imaging field, for image classification and segmentation tasks (Ronneberger, Fischer & Brox 2015). This has consequently led to the use of DL in remote sensing (Helber et al. 2019). Since DL can recognise contextual features through the built-in hierarchical structure of Convolutional Neural Networks (CNN), the popularity of DL, and more specifically, CNN for land cover classification has increased (Tong et al. 2018; Helber et al. 2019). In contrast to traditional classifiers, DL can automatically create and learn its own features (Ball, Anderson & Chan 2017). DL has therefore become more widely used in remote sensing for performing classification tasks, image fusion, image registration, scene classification, object detection and OBIA (Ma et al. 2019; Mahdianpari et al. 2018; Shendryk et al. 2019). CNNs

have proved effective in classifying high-resolution images into multiple land cover classes at accuracies over 85% (Maggiori et al. 2017; Yang, Rottensteiner & Heipke 2018).

Despite the advantages of DL and CNNs, the adoption of this classification method for a global land cover classification has been limited, and only as recently as 2021, have global land cover datasets, developed from DL models, become available (Brown et al. 2022; Karra et al. 2021). Remote sensing studies have predominantly applied DL to object extraction using very high spatial resolution satellite imagery since objects such as buildings show similarities to objects classified in CV (Mahdianpari et al. 2018).

Limitations to a global land cover classification have stemmed from intra-class heterogeneity of land cover classes around the world (Helber et al. 2019). In addition, satellite images are often affected by environmental conditions which can lead to reduction in spectral contrast, resulting in loss of detail. Image enhancement is the process of increasing the suitability of an image for a specific application (Shyam et al. 2017). Image enhancement algorithms have successfully been applied to imagery to improve visual representation (Demirel, Ozcinar & Anbarjafari 2010). Commonly used enhancement techniques include Histogram Equalisation (HE), Contrast Limited Adaptive Histogram Equalisation (CLAHE) and Global Contrast Normalisation (GCN) (Jadoon et al. 2017; Maini & Aggarwal 2010).

DL models require very large training datasets to incorporate variability within classes (Yu et al. 2017) and reduce overfitting, which would increase model generalisability (Shorten & Khoshgoftaar, 2019). Typical large DL training datasets, such as the ImageNet dataset (Krizhevsky, Sutskever & Hinton 2012), contain 15 million images, categorised into 22000 classes. Such large datasets have not yet been available in the remote sensing domain (Yu et al. 2017), however, the recently released global DynamicWorld (Brown et al. 2022) and Esri Land Cover (Karra et al. 2021) products were trained on over five billion Sentinel-2 pixel patches (Venter et al. 2022). One way to increase and enrich the training dataset for DL is to apply data augmentation to the initial training dataset. Several augmentation strategies have been successfully used to artificially increase the size of training sets, for example, geometric and colour transformations (Shorten & Khoshgoftaar 2019).

## **1.1 PROBLEM STATEMENT**

Global land cover datasets hold immense value for many applications, including natural resource management, and urban and agricultural monitoring. However, they are difficult to produce at a very high accuracy due to variable atmospheric and environmental conditions worldwide.

DL has shown promise in land cover classification as it combines the computational effectiveness of pixel-based techniques while still capitalizing on the inherent value of neighbourhood and contextual information. Little research currently exists which outlines the use of DL-based algorithms for medium-resolution (10 m), global tree classification. Two broad strategies are proposed to increase the accuracy of a DL model, namely image enhancement and normalisation, and training data augmentation.

Image enhancement and normalisation methods have been successfully used to improve the accuracies of DL models in the medical imaging domain, specifically those aiming to enhance the contrast of the target feature. Image normalisation techniques such as global contrast normalisation have shown to be effective at improving the performance and generalisability of DL models in natural image categorisation (Jadoon et al. 2017; Maini & Aggarwal 2010). To the best of the authors' knowledge, no comparison studies were found that experimented with the application of basic image enhancement and normalisation techniques for the case of tree classification.

One of the limiting factors in effectively training a DL model, is the lack of adequate training data, both in quantity and diversity to enable a model to generalise well on validation data. The difficulty is exacerbated with the task of pixelwise classification, where the production of such training data is very expensive. Data augmentation provides the capability of artificially inflating and increasing the diversity of a small training dataset (Shorten & Khoshgoftaar 2019). Data warping image transformations such as geometric and colour transformations are common data augmentation strategies. A research gap has been identified as no studies have thus far been done comparing data augmentation methods for a tree classification by using a DL model.

The classification of land cover over large areas is very challenging, particularly as the effects of differing ecological and climatic conditions, can negatively affect the classification accuracies derived by using classical machine learning methods (Woodcock et al. 2001). To reduce the variability caused by these factors, land cover classification can be processed in regions based on similar climatic, vegetation and ecological traits (Souza et al. 2020). Thus far, no studies have evaluated whether a DL model can generalise well across biomes. From these apparent research gaps, the following research questions were formulated:

1. Which image enhancement and normalisation methods will increase the accuracy of a deep convolutional neural network the most?
2. Can geometric and colour transformation data augmentation methods increase the generalisability of a deep convolutional neural network to train a single model for a global application?

3. Does a multi-model approach (a model per biome) yield a higher classification accuracy than a single-model approach?

## **1.2 AIMS AND OBJECTIVES**

The study aims to determine which image normalisation and enhancement algorithms or combinations thereof, applied to the input data of a convolutional neural network, can improve the accuracy of a land cover classification, and to determine whether data augmentation can be used to generalise a deep learning model to infer a global forest classification.

The following research objectives have been set to achieve the research aims:

1. Perform a literature review to cover topics surrounding forest mapping, deep learning, image enhancement and data augmentation.
2. Compile a labelled training and validation dataset of trees compiled for each biome that contains sufficient amounts of tree cover (Tropical Rainforest, Temperate Forest, Boreal Forest, Grassland, Savanna) from LuxCarta's expansive catalogue of land cover data.
3. Determine the enhancement/normalisation and combinations thereof which improves the accuracy of a model the most (Experiment 1).
4. Identify the most effective data augmentation methods and combinations to apply to the result from Experiment 1 as base data for improved classification accuracy (Experiment 2).
5. Compare and evaluate a local (model per biome) vs. global (single model) approach for the generalisability of forest classification to a global scale by applying enhancements (Experiment 1) and augmentations (Experiment 2) to training and test data (Experiment 3).

## **1.3 SIGNIFICANCE AND RATIONALE**

The recent availability of free, medium spatial resolution satellite data calls for the timely extraction of meaningful information from the imagery on a global scale. There is still a need for a high-resolution (10 m), global land cover dataset for radio propagation modelling and a realistic terrain dataset for virtual simulation environments (LuxCarta 2022). A high-speed, accurate and robust deep learning model is needed. Advances in hardware technology have led to CNNs becoming a popular choice to solve computer vision problems. Successful completion of this study can improve the overall classification accuracy and generalisability of a deep learning model by improving the quality and quantity of input data through various normalisation, enhancement and augmentation techniques. In this study, the aims and objectives have been specifically designed to test whether image enhancement, normalisation and augmentation methodologies can be used to

enable the training of a robust, universally applicable forest classification model. A low cost-time function is an important consideration, therefore a multi-temporal approach has been ruled out as has the use of more than three input channels (red, green, near infrared channels). Currently, no studies have been conducted to determine whether image enhancement and normalisation, along with data augmentation, can improve the accuracy and generalisability of deep convolutional models for a global forest classification with a spatial resolution of 10 m.

#### **1.4 RESEARCH METHODOLOGY AND DESIGN**

A research approach following a positivistic paradigm, employing primary quantitative data was adopted to conduct the research. An experimental approach consisting of three experiments was followed as outlined in Figure 1. The research agenda informs the workflow of this thesis and the contents of each chapter, with this chapter (Chapter 1) describing the proposal phase.

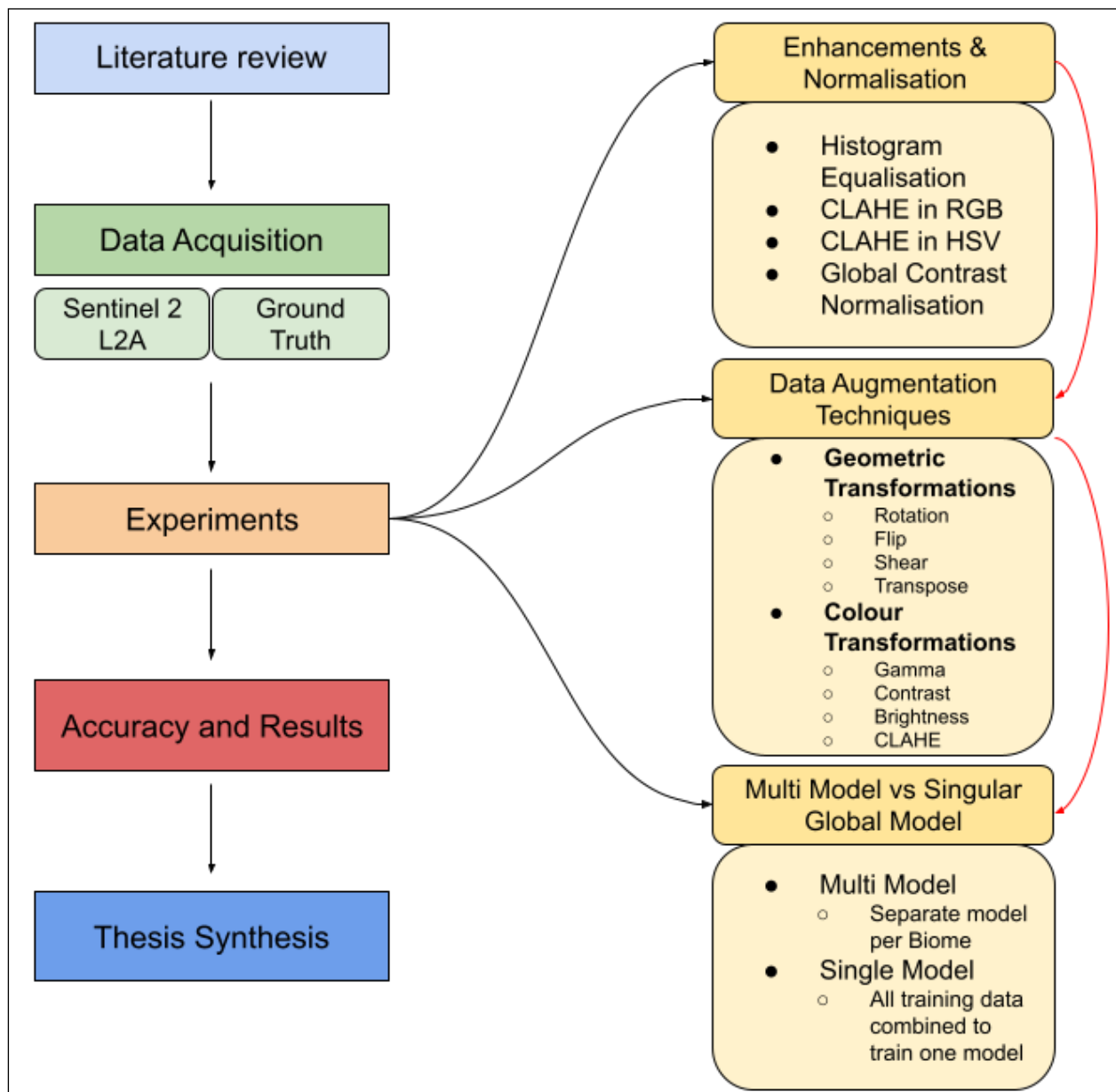


Figure 1 Graphical representation of the research design

The first experiment, described in Chapter 3, identifies image enhancement and normalisation methods which could improve classification accuracies. Experiment 2 uses the outcomes from Experiment 1 to improve model generalisability by testing a series of data augmentation methods in the model training phase. Finally, in Experiment 3, the optimum data enhancement and augmentation methods identified in Experiment 1 and Experiment 2, are used to train and compare classification accuracies of a multi-model (one model per biome) approach with a singular, all-inclusive global model trained on all combined training data from the multi-model approach. Chapter 4 addresses Experiment 2 and Experiment 3, comparing the accuracies of the model per biome approach, with the single globally representative model trained on data from all biomes combined. Chapter 5 provides a synthesis of the research and draws the thesis to a close. This chapter also highlights limitations of the study and makes recommendations for future research.

Having concluded Chapter 1, which introduced the research problem, aim and objectives, Chapter 2 provides background literature on remote sensing, deep learning, image enhancement and normalisation methods as well as data augmentation techniques.

## **2 CHAPTER 2: LITERATURE REVIEW**

### **2.1 INTRODUCTION**

Forests provide a range of important ecosystem services including timber production, biodiversity protection and climate change mitigation through carbon storing (Kornatowska & Sienkiewicz 2018). These resources are under severe pressure due to anthropogenic influences. Effective management of global forest resources requires the ability to efficiently extract a global tree cover dataset. Classifying forests accurately at a global scale and at a high resolution remains a major challenge (Sjöqvist, Långkvist & Javed 2020). Advances in computer vision (CV) have paved the way for the adoption of advanced image processing techniques, such as deep learning (DL), in other realms, including remote sensing (RS).

There has been growing use of DL algorithms within the RS field as they have shown to outperform more traditional machine learning classifiers. Specifically, DL algorithms have shown to be able to better extract meaningful information for RS imagery. However, DL models require large amounts of data to effectively train robust models, and there is still a major lack in sufficient amounts of accurately annotated training data. To overcome this problem, data augmentation techniques adopted from the CV domain have been applied in RS to enhance the size and diversity of existing training datasets. Few studies however have investigated the effectiveness of these augmentation methods within the RS domain.

### **2.2 REMOTE SENSING**

RS is the process of detecting the emitted electromagnetic radiation reflected by objects on the surface of the Earth. The data is typically captured at a distance, commonly from aircraft or satellite mounted sensors. The reflected radiation can be used to gain knowledge of the characteristics and spectral signatures of surface features (Campbell & Wynne 2011).

#### **2.2.1 ACTIVE AND PASSIVE REMOTE SENSING**

The source of the electromagnetic energy that is recorded by a sensor determines the type of RS taking place. Active remote sensing relies on the source energy be transmitted from the sensor itself, then recording the surface response of this energy. Light Detection and Ranging (LiDAR) is a common form of active RS. Passive remote sensing relies on the naturally occurring electromagnetic energy generated by the sun. The reflection of this light is subsequently captured and recorded by the sensor. Optical sensors rely on passive remote sensing (Campbell & Wynne 2011).



### **2.2.2 SENSOR RESOLUTION**

Sensor resolution refers to the amount of information a sensor is designed to capture. Resolution can be subdivided into four variables which characterises the amount detail that can be distinguished from an image. These are the spatial, temporal, spectral and radiometric resolutions (Campbell & Wynne 2011).

Spatial resolution refers to the size of the smallest object that is detected from a sensor (Chuvieco 2016). This generally relates to the pixel size. Spatial resolution is often referred to as the ground sampling distance and is measured in kilometres for coarse resolutions down to sub meter resolution for very high resolution (VHR) data.

Temporal resolution, also known as the revisit time, refers to the frequency of data acquisition of a given geographic point on the Earth's surface (Campbell & Wynne 2011). The main factors which affect temporal resolution are swath-width, satellite orbit altitude and speed. A high temporal resolution is desirable for capturing data more frequently, especially in tropical regions of the world where cloud cover can be a hindrance. Thus, a higher temporal resolution provides more opportunity of capturing cloud free data. High temporal resolution is favoured for applications requiring constant monitoring of rapidly changing environments such as those in natural disaster monitoring. A daily revisit is considered a high temporal resolution, whereas revisit periods of once every 28 days is considered low.

Spectral resolution generally refers to the number of bands a sensor is able to capture. The bands are defined by a range in specified electromagnetic wavelengths. The reflection of radiance off features on the Earth's surface interact at different wavelengths. Bands at different wavelengths can capture complimentary information and thus can help separate more difficult to distinguish classes, such as characterising different tree species (Immitzer, Atzberger & Koukal 2012).

The amount of data a sensor is able to capture with regards to luminance describes the radiometric resolution of the sensor. The radiometric resolution is indicated as a bit-depth level (Campbell & Wynne 2011). Typically, modern satellites sensors have a bit depth ranging from 8bit to 16bit. An image with a bit depth of 8bit can have a range of 256 different digital number (DN) values. 16bit data can have a range of up to 65 536 DN values.

## **2.3 DATA AND GLOBAL LAND COVER**

### **2.3.1 SENTINEL-2 OPTICAL SENSORS**

Recent technological improvements in satellite imaging sensors have resulted in a surge of freely available optical data at a medium spatial resolution (ten meters). Most notably has been the

introduction of the European Space Agency's (ESA) Copernicus programme. ESA provides global coverage of optically sensed data at a high frequency via their constellation of Sentinel-2 satellites (Roy et al. 2017). The Sentinel-2 mission consists of Sentinel-2A, which was the first of the two satellites launched on 23 June 2015, followed by an identical satellite; Sentinel-2B which was launched on 7 March 2017. The addition of Sentinel-2B resulted in a high temporal resolution, with a 5-day revisit period at the equator, and even more frequent revisit time towards further latitudes. The orbits are sun-synchronous at an altitude of 786 km (ESA 2014). Data is available as level-1-C (L1C) processed data (orthorectified with top-of-atmosphere reflectance values). ESA have also made all data captured from December 2018 onward available at level-2-A (L2A) processed data (atmospheric correction applied with bottom-of-atmosphere reflectance). The L1C and L2A products are made available in 100 x 100 km tiles (Phiri et al. 2020). The main objective of the Copernicus programme is to provide publicly available, high resolution data, for the use in land cover mapping, global monitoring of climatic and ecological trends, and the monitoring of natural disasters (ESA 2014). Table 1 details the characteristics of the spectral bands of the Sentinel-2 sensors.

Table 1: Spectral bands for the Sentinel-2 sensors (source: <https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument>)

	<b>S2A</b>		<b>S2B</b>			
<b>Band Number</b>	<b>Central wavelength (nm)</b>	<b>Bandwidth (nm)</b>	<b>Central wavelength (nm)</b>	<b>Bandwidth (nm)</b>	<b>Spatial resolution (m)</b>	<b>Band Description</b>
<b>1</b>	442.7	20	442.3	20	60	Coastal aerosol
<b>2</b>	492.7	65	492.3	65	10	Blue
<b>3</b>	559.8	35	558.9	35	10	Green
<b>4</b>	664.6	30	664.9	31	10	Red
<b>5</b>	704.1	14	703.8	15	20	Vegetation Red Edge
<b>6</b>	740.5	14	739.1	13	20	
<b>7</b>	782.8	19	779.7	19	20	
<b>8</b>	832.8	105	832.9	104	10	NIR
<b>8a</b>	864.7	21	864.0	21	20	Narrow NIR

9	945.1	19	943.2	20	60	Water Vapour
10	1373.5	29	1376.9	29	60	SWIR (Cirrus)
11	1613.7	90	1610.4	94	20	SWIR
12	2202.4	174	2185.7	184	20	

### 2.3.2 GLOBAL LAND COVER

Global land cover is a relatively new concept, with the first land observation satellite only being launched in 1972 (Gong et al. 2016). Currently, multiple global land use and land cover (LULC) datasets exist at 1 km, 300 m and most recently 30 m (Chen et al. 2015). However, the accuracy of these datasets has been assessed and found to be unsatisfactory for many applications (Chen et al. 2015; Gong et al. 2016). One of the most challenging aspects of global land cover is the ability to accurately classify heterogeneous landscapes (Gong et al. 2013).

#### 2.3.2.1 MULTICLASS GLOBAL LAND COVER

The Copernicus programme, recently in May 2019, has produced and released a set of global land cover layers, known as Collection 2 (Buchhorn et al. 2020). This dataset consists of a 20 class, global discrete land cover map at 100m spatial resolution with overall accuracy of approximately 80% (Buchhorn et al. 2020). The dataset is useful for a wide range of uses such as forest monitoring, crop monitoring, biodiversity and conservation and climate modelling. However, the uses are limited to applications where fine-grained detail is of less importance and more suited to deducing general global trends. Another attempt at producing a pixelwise land cover classification at the highest spatial resolution of the source Sentinel-2 imagery (10 m) on a large scale was undertaken by a joint effort between ESA, the Space Research Centre of Polish Academy of Science, Poland and Industrieanlagen-Betriebsgesellschaft mbH, Germany. The project, known as Sentinel-2 Global Land Cover (S2GLC), aimed to derive an automatic methodology based on scientific recommendations to infer a 13 class global land cover dataset using a multi-temporal test strategy (Lewinski et al. 2017). Thus far, the whole of Europe has been classified, with an overall accuracy of 82.9% (Lewinski et al. 2019).

#### 2.3.2.2 SINGLE-CLASS GLOBAL LAND COVER

A number of global coverage binary classification maps have been produced within recent years. Global Human Settlement (GHS) 2020 (Corbane et al. 2021), Forest/Non-forest (Shimada et al.

2014), Forest Cover Change (Hansen et al. 2013) and Global Surface Water (Pekel et al. 2016) are some of these single class global land cover maps. The only map to use DL has been the more recent GHS, which was based on the four 10m bands from Sentinel-2, namely the red, green, blue and near infrared (NIR) bands (Corbane et al. 2021). The other global datasets were produced using a wide variety of machine learning algorithms, using multiple input data sources (Bratic, Vavassori & Brovelli 2021).

### **2.3.3 LARGE-SCALE LAND COVER CHALLENGES**

The challenges of large-scale land cover are well known. Differences in illumination, atmospheric, terrain and seasonal variances can affect inter-class spectral separability negatively (McDermid, Franklin & LeDrew 2005; Townshend et al. 2012). Woodcock et al. (2001) found that overall accuracies were negatively affected by ecological and climatic differences when classifying over extensive geographical regions using classical machine learning methods. This was mainly due to variances in phenology and the structure of vegetation types (Olthof, Butson & Fraser 2005). The spectral separability of land cover classes generally has an inverse relationship with distance, whereby separability decreases with an increase in distance (Verhulp & Van Niekerk 2016). Higgs (2021) suggests that for forest genera mapping from Sentinel-2 data, training data should be collected within 500km of where a random forest (RF) classifier is applied and must be of similar rainfall patterns. Pax-Lenney et al. (2001) found that applying simple dark object subtraction as an atmospheric correction method enabled a neural network (NN) to generalize well temporarily. However, accuracies dropped by 8%-13% in mean accuracies when extending models spatially from regions where training data was derived.

### **2.3.4 BIOME APPROACH**

To address the issues faced by such large scale classification, Souza et al. (2020) designed a mapping framework of classifying homogenous areas based on biomes. A biome is a region based on homogenous physical characteristic such as vegetation structure and climatic patterns (Mucina & Rutherford 2006). The MapBiomass mapping protocol allowed for custom feature sets and training data for each biome to be created (Souza et al. 2020). The MapBiomass application relies on a RF classifier trained for each biome. A single RF classifier was not robust enough to include all LULC classes. A map integration protocol is also produced as part of the MapBiomass application, this manages the smooth merging of data along transitional ecotone zones (Souza et al. 2020).

## **2.4 IMAGE CLASSIFICATION**

### **2.4.1 PIXEL-BASED CLASSIFICATION**

Pixel-based image classification was one of the first methodologies to be implemented to perform automatic, rapid land cover classification from LANDSAT imagery (Shlien & Smith 1975). A pixel-based classification assigns a class label to every pixel based on the input bands' values of that specific pixel only. The drawback of pixel-based classification is that it is prone to produce a salt-and-pepper noise effect on the output classification (Weih & Riggan 2008). This problem seems to be exacerbated with finer resolution imagery. This has become noticeable with VHR imagery becoming more available, such as Worldview, thus, there has been a paradigm change to an object-based approach.

### **2.4.2 OBJECT-BASED IMAGE ANALYSIS**

To address some of the short comings of pixel-based classification, researchers developed an object-based approach, where pixels with similar spectral values are grouped together into separate clusters prior to classification. Classification algorithms are then applied to a whole cluster instead of only a single pixel. Object based image analysis (OBIA) allows for further feature sets to be created, encompassing spatial attributes such as shape and distance to neighbours, as well as texture based attributes (Blaschke 2010; Weih & Riggan 2008). OBIA has shown to outperform pixel-based tree classification tasks using Sentinel-2 optical imagery (Li et al. 2014; Wessel, Brandmeier & Tiede 2018). However, it is worth noting that OBIA approaches tend to be highly computationally expensive.

### **2.4.3 CLASSICAL MACHINE LEARNING**

Classification algorithms can be divided into two main categories, supervised and unsupervised classifiers. Unsupervised algorithms separate pixels into classes based on similar spectral reflectance values. No prior knowledge of the image is required. After clustering the user must interpret the clusters and assign class labels to the clusters. An unsupervised approach is less prone to human biases since no user inputs in terms of training data are required (Campbell & Wynne 2011). However, the disadvantage is that if intra-class variance is high, the model may shift away from the center of the class and this may lead to confusion between classes (Al-Doski et al. 2013).

Supervised classification relies on labelled training data to inform a classifier. The inputs are class labels corresponding to the associated pixel information. The classifier uses this data to infer a pre-defined class label to an unclassified pixel. The accuracy of a supervised classification is very dependent on the quality of the training dataset (Radoux et al. 2014; Tuia & Camps-Valls 2009).

A weak representation of a target class in the training data will yield detrimental results (Al-Doski et al. 2013).

The increasing computational power and decreasing cost of hardware has promoted the use of machine learning in RS. Machine learning is an effective empirical methodology for classification and regression problems. This yields well for use in RS, and more specifically for land cover classification (Lary et al. 2016). Commonly used types of machine learning algorithms are support vector machines (SVM), decision trees (DT) and random forest (RF) (Lary et al. 2016).

RF classifiers have proven effective in LULC classifications (Buchhorn et al. 2020) as well as specifically for forest classification (Betts et al. 2017). RF has proved to produce higher accuracies than that of standard decision trees (Rodriguez-Galiano et al. 2012) due to the ensemble architecture of RF where several classification trees are trained on a subgroup of training data (Breiman 2001).

SVM is a supervised machine learning algorithm which aims to find the optimal hyperplane which clusters training data into separate, predefined classes. It does this by mapping features to a high dimensional feature space by using a simple kernel function. A major advantage of SVMs is that it is able to handle data successfully with minimal input data samples. This makes SVMs an attractive classifier in the RS domain (Kulkarni & Lowe 2016). Noi & Kappas (2017) compared the performances of RF, SVM and kNN classifying algorithms on land cover classifications from Sentinel-2 optical data. SVM yielded the highest overall accuracy.

These traditional classifying approaches still lack very high accuracy and have a great processing time cost (Noi & Kappas 2017) in that each pixel in an image needs to be classified individually, or in the case of an object-based approach, segmentation needs to be performed first. These methods still require expert knowledge in developing and applying hand-crafted features sets (Scott et al. 2017). An alternative methodology which addresses these issues is required. There has been growing popularity for the use of DL, more specifically, Convolutional Neural Networks (CNN) for land cover classification (Helber et al. 2019)

#### **2.4.4 DEEP LEARNING**

DL is a subset of machine learning. It forms part of the class of algorithms of representation learning. Representation learning allows the input of raw values and will automatically find and create abstract feature extractors (Lecun, Bengio & Hinton 2015). DL learns low to high level features, and, by composing a multitude of these features, are able learn very complex, non-linear functions (Goodfellow, Bengio & Courville 2016; Lecun, Bengio & Hinton 2015). The premise of deep learning is inspired by the complex structure of neurons of the human brain and tries to

mimic its learning procedure (Zhang, Zhang & Du 2016). The applications of DL are constantly expanding and covers a multitude of domains. Applications of DL range from speech recognition, visual pattern recognition, stock market analysis, sentiment analysis and security (Ball, Anderson & Chan 2017; Lecun, Bengio & Hinton 2015). New research in DL is constantly being published with ever-improving architectures and novel applications (Chen et al. 2018; Zhang et al. 2022).

The artificial neural network (ANN) is the basic building block of deeper networks (Mohamed et al. 2015). A multi-layer perceptron (MLP) is a fully connected neural network with multiple layers. It has an input layer, a hidden layer and an output layer. Figure 2 illustrates the architecture of a basic hypothetical MLP with four inputs, five neurons in the hidden layer and one output neuron. The number of neurons in the output layer is determined by the possible number of output classes. The MLP is said to be fully connected, as each neuron in the input layer is connected to each neuron in the hidden layer, and similarly with respect to the connectedness of the hidden neurons with the output neuron. As soon as this structure has more than one hidden layer, it forms part of the realm of deep ANNs (Lecun, Bengio & Hinton 2015). To leverage DL for complex visual recognition tasks, CNNs were developed.

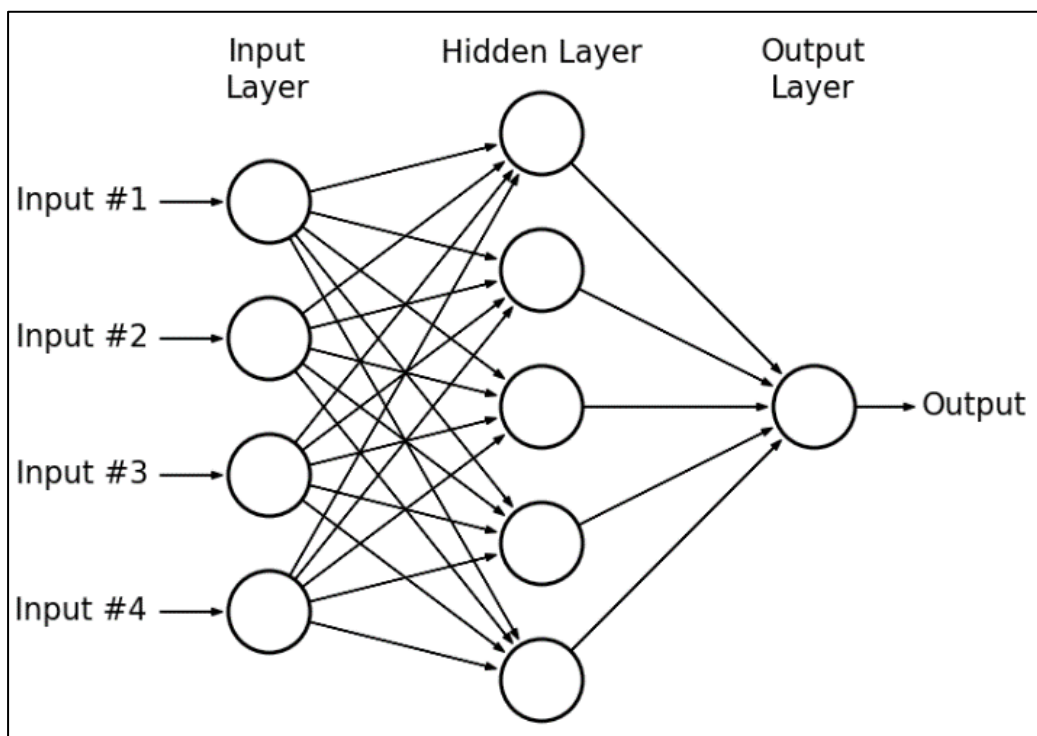


Figure 2 Illustration of a multilayer perceptron

The inspiration for the CNN was derived from the visual cortex of mammals. It is based on the way the visual cortex detects light in a receptive field (Gu et al. 2018). The basic CNN is comprised of three main components: the convolutional layers, pooling layers and fully-connected layers (Ball, Anderson & Chan 2017). CNNs learn feature representations in a hierarchical manner.



The hierarchical structure allows it to learn low and high level features (Zhang, Zhang & Du 2016). The convolutional layers extract feature maps from the input layer by performing a dot product on two arrays. The first array being learnable parameters (weights and biases) and the other being a kernel filter within the receptive field. The kernel acts as a sliding window across the width and height of the input image. The output of the filter produces a 2-D activation map (feature map) which is the response of the kernel at each stride of the image.

A convolutional block of a layer is normally preceded by a pooling layer. The pooling layer reduces the size dimensions of the feature maps thus reducing the number of dot product computations at each convolution layer (Ball, Anderson & Chan 2017). Pooling introduces shift invariance, meaning that objects can be recognised in any part of the image (Gu et al. 2018). Max pooling is a common pooling strategy which selects the maximum value to proceed with (Ball, Anderson & Chan 2017).

Activation functions are vital for the ability of a model to learn complex, non-linear features which are present in images (Zhang, Zhang & Du 2016). Several activation functions exist, including rectified linear unit (ReLU), sigmoid, tanh and softmax functions. The choice of activation function is normally driven by case specific application. Softmax is normally used as the output activation function for the final layer in a multiclass problem (Ball, Anderson & Chan 2017). Sigmoid is suited as the output activation to a binary classification task where only two possible outcomes are possible (Zhang, Zhang & Du 2016). ReLU is a very popular activation function for use within the convolutional blocks (Lu et al. 2020). ReLU sets negative neurons to zero and retains the positive. Due to the simplicity of ReLU, it allows a deep network to train very efficiently (Gu et al. 2018). The basic architecture of a CNN is shown in Figure 3.

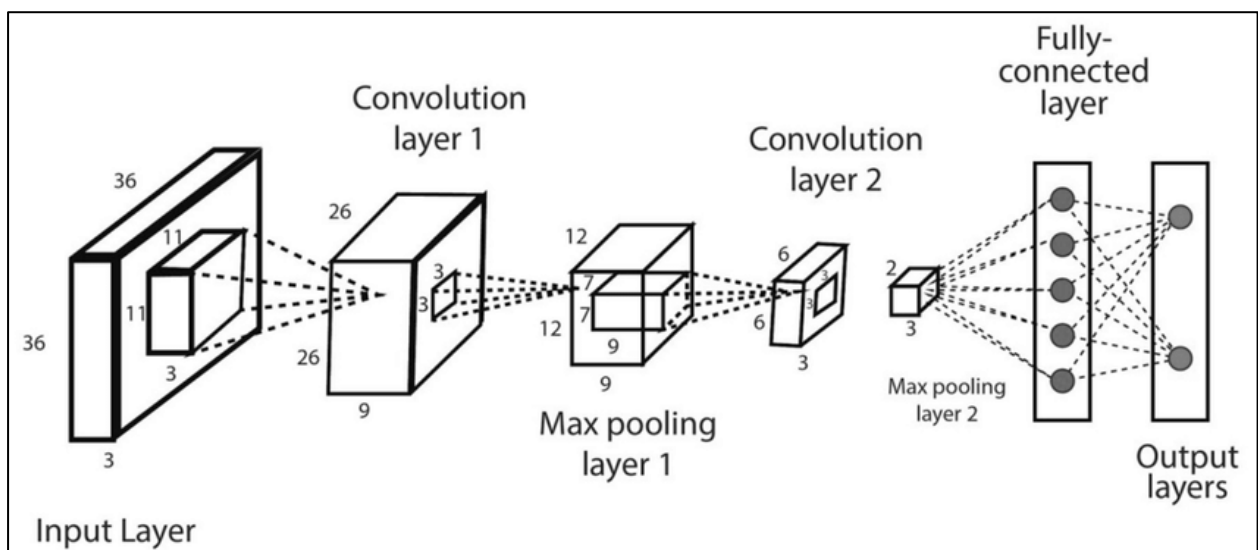


Figure 3 Visualising the architecture of a CNN (Sarker 2021)



CNNs are widely used for image categorisation tasks (assigning a label to an image) (Bar et al. 2015; Stivaktakis, Tsagkatakis & Tsakalides 2019). In frequent cases in remote sensing, such as with land cover classifications, scene classifications or other such image segmentation tasks, it is required to assign a class label to every pixel in an image. To achieve this, networks such as fully convolutional networks (FCN) and U-Net architectures have been developed. U-Net has shown to consistently outperform FCN for segmentation tasks (He, Fang & Plaza 2020; Ozturk, Saritürk & Seker 2020; Ronneberger, Fischer & Brox 2015).

A U-Net is split into two halves, the first half being a convolutional encoder which learns hierarchical feature maps and the second half being a deconvolutional decoder. Skip connections are used to map feature maps in the encoder to the deconvolutional layers in the decoder (Ronneberger, Fischer & Brox 2015). The up-sampling step enables a semantic segmentation of an image whereby all pixels are classified into target classes. Figure 4 shows the U-shaped structure of a U-Net.

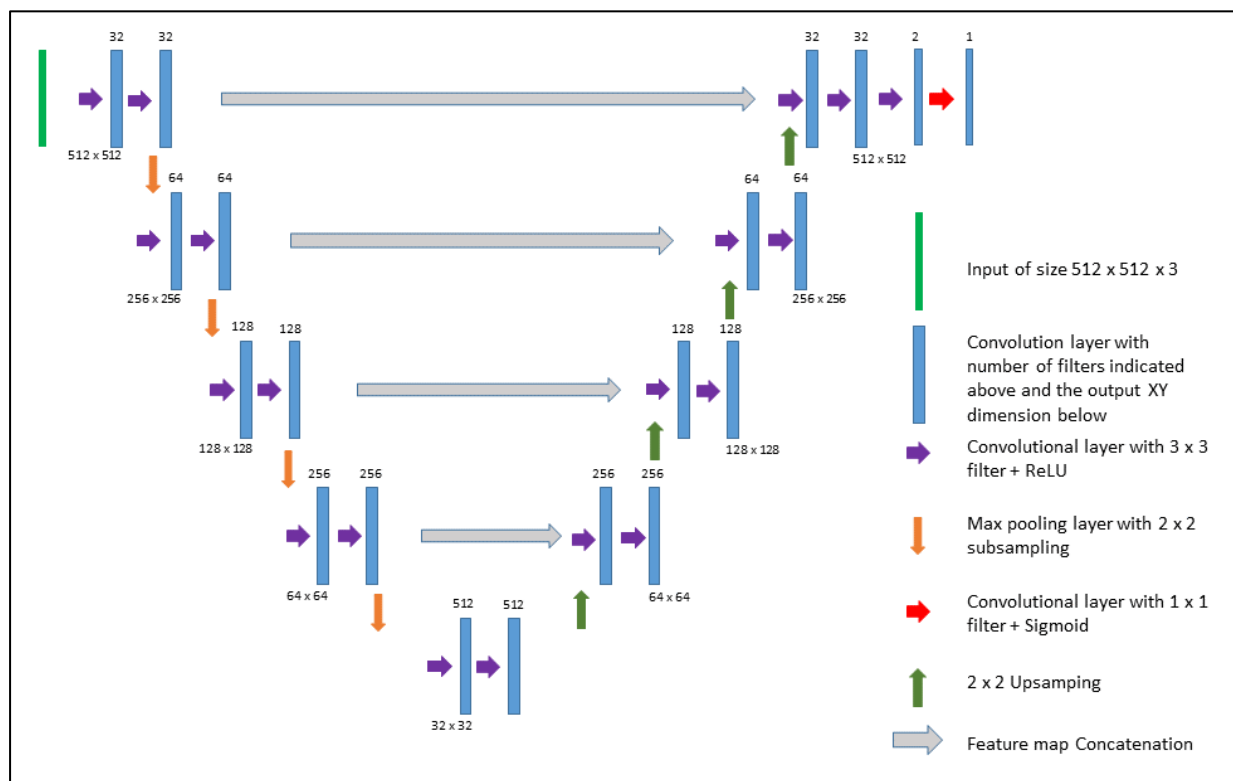


Figure 4 Visual depiction of a U-Net architecture

#### 2.4.4.1 APPLICATIONS OF U-NET IN REMOTE SENSING

The use of U-Nets in RS has been applied with much success. Ozturk et al. (2020) achieved 96.33% overall accuracy (OA) on validation data for road segmentation tasks from VHR imagery. In the same study, a comparison using the same data was performed against FCN but could only achieve 90.23% OA on the validation data. A U-Net produced impressive results for palm tree detection

from 40 cm pan-sharpened Worldview3 imagery (Freudenberg et al. 2019). It achieved an F1-score of 95.7%, outperforming previous methods by 10-13% (Freudenberg et al. 2019). Comparing U-Nets with traditional classifiers such as RF, Solórzano et al. (2021) found that deriving LULC using the four 10m resolution bands from Sentinel-2 (R,G,B,NIR) with Sentinel-1 synthetic aperture radar (SAR) data as an additional input, yielded a superior result compared to a RF classifier for all classes. Flood, Watson & Collett (2019) trained a U-Net model for the prediction of trees and shrub over an 827km<sup>2</sup> study site, based on 1m spatial resolution, RGB imagery. An OA of 89.91% was achieved. Pan et al. (2020) proposed a U-Net-based approach for accurately segmenting individual buildings in unplanned urban settlements located in Guangzhou City, China. The results of this study were compared against those obtained from RF and OBIA methods. The U-Net model demonstrated superior performance, with an overall accuracy of 87.2%. This result was significantly higher than the accuracy scores of 77.7% and 58.1% achieved by the RF and OBIA methods, respectively. Li et al. (2016) conducted a study to compare the performance of a Stacked Autoencoder (DL method) with traditional machine learning classifiers, including RF, SVM and ANN, for large scale land cover mapping of the African continent. The results showed that the Stacked Autoencoder outperformed the other classifiers, with an overall accuracy of 78.99%, compared to 76.03%, 77.74%, and 77.86% for RF, SVM, and ANN, respectively. All these studies attribute the contextual features learnt by a U-Net for their success. Most of the reviewed studies all had localised study areas, with Li et al. (2016) classifying on a continental scale; none have attempted global scale classifications.

## **2.5 IMAGE ENHANCEMENTS**

RS images are often polluted with undesirable environmental conditions which leads to the reduction in their contrast, which in turn causes loss of detail. Image enhancement is the process of making an image more suitable to a specific application than the original image (Shyam et al. 2017). To improve the quality of such images, enhancement algorithms can be applied to RS images with the aim to improve their visual representation (Demirel, Ozcinar & Anbarjafari 2010). There is a plethora of literature which exist on image enhancements, however, very few studies have been done on improving image enhancements in the RS domain (Lee et al. 2013; Lisani et al. 2016; Liu et al. 2014). The literature discussed here will focus on commonly used enhancement techniques in natural image processing. The following image enhancements will be discussed further: histogram equalisation (HE), contrast limited adaptive histogram equalisation (CLAHE) and global contrast normalisation (GCN) (Jadoon et al. 2017; Maini & Aggarwal 2010; Malvika 2017).

HE is one of the most frequently used image enhancement methods which aim to improve contrast in images. This is due to its simplicity in implementation (Fu et al. 2015). HE is a global contrast enhancement method used to spread out the most frequent intensity values across the available bit range. To compute HE a cumulative distribution function is calculated, based on the bit depth of the image and the width and height dimension of the image. The image is then scaled using the cumulative distribution function, resulting in an equalised histogram (Gupta & Kaur 2014). Outlying peaks in the histogram can lead to over enhancement which can cause saturation artifacts in the processed image, thus making HE unsuitable for some applications (Fu et al. 2015). HE has been shown to improve a CNN trained to detect and classify emotions via facial expressions with an accuracy of 89.18% versus the baseline accuracy of 61.81% when using raw data (Pitaloka et al. 2017).

CLAHE is a variation of histogram equalisation which reduces the over-enhancement and saturation artifacts caused by a basic histogram equalisation process. CLAHE utilizes a user-specified maximum contrast limit and tile size which prevents over-enhancement. The histogram for each user-defined tile is calculated independently, unlike HE which looks at the full image, and the pixel value is updated against the calculated histogram. CLAHE is a common enhancement technique used in medical imaging, particularly in the pre-processing of digital mammogram imagery (Jadoon et al. 2017; Maitra, Nag & Bandyopadhyay 2012), dermoscopy (Premaladha & Ravichandran 2016) and to improve CNN performance for the detection of COVID-19 in chest X-rays (Umri, Wafa Akhyari & Kusrini 2020).

The use of CLAHE in RS has been much less popular and little research exists concerning this topic. Ganesh & Ramesh (2017) tested the effectiveness CLAHE by applying it to both Landsat-8 and Sentinel-2 data. They concluded that CLAHE, when applied to the NIR and SWIR bands enhance the interpretability of an image well enough to identify wetland delineations without the need for calculating a normalised difference water index. The same study suggests that applying CLAHE in the NIR bands can enhance the boundaries of water bodies and improve contrast between changes in reflectance due to sediment content (Ganesh & Ramesh 2017). Although this study focused on improving image legibility for human data capturers, the results suggest that it may aid in improved scene understanding for DL models too.

Another simple yet effective normalisation technique is GCN. The purpose of GCN is to centre the image data around zero. Calculating GCN is comprised of two steps. Firstly, the data needs to be mean normalised and then it is standardised. Mean normalisation involves calculating the mean value across the feature vector and subtracting this value from each pixel (Pal & Sudeep 2016). The standard deviation of the feature vector is calculated after which the mean normalisation is

divided by the standard deviation (Coates, Lee & Ng 2011). This ensures the data is mean and variance normalised. GCN was successful in improving the results of facial emotion recognition from 61.81% on raw unprocessed images to 88.31% (Pitaloka et al. 2017). GCN has also been applied as a preprocess to DL training for the segmentation of retinal blood vessels (Liskowski & Krawiec 2016). Gudi (2015) found that using GCN as a preprocess model yielded an improvement in accuracy of 3% for semantic facial feature recognition by DL.

Image enhancements, specifically for improving the performance of DL models, are widely used in the CV and medical image analysis domains, however, less emphasis has been placed on the potential impacts of image enhancements in the RS domain. The possible reasoning or the lack of adoption of enhancements in RS could be that the spectral integrity of data in remote sensing is very important for downstream applications such differentiating land cover classes, especially when using traditional classification methodologies. By applying image enhancements, the spectral integrity is altered, and this could further increase intra-class spectral variation. Due to the nature of CNN based models, where contextual information is also learnt, image enhancements may prove beneficial in improving classification accuracies when combined with DL technologies.

## **2.6 DATA AUGMENTATION**

DL models require very large training datasets in order to incorporate variability within classes (Weinstein et al. 2020; Yu et al. 2017a). Although the problem of overfitting can be addressed by adding dropout functions and batch normalization layers in the network architecture (Zheng et al. 2018), it is important to address the root cause, namely the lack of a large enough dataset. The production of datasets is very costly and time consuming (Townshend et al. 2012). Consequently, the availability of image-level training datasets in the RS domain is limited, and even more so in the case of pixel-level datasets (Townshend et al. 2012). Large image-level datasets, such as the ImageNet dataset, contain 15 million images categorised into 22000 classes. Such large datasets are not yet available in the RS domain (Yu et al. 2017a). One method for introducing and increasing diversity in the training dataset is to enlarge the training dataset artificially by applying data augmentation to the existing training dataset (Shorten & Khoshgoftaar 2019). Data augmentation can be categorised into two main types: data warping augmentation and oversampling augmentation (Shorten & Khoshgoftaar 2019).

Oversampling refers to the generation of a completely new, synthetic image. Image mixing is an example of oversampling. It involves combining multiple images into one by taking the average pixel value for each image stack over N number of images. An example of oversampling that has been applied successfully is a Generative Adversarial Network (GAN) used to create additional

training images (Loey, Smarandache & Khalifa 2020). GAN-based data augmentation techniques have been especially useful in the medical imaging domain, for which there is also a scarcity of training datasets. In the case of training a robust model for liver lesion identification, GAN-based data augmentations outperformed models trained with basic data warping, with model performance increasing from 88.4% using data warping, to 92.4% using GAN-based augmentations (Frid-Adar et al. 2018). Data augmentation via adversarial training has been applied in the RS domain by training a U-Net model for building detection (Lv et al. 2021). The dataset used was the Mnih Massachusetts building dataset (Mnih 2013). All accuracy metrics including F1-score, OA and intersection over union (IoU) showed improvements (Lv et al. 2021).

Oversampling and data warping are not mutually exclusive as data warping can be applied to synthetically created images from GAN's to further enhance and diversify the relevant datasets (Shorten & Khoshgoftaar 2019). Geometric and colour transformation fall into the broad category of data warping. Geometric transformation modifies the geometric properties of an image. This type of augmentation is suitable for satellite data as semantic features are orientation invariant (Scott et al. 2017). Colour transformation changes the pixel intensity values, thereby mimicking different illuminating conditions and atmospheric differences, but retain the geometric properties of images.

Geometric transformations refer to changes to the affine transformation matrix of the pixel data. Popular geometric transformations include flip, rotation and translation (Yu et al. 2017a). A flip can be performed on the X or Y axis of an image, or both. Rotations involve rotating an image anywhere between zero and 360 degrees. Shear translation performs a shift in both X and Y directions. These methods increase the diversity in the training dataset without changing the spectral or topological properties in the data (Yu et al. 2017a). Yu et al. (2017) evaluated the effectiveness of basic flip, rotation and translation augmentations on scene classification using four datasets. Applying augmentations improved the Kappa index of the models of all four datasets, with the UC Merced Land Use dataset (Yang & Newsam 2010) seeing the greatest improvement in Kappa index (0.48 to 0.87) (Yu et al. 2017a).

Colour transformation is another method which has seen success in the CV realm (French & Mackiewicz 2022). Applying gamma and contrast variation to augment training data has been successfully used to improve the accuracy of DL models, both in self-portrait segmentations (Shen et al. 2016) and melanoma classification (Perez et al. 2018). Colour transformations preserve the geometric characteristics of objects, therefore in some cases it would be more useful to apply augmentations which vary pixel intensity than geometric augmentations as in a text/number recognition task since a value six (6) when flipped in the vertical axis can be interpreted as a value

nine (9) which would destroy the integrity of the training dataset. Conversely in cases where the colour property is a distinctive feature, it would be desirable to only apply geometric augmentations so as to preserve spectral and topological properties as in the case of optical satellite imagery. However, Robinson et al. (2019) found that although colour was a very predictive feature within a small geographic region, this became less consistent across larger regions where intra-class colour variability was high. Thus, introducing colour augmentation to account for image variability across geographical regions may be useful. Liu et al. (2020) applied image brightness augmentation to increase the dataset size for training a maize plant detection DL model from VHR, RGB data. However, no analysis was done on the effect that this augmentation had on the performance of the model.

Data augmentation, including both data warping and oversampling methodologies, have shown to be very effective in training DL models. From the literature, the improvements observed from data augmentation are generally amplified in cases where the original training dataset size is limited. Geometric augmentations are the preferred augmentations applied in the RS domain, with little literature available regarding the use of colour augmentations.

## **2.7 SAMPLING METHODOLOGY**

In order to infer metrics from an experiment, a subset of the population data is used to create statistical inferences. The design of the sampling method is dependent on the data characteristics of the of the population dataset. Two main sampling methodologies exist; probability sampling and non-probability sampling (Showkat & Parveen 2017). Probability sampling is a method where each member of a population has an equal opportunity of being sampled. It is based on a theory of probability that the sample is representative of the population (Showkat & Parveen 2017). Non-probability sampling is not based on random sampling. Samples are chosen with some form of bias, thus, not all members of a population have an equal chance of being chosen (Showkat & Parveen 2017).

Simple random sampling is an example of a probability sampling method. As the name suggests, this form of sampling is very simple whereby every member of the population has an equal chance of being selected. However in the case of land cover classification, population datasets are not always balanced (Ottosen et al. 2020; Ramezan, Warner & Maxwell 2019). To overcome the problem with simple random sampling, stratified random sampling is suggested (Ramezan, Warner & Maxwell 2019). Rather than selecting a sample at complete random to the population, stratified random sampling randomly selects a proportionate number of samples from each stratum in the population (Ramezan, Warner & Maxwell 2019; Showkat & Parveen 2017).

Ramezan, Warner & Maxwell (2019) suggest that non-probability sampling must be avoided for large scale classifications. The majority of studies involving land cover classification reviewed in the literature has employed a stratified random sampling methodology for drawing samples to infer statistics (Abdi 2020; Karra et al. 2021; Noi & Kappas 2017; Sekertekin, Marangoz & Akcin 2017; Souza et al. 2020).

Training a DL model requires a training/test split in the input dataset. The test data is separated from the training data and is used solely for calculating a validation score at the end of each epoch of training. Karra et al. (2021) used a training/test split of 85/15 for their model evaluation of global land cover. Abdi (2020) utilised a 70/30 split in their data for comparing land cover data derived from different classification algorithms.

## 2.8 ACCURACY ASSESSMENT

The statistical analysis of a confusion matrix is often done using the F1 score and overall accuracy (OA) to infer results, depending on the goal of the research. However, these metrics can overstate results in unbalanced datasets (Chicco & Jurman 2020). The Matthews Correlation Coefficient (MCC) score is a well-suited performance metric in analysing results from binary classifications (Chicco & Jurman 2020), particularly for unbalanced datasets. MCC considers True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) results and will only generate a high score if the majority of both positive and negative test cases were correctly predicted. MCC ranges between from -1.0 to 1.0, with -1.0 indicating complete misclassification, 0.0 no better than random, and +1.0 a perfect classification.

MCC is defined in Equation 1 as:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \quad \text{Equation 1}$$

Where	TP	True positive: correct prediction of the positive class;
	TN	True negative: correct prediction of the negative class;
	FP	False positive: incorrect prediction of the positive class; and
	FN	False negative: incorrect prediction of the negative class.

Precision and recall are two metric which can also be used to measure the effectiveness of model performance. Precision is the fraction of TP cases over the sum of all positive predictions and is often referred to as confidence (Powers 2007). It measures the probability that a positive test case is truly positive. Recall, also known as the true positive rate or sensitivity of a model, measures



the proportion of positive cases which have been correctly classified as positive. Precision and recall should always be analysed inclusively of each other since a high recall can be achieved even with a low precision and vice versa. F1 score attempts to combine these scores to produce a single measure to represent how well a model has performed. F1 score is the harmonic mean of precision and recall (Powers 2007).

The F1 score can be misleading in that the true negative cases are not considered. This can be seen from the F1 score equation (Equation 2):

$$F1\ Score = \frac{2TP}{2TP + FN + FP} \quad \text{Equation 2}$$

Where	TP	True positive: correct prediction of the positive class;
	FP	False positive: incorrect prediction of the positive class; and
	FN	False negative: incorrect prediction of the negative class.

The F1 score can give false truth values when the target class is in the minority and gets used as the TP case. Thus, in the case when the target class is the minority class, it is preferred to switch the positive and negative labels such that the target class presents the TN case in the error matrix (Chicco & Jurman 2020). F1 score is thus useful when class labels are balanced.

Cohens kappa statistic is a commonly used metric for evaluating performance metric. It is widely used in RS for multiclass land cover classification (Jin et al. 2019; Mellor et al. 2013; Sekertekin, Marangoz & Akcin 2017; Sharma, Hara & Tateishi 2017). However, the kappa index is very sensitive to the distribution of marginal totals, and this can cause misleading outcomes (Flight & Julious 2015).

## 2.9 SUMMARY OF LITERATURE

Chapter 2 provided a review of the literature on the use of deep learning applied in remote sensing for the classification of trees on a global scale. The main findings suggest that DL algorithms consistently outperform traditional classifiers for LULC tasks. This is true for both VHR and medium (10m to 30m) resolution imagery. The unsupervised feature learning ability of DL reduces the need for constructing complicated, handcrafted features. The U-Net architecture is well suited to a pixelwise classification due to the upscaling of feature maps in the decoder, by using skip



connections to link the feature maps in the encoder to features in the decoder. Basic image enhancements have shown positive results when training DL models outside in other domains. HE, CLAHE and GCN have been identified as enhancements worth testing for a tree classification using a U-Net. Large, accurate and annotated training datasets remain a limitation within the RS domain, however basic data augmentation methods including geometric and colour augmentations have shown to improve the performance of DL models in other applications. Geometric augmentations have been successfully used in the RS domain, however, colour augmentations have been less studied in this domain. Random stratified sampling has been identified as a reliable sampling methodology for inferring statistics for land cover classification. MCC is suggested to be a reliable metric to compare model performances, specifically for binary classification tasks as it takes into account the prediction of both positive and negative test cases and will only produce a high score if both these cases have scored well.

### **3 CHAPTER 3: IMAGE ENHANCEMENT METHODS FOR IMPROVING THE RESULTS YIELDED BY A CNN BASED GLOBAL FOREST CLASSIFICATION MODEL USING SENTINEL-2 OPTICAL DATA**

#### **3.1 ABSTRACT**

Effectively managing global forest resources, under threat from climate change, deforestation and fragmentation, requires the efficient extraction of a global tree cover dataset. In this chapter we aim to improve the accuracy of a deep learning model for global forest classification using Sentinel 2 optical data. We present several image enhancement methods widely used in natural image classification and biomedical imaging domains, including histogram equalisation (HE), contrast limited adaptive histogram equalisation (CLAHE) and global contrast normalisation (GCN), as pre-processing steps. The enhancement methods were compared with each other on a per biome basis, and both training and validation regions were selected to represent the heterogeneity within biomes. Selected images were captured within the local optimal foliage growing season and contained minimal or no clouds. A UNet convolutional neural network model was trained for each enhancement per biome and used to perform inference on validation images for each of the corresponding biomes and enhancements. Random stratified samples were collated for all validation images per biome per enhancement for statistical analysis. Only GCN and CLAHE RGB returned higher means than the baseline dataset. The results showed that GCN most consistently improved classification results for tree cover across biomes, possibly due to the standardization of contrast levels of the training and validation images.

#### **3.2 INTRODUCTION**

Forests provide a range of important ecosystem resources, including timber production, biodiversity protection and climate change mitigation through carbon storing (Kornatowska & Sienkiewicz 2018; Schulze, Katharina; Malek, Ziga; Verberg 2019; Timothy R.H. Pearson Sandra L. Brown Richard A. Birdsey 2007). These resources are under severe pressure due to anthropogenic influences. The effective management of global forest resources requires extracting an efficient global tree cover dataset. Accurately classifying land cover, such as forests, on a global scale and at a high resolution, remains a major challenge (Sjöqvist, Längkvist & Javed 2020). Recent technological advancements in satellite imaging sensors, as well as developments in the field of artificial intelligence, offer opportunities for further research on accurate and efficient methods of classification on a global scale.

### 3.2.1 GLOBAL FOREST MAPPING

Global land cover is a relatively new concept, with the first land observation satellite being launched only in 1972 (Gong et al. 2016). Currently, there are multiple global datasets at spatial resolutions of 1 km, 300m, 30m and most recently 10m (Chen et al. 2015). However, assessments have shown the accuracy of these datasets to be unsatisfactory for many applications (Chen et al. 2015; Gong et al. 2016).

Recent improvements in satellite imaging sensors have yielded a wealth of freely available optical data at a high spatial resolution (10m). Of particular note has been the recent introduction of the European Space Agency (ESA) Copernicus programme. ESA provides global coverage of optically sensed data at a high frequency (3–5 day revisit period) via their constellation of Sentinel-2 satellites (Roy et al. 2017). Consequently, there has been an increased need for accurate algorithms to extract meaningful data from the vast amount of available satellite imagery (Shendryk et al. 2019).

The Copernicus programme has released a set of global land cover layers known as Collection 2 (Buchhorn et al. 2020). The dataset consists of a 20-class global discrete land cover map at 100m spatial resolution with overall accuracy reaching around 80% (Buchhorn et al. 2020). The dataset benefits a wide range of services, such as climate modelling, forest and crop monitoring, biodiversity and conservation and natural resource management (Bratic, Vavassori & Brovelli 2021; Buchhorn et al. 2020). However, as these applications are less reliant on fine-grained detail, this dataset is more suited to informing general global trends. In contrast, Sentinel 2 Global Land Cover (S2GLC) (Lewinski 2017), delivers a 10m resolution 13-class global land cover dataset by means of an automatic pixel-wise land cover classification methodology using a multi-temporal test strategy. To date, the whole of Europe has been classified, with an overall accuracy of 82.9% (Lewinski 2019). More recently, ESA produced a global land cover layer (WorldCover) at 10m spatial resolution but with a minimum overall accuracy of only 74.4%. The data layer was derived from both the Sentinel-1 Synthetic Aperture Radar (SAR) and the Sentinel-2 multispectral data with a minimum of 10 land cover classes (ESA 2021).

The most notable previous work on global forest classification and measurement of forest gains and losses was produced by Hansen et al. (2013), at a spatial resolution of 30m. The data was based on a time-series approach using data derived from a range of Landsat missions on multiple dates (Hansen et al. 2013). However, there is still a gap in available literature for the classification of trees at 10m resolution from a single date image. The single date aspect is important in that it enables an operationally efficient classification. Increasing the accuracy of forest classification through a multi-temporal approach is limited (Wittke et al. 2019). In some cases, such as

vegetation height regression from Sentinel-2 (Lang, Schindler & Wegner 2019), a multi-temporal approach was found to have reduced accuracy compared with a single date regression. A critical aspect for effective tree cover classification is that the source imagery is acquired in the middle of the leaf-on vegetation season, when leaf coverage is at its most dense. (Mirończuk & Hościło 2017).

Land cover classification utilising traditional machine learning classifiers, such as random forest decision trees, are more prone to the “salt and pepper” effect, or speckling, caused by dissimilar pixels within close proximity. Deep learning (DL) based classification has been shown to outperform traditional machine learning methods for tree classification at high resolution this is because a convolutional neural network (CNN), a subclass of deep learning, learns contextual information about a pixel and does not classify each pixel based purely on its spectral data, but rather a combination of spectral features and contextual information (Flood, Watson & Collett 2019; Freudenberg et al. 2019; Korznikov et al. 2021)

### **3.2.2 MACHINE AND DEEP LEARNING**

Increased computational power and lower hardware costs have facilitated the use of machine learning for remote sensing. Machine learning is an effective empirical methodology for dealing with classification and regression, and thus lends itself to remote sensing, and more specifically to land cover classification (Lary et al. 2016). The most commonly used types of machine learning algorithms are Support Vector Machine (SVM), Decision Trees (DT) and Random Forest (RF) (Lary et al. 2016)

RF has proven effective for land use and land cover classifications (Buchhorn et al. 2020, Kulkarni & Lowe. 2016, Rodriguez-Galiano et al. 2012) as well as for forest classification (Akar & Güngör 2012; Betts et al. 2017; Mellor et al. 2013). RF produced higher accuracies than those of standard decision trees (Rodriguez-Galiano et al. 2012) due to the ensemble architecture of RF whereby several classification trees are trained on a subgroup of training data (Breiman 2001).

Gong et al, 2013 evaluated four different machine learning classifiers for producing a global land cover classification at a spatial resolution of 30m using Landsat multispectral imagery. Maximum likelihood classifier (MLC), J4.8 DT, RF and SVM were tested for performance. MLC returned the lowest overall accuracy (53.88%), followed by J4.8 decision trees (57.88%). Random forests and SVM returned the highest OA with 59.83% and 64.89% respectively (Gong et al, 2013). These results align with the Noi & Kappas (2017) findings that SVMs’ and RFs’ perform well for land cover classification tasks. The overall accuracies, however, leave room for improvement,

particularly in respect of a global land cover dataset. This leaves a gap for significant improvement that DL could fill because of its ability to learn complex, non-linear features.

Advances in computing technology have accelerated the use of advanced DL algorithms. DL achievements in the field of visual recognition (Krizhevsky, Sutskever & Hinton 2012) have paved the way for image categorisation and segmentation tasks in the remote sensing domain (Korznikov et al. 2021; Neupane, Horanont & Aryal 2021; Yang, Rottensteiner & Heipke 2018). There has been growing popularity for using DL, specifically CNN, for land cover classification, as it has been shown to outperform non-deep learning classification algorithms (Castelluccio et al. 2015; Marmanis et al. 2016). Semantic segmentation using deep learning has achieved success in precise tree recognition from very high-resolution imagery (Korznikov et al. 2021) as well as for multi-class land cover mapping from Sentinel-2 data (Karra et al. 2021).

Despite the successes of DL, and specifically CNN, the adoption of this classification method for global land cover has been limited. Most studies have looked at very high spatial resolution satellite imagery due to objects, such as buildings, having similarities to objects classified using traditional computer vision tasks (Mahdianpari et al. 2018). To improve the delineation of buildings using CNN, Xu (2018) edge enhancements were applied as a pre-processing step for the imagery. Adding edge enhancements improved the overall accuracy compared to no enhancement (Xu 2018). Biological diversity patterns vary across the world and are generally driven by climatic conditions (Silva-Flores, Pérez-Verdín & Wehenkel 2014). The limitations of current global land cover classification methodologies stem from intraclass heterogeneity of land cover classes around the world (Helber et al. 2019). For example, a forest in a boreal biome is not as clearly defined as a forest in a temperate forest biome. Thus, it follows that it may be useful to use a biome-based approach within which foliage patterns and species are more homogenous. This approach in conjunction with image enhancement as a pre-process could potentially reduce the negative effect of intraclass variability for global tree classification.

### **3.2.3 IMAGE ENHANCEMENT**

Remote sensing images are often polluted by artefacts from undesirable environmental conditions, leading to a reduction in their contrast and consequent loss of detail (Fu et al. 2015). The goal of image enhancement is to process an image in a way that the resulting image emphasizes the objects of interest, suppresses information of lesser importance, and increases interpretability by humans (Demirel et al. 2010; Maini & Aggarwal 2010).

Enhancement techniques that have been used to improve image data successfully include histogram equalisation (HE), contrast limited adaptive histogram equalisation (CLAHE) (in both

the red, green and blue colour space as well as on the value channel in the hue, saturation and value colour space), and global contrast normalisation (GCN) (Jadoon et al. 2017; Maini and Aggarwal 2010). These enhancements and normalisation techniques have proved to be effective in various digital imaging domains, specifically in digital medical image processing (segmentation tasks) (Premaladha & Ravichandran 2016), and computer vision problems, such as automatic emotion recognition (categorisation tasks) (Pitaloka et al. 2017). To date there has been limited investigation of the use of these image enhancement methods in the remote sensing field. It appears they could be utilized as pre-processing steps to input data for training a CNN-based forest classification from optical satellite imagery.

Image enhancement used as a pre-processing step of inputs for a CNN has proved to be effective for facial emotion classification, natural image classification and biomedical imaging classification (Koo & Cha 2017; Lu et al. 2020; Pontalba et al. 2019; Umri, Wafa Akhyari & Kusrini 2020). GCN was successful in improving the results of facial emotion recognition from 61.81% on raw unprocessed images to 88.31% (Pitaloka et al. 2017), HE too improved results to 89.18% (Pitaloka et al. 2017). CLAHE has been used as an enhancement on greyscale training and validation images from chest X-rays where a CNN was trained to detect automatically whether a patient had been infected with the Covid-19 virus (Umri, Wafa Akhyari & Kusrini 2020). CLAHE as a pre-process has been shown to improve the effectiveness of a CNN to detect subtle features and micro aneurysms in fundusoscopic images for early-stage detection of diabetic retinopathy, whereas a CNN fed raw images were unable to detect these subtle features (Lam et al. 2018). Similarly, these enhancements may be useful for a remote sensing application by exposing subtle differences in vegetation (for example where the border of trees and grasses are less pronounced) thereby potentially improving the ability of a deep learning model to learn these exposed features and improve the model performance.

Basic image enhancement techniques, commonly used in the medical imaging domain and photography, have not been adopted widely for use in satellite image pre-processing for image classification. However, CLAHE has shown to be more useful than global HE by improving the legibility of an interpreter to delineate water bodies, shorelines and wetlands (Ganesh & Ramesh 2017).

This chapter demonstrates the effectiveness of applying image enhancement and normalisation as pre-processing steps for improving deep learning models for the global classification of trees within different biomes. The enhancements, CLAHE, HE and GCN, which have successfully been used in the medical research and computer vision domains, are applied to Sentinel-2 optical

imagery. To reduce the influence of differing environmental conditions, enhancements are applied for five different biomes with models trained for each enhancement per biome.

The remainder of the paper is organized as follows. Section 2 details data characteristics, the selection of training data, and the work environment. Experimental flows, brief reviews of each of the enhancements and normalisations applied as pre-processing to deep learning models, and accuracy measures are described in Section 3. Section 4 details results achieved and compares the impact of the enhancements on the performance of the DL model. Variations per biome are discussed in this section followed by the conclusion of the paper in Section 5.

### **3.3 DATA AND PROCESSING CHARACTERISTICS**

#### **3.3.1 SENTINEL-2 MULTI-SPECTRAL DATA**

Sentinel-2 data provides 13 optical bands with spatial resolutions ranging from 10m to 60m, wavelengths ranging from 442.2nm to 2185nm across the electromagnetic spectrum, and covering visible, near infrared and shortwave infrared ranges. Sentinel-2 optical data were retrieved from the open access AWS S3 bucket. Only bands four (red), three (green) and eight (near infrared) were used in this study. The blue band was excluded because it is very sensitive to atmospheric changes, making it less useful for any large-scale mapping of trees.

Image data is provided at two processing levels, Level-1-C (L1C) and Level-2-A (L2A). L1C is atmospherically corrected to Top of Atmosphere (TOA) reflectance values and L2A is further processed to Bottom of Atmosphere (BOA) reflectance. This methodology should be operationally efficient as the application is based on the classification of trees on a global scale. In order to reduce inefficiencies, ESA data already atmospherically corrected in the form of L2A processed data was used.

#### **3.3.2 BASE IMAGE**

The base image, used as the reference image to which enhancements were applied, was derived from LuxCarta's BrightEarth global Sentinel-2 L2A (<https://brightearth.ai/>) cloud-free mosaic (Swaine et al. 2020) at an 8-bit depth.

#### **3.3.3 TRAINING AND VALIDATION DATA SOURCING**

LuxCarta (<https://www.luxcarta.com/>) houses land cover data with coverage across all continents with the exception of Antarctica. Training and validation datasets were derived from LuxCarta's archive datasets for each of the following biomes: Tropical Rainforest, Temperate Forest, Boreal Forest, Grassland and Savanna as determined by the World Wildlife Foundation (Olson et al. 2001)

2001). Desert and Tundra were not considered for the experiment since trees are not a common feature in these biomes (UCMP 1996).

The ground truth and training labels for our model were derived from LuxCarta's extensive archive of global land cover data. To ensure the highest level of accuracy, this data was manually refined to align with the corresponding multispectral images. We made every effort to maintain the integrity of the data in relation to the images used for training. Despite these efforts, it is possible that some errors may exist in certain areas due to the limited time constraints to perform rigorous quality checks on such a large volume of data. A total of four training regions were selected for each biome. As far as possible, the regions were selected for global representative coverage to include heterogeneity within biomes. The image selection criteria applied were to select images that were captured within the local optimal foliage growing season and which contained minimal or no clouds.

Sentinel-2 images were projected in their native Universal Transverse Mercator (UTM) projection. The near infrared, red and green bands all had a ground sampling distance of 10 metres. To ensure the compilation of balanced training data, all training images were clipped to 3000 by 3000 pixels. This equates to 900 square kilometres per training image.

Six validation datasets were selected for each biome, and were also globally representative. Similar climatic conditions and vintage as the respective training datasets were used for the validation dataset to ensure standardization across training and validation datasets. Validation images were 2000 by 2000 pixels in size, which equated to a validation area of 400 square kilometres per validation image. Considering the time limitations for dataset creation, it was decided to focus on smaller regions for the validation dataset, but with an increased number of geographic locations to ensure broader coverage. The compromise of having broader geographic coverage with more validation images was preferred as it was a way in which to reduce data creation time and to have a globally representative result. Figure 5 shows the global distribution of training and validation datasets for all five selected biomes.



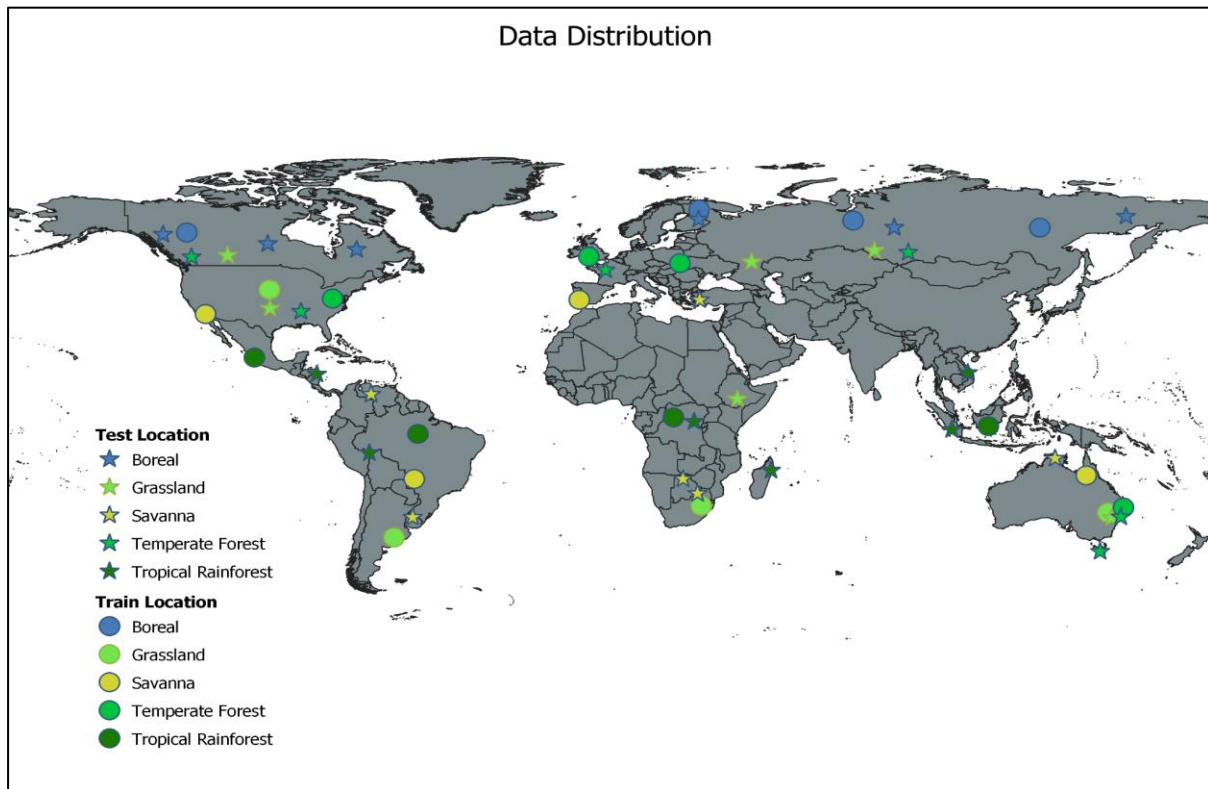


Figure 5 Global distribution of training and validation locations

As training a CNN model requires image patches of 512 by 512 pixels, with three colour channels and a corresponding label with a matching extent of 512 by 512 pixels, all images and labels were clipped into the respective patches. All patches had overlaps of 100 pixels on all sides. Training patches were split into an 80/20 training/validation split for model training. It is important to note the difference between validation patches (training patches used directly in model training to evaluate model accuracy after each epoch of training), and validation data (image and corresponding ground truth from which the results would be analysed, and predictions would be run). Validation patches (for model training purposes) were derived from the training image dataset, whereas validation data (for testing our trained models) were located in entirely different regions of the world from the training data, and specifically used to measure model performance. A stratified random sampling method was used to choose 40 000 sampling points per validation image, equating to 0.25% of the total number of pixels in each image.

### 3.3.4 COMPUTING RESOURCES

Data processing was run on a CPU and model training was run on a GPU. The system comprised an Intel i9 9900K, 16 core CPU, and an Nvidia GTX1080Ti GPU with 11Gb of memory and system memory of 64Gb RAM. The Python libraries used were OpenCV, Numpy and GDAL for data pre-processing, and Keras and Tensorflow DL libraries for training and inference. Numpy, SciPy, Matplotlib and Pandas libraries were used for validation sampling and statistical analyses.

### 3.4 METHODS

Figure 6 illustrates the general steps taken to classify tree cover using the DL model. Data selection and sourcing of training data were discussed in the previous section, while details of enhancements, a deep learning model and accuracy assessment will be provided in subsequent sections.

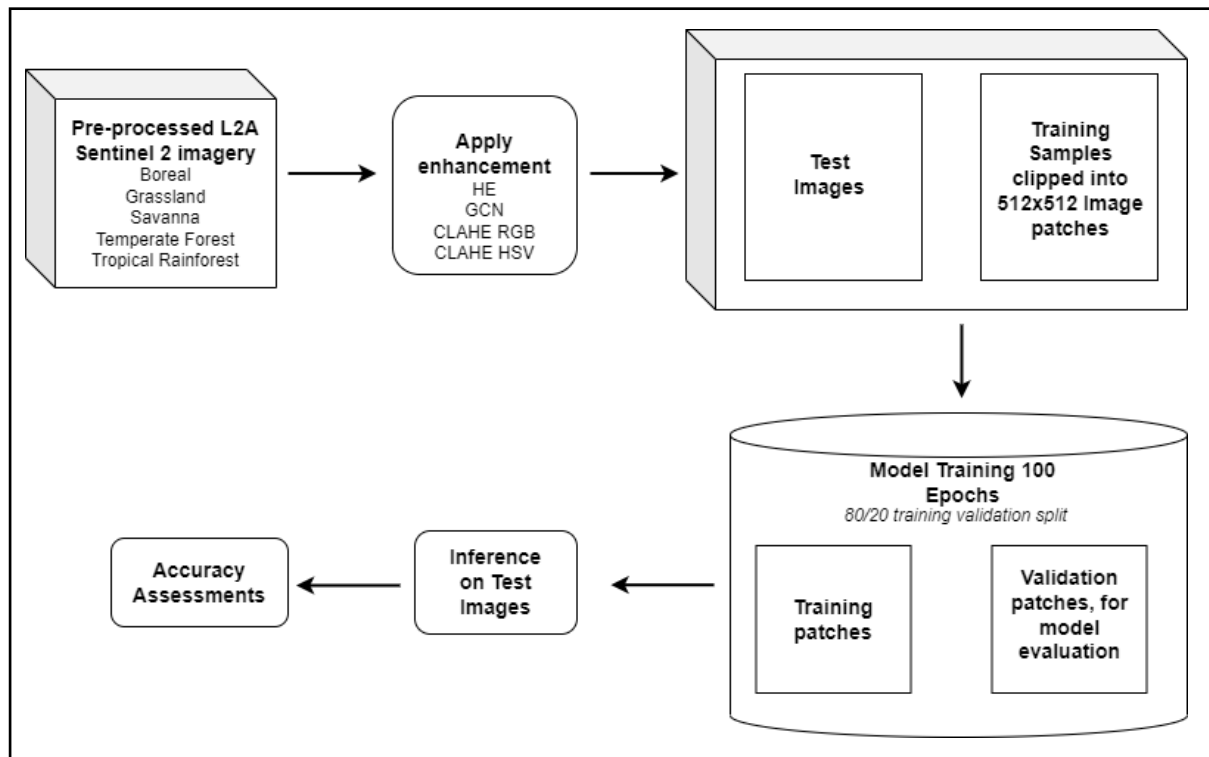


Figure 6 Experiment workflow

The entire workflow was applied for each enhancement method, per biome; thus, the workflow was repeated 20 times, and 20 models were trained. Each model was used to perform inference on six validation images for each of the corresponding biomes and enhancements. Random stratified sampling was performed on each validation image. Samples were collated for all six validation images per biome per enhancement and statistics were derived from these results.

#### 3.4.1 IMAGE ENHANCEMENTS

##### 3.4.1.1 HISTOGRAM EQUALISATION

Histogram equalisation (HE) is a common contrast enhancement method used for basic images such as human portrait images (Ma et al. 2018) as well as in medical image analysis. HE is a global contrast enhancement method which distributes the most frequent intensity values across the bit range. To visualise the effect of HE on a histogram, Figure 7 a) Histogram of image with no enhancement applied b) Histogram equalisation applied. HE is prone to over amplification and

distortion in some areas of the image, particularly in areas of homogenous intensities (Ma et al. 2019).

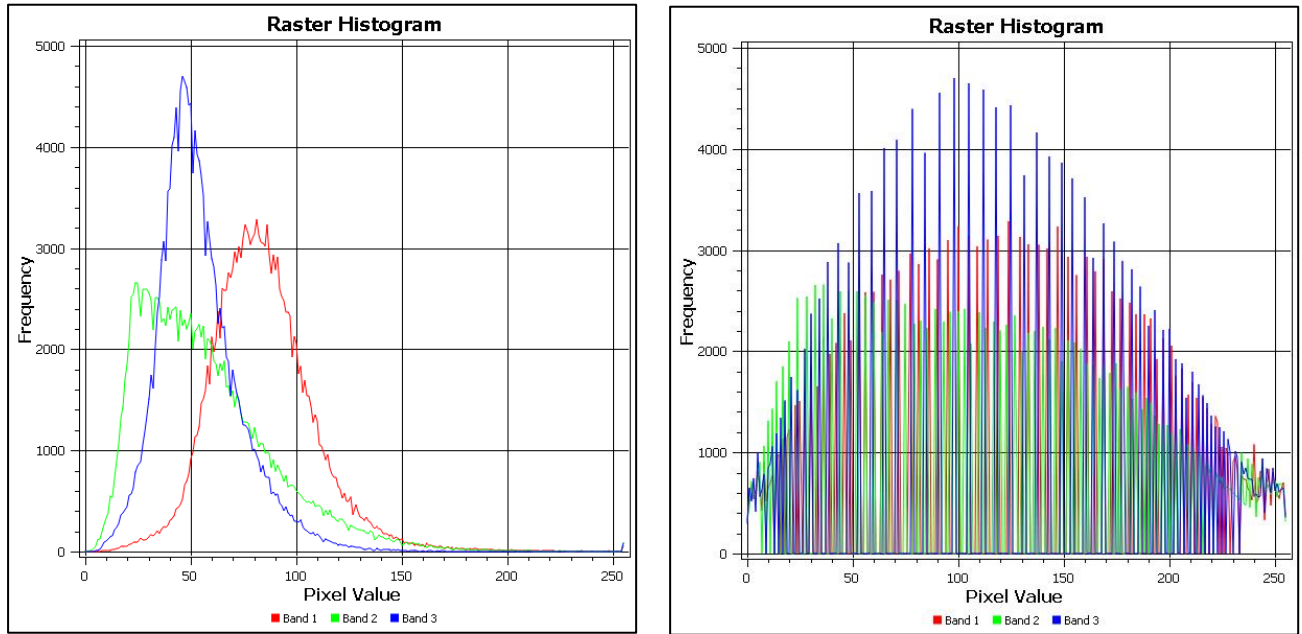


Figure 7 a) Histogram of image with no enhancement applied b) Histogram equalisation applied

To calculate the HE, a probability function (Equation 3) first needs to be defined.

For an image  $f$  of size  $i$  by  $j$ , forming a matrix ( $N$ ) of integer pixel intensities with values in the range 0 to  $L - 1$ , where  $L$  is the bit depth of values (256 for 8bit), let  $p$  represent the normalized histogram of  $f$  with a bin for each possible intensity.

$$p_n = \frac{\text{pixels of intensity } n}{N} \quad \text{Equation 3}$$

Where

$p_i$  Probability that a pixel has an intensity value  $n$

$n$  Number of pixels with intensity  $n$

$N$  Total number of pixels

Following from this, the equalized histogram (Figure 7 b) can thus be defined by  $g$  as:

$$g_{i,j} = \text{floor} \left( (L - 1) \sum_{n=0}^{f_{i,j}} p_n \right) \quad \text{Equation 4}$$

Where

$g_{i,j}$  Equalized histogram

$L$  Bit depth (256 for 8bit)

$p_n$  Probability that a pixel has an intensity value  $n$

$f_{i,j}$  Pixel intensity at position  $i, j$

### 3.4.1.2 GLOBAL CONTRAST NORMALISATION

GCN is a method that endeavours to normalise the contrast of an image based on the statistics of the entire training image stack. GCN is calculated by subtracting the mean pixel value of each colour band in the image stack from each pixel value in its corresponding colour band and dividing it by the standard deviation calculated across the full training image stack (Coates, Lee & Ng 2011; Pitaloka et al. 2017). GCN is calculated using the following equation:

$$x' = \frac{x - \mu}{\sigma} \quad \text{Equation 5}$$

Where

$x$  Input pixel intensity value

$\mu$  Mean intensity value across full training image stack per band

$\sigma$  Standard deviation of pixel values in full training image stack per band

When comparing the histograms of the base image and the GCN image, it is clear that the differences are very subtle (Figure 8). The aim of GCN is to reduce the amount of contrast within a set of images and to have the same intensity range across all images. This is beneficial when training a deep learning model as in the training process weights and biases are multiplied and summed respectively to the input data in order to produce activations that are then back propagated with gradients to train the model. Ideally the input data should all have a similar range which in turn reduces the possibility of an exploding or vanishing gradient during model training.

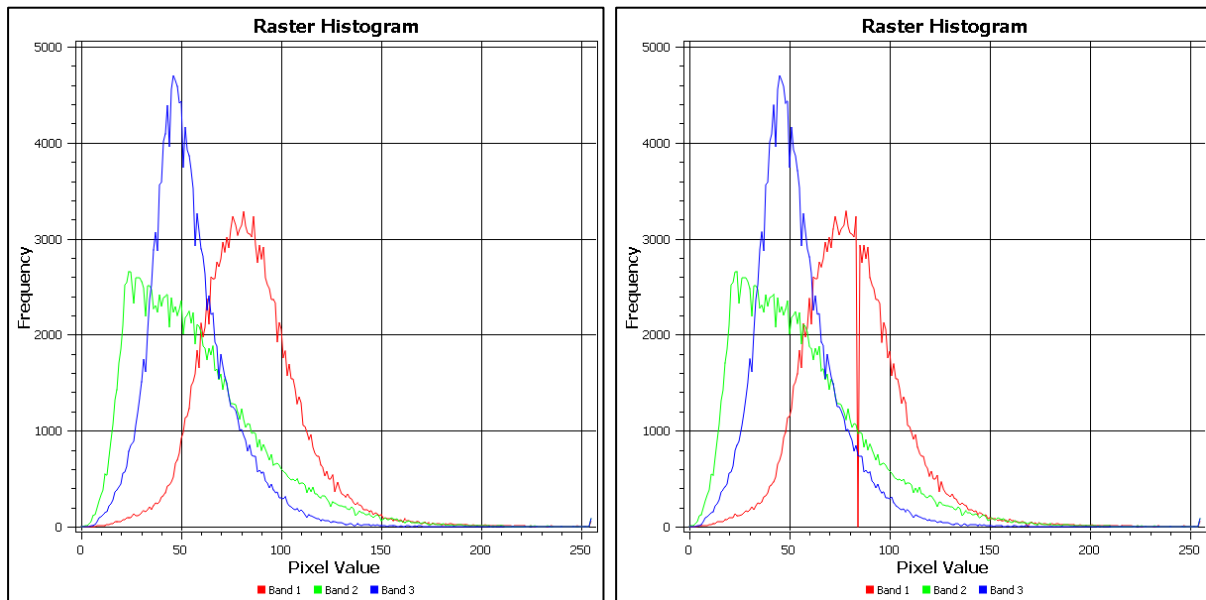


Figure 8 a) Histogram of image with no enhancement applied b) Histogram of image with global contrast normalization applied.

### 3.4.1.3 CONTRAST LIMITED ADAPTIVE HISTOGRAM EQUALISATION

To overcome the over-amplification caused by a global histogram stretch, CLAHE is proposed. CLAHE is processed on smaller tiles of the image, therefore localizing the HE process. Multiple histograms are computed for the full image, each referring to a different subsection of the image. These are all used to redistribute values across the full image using the cumulative distribution function (Ma et al. 2018). Two colour spaces were tested for the implementation of CLAHE, namely the red, green and blue (RGB), and the hue, saturation and value (HSV).

A colour space is an abstract model to describe colours in terms of intensity. The most common colour space to represent colours through digital media is the RGB colour space, which describes colours with linear intensity values for the red, green and blue channels. A mixture of these intensities on each colour channel displays the desired colour.

A popular colour space to perform image enhancement is the HSV colour space, which relies on hue, saturation and intensity values/ Hue is an angle that directly describes colour, saturation measures how close the hue is in relation to a white reference. (Saturation values are described as the radial distance from a central axis to the outer perimeter.) The value channel describes the illumination of the colour. (The higher the value, the brighter the corresponding pixel when viewed in RGB.) Since digital images are all displayed using the RGB colour space, images need to be transformed into the desired colour space first, processed, and then converted back to the RGB colour space to be interpreted in their true colour (Bora 2017).

### 3.4.1.3.1 CLAHE RGB

Applying CLAHE in the RGB colour space uses a scalar technique whereby CLAHE is processed independently on each colour channel, thereafter each processed channel is concatenated. A notable drawback is that the chroma of the image deviates from the original input. As seen in Figure 9, the histogram has a shift to the right, indicating a brighter image. The histogram with CLAHE applied in RGB is far more stratified showing less normalization within each colour band.

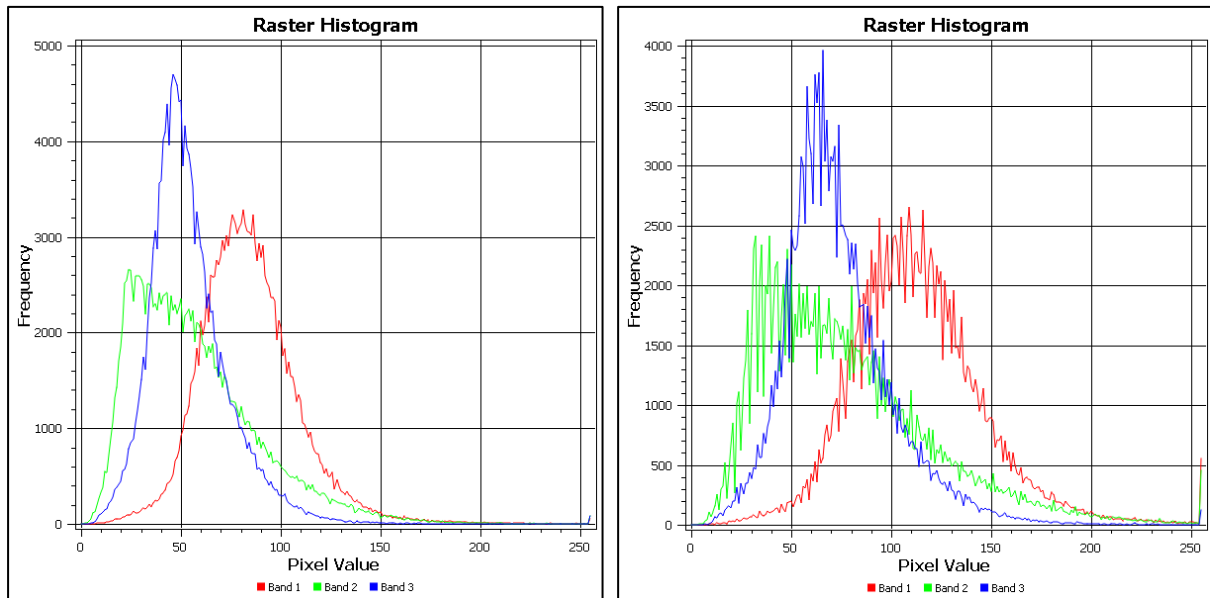


Figure 9 a) Histogram of image with no enhancement applied b) Histogram of image with CLAHE applied on each channel in the RGB colour space

### 3.4.1.3.2 CLAHE HSV

To retain the original chroma properties of an image, the image is first transformed into the HSV colour space. CLAHE is applied only to the V channel, thereby preserving the hue and saturation characteristics of the image. The image is then transformed back into the RGB colour space. This is a more efficient method of applying CLAHE, as it is applied only to a single channel in contrast with applying CLAHE in the RGB colour space where the number of processes depends on the number of channels in the image. As can be seen in Figure 10 b, the histogram is stretched across the full bit range whilst still maintaining a similarity in the shape of the peaks with those of the original histogram (Figure 10 a). This leads to a larger contrast between the dark and bright areas within the image.

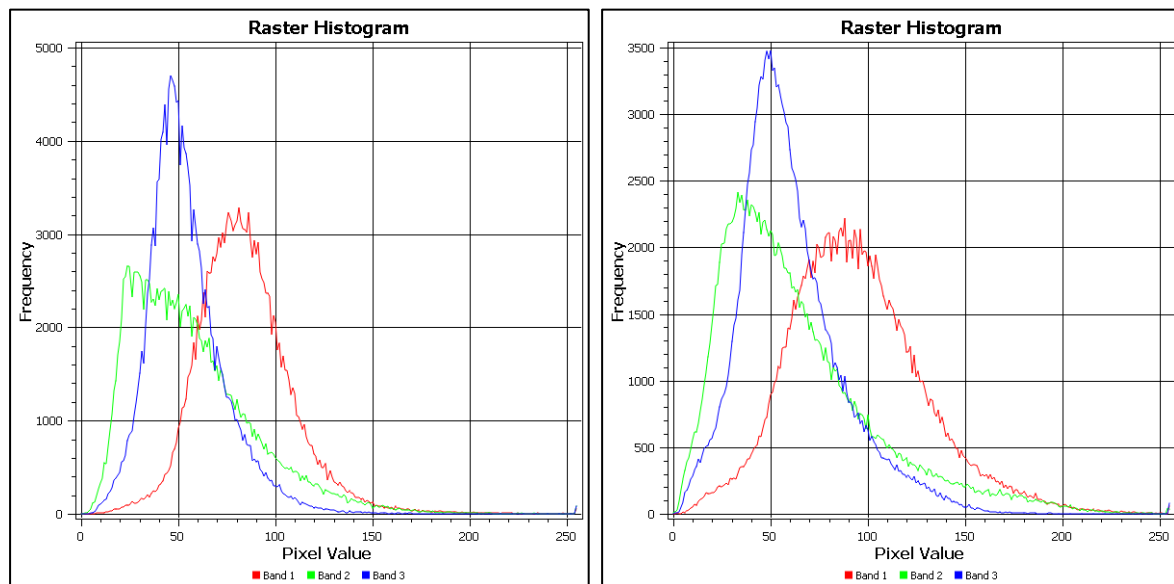


Figure 10 a) Histogram of image with no enhancement applied b) Histogram of image with CLAHE applied on V channel in HSV colour space

### 3.4.2 DEEP LEARNING

A UNet (Ronneberger, Fischer & Brox 2015) was the preferred architecture (Figure 11) because it uses skip connections between the convolutional layers in the encoder and the corresponding deconvolutional layers in the decoder. This enables the network to up-sample results at the same spatial resolution as the input data. The naming of the architecture is derived from the U-shaped symmetry of the architecture with a convolutional encoder and a corresponding deconvolution decoder (Figure 11).

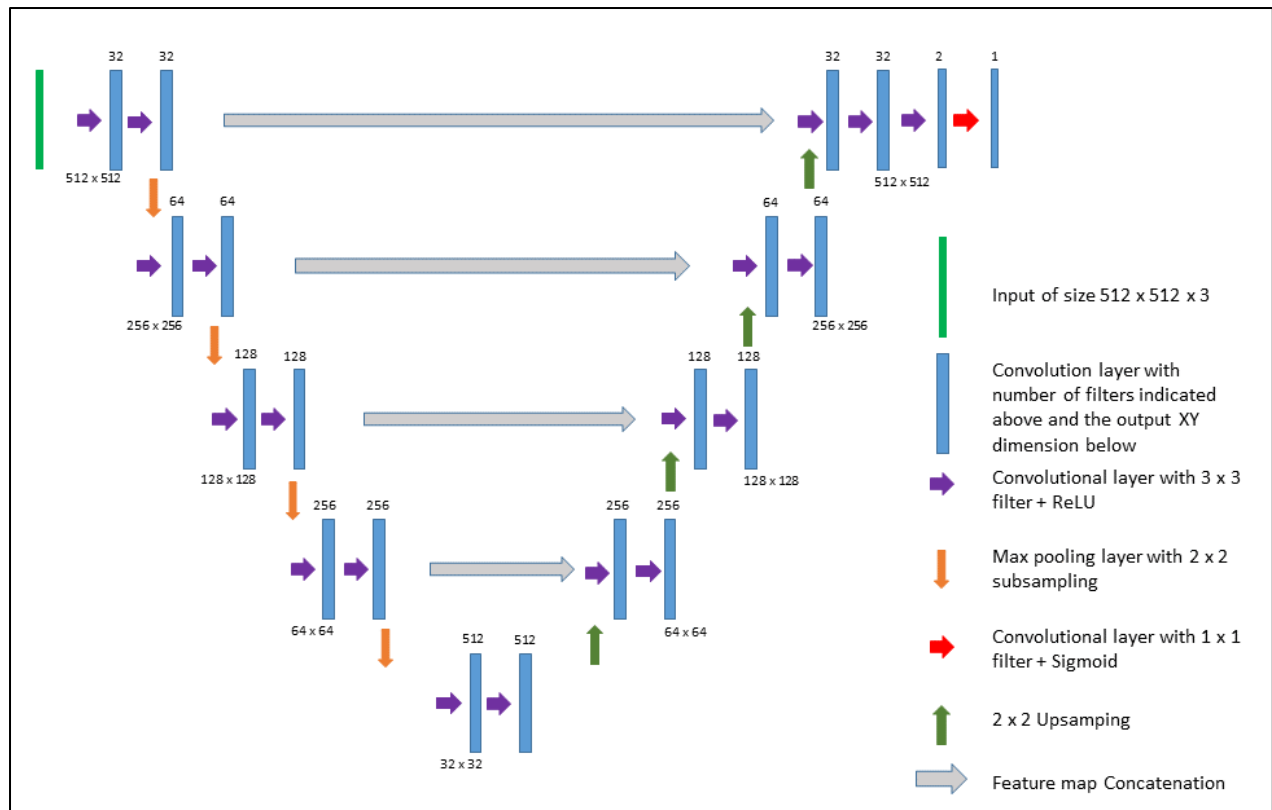


Figure 11 Graphical representation of the UNet architecture

Training a CNN can be difficult, and at times unstable, due to the initialization of random weights in the model. After each batch, the differences between inputs change when the weights are updated, thereby always changing the target. To counter this, batch normalisation was applied before each pooling layer to stabilise the learning process and reduce the number of epochs required to train the model effectively (Ioffe & Szegedy 2015). Dropout was applied at the end of the fourth and fifth convolutional blocks to reduce the chances of overfitting the model and to improve model generalisation. A probability of 0.5 was set for dropout for each batch processed by the model. **Error! Reference source not found.** outlines the hyper parameters used for this study.

Table 2 Hyper parameters used by the UNet used in this study

Hyper parameter	Type
Activation function	ReLu
Optimizer	Adam
Learning rate	0.0001
Loss function	Binary cross-entropy
Batch size	4



Epoch	100
-------	-----

The Rectified Linear Unit (ReLU) was used as the activation function for each convolutional layer in the encoder and decoder. The activation function allowed the model to learn more complex features by introducing non-linearity to the model and thereby improve its performance (Agarap 2018). ReLU also helps solve the issue of vanishing gradients where the gradients become too small to effectively update the model parameters during training (He et al. 2015). The final output layer was a sigmoid function with output probability values between zero and one.

Binary cross-entropy was employed as the loss function. It measures the difference between the predicted probability and the true binary label. Binary cross-entropy punishes the model heavily for predictions that are far from the target, making it a good choice for binary classification tasks like image segmentation, where the objective is to accurately classify pixels in a target class background. (Jadon 2020).

The Adam optimizer was chosen based on its ability to adaptively adjust the learning rate for each parameter based on the previous gradient information. This can lead to faster model convergence and improved performance compared to other optimization algorithms (Ronneberger, Fischer & Brox 2015; Xie et al. 2017).

The batch size has a significant effect on the training of DL models. It influences the trade-off between the stability of the learning process and the computational cost. A smaller batch size results in more frequent weight updates and can create a noise-like fluctuation in the gradients that can help the model escape from suboptimal solutions and converge to a better optimum. However, the computation is more expensive, and the convergence is slower. A larger batch size leads to smoother gradients and faster convergence, as the weight updates are based on a more accurate estimate of the gradients. However, the optimization may be less likely to escape from a local minimum (Keskar et al. 2017). A batch size of four was selected due to the computational resource limitations, however, it was still small enough to prevent the model from falling into a sub-optimal minima.

### 3.4.3 ACCURACY AND STATISTICS

To quantify the differences achieved by applying an enhancement to an image, the mean absolute difference (MAD) across the near-infrared, red and green bands in all training and validation images were calculated. The MAD evaluates whether the quantitative difference between an enhanced image and a base image may be indicative of the performance of a DL model. The MAD was calculated for each validation and training image by subtracting the enhanced image from the base image and then calculating the absolute mean value per image. The smaller the MAD, the better the matching of the baseline and the enhanced image.

The chosen metric for describing model performance was Matthews Correlation Coefficient (MCC), defined as (Matthews 1975):

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \quad \text{Equation 6}$$

Where

T True positive: correct prediction of  
P the positive class

T True negative: correct prediction of  
N the negative class

FP False positive: incorrect prediction  
of the positive class

F False negative: incorrect prediction  
N of the negative class

MCC considers true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). The MCC metric is specifically designed for binary classification, and it will generate a high score only if most of both positive and negative test cases are correctly predicted. Furthermore, MCC is especially useful for unbalanced datasets (Chicco & Jurman 2020). MCC ranges from -1 to 1, with -1 being a complete misclassification, 0 being no better than random and +1 being a perfect classification.

The overall accuracy (OA) is the proportion of true results measuring the degree of reliability of a classification where  $OA = (TP+TN) / (TP+FP+TN+FN)$ . OA's were calculated for each classification as an affirmation of the results achieved. OA cannot be used as a standalone statistic as it may not represent the true performance of a model due to the dataset being unbalanced.

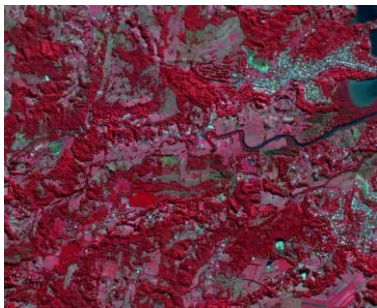

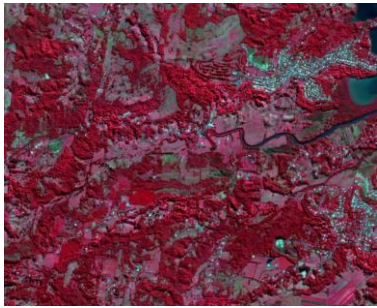
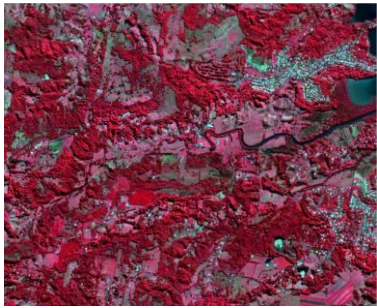
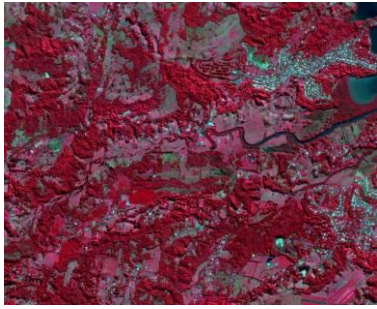
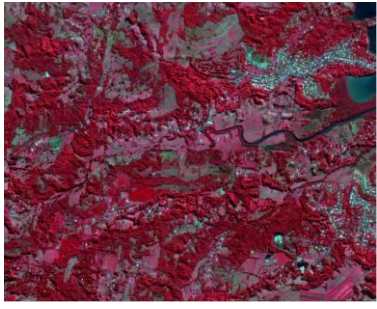
The results were based on 40 000 sampling points per validation image, equating to 1% of the total number of pixels in each image. A stratified random sampling method was used to choose samples.

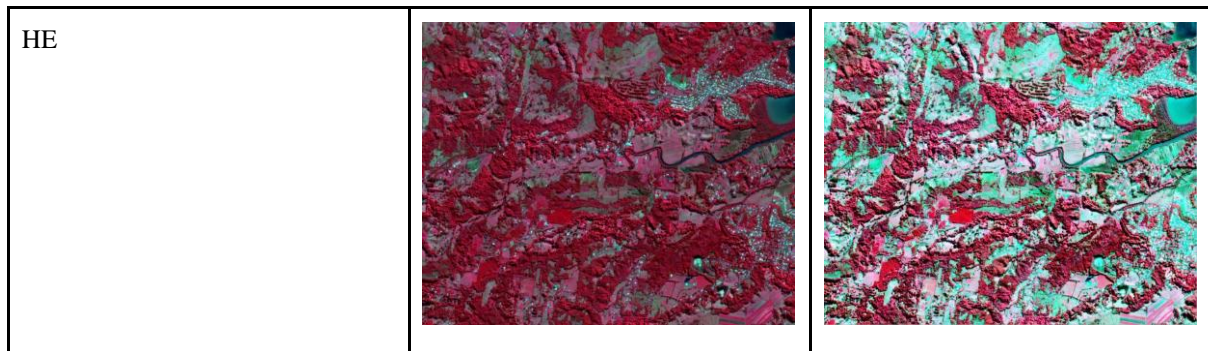
### 3.5 RESULTS & DISCUSSION

#### 3.5.1 EXAMPLE OF ENHANCEMENTS

Table 3 illustrates the enhancements applied to an image of a temperate forest biome. The images are in a false colour view with a band order of near-infrared, red and green as this was the combination passed through the DL model.

Table 3 Example of enhancements applied to a false colour composite

	Base Image	Enhanced Image
CLAHE RGB		
CLAHE HSV		
GCN		



Visually the GCN and CLAHE HSV outputs have almost unnoticeable differences (Table 3) between the base images and the enhanced images. The histograms of GCN (Figure 8) and CLAHE HSV (Figure 10) agree with the visual representations. GCN reduced the contrast of all images in a training set by subtracting the global mean value for each band in the stack and dividing that by the global standard deviation. Whereas the visual differences of the images may be subtle, the model's ability to learn features tended to improve when inputs were normalized mean centred. The same mean and standard deviation factors were applied to all validation data prior to prediction. Visually the CLAHE RGB and HE outputs showed significant differences from the base images, with vegetation being accentuated. CLAHE HSV increased contrast whilst preserving the hue and saturation of the base images. The HE outputs indicated a risk of over exposing images. In the example in Table 3, HE enhanced the contrast, especially over vegetated areas, resulting in vegetation being identified more easily by the naked eye. The differences between the HE images, and the base images were much more prominent, whereas CLAHE in HSV and GCN displayed very subtle differences.

### 3.5.2 EFFECT OF ENHANCEMENT (MAD)

Figure 12 Mean absolute residual difference (MAD) for each test and training image illustrates the MAD of each baseline training/validation image compared with the respective enhanced training/validation image. The MAD results across all images per enhancement, along with their respective standard deviations, are further outlined in Table 4.

Figure 12 Mean absolute residual difference (MAD) for each test and training image

Table 4 Overall MAD score and standard deviation across all training and test images per enhancement

	<b>Mean Absolute Difference</b>	<b>Standard Deviation</b>
CLAHE RGB	0.1020	0.0205
GCN	0.0162	0.0125
HE	0.2304	0.0546
CLAHE HSV	0.0345	0.0094

GCN had the lowest MAD measure and the lowest standard deviation. HE had the highest MAD measure, namely the largest differences between the baseline and enhanced images. HE also had the largest variation across images. The resultant accuracy for each UNet prediction was determined based on the MCC score. The mean MCC scores for each enhancement technique per biome are shown in Figure 13.

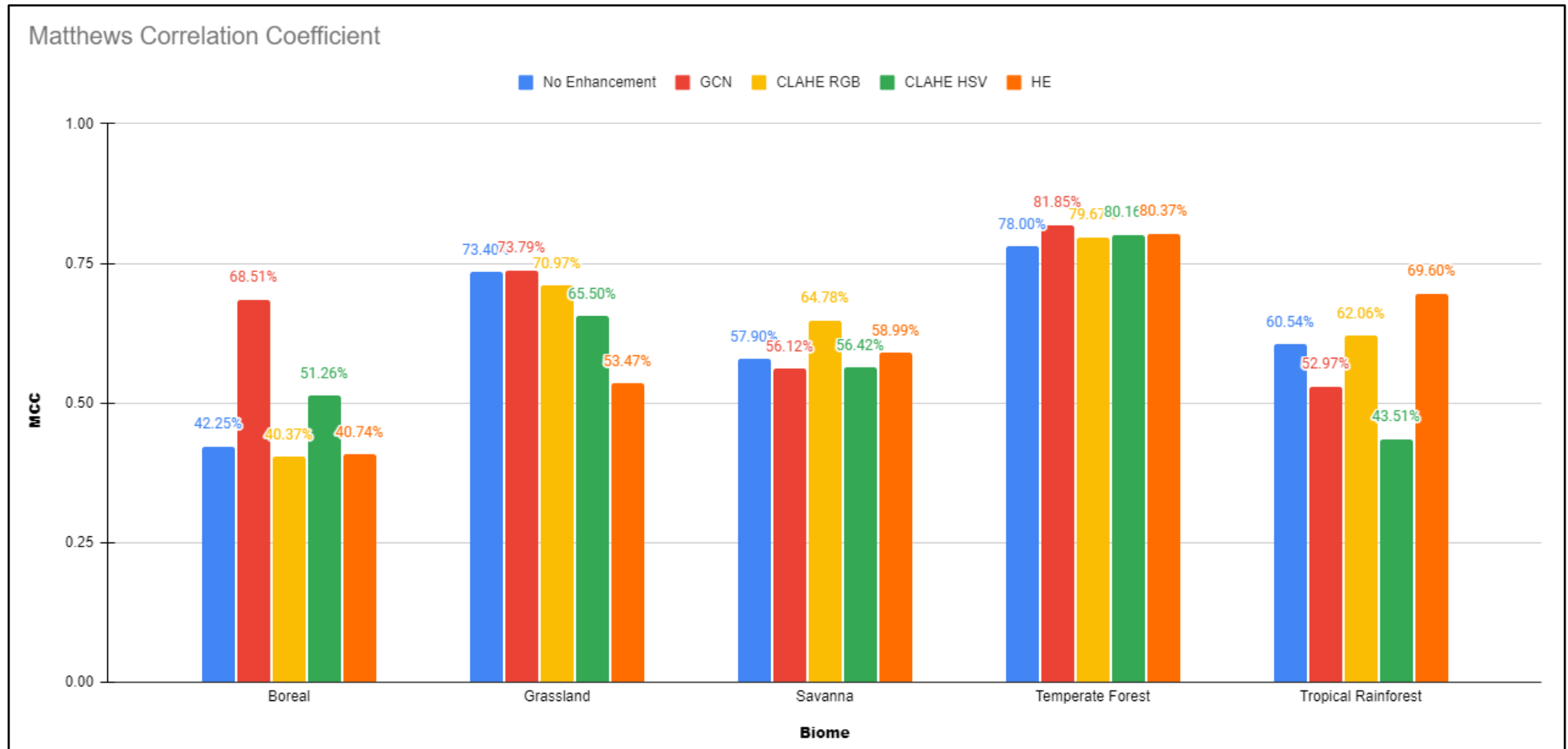


Figure 13 Mean MCC score per enhancement per biome



Among all enhancement techniques, GCN demonstrated the most consistent performance across various biomes, with the exception of Tropical Rainforests. GCN also had improved MCC scores when measured against imagery with no enhancement technique in three of the five biomes. Most notably, in the Boreal biome, GCN outperformed base imagery by more than 26%, with GCN producing an MCC (expressed as a percentage) of 68.51% and no enhancement producing an MCC of 42.25%. GCN was least effective when applied to images in the Tropical Rainforest biome. This was possibly due to the similarities of the histograms for images in Tropical Rainforests. GCN showed the highest MCC of all the enhancements in three of five biomes.

CLAHE applied in the RGB colour space showed mixed results across biomes. It was the best performing enhancement in the Savanna biome with an MCC of 64.78%. HE was the second best enhancement in this biome with a meagre MCC of 58.99%. CLAHE in RGB was the worst performing enhancement in the Boreal biome with an MCC of 40.37%. CLAHE in RGB did not outperform any enhancement in three of five biomes. CLAHE in RGB performed better than CLAHE in HSV in three of five biomes. However, of all the enhancements, CLAHE in HSV did not rank highest in any biome. CLAHE in HSV outperformed no enhancement in only two of the five biomes. CLAHE in HSV had the lowest MCC score of all enhancements in the Tropical Rainforest biome. Along with GCN, CLAHE in the HSV was the only other enhancement to have a higher MCC than no enhancement in the Boreal biome.

HE had mixed results across biomes. HE had higher MCC scores in three of the five biomes compared with no enhancements. HE was the best enhancement in the Tropical Rainforest biome with an MCC score of 69.60 compared with the second highest MCC being CLAHE in RGB with an MCC of 62.06%. HE also did well in the Temperate Forest biome with an MCC of 80.37%. In the Grassland biome, HE was by far the worst performing enhancement, with the difference between the highest MCC scoring enhancement (GCN) and the lowest scoring enhancement (HE) being more than 20%. The biomes in which HE did well, specifically the Temperate Forest and Tropical Rainforest biomes, were predominantly covered with vegetation. As HE aims to spread the most frequent intensities across the full bit range, and if the majority of intensity values represent forests, those pixels would be centred in the histogram. Conversely, for Grassland where HE performed the worst, the most frequently occurring intensity values were higher in the bit range bearing in mind that much of the land area was covered by crops and grassland as opposed to darker forests.

It is interesting to note that in the Tropical Rainforest biome, the two worst performing enhancements; GCN and CLAHE in HSV, were also the two enhancements with the least residual differences when compared with their baseline images. Conversely, in the Boreal biome, these two

enhancements produced the highest MCC scores of all the enhancements as well as against no enhancement.

GCN had the highest mean MCC across all biomes with an MCC rating of 0.67. CLAHE applied in the HSV colour space had the lowest mean MCC of 0.59. Only GCN and CLAHE RGB had higher means than the baseline dataset. Table 5 highlights the MCC and OA scores of the enhancements which outperformed no enhancement in **bold**.

Table 5 Mean MCC and OA per biome per enhancement, bold indicating cases outperforming no enhancement

	No Enhancement		GCN		CLAHE RGB		CLAHE HSV		HE		MCC		OA	
	MCC	OA	MCC	OA	MCC	OA	MCC	OA	MCC	OA	Mean	STD	Mean	STD
Boreal	0.42	0.71	<b>0.69</b>	<b>0.84</b>	0.40	0.70	<b>0.51</b>	<b>0.76</b>	0.41	0.69	0.49	0.12	0.70	0.06
Grassland	0.73	0.90	<b>0.74</b>	<b>0.90</b>	0.71	0.89	0.65	0.86	0.53	0.84	0.67	0.08	0.84	0.03
Savanna	0.58	0.79	0.56	0.79	<b>0.65</b>	<b>0.83</b>	0.56	0.75	<b>0.59</b>	<b>0.80</b>	0.59	0.04	0.76	0.03
Temperate Forest	0.78	0.89	<b>0.82</b>	<b>0.91</b>	<b>0.80</b>	<b>0.90</b>	<b>0.80</b>	<b>0.90</b>	<b>0.80</b>	<b>0.90</b>	0.80	0.01	0.88	0.01
Tropical Rainforest	0.61	0.82	0.53	0.79	<b>0.62</b>	<b>0.83</b>	0.44	0.75	0.70	0.86	0.58	0.10	0.77	0.04
Mean	0.62	0.82	0.67	0.85	0.64	0.83	0.59	0.80	0.61	0.82				
STD	0.14	0.08	0.12	0.06	0.15	0.08	0.14	0.07	0.15	0.08				

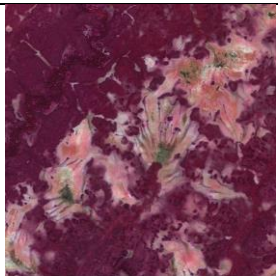
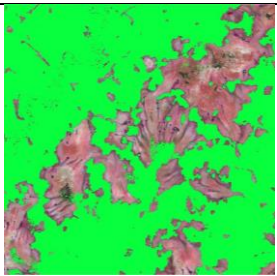
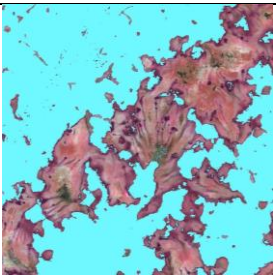
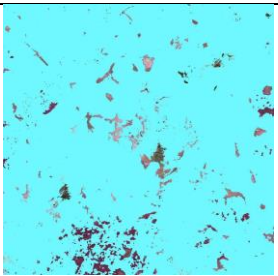
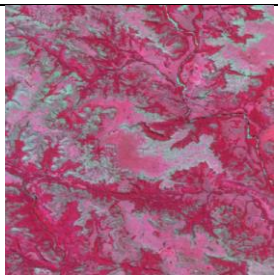
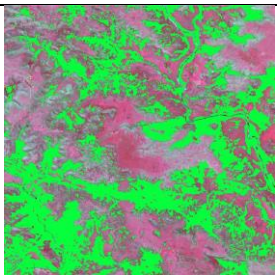
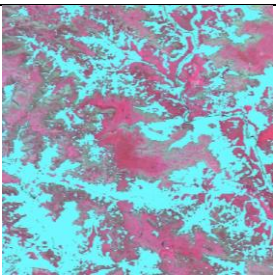
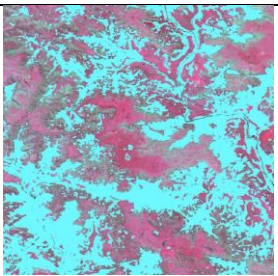
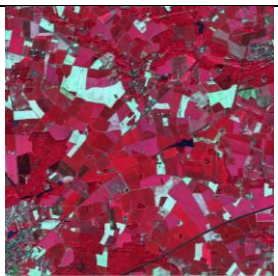
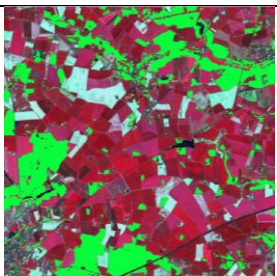
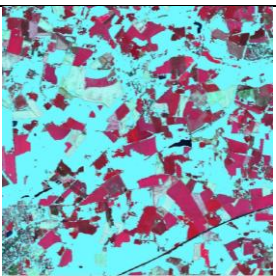
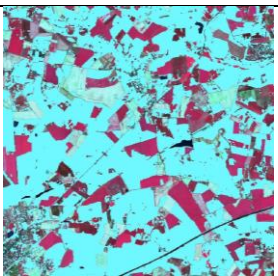
Temperate Forest was the only biome in which all enhancements outperformed the baseline dataset. This may suggest that enhancements are useful in cases of the objects of interest having well delineated boundaries, as is the case of trees in the Temperate Forest biome. This finding mimics the results found of a previous study in which the classification of melanoma and benign skin lesions was improved (Özbay & Özbay 2021). In that case, the objects of interest also had clear delineated boundaries (Özbay & Özbay 2021). There was high variability in the performance of enhancements in the Boreal biome, with standard deviations of 0.12 and 0.06 in the MCC and OA scores respectively.

Table 6 illustrates examples of results for each enhancement in the biome in which an enhancement performed the best in relation to the others. It's important to keep in mind that these results are specific to the image presented in the table and may not necessarily align with the findings presented in Table 5. Precision and recall for each of the images were calculated. Precision is the ratio of true positives to the sum of true positive and false positive cases. A higher precision score



indicates a lower rate of false positive cases. Recall is the ratio of true positives to the sum of true positives and false negative cases. A higher recall rating indicates a lower rate of false negative cases. GCN in the Boreal biome achieved much higher precision and recall metrics of 0.99 and 0.84 respectively, compared with the baseline prediction which had a high precision of 0.95. But a high number of false negative tests led to a low recall of 0.42. This is reflected in the MCC score with GCN attaining an MCC of 0.79 and no enhancement scoring an MCC of 0.20. The overall accuracy aligns with these results, but the high number of false positive test cases are not reflected in this metric. Although the table below is not indicative of the results trends for each enhancement and biome, it nonetheless illustrates the importance of using the MCC score as a metric for evaluating model performance. For example, CLAHE in HSV and no enhancement in the Temperate Forest biome exhibits recalls of 0.99 and 1.0 respectively, however they have low precision ratings of 0.59 and 0.5, with a deceptively high OA recording 0.8 and 0.72.

Table 6 Example result of best performing enhancement for each biome

	Image	Ground Truth	Prediction	No Enhancement Prediction	Precision & Recall			
GCN (Boreal)						No enhancement	GCN	
					Precision	0.95	0.99	
					Recall	0.42	0.84	
					MCC	0.20	0.79	
					OA	0.69	0.88	
CLAHE RGB (Savanna)						No enhancement	CLAHE RGB	
					Precision	0.81	0.83	
					Recall	0.97	0.97	
					MCC	0.81	0.83	
					OA	0.91	0.92	
CLAHE HSV (Temperate Forest)						No enhancement	CLAHE HSV	
					Precision	0.50	0.59	
					Recall	1.00	0.99	
					MCC	0.55	0.65	
					OA	0.72	0.80	



Mean MCC scores for all the enhancements across all biomes (Figure 14) were calculated and can be directly compared with one another as the same validation images for all biomes were used for each of the enhancements. GCN and CLAHE in RGB were the only enhancements to have higher mean MCC's than no enhancement. The application of GCN as a normalisation technique has been shown to improve classification accuracies of deep learning models outside the remote sensing domain (Pitaloka et al. 2017). The accuracy of classifying emotion based on facial expression improved from 61.81% to 89.18% after using GCN as a pre-processing step in a CNN (Pitaloka et al. 2017). Although the study was focused on image categorization, not segmentation classification.

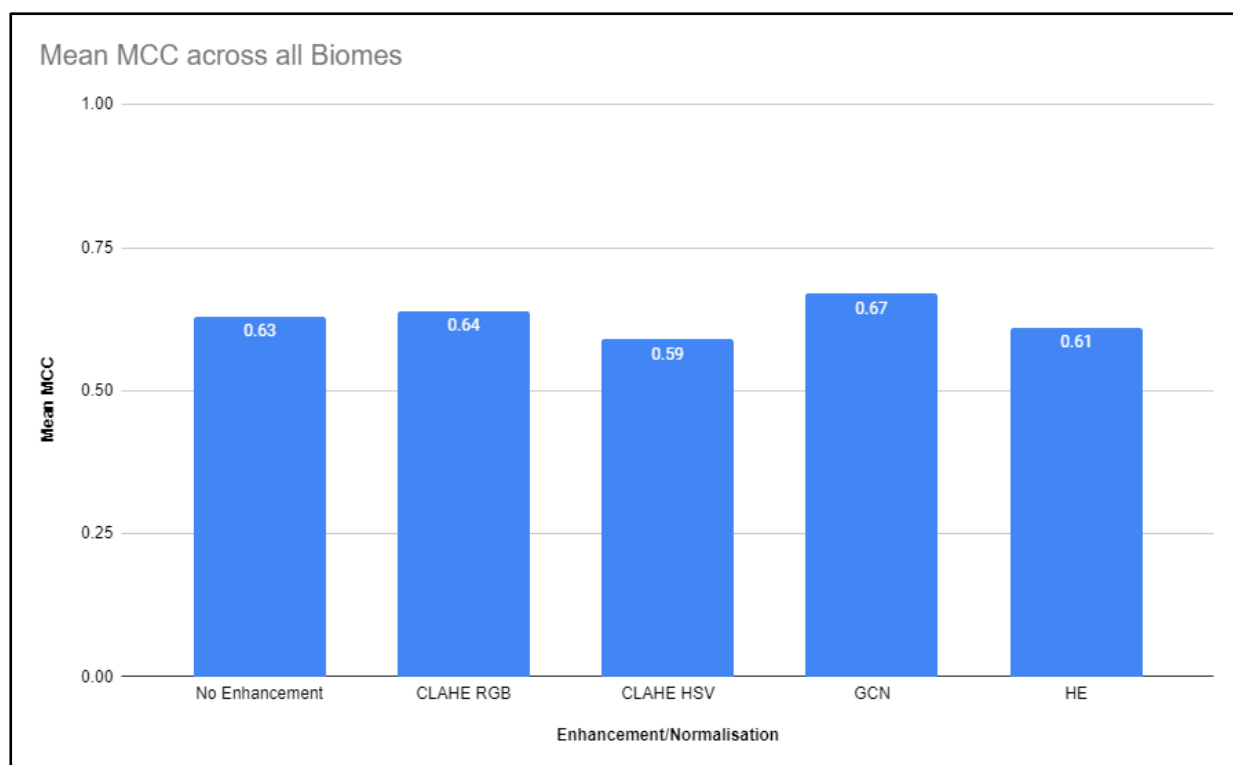


Figure 14 Mean MCC across all biomes

Image enhancements have been shown to improve the accuracy of natural image classifications when the object of interest is clearly defined, such as the segmentation of motor vehicles, pedestrians and street signs in the case of self-driving cars. The segmentation model is less dependent on the spectral values and ratios between colour channels than the geometric characteristics of the objects. The objects in these cases have clear edges. In contrast, in the context of tree segmentation from medium resolution satellite imagery, some trees may be smaller than the size of the relevant pixel. It is difficult to characterize the texture of a single tree from 10m resolution images thus the more important characteristics are the spectral values and the ratios between spectral values of pixels as well as between signal bands. This could be why contrast

enhancements such as CLAHE and HE, had minor impacts (Table 5) on the performance of the DL model. Enhancement methods may be useful for problem-specific cases, as an enhancement may work well for one case but yield poor performance for another case. A more suitable use case for image enhancements in remote sensing may be for the segmentation of buildings and/or trees in very high-resolution imagery (0.5m). This will result in the object of interest being much larger in relation to the pixel size, and will enhance the available semantic information (i.e. regular shapes of footprints, roof shapes, driveways, yards etc). GCN aims to normalise the contrast of all images in a training stack and subsequent image predictions. The normalisation results in all images having similar contrast level. It might be possible that the model has learned pixel ratio features more effectively with GCN as the ratios between bands are normalised. Trees in the Boreal biome are more difficult to delineate visually in comparison to other biomes, such as Savanna or Temperate Forests, in which the boundaries between trees and no trees are quite clearly defined, thus assisting the model to learn the geometric and semantic properties of trees more effectively. This could explain why enhancements such as CLAHE and HE, which promote the learning of geometric features, are more effective than other enhancements in biomes where trees are more easily delineated, such as in Savanna and Tropical Rainforest (Table 5).

### 3.6 CONCLUSION

This chapter outlines how various image enhancement techniques were tested in order to determine which would perform best for global forest classification from Sentinel-2 imagery using a CNN. Testing was done per biome to reduce intraclass variability and improve model performance. GCN, HE, CLAHE, both in the RGB and HSV colour spaces, were tested against a baseline image to which no enhancement had been applied. On the basis of MCC scores, GCN performed the best against the baseline images used in most of the biomes. GCN was especially effective for the Boreal biome with a MCC score 0.1725 higher than the next best performing enhancement. GCN was also one of the more consistent techniques, outperforming the baseline prediction in three of the five biomes, and having the highest mean MCC (0.67) across all the biomes. The next best performing technique was CLAHE applied in the RGB colour space with a mean MCC of 0.64. CLAHE in the RGB colour space was the only other technique that outperformed baseline predictions on average. Given that only two of the tested enhancements improved model MCC scores, it is recommended that image enhancements are critically evaluated for their suitability to the context of their application.

A limitation of the study was the exclusive use of LuxCarta's proprietary 8bit scaled data, which is not openly accessible. Although best efforts were taken to limit any inherent biases in the training and validation labels datasets, these do still exist. Images were selected on summer

imagery of the locale of each image; however, some seasonal variations still exist in both training and validation datasets.

To the best of the author's knowledge, this is the first paper critically comparing enhancement techniques specifically for the improvement of a CNN for global forest classification. The paper serves as an important link with results obtained in other computer vision and medical research fields and demonstrates that some of the techniques examined have potential in the remote sensing domain. It is intended that this work be further developed by assessing image normalization via DL domain adaptation techniques, and by the use of state-of-the-art pre-trained networks with the aid of transfer learning. There is potential for these activities to yield further improved results and pave the way to training a more accurate, global forests model.



## **4 CHAPTER 4: DATA AUGMENTATION METHODS FOR IMPROVING THE RESULT OF A CNN BASED GLOBAL FOREST CLASSIFICATION MODEL FROM SENTINEL-2 OPTICAL DATA**

### **4.1 ABSTRACT**

In the absence of accurately annotated training data for remote sensing, training a robust, deep learning model for global tree cover classification remains a challenge. As its first objective, this study evaluated basic data augmentation methods and prediction frameworks that might lead to achieving an accurate, global tree cover classification. A training dataset was artificially inflated using common geometric and colour data augmentation methods borrowed from the computer vision domain. Their effectiveness in improving the generalisability of a U-Net model for tree classification was tested. Both geometric and colour augmentations, when applied individually, showed improvements in model accuracy. When applied together, the combined augmentations showed only marginal improvements over the individually applied augmentations. The second objective was to test two approaches towards achieving a global tree classification. The first was a model per biome approach, whereby a model was trained with data derived only from the respective biome. The second involved training a single globally representative model with training data from all biomes combined. This resulted in higher MCC scores than the multi-model approach. The diversity in training data appeared to increase model robustness. Thus, it was found that training a single, globally representative model with a combination of colour and geometric augmentations led to an effective framework to infer a global tree classification.

### **4.2 INTRODUCTION**

Information about the spatial distribution of global forests is important for commercial purposes, ecological monitoring and also indirectly impacts on human and environmental health (Ottosen et al. 2020). Due to anthropogenic influences, these resources are under severe pressure (Drummond & Loveland 2010; Kouassi et al. 2021). The effective management of global forest resources requires an efficient and robust classification method to extract a global tree cover dataset. Achieving this remains a major challenge, due to the computational cost of processing a global dataset at a high spatial resolution. There has been a proliferation of open access to medium resolution (10m), multispectral imagery from the Sentinel-2 constellation of satellites, as well as technical advancements in the field of artificial intelligence, particularly the use of deep learning (DL) for medical imaging analyses and computer vision tasks (Krizhevsky, Sutskever & Hinton 2012; Ronneberger, Fischer & Brox 2015). The combination of these advances offers an opportunity for further research on the transferability of concepts from computer vision to remote sensing as potentially efficient methods for tree classification on a global scale.

### 4.2.1 CLASSIFICATION APPROACHES

Common image classification categories in the computer vision domain rely on image-level classification, whereby a single label is assigned to describe a full image. For remote sensing, image level classification is susceptible to a loss in spatial resolution and patch-based classification is preferred. Pixel based classification is a common classification category for remote sensing (Sekertekin, Marangoz & Akcin 2017), with each pixel being assigned a class. Pixelwise classification can be achieved by DL within convolutional neural networks (CNN) through the use of fully convolutional architectures (Maggiori et al., 2017). Remote sensing imagery can be noisy and result in visually unpleasing classification maps (Boonprong et al. 2018). Object based image analysis (OBIA) is a form of image classification that overcomes this problem by grouping pixels into homogenous zones based on colour and shape, and then assigning a class label to each pixel within the segment (Lu & Weng 2007). However, OBIA is a computationally expensive process as the segments need to be computed prior to classification. Image segmentation in DL, in the form of semantic segmentation, provides the benefits of pixel-based classification with the smoothness of OBIA. Pixels are assigned a class using both spectral values and contextual information to group them into classes based on semantic features. A pixelwise classification approach is preferred as the spatial resolution is retained.

### 4.2.2 DEEP LEARNING

DL models, especially when trained with relatively small datasets, are prone to overfitting, meaning the model does not generalise well in respect of unseen data (Cogswell et al. 2016). To reduce overfitting, DL models require very large training datasets in order to incorporate variability within classes (Weinstein et al. 2020; Yu et al. 2017b). Although the problem of overfitting can be addressed by adding dropout functions and batch normalization layers in the network architecture (Zheng et al. 2018), it is important to address the root cause, namely the lack of a large dataset. The production of datasets is very costly and time consuming (Townshend et al. 2012). Consequently, the availability of image-level training datasets in the remote sensing domain is limited, and even more so in the case of pixel-level datasets (Townshend et al. 2012). Large image-level datasets, such as the ImageNet dataset, contain 15 million images categorised into 22000 classes. Such large datasets are not yet available in the remote sensing domain (Yu et al., 2017b). One method for introducing and increasing diversity in the training dataset is to enlarge the training dataset artificially by applying data augmentation to the existing training dataset (Shorten & Khoshgoftaar, 2019). Data augmentation can be categorised into two main types: data warping augmentation and oversampling augmentation (Shorten & Khoshgoftaar, 2019).



#### 4.2.2.1 OVERSAMPLING

Oversampling refers to the generation of a completely new, synthetic image (Shorten & Khoshgoftaar, 2019). Image mixing is an example of oversampling which involves combining multiple real images into one by taking the average pixel value for each image stack over N number of images. An example of oversampling that has been applied successfully is a Generative Adversarial Network (GAN) used to create additional training images (Loey, Smarandache & Khalifa 2020). GAN-based data augmentation techniques have been especially useful in the medical imaging domain, for which there is also a scarcity of training datasets. In the case of training a robust model for liver lesion identification, GAN-based data augmentations outperformed models trained with basic data warping, with model performance increasing from 88.4% using data warping, to 92.4% using GAN-based augmentations (Frid-Adar et al. 2018).

#### 4.2.2.2 DATA WARPING

Oversampling and data warping are not mutually exclusive as data warping can be applied to synthetically created images from GAN's to further enhance and diversify the relevant datasets (Shorten & Khoshgoftaar, 2019). Geometric and colour transformation fall into the broad category of data warping. Geometric transformation modifies the geometric properties of an image. This type of augmentation is suitable for satellite data as semantic features are orientation invariant (Scott et al., 2017). Colour transformation changes the pixel intensity values, thereby mimicking different illuminating conditions and atmospheric differences, but retain the geometric properties of images. This study will focus on data warping augmentation.

##### 4.2.2.2.1 GEOMETRIC TRANSFORMATIONS

Geometric transformation refers to changes to the affine transformation matrix of the pixel data. Popular geometric transformation applications include flip, rotation and transposition (Yu et al., 2017a). A flip can be performed on the X or Y axis of an image, or both. Rotations involve rotating an image anywhere between zero and 360 degrees. Shear translation performs shifts in both the X and Y axes and thus distorts the shapes of objects in an image. Geometric transformations are very useful in classification tasks where images are rotation invariant. These methods increase diversity in the training dataset without changing the spectral or topological properties of the data (Yu et al. 2017a).

##### 4.2.2.2.2 COLOUR TRANSFORMATIONS

Colour transformation is another method that has been successful in computer vision (French & Mackiewicz 2022). Applying gamma and contrast variations to augment training data has been

used to improve the accuracy of DL models, both in self-portrait segmentations (Shen et al. 2016) and melanoma classifications (Perez et al. 2018). Colour transformation preserves the geometric characteristics of objects. In some cases it is therefore more useful to apply augmentations that vary pixel intensity (colour) than using geometric augmentations, as in a text/number recognition task. This is because a value six (6) when flipped in the vertical axis can be interpreted as a value nine (9) and this would destroy the integrity of the training dataset. The nature of land cover classification is rotation-invariant, meaning that a satellite image can be interpreted correctly regardless of its orientation. Conversely in cases when colour properties are distinctive features, it would be desirable to apply only geometric augmentations so as to preserve spectral and topological properties, as in the case of optical satellite imagery. However, Robinson et al. (2019) found that although colour was a very predictive feature within a small geographic region, it became less consistent across larger regions with high intra-class colour variability. Thus, using colour augmentation to illustrate image variability across geographical regions may be useful. The relationship between the colour input bands can be a distinctive feature. Chatfield et al. (2014) showed that when all input RGB images were converted to grayscale and trained again, model accuracy reduced by 3% in the case of image labelling tasks. Pixel intensity and basic geometric augmentations have been shown to be effective in increasing the size and improving the robustness of CNN's in domains with very limited training dataset sizes, as in the case of brain tumour detection in the medical imaging domain (Khan et al. 2020).

Despite the abundance of research papers on refining model architectures to improve model performance, there has been less focus on how data augmentation methods affect and may improve model performance (Lei et al. 2019). This constitutes a significant gap in literature, despite the existing limited studies comparing the effects of different data augmentation methods on the performance and robustness of a deep learning model, specifically in the remote sensing domain.

#### **4.2.3 BIOME BASED CLASSIFICATION**

Currently, there is no research on directly comparing the performance of a DL model trained exclusively with training data derived from a single biome, with that of a model trained with data derived from multiple biomes. However, it should be noted that the effects of differing ecological and climatic conditions on classification via classical machine learning, has been studied. Woodcock et al. (2001) found that overall accuracies were negatively affected by ecological and climatic differences, when classifying over extensive geographical regions using classical machine learning methods. This was mainly due to variances in phenology and the structure of vegetation types (Olthof et al., 2005). In addition, temporal and illuminating conditions can cause differences of spectral signatures of the same land cover class (McDermid, Franklin & LeDrew 2005). The

spectral separability of land cover classes generally has an inverse relationship with distance, whereby separability decreases with an increase in distance (Verhulp & Van Niekerk, 2016). Higgs (2021) suggest that for forest genera mapping from Sentinel-2 data, training data should be collected within 500km of where a RF model is applied and must be of similar rainfall patterns. Pax-Lenney et al. 2001 found that applying simple dark object subtraction as an atmospheric correction method enabled a neural network to generalize well temporarily. However, accuracies dropped by 8%-13% in mean accuracies when extending models spatially from regions where training data was derived. Thus, the hypothesis is that a model trained with data from one biome will increase accuracy when applied to the same biome on a different continent, However, the lack of diversity in the training set of a single biome, may limit the model's ability to detect all trees, particularly in areas close to transition zones between different biomes, where phenology, rainfall patterns and vegetation structure are less similar to the training data derived for the respective biome.

For use in an operational environment, a targeted model trained per biome may cause unwanted edge matching artefacts in bordering biomes in which a different model is applied. In contrast, a single model trained using a balanced training dataset derived from all biomes, would not display edge matching issues when applied to bordering areas of biomes. However, it is not known whether a single model trained with data from all biomes, with data augmentation applied, would be sufficiently robust to outperform a targeted model.

#### **4.2.4 OVERVIEW**

This study aims to determine whether the application of geometric or colour augmentation (or a combination of both) to training data, would improve the generalisability of a deep convolutional neural network (U-Net model) in respect of a pixel level tree classification. The secondary aim is to determine whether a single globally trained model (hereinafter referred to as single model) is sufficiently robust to outperform a targeted biome-specific model (hereinafter referred to as a multi-model).

The remainder of the paper is organized as follows: Section 2 details the experimental flow and data characteristics, reviews geometric augmentations and colour augmentations, provides details of the deep learning model used, and the approach taken to test multi-model versus a singular global model application. Sampling methodology, accuracy metrics and computing resources are also covered. Section 3 discusses the results of the models trained with colour augmentation only, geometric augmentation only and a combination of both geometric and colour augmentations. Section 3.3 highlights the results of a multi-model approach versus a single model approach.

Section 4 outlines the limitations of the experiment and provides recommendations for future work, followed by concluding remarks.

### 4.3 METHODS

This research paper is framed by a positivistic research paradigm using primarily quantitative data to determine whether data augmentation can improve the performance of a U-Net DL model to infer a global tree classification, and whether a multi-model approach or single model yields higher classification accuracies. The experiment is thus in two parts. The first focuses on identifying and analysing performance increases by applying geometric, colour and a combination of geometric and colour augmentations. The second part focuses on developing an optimal strategy to infer a global tree classification. It does this by comparing the performance of a model trained only with data derived from the specific biome with data augmentation, with the performance of a model trained with data derived from multiple biomes with data augmentation applied.

#### 4.3.1 DATA

Multi-spectral data was derived from Sentinel-2 Level-2A processed data via LuxCarta's BrightEarth Mosaic (<https://brightearth.ai/>), which is a global, eight bit, cloud free base map with four available bands (Red, Green, Blue, Near Infrared) at a spatial resolution of 10m (Swaine et al. 2020). Only three input bands were selected, the near infrared, the red and the green bands. Adding in the blue band as a fourth input was tested beforehand, but the results showed a deterioration in model performance.

Training and test images were normalised using Global Contrast Enhancement (GCN). When applied across the training set, CGN produced the highest Matthews correlation coefficient (MCC) scores for tree classification using the same DL architecture (Chapter 3). GCN applied as a normalisation method has been shown to improve DL models in natural image recognition tasks such as emotion recognition (Pitaloka et al. 2017). Labelled datasets were derived from LuxCarta's expansive archive of highly accurate data and were further manually corrected to match the corresponding Sentinel-2 images.

All major biomes (Grassland, Savanna, Temperate Forest, Tropical Rainforest and Boreal Forest) which contain trees or forests, were included in the training and test datasets to ensure globally representative data. A total of 20 images of size 3 000x3 000 pixels (900 square kilometres per image) were used as training images (five biomes with four images per biome). The images were clipped into smaller 512x512 image patches with 100pixel overlaps on all edges. The same clips

were applied to the corresponding binary raster labels. This combination of patch size and overlap resulted in a baseline dataset of 1 280 patches, which were split into an 80:20 ratio between training and test data in the DL model.

The unseen validation dataset was derived using the same method as for creating the training dataset. Each validation region covered an area of 2 000x2 000 pixels (400 square kilometres). The number of validation regions was increased to six test regions per biome to include more diverse representations of each biome, to test the robustness of the trained models. Thus, in total, 30 validation areas were produced to test the models. Validation images were chosen with global representation in mind. A minimum of one image per biome was chosen for each continent where the biome occurs, and at minimum, validation images were more than 500km apart geographically from any training images.

### **4.3.2 EXPERIMENTAL FLOW**

The experiment design in Figure 15 details the steps for which each augmentation strategy was applied. To test which type of data augmentation most improves the performance of a deep learning model, a baseline model was trained. The baseline dataset comprised all training patches from the different biomes based on the training dataset from Chapter 3, with GCN applied as an enhancement. For the baseline model, a DL model was trained over 100 epochs on the dataset with no data augmentation applied. This baseline model was used to predict trees in each of the 30 validation images (Chapter 3). Four scenarios were developed for comparing the effects of the various augmentation strategies with the baseline model, namely: geometric only transformation, colour-only transformation, geometric-with-shear, and geometric and colour combined augmentations. The training process was repeated for each of the four data augmentation scenarios. Each model was compared with the baseline results. Consistency in the base training/test patches was maintained to ensure that a direct comparison between trained models could be made. A global MCC score was calculated for all images as well as for a local MCC per validation image.

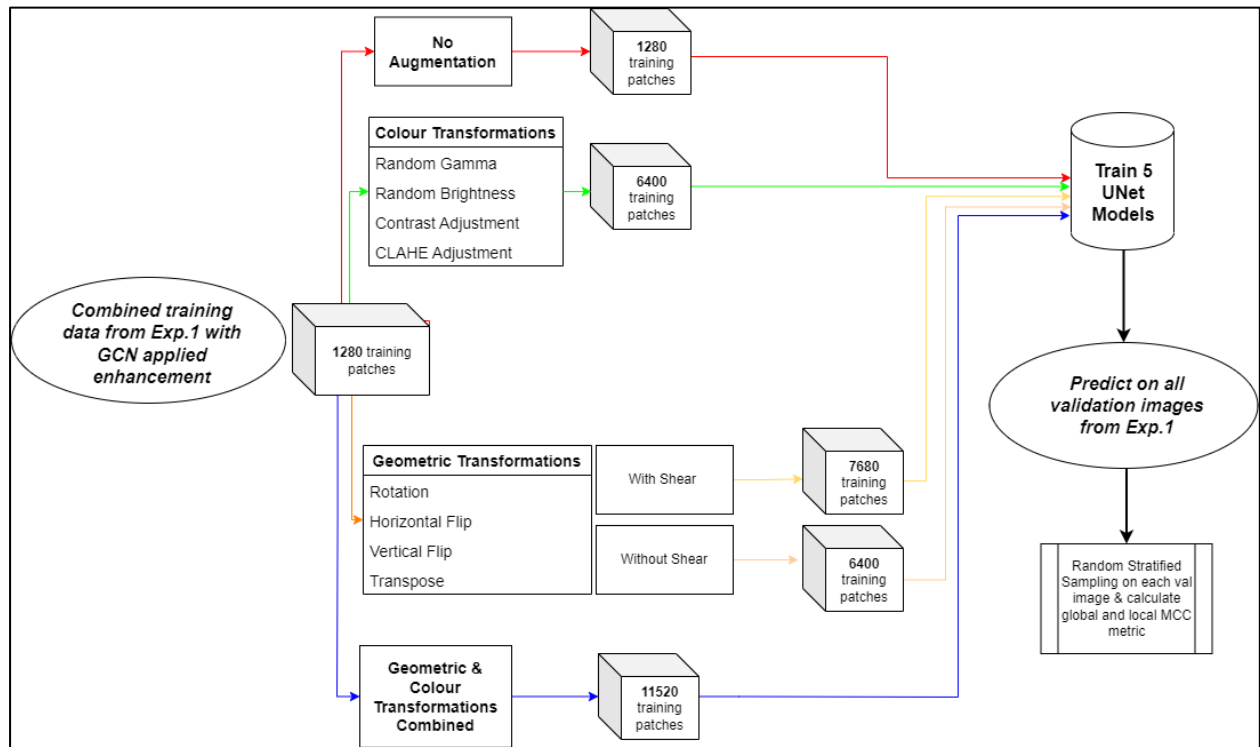


Figure 15 Experiment flow, part one

### 4.3.3 DATA AUGMENTATION

#### 4.3.3.1 SCENARIO 1 – GEOMETRIC TRANSFORMATIONS ONLY

For scenario one, geometric transformations including vertical flip, horizontal flip, rotation and transpose, were applied to each training and validation image patch. The resulting number of training patches for the DL model was 6 400 patches. A vertical flip flips an input image vertically around the x-axis (Figure 16 b) (Khalifa, Loey & Mirjalili 2022). Similarly, a horizontal flip, flips an input image horizontally around the y-axis (Figure 16 c). A transpose performs a swap of the X and Y axes (Figure 16 d). A rotation can be performed by a specified degree (0-360) from the centre of the image. The angle of rotation was randomly assigned using only multiples of 90 degrees to retain the image aspect ratios (Figure 16 e, f).



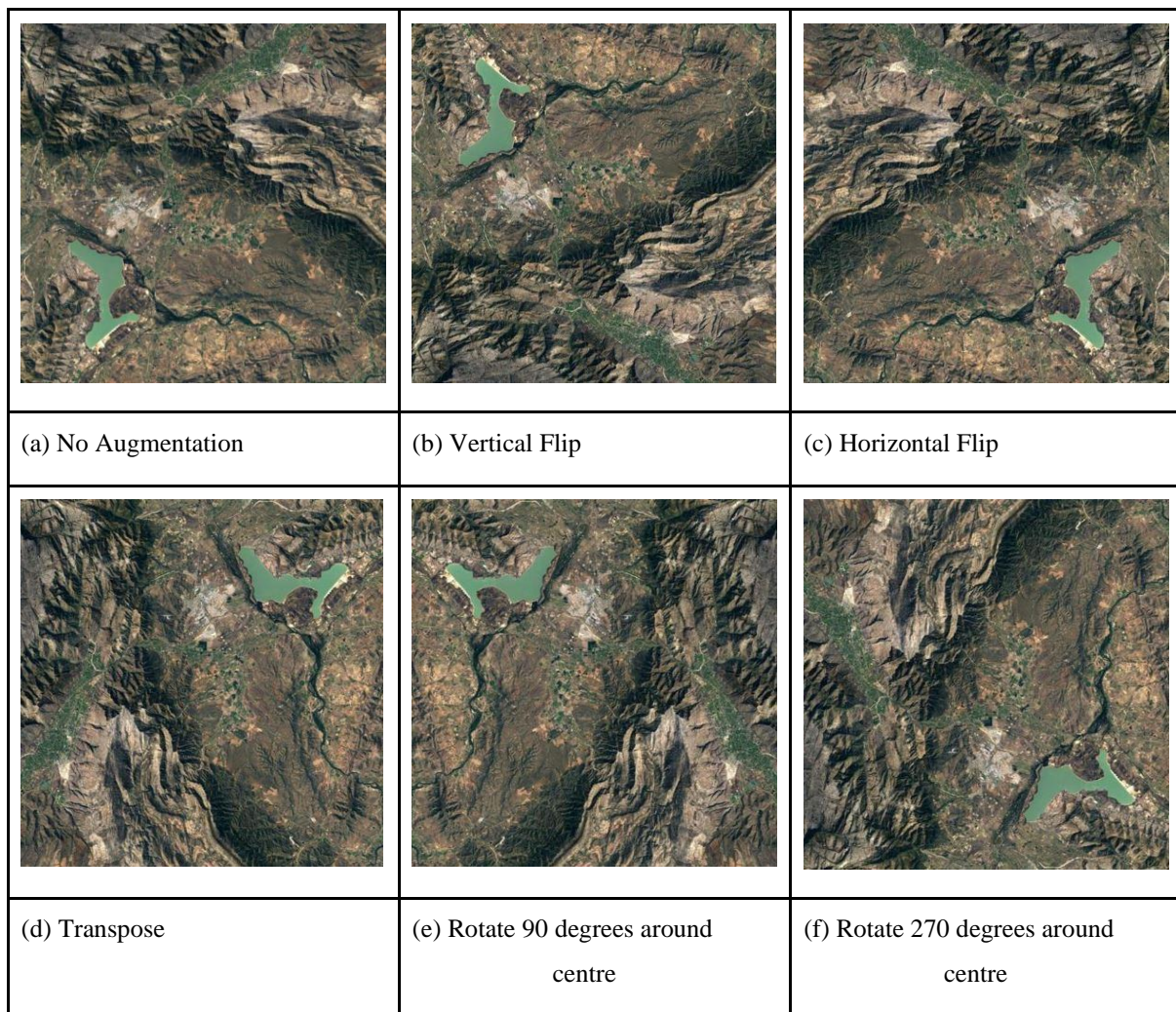


Figure 16 Geometric transformations

#### 4.3.3.2 SCENARIO 2 – COLOUR TRANSFORMATION ONLY

Colour transformations include changes in brightness, contrast, gamma and contrast limited adaptive histogram equalization applied in the hue, saturation, value colour space (CLAHE) and were applied to the training patches in scenario two. Random changes in image brightness were applied by multiplying a random gain/loss of between -0.2 and +0.2 to each image patch (Figure 17 b, c). A random contrast with an adjustment limit of 0.2 was applied. (Figure 17 g illustrates a random contrast adjustment of 0.11). Random gamma changes ranging from 80 (Figure 17 e) and 120 (Figure 17 h) were applied to image patches (100 being no change). CLAHE (Figure 17 f, i) with a random clip limit of between 0 and 2.5 and a grid size of 8x8 was applied to each image patch. The resulting number of training patches for colour augmentations was 6 400, to serve as input for the DL model. Manual testing was conducted to determine the range limits for parameters based on random values. Results outside the selected range produced unrealistic images unlikely to occur in reality.



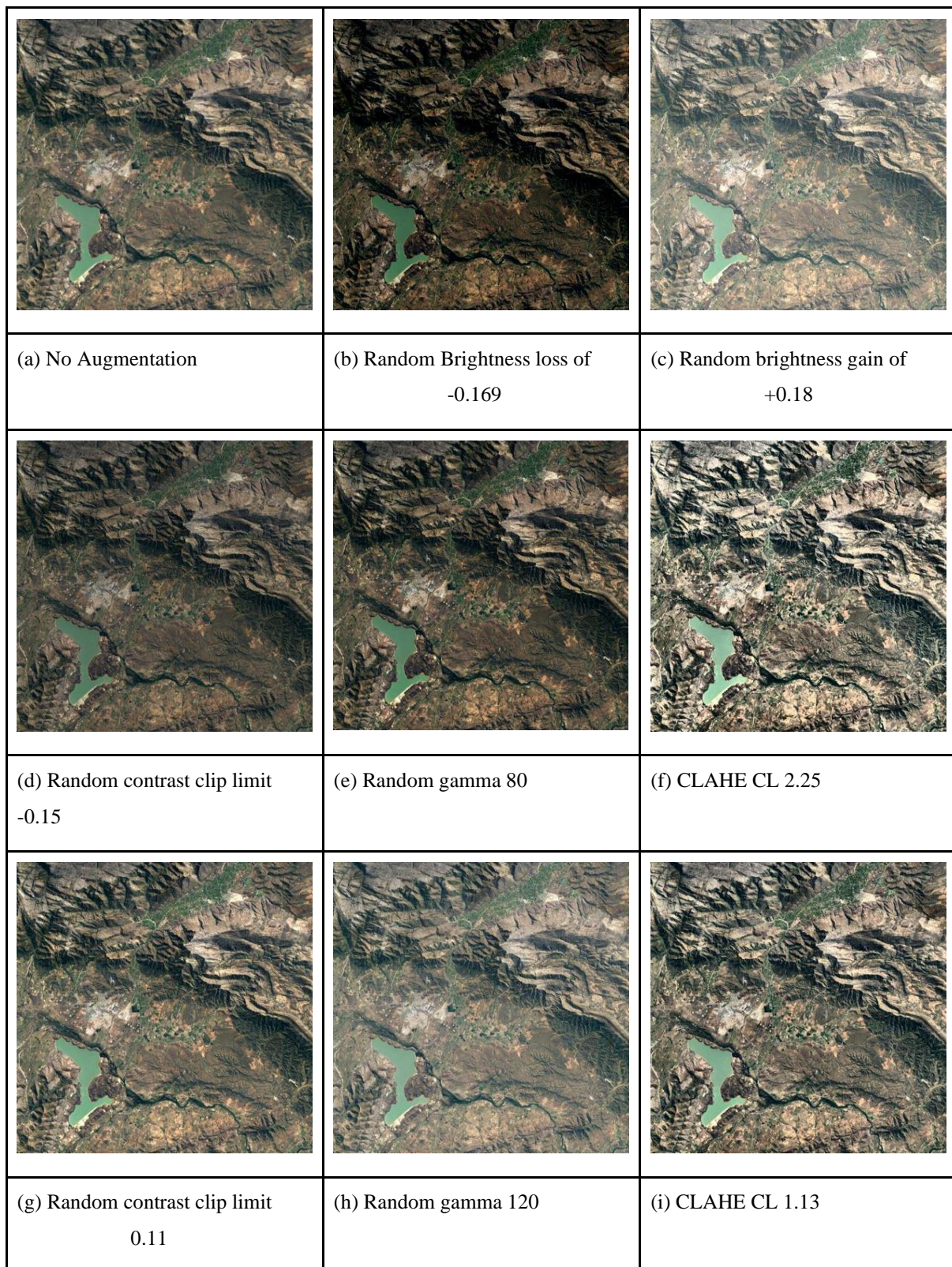


Figure 17 Colour transformations

#### 4.3.3.3 SCENARIO 3 – SHEAR TRANSFORMATION

Scenario three involved adding random shear as an additional transformation to the geometric transformations for each image patch. As opposed to naive augmentation, such as random flips, rotations and mirroring which change the orientations of features in an image, shear warps the



shapes of features. A random value of between -30 and +30 was applied in both the X and Y planes to perform the shear transformations of each training patch (Figure 18). This resulted in a total of 7 680 image patches used to train the model.

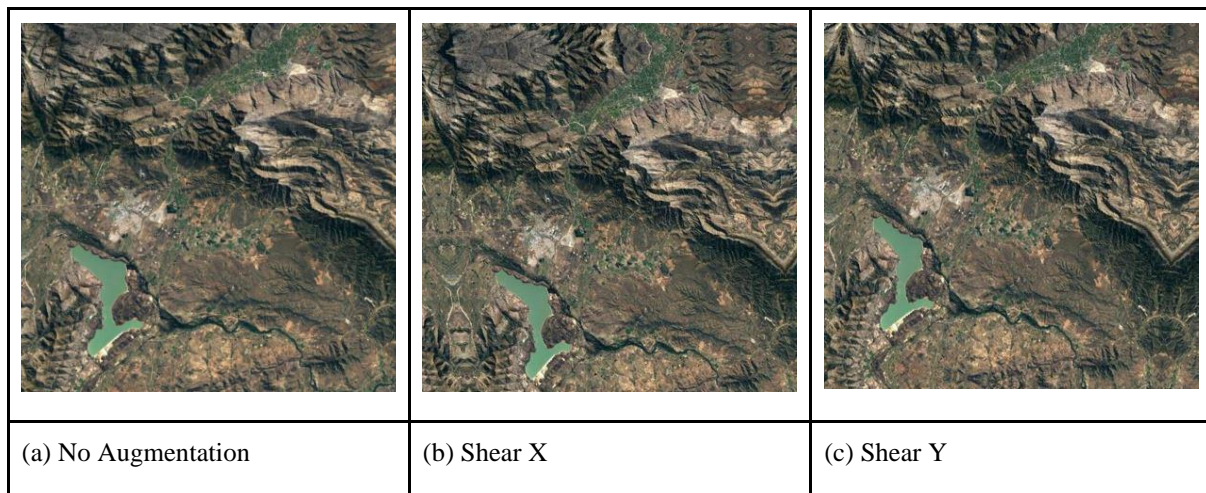


Figure 18 Shear transformation

#### 4.3.3.4 SCENARIO 4 – COMBINATION OF COLOUR AND GEOMETRIC TRANSFORMATIONS

A combination of both colour and geometric transformations was used to train a DL model to determine whether adding combined augmentation types would improve performance. Combining all augmentation types effectively doubled the number of augmented training patches (11 520 patches). Additionally, the scenario tested whether the increase in the number of training patches would result in a more robust model compared with applying one method of augmentation.

#### 4.3.4 DEEP LEARNING MODEL

A vanilla U-Net architecture was preferred because it uses skip connections between the convolutional layers in the encoder and the corresponding deconvolutional layers in the decoder (Ronneberger, Fischer & Brox 2015). This enables the network to up-sample results at the same spatial resolution as the input data. The naming of the architecture is derived from the U-shaped symmetry of the architecture with a convolutional encoder and a corresponding deconvolution decoder (Figure 4).

Batch normalisation was applied before each pooling layer to stabilize the learning process and thus reduce the number of epochs required to train a model effectively (Ioffe & Szegedy 2015). Dropout was applied at the end of the fourth and fifth convolutional blocks to reduce the chances of overfitting the model and to improve model generalisation. A probability of 0.5 was set for dropout to occur for each batch passed through the model. The Rectified Linear Unit (ReLU) was used as an activation function for each convolutional layer in the encoder and the decoder. This

activation function enabled the model to learn more complex features such as non-linearity to improve model performance (Agarap 2018). The final output layer implemented a sigmoid function with output probability values between zero and one. Binary cross-entropy was employed as the loss function. Binary cross-entropy measures the difference between the output prediction and the label and is very effective for binary segmentation tasks and pixel level classification (Jadon 2020).

### 4.3.5 MULTI-MODEL VS. SINGULAR GLOBALLY TRAINED MODEL

A significant challenge in producing a global tree dataset relates to the vast variations in the world's atmospheric and geophysical environments. Two approaches were followed to test which would be more suitable for achieving an accurate, global forest classification. Whereas a multi-model approach requires training a separate model for each biome with training data derived from the respective biome, a global, singular model is trained on all training data from all biomes (Figure 19). The Savanna and Temperate Forest biomes were chosen for the multi-model approach, respectively having the lowest (0.56) and the highest (0.82) MCC scores in Experiment 1 (Chapter 3) where GCN was applied as an enhancement. This experiment measured the extent to which data augmentation could increase the performance of a model trained with data from a biome (Savanna) in which trees are less distinguishable in an image, compared with a model trained with data from the Temperate Forest biome in which trees are more pronounced in an image. This approach was then compared with the alternative, which was to train one model, with training data compiled from a globally representative dataset covering all biomes.

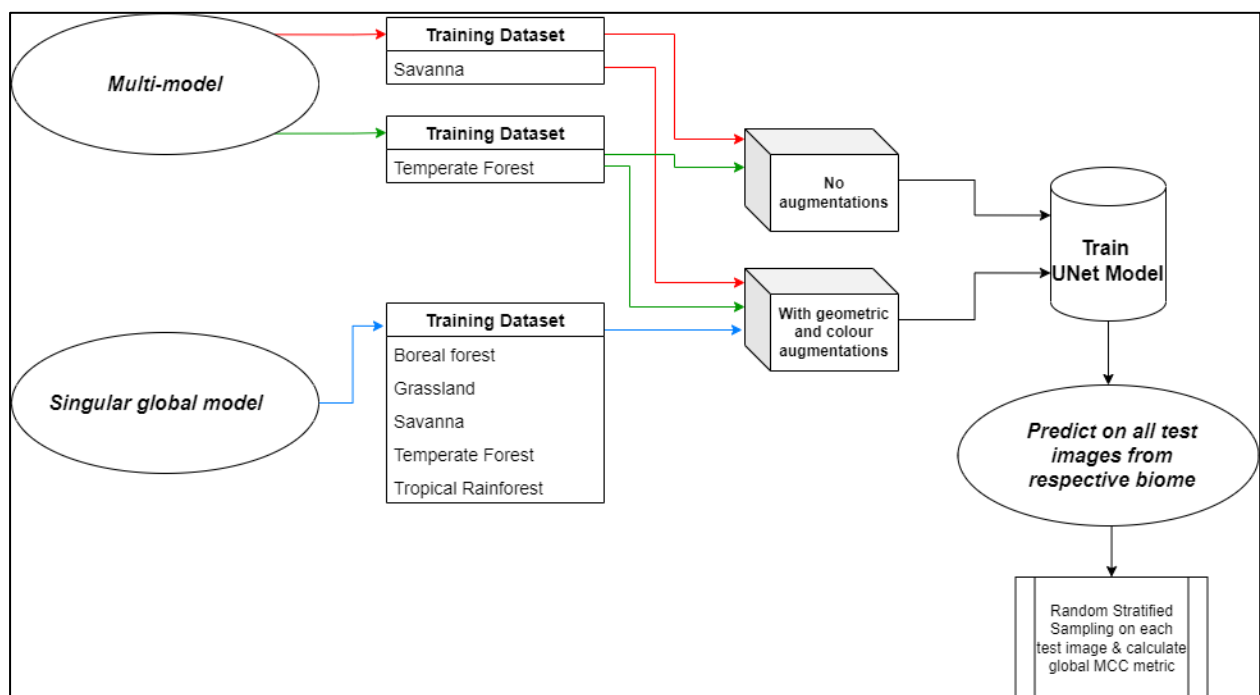


Figure 19 Experiment flow, inference strategy

For the baseline results, no data augmentation was applied to the training data (Chapter 3). Each training dataset contained 256 image/label training pairs and were trained over 100 epochs. Following the multi-model approach, models for both Savanna and Temperate Forest biomes were trained with geometric and colour augmentations applied to the training dataset, thereby inflating the training set of each biome to 2 304 image/label training pairs. The single model was trained by combining all training data from all biomes and applying geometric and colour augmentations to enhance and diversify the dataset, thus inflating the dataset to 11 520 image/label pairs. The biome-specific models were used to predict trees from test images derived only from the respective biomes. The singular global model was used to detect forests in both Savanna and Temperate Forests test data for comparison. A random stratified sampling strategy (Section 2.6) was employed to produce a confusion matrix to infer the MCC score for direct model performance comparisons.

#### **4.3.6 ACCURACY AND STATISTICS**

A random stratified sampling methodology was followed to generate sample points. With validation image dimensions of 2 000x2 000 pixels (total of 4 000 000 pixels), 1% of the pixels of each validation image were sampled. Therefore 40 000 points were selected for each image. The number of positive test samples were based on the proportion of pixels labelled as tree in the corresponding ground truth dataset. For example, if 10% of pixels of the ground truth dataset were trees, then 4 000 of the 40 000 samples extracted were trees. A confusion matrix of the samples was created for each predicted image. To analyse the confusion matrix statistically, researchers often utilise the F1 score and overall accuracy (OA) to infer results, depending on the goal of the research. However, these metrics can overstate results in unbalanced datasets (Chicco & Jurman, 2020). To ensure a more accurate representation of the results, the Matthews Correlation Coefficient (MCC) was chosen as the comparison metric. MCC is a well-suited performance metric in analysing results from binary classifications (Chicco & Jurman, 2020), particularly for unbalanced datasets. MCC considers True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) results and will generate a high score only if the majority of both positive and negative test cases were correctly predicted. MCC ranges from -1 to 1, with -1 indicating complete misclassification, 0.0 no better than random, and +1.0 a perfect classification.

MCC is defined in Equation 7 as:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \quad \text{Equation 7}$$

Where	TP	True positive: correct prediction of the positive class;
	TN	True negative: correct prediction of the negative class;
	FP	False positive: incorrect prediction of the positive class; and
	FN	False negative: incorrect prediction of the negative class.

Precision and Recall are two metrics that can also be used to measure the effectiveness of model performance. Precision is the fraction of TP cases over the sum of all positive predictions and is often referred to as confidence (Powers, 2007). It measures the probability that a positive test case is truly positive. Recall, also known as the true positive rate or sensitivity of a model, measures the proportion of positive cases that have been correctly classified as positive. Precision and Recall should always be analysed inclusively of each other as a high Recall may be achieved even with a low precision, and vice versa. The F1 score attempts to combine those scores to produce a single measure to represent how well a model has performed. The F1 score is the harmonic mean of Precision and Recall (Powers, 2007).

The F1 score can be misleading in that the true negative cases are not considered. This can be seen from the F1 score equation (Equation 8):

$$F1\ Score = \frac{2TP}{2TP + FN + FP} \quad \text{Equation 8}$$

Where	TP	True positive: correct prediction of the positive class;
	FP	False positive: incorrect prediction of the positive class; and
	FN	False negative: incorrect prediction of the negative class.

The F1 score can give false truth values when the target class is in the minority and is used as the TP case. Therefore, when the target class is the minority class, it is preferable to switch the positive and negative labels so that the target class presents the TN case in the error matrix (Chicco & Jurman, 2020). Thus, the F1 score is useful when class labels are balanced.

The advantage of using MCC over the F1 score in this study, was that the test datasets were unbalanced. In the case of tree classification, the tree class was likely to be the majority class in an image of tropical rainforests. However, it was almost certain to be the minority class in an image

of the Savanna biome. Hence, it was not a certainty that the tree class would be the minority class in all test cases. Taking these factors into account, inverting the labels could have resulted in the F1 score giving truthful results. MCC was thus the chosen metric for this study as it gave a better representation of the result regardless of the data being unbalanced.

#### **4.3.7 COMPUTING RESOURCES AND SOFTWARE LIBRARIES**

The system used for processing consisted of an Intel i9 9900K, 16 core CPU, and an Nvidia GTX1080Ti GPU with 11Gb of memory and system memory of 64Gb RAM. Data processing was run on a Central Processing Unit (CPU) while model training was run on a Graphical Processing Unit (GPU). OpenCV, Numpy and GDAL Python libraries were used for data pre-processing, Keras and Tensorflow DL libraries for training and inference, and Albumentations for data augmentation. Numpy, SciPy, Matplotlib and Pandas libraries were used for test sampling and statistical analyses.

The Albumentations (Buslaev et al., 2019) (<https://albumentations.ai/>) Python library was used to apply all colour, geometric and affine transformations to the training patches. The advantage of Albumentation is that the transformation is applied to the image patch while preserving the label by applying the same transformation to the corresponding label patch. This was applicable only in the case of geometric and affine transformations, as the label remained unchanged for colour augmentations.

### **4.4 RESULTS & DISCUSSION**

#### **4.4.1 EFFECTS OF AUGMENTATION**

Results from the model trained with no augmentation were the least robust with an MCC score of 0.668 (Figure 20). The MCC from this base model served as the baseline data against which all augmented models were compared. The low MCC score was an expected result as the model was trained on only 1 280 training patches, the least number of training image/label pairs of all the scenarios.

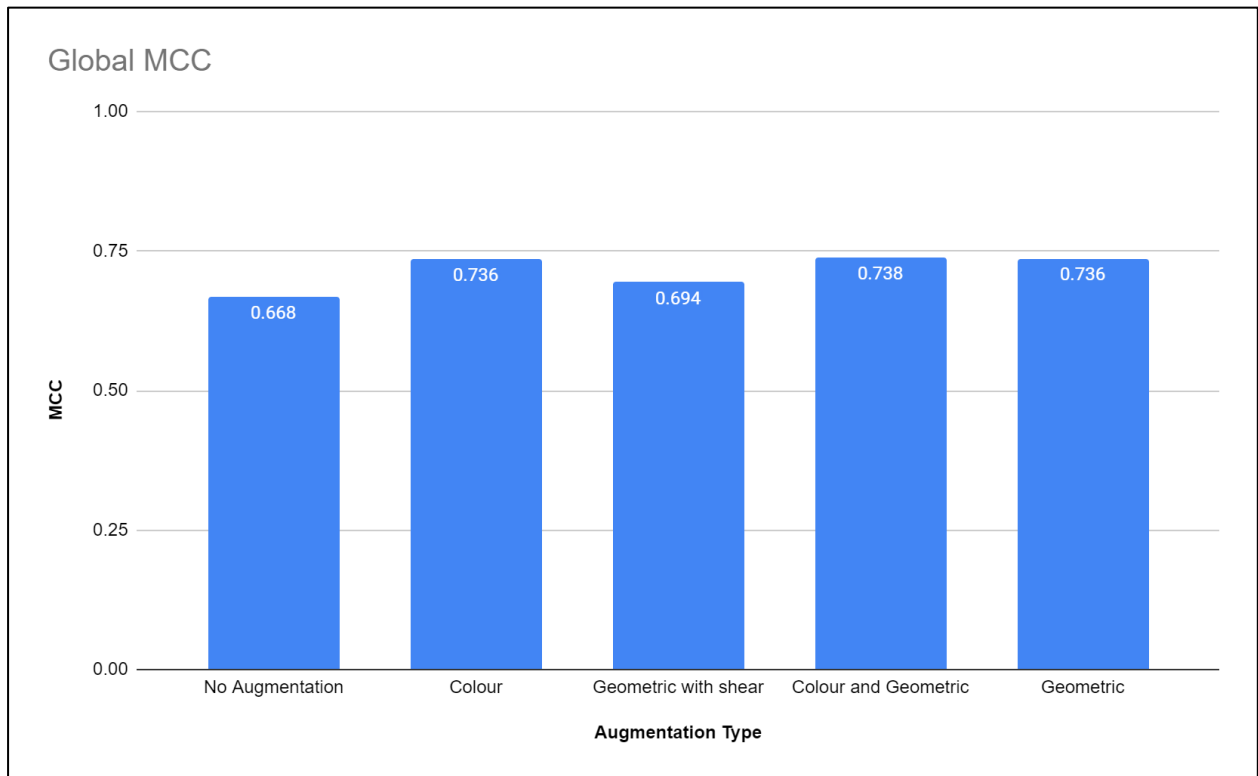


Figure 20 Global MCC by augmentation method

The individual application of geometric and colour augmentations increased the MCC to 0.736 (Figure 20), which equates to a 7% increase on the baseline model MCC. The augmentation scenario that produced the highest MCC score of 0.738, was the combination of geometric and colour augmentation. Although this represents a significant increase in performance compared with the baseline model, it is only slightly better than only geometric or only colour augmentation. However, the combined geometric and colour augmentation comes at a much greater computational cost in terms of training time, in relation to the gain in MCC. The latter model was trained on 11 520 patches whereas the colour and geometric models individually were trained on only 6 400 image patches. This result aligns with the findings of Lei et al. (2019) that the addition of multiple augmentation methods did not equate to a linear increase in model accuracy.

The model trained with geometric augmentation with shear recorded an MCC of 0.694. It performed only marginally better than the baseline with no augmentation. For geometric transformation the addition of shear as an augmentation resulted in a reduced MCC score, compared with a model trained with geometric transformations without shear (Figure 20).

The MCC was calculated for each of the 30 test images to give insight into the effects of different augmentation strategies on a per-tile basis (Figure 21).

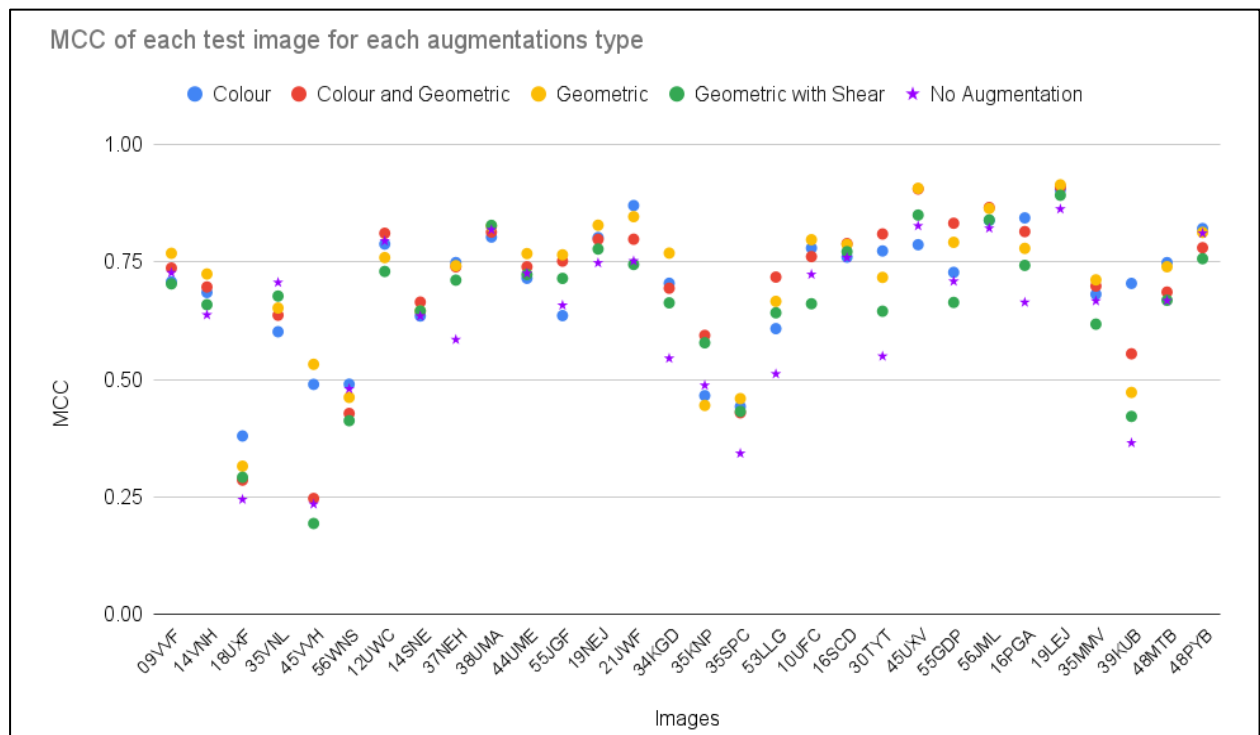


Figure 21 MCC of each test image per augmentation method

The worst performing scenario was that of no augmentation applied during model training (15 images). Nine images showed that geometric with shear applied was the worst performing augmentation scenario. Colour-only augmentation performed the worst in five images while geometric-only augmentation was the worst in only one image. When considering the individual results on a per image basis, the combination of geometric and colour augmentations was the only scenario that consistently scored higher than the other augmentation types in the 30 test images.

Figure 22 visually depicts the difference between a model trained with colour augmentation (c) compared with a RGB image (a), the ground truth (b) and the result from the baseline model with no augmentation applied (d). The associated Precision and Recall statistics are provided below the model output.



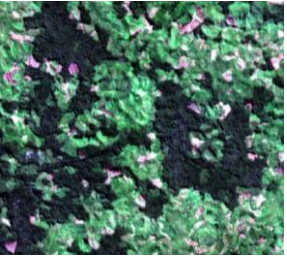
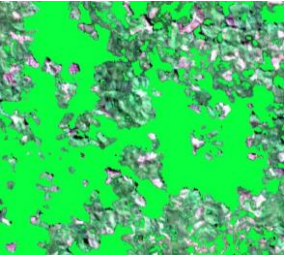
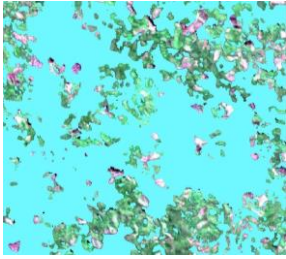
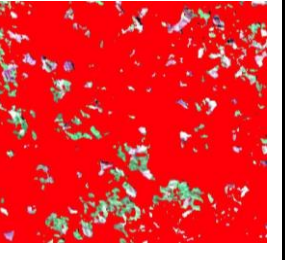

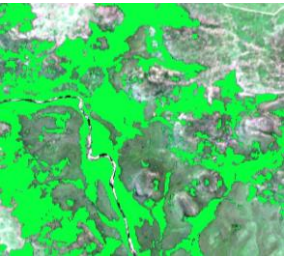
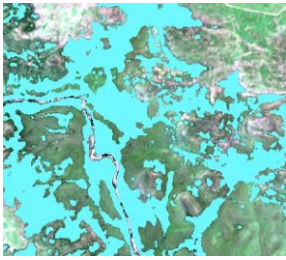
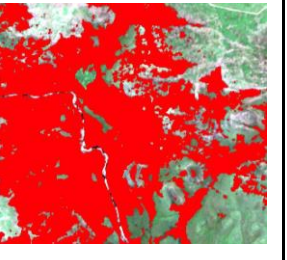
(a) Image: 39KUB	(b) Ground truth	(c) Colour	(d) No Augmentations
			
MCC		0.70	0.37
Precision		0.72	0.47
Recall		0.96	0.99

Figure 22 Visual effect of model applied with colour augmentations (c) vs. no augmentations (d)

It is very clear that the baseline model (Figure 22 d) was not sufficiently robust to learn clear features for trees, as the model resulted in excessive false positive cases. This was confirmed by a precision of only 0.47 and an MCC score of 0.37 for the specific image. The colour augmented model (Figure 22 c) achieved a higher MCC of 0.70, which was the highest recorded for this specific image when comparing it with other augmentation strategies.

Figure 23 shows the difference in performance between training on a colour augmented model (c), compared with a model involving geometric augmentation with shear transformation being applied to the datasheet (d). Geometric-with-shear was the worst performing strategy, particularly for the image in Figure 23 d. As was the case for colour augmentations against no augmentation (Figure 22), colour augmentations (Figure 23 c) produced both a high Precision and Recall, as well as the high MCC Figure 23 c.

(a) Image: 21JWF	(b) Ground truth	(c) Colour	(d) Geometric with shear
			
MCC		0.87	0.74
Precision		0.89	0.72



Recall		0.94	0.99
--------	--	------	------

Figure 23 Visualising the difference in model trained with colour augmentations (c) vs. geometric-with-shear augmentations (d)

A comparison of test images between a model trained with basic geometric augmentations (Figure 24 c) and a model trained with the same augmentations with shear added (Figure 24 d), showed a tendency by the latter model to give false negative predictions. This is evident by the lack of trees classified in the north-east quadrant of the image.


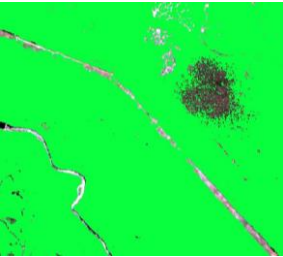
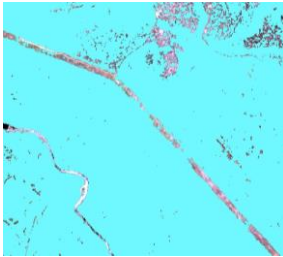
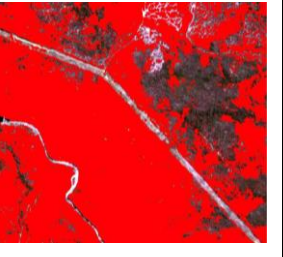
(a) Image: 55GDP	(b) Ground truth	(c) Geometric	(d) Geometric-with-shear
			
MCC		0.79	0.66
Precision		0.95	0.94
Recall		0.95	0.89

Figure 24 Visualising the difference between model trained with basic geometric augmentations (c) and shear with geometric augmentations (d)

The false negative tendency was substantiated by the higher Precision over Recall values recorded. It appears that applying shear to the training data pool may have skewed the model into learning semantics that are not realistically found in satellite data.

The MCC produced for each augmented scenario was reported relative to the MCC of the baseline MCC, thus normalising the scale of the results. The difference (normalisation) between the MCC scores of the predictions of each of the augmented models and the baseline predictions were calculated for each of the 30 test images (Figure 25).

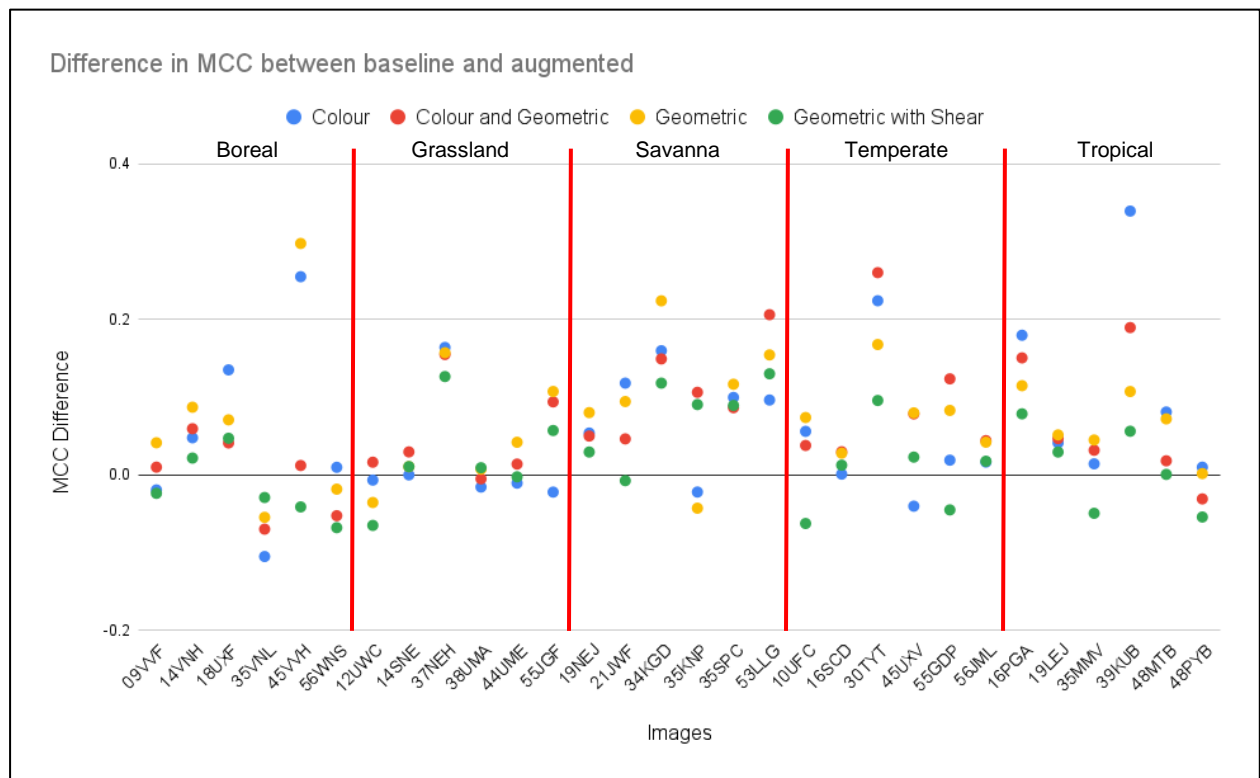


Figure 25 Normalised difference in MCC between no augmentation and applied augmentations

For most augmentation scenarios, predictions performed better with augmentation than for the baseline. All data points above zero (black line) indicate a positive increase in MCC against the baseline, while all data points below the zero line, performed worse than the baseline. In only one image (35VNL), all augmentation scenarios performed worse than the baseline. This could be an indication that the ground truth data for that particular test image may not have been up to standard. Of all the augmentation scenarios, the geometric augmentation means line (yellow) showed the highest mean difference from the baseline. Adding shear to geometric augmentation resulted in a large drop in performance and was only marginally better than no augmentation.

This indicates that shear may not be a suitable augmentation method for tree classification at a 10m resolution. However, shear as a data augmentation method, has successfully been used in many other DL domains, including other binary categorisation problems such as skin lesion analyses for melanoma classification (Perez et al, 2018). Applying shear as a data augmentation to one of the most widely used datasets for machine learning research, the CIFAR-10 dataset (Canadian Institute for Advanced Research), outperformed some other commonly used augmentations, such as rotations and colour augmentations (Lei et al., 2019). In contrast, however, geometric augmentations outperformed models applied with colour augmentations when trained on the Caltech101 dataset (Li et al. 2022) for image label classification (Taylor & Nitschke 2017). Any decision to use shear as an augmentation method should thus be carefully considered, and take into account the domain application. Notably, using shear in a medical application for

diagnosing Covid-19 from lung X-rays was possibly detrimental, in that the transformed images would probably be images that cannot clinically exist (Elgendi et al. 2021). Similarly, the impact shear may have on Sentinel-2 images is the distortion of the original images to such a degree that the DL model cannot effectively learn the semantic features of trees.

Of the 30 test images (Figure 25), the geometric only and the combined colour and geometric augmentation scenarios each returned 26 images that outperformed the baseline scenario, while the colour only and geometric with shear augmentations respectively returned 22 and 19 test images with positive MCC differences from the baseline.

#### 4.4.2 COMPUTATIONAL COST

When training deep learning models, it is useful to know the threshold at which additional training data will not increase performance. This reduces training time, which increases linearly with the increase in the number of training patches (Lei et al, 2019). Figure 26 graphically shows training time vs. performance as measured by MCC scoring.

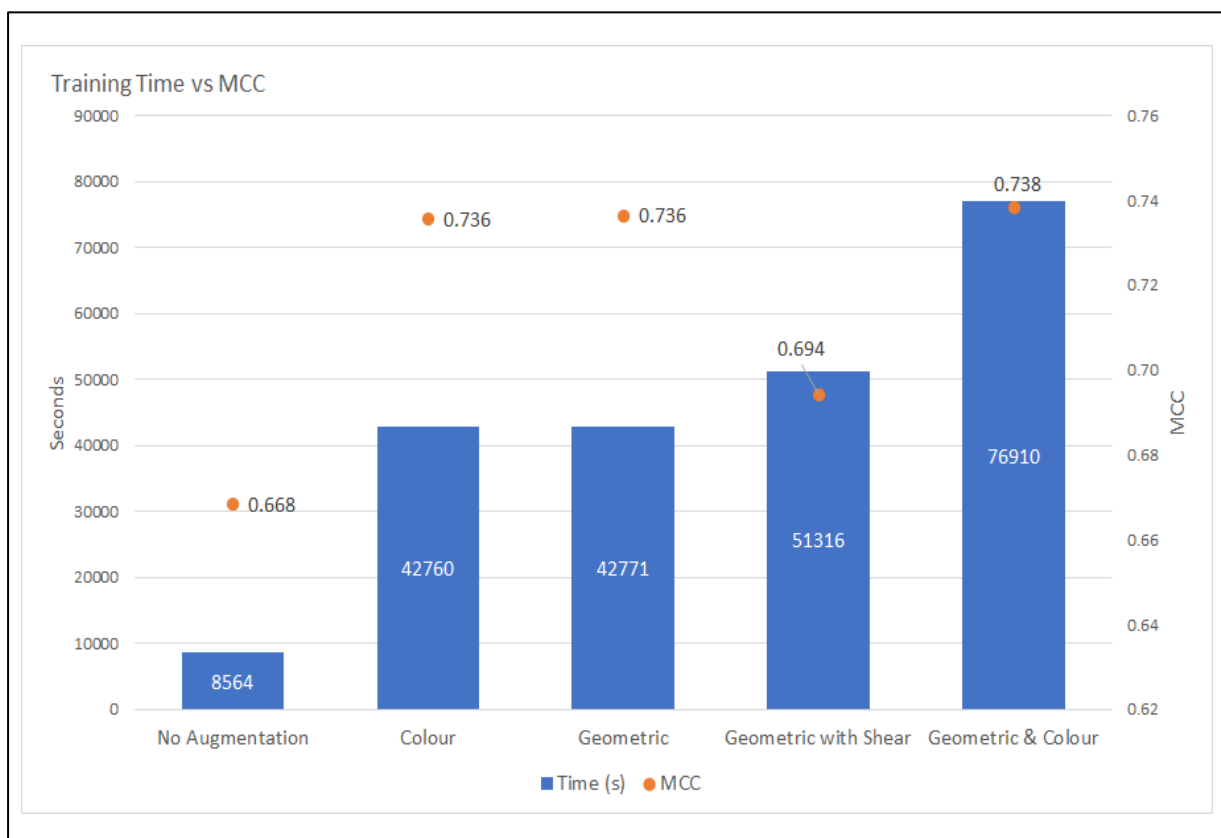


Figure 26 Computation cost vs MCC

The baseline scenario, which had no augmentation, had the smallest number of training patches and trained the quickest, but had the lowest MCC score. The geometric only and colour only scenarios had augmentation rates five times that of the baseline and with 6 400 training/label pairs both augmentation methods improved the MCC's of the respective models by 10.2%. Training

time increased by about 400% from 8 564 seconds to 42 760 and 42 771 seconds for the colour and geometric augmentation models respectively. Combining the geometric and the colour augmentations and thereby effectively doubling the training dataset size and the training time, resulted in an increase in MCC of only 0.002 (+0.3%). Other additional training data did not guarantee any increase in model performance. It was shown that while the addition of shear as an added geometric augmentation increased the number of training patches, the outcome was an MCC of only 0.694 (+3.9%). This was in line with experimental results from a range of state-of-the-art deep learning models trained on CIFAR-10 and Fashion-MNIST (Xiao, Rasul & Vollgraf 2017) datasets that showed that higher augmentation rates did not always translate into higher model accuracies after two to three times augmentation rates (Lei et al., 2019). Furthermore, data augmentation had the greatest positive effect in cases when the number of training samples was limited (Lei et al., 2019).

#### 4.4.3 MULTI-MODEL VS. SINGULAR GLOBAL MODEL

Part two of this experiment compared two approaches for inferring a global tree classification: a model trained per biome approach and a single, global model approach. Comparisons were made with the baseline models which had no augmentation applied. Baseline MCC scores of 0.56 and 0.82 were produced for Savanna and Temperate Forests respectively (Figure 27).

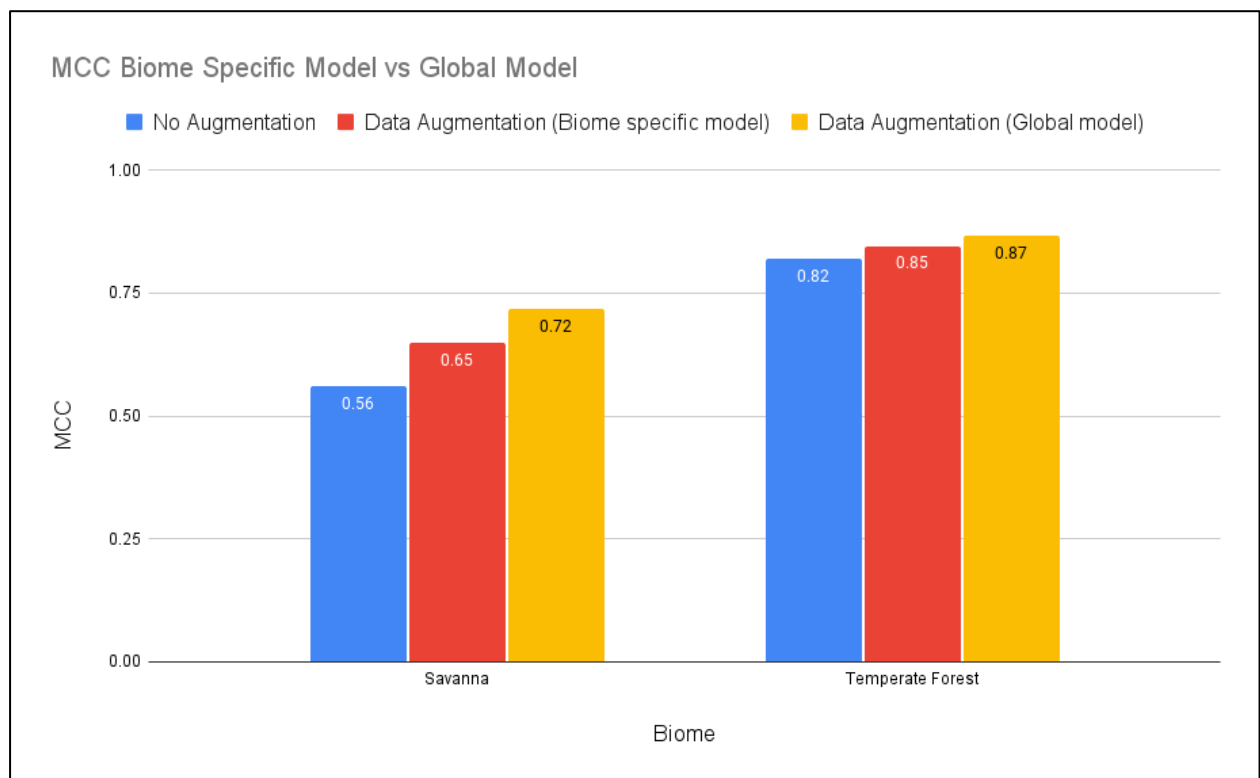


Figure 27 MCC of biome-specific model approach vs globally-trained model approach

Applying data augmentation to the biome-specific models improved results for both the biomes. However, the increase in performance was less for the Temperate Forest biome (3.7%) than for the Savanna biome (16%). This was expected as the potential for improvement from an MCC score of 0.82 is lower than the potential for improvement on a model with a baseline MCC of only 0.56. In both biomes, the singular, global model outperformed the biome-specific models by 10.8% in the Savanna biome and by 2.4% in the Temperate Forest biome. The diversity of training patches with augmentations applied in the global model seemed to offer more robustness in recognising trees, even when the biome-specific models could not perform as accurately. Figure 28 shows an example of a typical image of the Savanna biome in which the global model outperformed the biome-specific model. Precision and Recall values confirmed the visual depiction of a higher number of false positive cases occurring in the Savanna-specific model, whereas the global model performed better in all metrics.


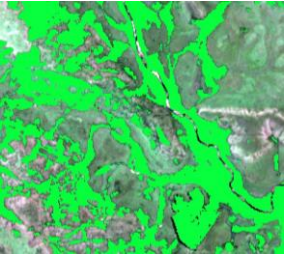
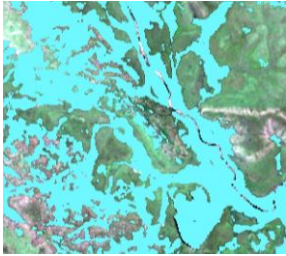
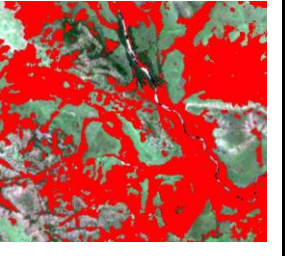
Image: 21JWF	Ground truth	Global model	Savanna only model
			
MCC		0.80	0.64
Precision		0.78	0.66
Recall		0.98	0.94

Figure 28 Visual prediction of globally trained model on Savanna biome vs Savanna-specific model prediction

It was expected that the biome-specific model would outperform the globally trained model as the intraclass variability within each biome was expected to be lower. It was assumed the biome-specific model would learn the finer features of trees specific to each biome more effectively. From an operational viewpoint, a single, globally representative model is more desirable for the production of a global tree dataset as it may alleviate the impact of artefacts near the ecotone zones from one biome to another. The additional variability introduced in the training data of a globally representative model is more robust to variations in features within specific biomes, whereas a model trained with biome-specific data did not perform well in areas where geographical, ecological and contextual information diverge from the training dataset.

It is well known that data augmentation for DL generally improves model robustness. Often such data augmentations are applied with little regard of the effects each augmentation type may have on the trained model. Whereas a few studies in the medical imaging domain have analysed the effects augmentation types may have on the performance of a DL model (Chlap et al. 2021; Elgendi et al. 2021), these studies have not yet been expanded to remote sensing for tree classification. For some tasks it is obvious which augmentations should be avoided, for example, it would be detrimental to apply image rotations for numeric character recognition tasks. In the case of tree classification from Sentinel 2 data, applying basic geometric augmentations, which change image orientations, showed significant improvement in model performance (Section 3.1). Adding shear transformations into the pool of geometric augmentations reduced the performance of the DL model (Figure 20). Applying basic geometric and colour augmentation methods to a limited training dataset, to enhance the size and complexity of the data, may increase model robustness and be an effective way of mitigating the risk of overfitting. The results suggest that it is not always necessary to apply both geometric and colour augmentations as the marginal gain in MCC may not be worth the increased computational cost of training using double the amount of data. However, if model training time were not a limiting factor, the combination of both geometric and colour augmentation would lead to an improvement in MCC by 0.2% and would be a recommended step as the computational cost of inference is independent of the size of the training data.

## 4.5 CONCLUSION

Achieving a global tree classification at 10m spatial resolution is a challenging task. Atmospheric and landscape conditions vary greatly globally, and thus training a one-size-fits-all model would be a big challenge. To achieve a global tree classification, two possible approaches were tested. The first was a multi model approach, whereby a model was trained with data specific to each biome. The premise was that this would lead to a refinement of learned biome-specific features. The second approach relied on a single, globally trained model for which all training data from all biomes was combined. When assessing the performance of the DL models trained per biome, performance as measured through the MCC score was reduced, whereas training a model on the global dataset resulted in more accurate outputs. Thus, a globally trained model, essentially has been given representation of more intra class variation and diversity and performed better across the validation images.

Inherent data biases exist in the training and test datasets, stemming from the way in which the datasets were produced. As these biases carry over into the trained models, their impact needs to be considered when interpreting results. Improvements to the input dataset could be made by using

Sentinel 2 L2A image data in its original bit depth of 12bit instead of the reduced bit depth of 8bit as was used in this study. Although best attempts were made to acquire imagery based on similar seasonality, this aspect could not be guaranteed. There is scope for further experiments using oversampling augmentation methods, specifically artificially inflating, and adding diversity into training datasets using GAN's. Furthermore, tests could be done to analyse the effects of data augmentation when applied during the prediction phase, by running multiple predictions per patch, and merging output probability. This may lead to improved overall model performance as features associated with specific orientations could be activated using slightly different visual input. Further pursuing a model per biome approach would entail training a global model with all combined data, and thereafter fine-tuning the model for each biome, thus learning both global features and refined features from the fine-tuned data for each model. These optimizations could help achieve a very accurate, global, 10m resolution, tree classification.



## 5 CHAPTER 5: DISCUSSION & CONCLUSION

This chapter summarises the primary findings of the research. Section 1 recaps the aims and objectives. Section 2 examines the main findings and the value of the research. Section 3 highlights the limitations of this study, and the potential and scope for further relevant research. The final section draws conclusions from this study.

### 5.1 REVISITING THE AIMS AND OBJECTIVES

The aims of the research were to investigate which image normalisation and enhancement algorithms, or combinations thereof, applied to the input data of a convolutional neural network could improve the accuracy of a land cover classification, as well as to evaluate whether data augmentation could be used to generalise a deep learning model and thereby infer a global forest classification. Accurate mapping of global tree resources is vital for effectively managing ecosystem services. Timber production, biodiversity protection and climate change mitigation through carbon sequestration, are some of the main functions, ecosystems provide. This research was thus conducted in the context of developing an operationally efficient and accurate methodology for intercontinental tree classification in support of global forest monitoring.

The literature review (Objective 1) provided a plethora of research extending across existing global land cover, Sentinel-2 data, image classification, ML, DL, image enhancement and image augmentation. Existing research shows that image enhancements can improve DL model accuracies, but that these have not been applied in the remote sensing domain. Data augmentation has been commonly used in studies for which the size of the training dataset is limited, and models are prone to overfitting. Despite the abundance of research papers focusing on improving model architectures to increase model performance, there has been less focus on investigating the suitability and effects of various enhancement and augmentation methods.

The collection of training and validation data (Objective 2) was derived from LuxCarta's BrightEarth (<https://brightearth.ai/>) product; a global, cloud-free mosaic at a bit depth of 8bits. Training labels and ground truth validation data were derived from LuxCarta's expansive archive of accurate, global land cover data, and further refined via desktop edits to match the training and validation data with the imagery.

Chapter 3 dealt with determining the suitability of common image enhancement and normalisation techniques for improving a DL model for tree classification. This was done to satisfy Objective 3. The image enhancements tested were HE, GCN, CLAHE in RGB and HSV colour spaces. The enhancements were applied and trained for each biome. GCN was the most effective enhancement overall, with CLAHE in the RGB colour space being the only other enhancement that



outperformed the baseline. It was thus determined that GCN was an effective enhancement for improving tree classification accuracies using a DL model for segmentation tasks. All the other enhancements were disregarded subsequent to Chapter 3.

Chapter 4 addressed Objective 4, which was to determine the most effective data augmentation methods for improving a CNN for a global forest classification. Five models were trained, each with a different augmentation strategy, namely: 1. no augmentation, 2. geometric augmentations (rotations, translations, horizontal and vertical flips), 3. colour augmentations (brightness, contrast, gamma, CLAHE), 4. geometric augmentations with shear included, and 5. combinations of geometric and colour augmentations. All the augmentation techniques improved model generalisability against the baseline, and the combination of geometric and colour augmentations showing the biggest improvement in MCC score (+ 9.2%).

The high variability of climatic, phenological and geographical conditions around the world make it difficult to produce a global tree map. Objective 5 sought to find the most operationally effective manner in which to infer a global tree classification (in both time and accuracy). Two approaches were tested in this experiment. Firstly, a model per biome approach, for which training, and validation data were derived from the respective biomes. The Savanna and Temperate Forest biomes were chosen to assess the effectiveness of this approach. The second approach was to train a globally representative model that included training data from all biomes. The global model was validated against the Savanna and Temperate Forest biomes. The global model achieved the highest MCC scores in both biomes, with 0.72 in the Savanna biome and 0.87 in the Temperate Forest biome. Thus, it was shown that a single, globally representative model was more robust than a model trained with less data, specific to a biome.

## **5.2 SYNTHESIS: MAIN FINDINGS AND VALUE OF RESEARCH**

The main findings of this research are as follows:

- Applying GCN as an image enhancement prior to training a DL model for tree segmentation was an effective way of improving model accuracy.
- There were no significant trends in respect of image enhancement methods and the biomes to which they are applied.
- Both colour and geometric data augmentations were effective for improving the generalisability of DL models for tree segmentation. Combining the augmentation methods used resulted in the highest accuracies achieved.

- The addition of shear as a geometric augmentation had a negative impact on model MCC score compared with excluding shear from geometric augmentations.
- Training a single, globally representative DL model resulted in a more robust model than training a model specific to each biome.

The experiment in Chapter 3 showed that GCN was the most consistent enhancement across all biomes, scoring higher than the baseline in three of the five biomes tested. The MCC score calculated across all biomes showed that GCN gave the highest improvement of all the enhancements tested, with an increase in MCC from 0.61 (baseline) to 0.67. Although there are no studies comparing how image enhancements affect the accuracy of a DL model in a remote sensing context, image enhancements have been utilised in the natural image processing domain. Available literature has shown that GCN improves the accuracy of deep learning models when used as a pre-process for image categorisation tasks (Goodfellow et al. 2013), face detection tasks (Gudi et al. 2015) and emotion recognition (Pitakola et al. 2017). GCN aims to improve accuracy by decreasing contrast differences in training images. The contrast difference is subtracted from validation images as well. In the experiments, histogram equalisation and CLAHE in the HSV colour space, led to a reduction in model performance compared with the baseline. CLAHE in the RGB colour space was the only other test enhancement that resulted in an MCC score higher than the baseline. While each enhancement was applied to data from the selected biomes, to determine whether any of the enhancements had a predisposition to improve models trained on specific biomes, no trends were detected regarding enhancement methods and biomes.

Data augmentations are often applied to data for training DL models, but without regard for any unintended consequences. Limited research is available on the effects different data augmentation methods have on the performance of DL models in the remote sensing domain, especially for segmentation tasks relating to tree cover from Sentinel-2 data. Chapter 4 focused on testing different data augmentation methods to improve the generalisability of a DL model for tree classification. Data augmentation is a commonly used method for artificially increasing the size and diversity of training datasets. It is applied particularly in domains for which training datasets are limited (Shorten & Khoshgoftaar, 2019). GCN was image enhancement of choice to apply to the input data for Objective 4. Basic colour and geometric augmentations, applied either individually or combined, improved the MCC score of models from 0.668 (no augmentations) to 0.736 for both geometric and colour augmentations applied individually, and to 0.738 when geometric and colour augmentations were combined. The combination of colour and geometric augmentations was the most consistent augmentation method that improved model generalisability.

Adding shear as an additional geometric augmentation, led to a reduced model performance (0.694) compared with applying only basic geometric augmentations. Nonetheless, it still performed better than applying no augmentation. The model trained with shear added displayed apparent confusion relating to the learning process. Shear applies an affine warp to an image, and this can skew the image features to the extent that they represent features in an unrealistic way that does not exist in the real world. It is not recommended that shear be utilised as a data augmentation method for training a DL model for tree segmentation.

In Chapter 4 it was shown that training a single, globally representative model, combining training data from all biomes, produced a higher MCC score than a targeted model trained per biome. The experiment was run on two biomes, the Savanna biome and Temperate Forest biome. A combination of both colour and basic geometric augmentations were applied during training of all the models in this experiment. In both cases, the single model outperformed the biome specific model with the MCC improving from 0.65 to 0.72 in the Savanna biome and from 0.85 to 0.87 in the Temperate Forest biome. The results demonstrated that a single model trained on a very diverse training dataset stemming from different climatic and environmental conditions, had the ability to generalise well across multiple biomes with different climatic and environmental conditions. The findings contrasted with those of Pax-Lenney et al. (2001) that the generalisability of a neural network could only be extended in a temporal dimension when dark object subtraction was applied as a pre-process. However, when attempting to generalise across geographic regions, a decrease of 8%-13% was found in mean accuracy (Pax-Lenney et al. 2001). There is currently no research comparing results of a DL model trained per biome or ecoregion with those of a model trained with very diverse training data.

In summary, the findings from Chapter 3 and Chapter 4 provide strong evidence that an operationally efficient pipeline to infer a global tree classification at 10m resolution is achievable. By applying GCN as an image enhancement process on training and validation data sampled with globally representative data, and further applying basic geometric and colour data augmentations, it is possible to train a single deep learning model that is able to generalise across geographic regions and infer an accurate global tree cover classification at 10m spatial resolution.

### **5.3 STUDY LIMITATIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH**

Several limitations were identified in this research. The source data was derived from LuxCarta cloud-free mosaic. The data type of the imagery was already reduced to a bit depth of 8bits from the original 12bit information from Sentinel-2 L2A data. Although care was taken to choose

summer imagery of the locales of all training and validation images, there were still some seasonal variations in the datasets. Furthermore, the creation of training data for DL models was prone to inherent biases stemming from the technician's ability to interpret features in imagery. The data used in this research was thus not indifferent to those biases, and careful scrutiny of the data indeed revealed some errors.

Only a few commonly used image enhancement and augmentation techniques were tested. It is recommended that future studies should test more sophisticated image normalisation methods, including DL based methods such as GAN's targeting domain adaptation. Over-sampling methods for data augmentation have shown promise in artificially inflating the size of a limited training dataset. Applying this method could produce completely unique training patches by creating fake, but plausible image / label pairs. Furthermore, future studies should test the effects of applying data augmentation at the prediction phase, by performing an inference on a few variations of the image patch, and then combining the probability maps. The final recommendation is to test the concept of transfer learning. Transfer learning inherits the well-trained weights from models trained on large datasets. This concept could be used to transfer the weights of a model trained with globally representative data, and thereafter fine-tuning the model for each biome, thus learning global features as well as refined features are also learnt for each model with the fine-tuned data.

## **5.4 CONCLUSION**

The mapping of global forests plays an important role in managing the ecosystem services these forests provide. Open access to high spatial and temporal resolution data provided by the Sentinel-2 constellation of satellites has underscored the need for an operationally efficient and accurate tree cover classifier.

This study assessed the effectiveness of various image enhancement and augmentation methods borrowed from the computer vision and medical imaging domain of DL to train U-Net models for tree classification. The combination of GCN as an image enhancement with basic geometric and colour data augmentation techniques, was identified as the most effective method to improve the performance of a U-Net model for tree classification. A single, globally representative model trained with data derived from all biomes generalises better over large areas, than models trained with more homogenous data derived from a single biome and applied to the respective biome.

The improvements identified in this study lay a framework for an operationally efficient, global forest classification. Up-to-date global forest maps serve as useful tools for identifying deforestation shift and re-forestation efforts in support of global forest monitoring.

## REFERENCES

- Abdi AM 2020. Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GIScience and Remote Sensing* 57, 1: 1–20. [online]. Available from: <https://www.tandfonline.com/doi/full/10.1080/15481603.2019.1650447> [Accessed 30 May 2020].
- Agarap AF 2018. Deep Learning using Rectified Linear Units (ReLU).
- Akar Ö & Güngör O 2012. Classification of multispectral images using Random Forest algorithm. *Journal of Geodesy and Geoinformation* 1, 2: 105–112.
- Al-Doski J, Mansor SB, Zulhaidi H & Shafri M 2013. Image Classification in Remote Sensing. 3, 10. [online]. Available from: [www.iiste.org](http://www.iiste.org) [Accessed 4 November 2022].
- Ball JE, Anderson DT & Chan CS 2017. Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. *Journal of Applied Remote Sensing* 11, 04: 1.
- Bar Y, Diamant I, Wolf L, Lieberman S, Konen E & Greenspan H 2015. *Chest pathology detection using deep learning with non-medical training*. 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI): 294–297.
- Betts MG, Wolf C, Ripple WJ, Phalan B, Millers KA, Duarte A, Butchart SHM & Levi T 2017. Global forest loss disproportionately erodes biodiversity in intact landscapes. *Nature* 547, 7664: 441–444. [online]. Available from: <http://dx.doi.org/10.1038/nature23285>
- Blaschke T 2010. ISPRS Journal of Photogrammetry and Remote Sensing Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 65, 1: 2–16. [online]. Available from: <http://dx.doi.org/10.1016/j.isprsjprs.2009.06.004>
- Boonprong S, Cao C, Chen W, Ni X, Xu M & Acharya BK 2018. The classification of noise-afflicted remotely sensed data using three machine-learning techniques: Effect of different levels and types of noise on accuracy. *ISPRS International Journal of Geo-Information* 7, 7.
- Bratic G, Vavassori A & Brovelli MA 2021. *REVIEW OF HIGH-RESOLUTION GLOBAL LAND COVER*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. [online]. Available from: <https://doi.org/10.5194/isprs-archives-XLIII-B4-2021-175-2021> [Accessed 7 July 2021].

- Breiman L 2001. ST4\_Method\_Random\_Forest. *Machine Learning* 45, 1: 5–32.
- Brown CF, Brumby SP, Guzder-Williams B, Birch T, Hyde SB, Mazzariello J, Czerwinski W, Pasquarella VJ, Haertel R, Ilyushchenko S, Schwehr K, Weisse M, Stolle F, Hanson C, Guinan O, Moore R & Tait AM 2022. Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data* 9, 1: 251. [online]. Available from: <https://doi.org/10.1038/s41597-022-01307-4>
- Buchhorn M, Lesiv M, Tsendbazar NE, Herold M, Bertels L & Smets B 2020. Copernicus global land cover layers-collection 2. *Remote Sensing* 12, 6: 1–14.
- Campbell J & Wynne R 2011. *Introduction to Remote Sensing, Fifth Edition*. [online]. Available from: [https://books.google.co.za/books?hl=en&lr=&id=NkLmDjSS8TsC&oi=fnd&pg=PP1&dq=Campbell+JB+%26+Wynne+RH.+Introduction+to+remote+sensing&ots=s2Kr3s01E&sig=m\\_grJYXV7ag7EfQOFCXN2SOqNfA#v=onepage&q=Campbell+JB+%26+Wynne+RH.+Introduction+to+remote+sensing&f=false](https://books.google.co.za/books?hl=en&lr=&id=NkLmDjSS8TsC&oi=fnd&pg=PP1&dq=Campbell+JB+%26+Wynne+RH.+Introduction+to+remote+sensing&ots=s2Kr3s01E&sig=m_grJYXV7ag7EfQOFCXN2SOqNfA#v=onepage&q=Campbell+JB+%26+Wynne+RH.+Introduction+to+remote+sensing&f=false) [Accessed 4 November 2022].
- Castelluccio M, Poggi G, Sansone C & Verdoliva L 2015. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. : 1–11. [online]. Available from: <http://arxiv.org/abs/1508.00092>
- Chatfield K, Simonyan K, Vedaldi A & Zisserman A 2014. Return of the devil in the details: Delving deep into convolutional nets. *BMVC 2014 - Proceedings of the British Machine Vision Conference 2014*: 1–11.
- Chen Jun, Chen Jin, Liao A, Cao X, Chen L, Chen X, He C, Han G, Peng S, Lu M, Zhang W, Tong X & Mills J 2015. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*.
- Chen LC, Papandreou G, Kokkinos I, Murphy K & Yuille AL 2018. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, 4: 834–848.
- Chicco D & Jurman G 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 21, 1: 6. [online]. Available from: <https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-019-6413-7> [Accessed 10 April 2021].
- Chlap Phillip, Min H, Vandenberg N, Dowling J, Holloway L, Haworth A, Chlap P, Phd M, Vandenberg ; N, Dowling M; J & Holloway ; L 2021. A review of medical image data

- augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology* 65, 5: 545–563. [online]. Available from: <https://onlinelibrary-wiley-com.ez.sun.ac.za/doi/full/10.1111/1754-9485.13261> [Accessed 2 April 2022].
- Chuvieco E 2016. *Fundamentals of Satellite Remote Sensing*. 2nd ed. CRC Press. [online]. Available from: <https://www.taylorfrancis.com/books/9781498728072>
- Coates A, Lee H & Ng AY 2011. *An analysis of single-layer networks in unsupervised feature learning*. *Journal of Machine Learning Research*: 215–223.
- Cogswell M, Ahmed F, Girshick R, Zitnick L & Batra D 2016. Reducing overfitting in deep networks by decorrelating representations. *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, October 2017.
- Corbane C, Syrris V, Sabo F, Politis P, Melchiorri • Michele, Pesaresi M, Soille P & Kemper • Thomas 2021. Convolutional neural networks for global human settlements mapping from Sentinel-2 satellite imagery. *Neural Computing and Applications* 33. [online]. Available from: <https://doi.org/10.1007/s00521-020-05449-7> [Accessed 5 January 2022].
- DeFries R, Foley J & Asner G 2004. Land-use choices: balancing human needs and ecosystem function. *Frontiers in Ecology and the Environment*, 2: 249–257.
- Demirel H, Ozcinar C & Anbarjafari G 2010. Satellite image contrast enhancement using discrete wavelet transform and singular value decomposition. *IEEE Geoscience and Remote Sensing Letters* 7, 2: 333–337.
- Drummond MA & Loveland TR 2010. Land-use pressure and a transition to forest-cover loss in the Eastern United States. *BioScience* 60, 4: 286–298.
- Elgendi M, Nasir MU, Tang Q, Smith D, Grenier JP, Batte C, Spieler B, Leslie WD, Menon C, Fletcher RR, Howard N, Ward R, Parker W & Nicolaou S 2021. The Effectiveness of Image Augmentation in Deep Learning Networks for Detecting COVID-19: A Geometric Transformation Perspective. *Frontiers in Medicine* 8, March: 1–12.
- ESA 2014. Sentinel Online [online]. Available from: <https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/level-2a-processing> [Accessed 5 November 2021].
- Flight L & Julious SA 2015. The disagreeable behaviour of the kappa statistic. *Pharmaceutical statistics* 14, 1: 74–78.
- Flood N, Watson F & Collett L 2019. Using a U-net convolutional neural network to map woody



- vegetation extent from high resolution satellite imagery across Queensland, Australia. *International Journal of Applied Earth Observation and Geoinformation* 82, February: 101897. [online]. Available from: <https://doi.org/10.1016/j.jag.2019.101897>
- French G & Mackiewicz M 2022. Colour Augmentation for Improved Semi-supervised Semantic Segmentation. : 356–363.
- Freudenberg M, Nölke N, Agostini A, Urban K, Wörgötter F & Kleinn C 2019. Large scale palm tree detection in high resolution satellite images using U-Net. *Remote Sensing* 11, 3: 1–18.
- Frid-Adar M, Klang E, Amitai M, Goldberger J & Greenspan H 2018. *Synthetic data augmentation using GAN for improved liver lesion classification*. Proceedings - International Symposium on Biomedical Imaging: 289–293.
- Fu X, Wang J, Zeng D, Huang Y & Ding X 2015. Remote Sensing Image Enhancement Using Regularized-Histogram Equalization and DCT. *IEEE Geoscience and Remote Sensing Letters* 12, 11: 2301–2305.
- Ganesh V & Ramesh H 2017. Effectiveness of Contrast Limited Adaptive Histogram Equalization Technique on Multispectral Satellite Imagery. *Proceedings of the International Conference on Video and Image Processing*. [online]. Available from: <https://doi.org/10.1145/3177404.3177409> [Accessed 16 October 2021].
- Gong P, Wang J, Yu Le, Zhao Yongchao, Zhao Yuanyuan, Liang L, Yu Liang, Wang L, Liu X, Shi T, Zhu M, Chen Y, Yang G, Tang P, Xu B, Giri C, Clinton N, Zhu Z, Chen Jin & Chen Jun 2013. Finer resolution observation and monitoring of global land cover : first mapping results with Landsat TM and ETM + data. 1161.
- Gong P, Yu L, Li C, Wang J, Liang L, Li X & Ji L 2016. Annals of GIS A new research paradigm for global land cover mapping. 5683.
- Goodfellow I, Bengio Y & Courville A 2016. *Deep Learning*. MIT Press. [online]. Available from: <http://www.deeplearningbook.org>
- Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J & Chen T 2018. Recent advances in convolutional neural networks. *Pattern Recognition* 77: 354–377.
- Gudi A 2015. Recognizing Semantic Features in Faces using Deep Learning. *CoRR* abs/1512.0. [online]. Available from: <http://arxiv.org/abs/1512.00743>
- Gupta S & Kaur Y 2014. *Review of Different Local and Global Contrast Enhancement Techniques for a Digital Image*. 18. *International Journal of Computer Applications* 100



- Hansen M., Potapov P, Moore R, Hancher M, Turubanova S, Tyukavina A, Thau D, Stehman S & Goetz, S.J.; Loveland T. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. 850, November: 850–854.
- He K, Zhang X, Ren S & Sun J 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *IEEE International Conference on Computer Vision (ICCV 2015)* 1502.
- He N, Fang L & Plaza A 2020. Hybrid first and second order attention Unet for building segmentation in remote sensing images. *Science China Information Sciences* 63, 4: 140305. [online]. Available from: <https://doi.org/10.1007/s11432-019-2791-7>
- Helber P, Bischke B, Dengel A & Borth D 2019. EuroSAT : A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification. : 1–9.
- Higgs C 2021. Discriminating Between Forest Plantation Genera Using Remote Sensing and Machine Learning Algorithms. Stellenbosch University. [online]. Available from: <https://scholar.sun.ac.za>
- Immitzer M, Atzberger C & Koukal T 2012. Tree species classification with Random forest using very high spatial resolution 8-band worldView-2 satellite data. *Remote Sensing* 4, 9: 2661–2693.
- Ioffe S & Szegedy C 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *CoRR* abs/1502.0. [online]. Available from: <http://arxiv.org/abs/1502.03167>
- Jadon S 2020. A survey of loss functions for semantic segmentation. *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2020*.
- Jadoon MM, Zhang Q, Haq IU, Butt S & Jadoon A 2017. Three-Class Mammogram Classification Based on Descriptive CNN Features. [online]. Available from: <https://doi.org/10.1155/2017/3640901> [Accessed 1 April 2020].
- Jin B, Ye P, Zhang X, Song W & Li S 2019. Object-Oriented Method Combined with Deep Convolutional Neural Networks for Land-Use-Type Classification of Remote Sensing Images. *Journal of the Indian Society of Remote Sensing* 47, 6: 951–965.
- Karra K, Kontgis C, Statman-Weil Z, Mazzariello JC, Mathis M & Brumby SP 2021. *GLOBAL LAND USE/LAND COVER WITH SENTINEL 2 AND DEEP LEARNING*. International Geoscience and Remote Sensing Symposium (IGARSS): 4704–4707.

- Keskar NS, Mudigere D, Nocedal J, Smelyanskiy M & Tang PTP 2017. *On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima*. International Conference on Learning Representations. [online]. Available from: <https://openreview.net/forum?id=H1oyRIYgg>
- Khalifa NE, Loey M & Mirjalili S 2022. A comprehensive survey of recent trends in deep learning for digital images augmentation. *Artificial Intelligence Review* 55, 3: 2351–2377. [online]. Available from: <https://doi.org/10.1007/s10462-021-10066-4>
- Khan HA, Jue W, Mushtaq M & Mushtaq MU 2020. Brain tumor classification in MRI image using convolutional neural network. *Mathematical Biosciences and Engineering* 17, 5: 6203–6216.
- Khatami R, Mountrakis G & 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: 89–100. [online]. Available from: <https://www.sciencedirect.com/science/article/pii/S0034425716300578>
- Koo KM & Cha EY 2017. Image recognition performance enhancements using image normalization. *Human-centric Computing and Information Sciences* 7, 1.
- Kornatowska B & Sienkiewicz J 2018. Forest ecosystem services-assessment methods. *Folia Forestalia Polonica, Series A* 60, 4: 248–260.
- Korznikov KA, Kislov DE, Altman J, Doležal J, Vozmishcheva AS & Krestov P V. 2021. Using u-net-like deep convolutional neural networks for precise tree recognition in very high resolution rgb (Red, green, blue) satellite images. *Forests* 12, 1: 1–17.
- Kouassi CJA, Khan D, Achille LS, Omifolaji JK & Kebin Z 2021. Forest Resources Depletion: An Ecological Model for Biodiversity Preservation and Conservation in Cote D’Ivoire. *Open Journal of Ecology* 11, 12: 870–890.
- Krizhevsky A, Sutskever I & Hinton GE 2012. *ImageNet classification with deep convolutional neural networks*. 6. *Communications of the ACM* 60 [online]. Available from: <http://code.google.com/p/cuda-convnet/> [Accessed 25 March 2020].
- Kulkarni AD & Lowe B 2016. Random Forest Algorithm for Land Cover Classification. *International Journal on Recent and Innovation Trends in Computing and Communication* 4, 3: 58–63.
- Lang N, Schindler K & Wegner JD 2019. Country-wide high-resolution vegetation height mapping with Sentinel-2. *Remote Sensing of Environment* 233, 2016.

- Lary DJ, Alavi AH, Gandomi AH & Walker AL 2016. Machine learning in geosciences and remote sensing. *Geoscience Frontiers* 7, 1: 3–10.
- Lecun Y, Bengio Y & Hinton G 2015. Deep learning. *Nature* 521, 7553: 436–444.
- Lee E, Kim S, Kang W, Seo D & Paik J 2013. Contrast enhancement using dominant brightness level analysis and adaptive intensity transformation for remote sensing images. *IEEE Geoscience and Remote Sensing Letters* 10, 1: 62–66.
- Lei C, Hu B, Wang D, Zhang S & Chen Z 2019. A Preliminary Study on Data Augmentation of Deep Learning for Image Classification. *Internetware*. [online]. Available from: <https://doi.org/10.1145/3361242.3361259> [Accessed 2 April 2022].
- Lewinski S, Malinowski R, Rybicki M, Gromny E, Nowakowski A, Jenerowicz M, Krupiński Michał, Krupiński Marcin, Krätzschmar E & Guenther S 2019. Automatic Land Cover Classification of Europe with Sentinel-2 Imagery. : 8216.
- Lewinski S, Nowakowski A, Rybicki M, Kukawska E, Malinowski R, Krupiński M, Krätzschmar E, Hofmann P, Bielski C, Dejon D, Prieto DF, Industriananlagenbetriebsgesellschaft P, Ug GE, Jena GF-U & Esa-esrin G 2017. Towards Automatic Global Land Cover Classification on Sentinel-2 Data. : 8216.
- Li, Andreeto, Ranzato & Perona 2022. Caltech 101.
- Li C, Wang J, Wang L, Hu L & Gong P 2014. Comparison of classification algorithms and training sample sizes in urban land classification with landsat thematic mapper imagery. *Remote Sensing* 6, 2: 964–983.
- Li W, Fu H, Yu L, Gong P, Feng D, Li C & Clinton N 2016. Stacked Autoencoder-based deep learning for remote-sensing image classification: a case study of African land-cover mapping. *International Journal of Remote Sensing* 37, 23: 5632–5646. [online]. Available from: <https://doi.org/10.1080/01431161.2016.1246775>
- Lisani JL, Michel J, Morel JM, Petro AB & Sbert C 2016. An Inquiry on Contrast Enhancement Methods for Satellite Images. *IEEE Transactions on Geoscience and Remote Sensing* 54, 12: 7044–7054.
- Liskowski P & Krawiec K 2016. Segmenting Retinal Blood Vessels With Deep Neural Networks. *IEEE Transactions on Medical Imaging* 35, 11: 2369–2380.
- Liu H, Sun H, Li M & Iida M 2020. Application of color featuring and deep learning in maize plant detection. *Remote Sensing* 12, 14.

- Liu J, Wang X, Chen M, Liu S, Shao Z, Zhou X & Liu P 2014. Illumination and contrast balancing for remote sensing images. *Remote Sensing* 6, 2: 1102–1123.
- Loey M, Smarandache F & Khalifa NEM 2020. Within the Lack of Chest COVID-19 X-ray Dataset : A Novel Detection Model Based on GAN and Deep. *Symmetry* 12, 4.
- Lu D & Weng Q 2007. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing* 28, 5: 823–870.
- Lu L, Shin Y, Su Y & Karniadakis G 2020. Dying {ReLU} and Initialization: Theory and Numerical Examples. *Communications in Computational Physics* 28, 5: 1671–1706. [online]. Available from: <https://doi.org/10.4208%2Fcicp.oa-2020-0165>
- LuxCarta 2022. LuxCarta [online]. Available from: <https://www.luxcarta.com/> [Accessed 25 April 2021].
- Lv N, Ma H, Chen C, Pei Q, Zhou Y, Xiao F & Li J 2021. Remote Sensing Data Augmentation through Adversarial Training. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14: 9318–9333.
- Ma L, Liu Y, Zhang X, Ye Y, Yin G & Johnson BA 2019. ISPRS Journal of Photogrammetry and Remote Sensing Deep learning in remote sensing applications : A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing* 152, November 2018: 166–177. [online]. Available from: <https://doi.org/10.1016/j.isprsjprs.2019.04.015>
- Maggiori E, Tarabalka Y, Charpiat G, Neu- PAC, Maggiori E, Tarabalka Y, Charpiat G & Alliez P 2017. Convolutional Neural Networks for Large-Scale Remote Sensing Image Classification To cite this version : Convolutional Neural Networks for Large-Scale Remote Sensing Image Classification.
- Mahdianpari M, Salehi B, Rezaee M, Mohammadimanesh F & Zhang Y 2018. Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery. *Remote Sensing* 10, 7: 1119. [online]. Available from: <http://www.mdpi.com/2072-4292/10/7/1119> [Accessed 27 March 2020].
- Maini R & Aggarwal H 2010. *A Comprehensive Review of Image Enhancement Techniques*. 2
- Maitra IK, Nag S & Bandyopadhyay SK 2012. Technique for preprocessing of digital mammogram. *Computer Methods and Programs in Biomedicine* 107, 2: 175–188.
- Malvika B 2017. COMPARISON OF IMAGE ENHANCEMENT ALGORITHMS FOR IMPROVING THE VISUAL QUALITY OF POST-DISASTER SATELLITE IMAGES

Landsat data.

- Marmanis D, Datcu M, Esch T & Stilla U 2016. Deep learning earth observation classification using ImageNet pretrained networks. *IEEE Geoscience and Remote Sensing Letters* 13, 1: 105–109. [online]. Available from: [http://www.ieee.org/publications\\_standards/publications/rights/index.html](http://www.ieee.org/publications_standards/publications/rights/index.html) [Accessed 7 January 2020].
- Matthews BW 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure* 405, 2: 442–451.
- McDermid GJ, Franklin SE & LeDrew EF 2005. Remote sensing for large-area habitat mapping. *Progress in Physical Geography: Earth and Environment* 29, 4: 449–474. [online]. Available from: <https://doi.org/10.1191/0309133305pp455ra>
- Mellor A, Haywood A, Stone C & Jones S 2013. The performance of random forests in an operational setting for large area sclerophyll forest classification. *Remote Sensing* 5, 6: 2838–2856.
- Mironczuk A & Hościło A 2017. Mapping tree cover with Sentinel-2 data using the Support Vector Machine (SVM). *Geoinformation Issues* 9, 1: 27–38. [online]. Available from: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2>
- Mnih V 2013. Machine Learning for Aerial Image Labeling. University of Toronto.
- Mohamed H, Negm A, Zahran M & Saavedra OC 2015. Assessment of Artificial Neural Network for bathymetry estimation using High Resolution Satellite imagery in Shallow Lakes : Case Study El Burullus Lake . *International Water Technology Conference* , March: 434–444.
- Mucina L & Rutherford MC 2006. *The vegetation of South Africa, Lesotho and Swaziland*.
- Neupane B, Horanont T & Aryal J 2021. Deep learning-based semantic segmentation of urban features in satellite images: A review and meta-analysis. *Remote Sensing* 13, 4: 1–41.
- Noi P & Kappas M 2017. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors* 18, 2: 18. [online]. Available from: <http://www.mdpi.com/1424-8220/18/1/18> [Accessed 10 March 2020].
- Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GVN, Underwood EC, D'amico JA, Itoua I, Strand HE, Morrison JC, Loucks CJ, Allnutt TF, Ricketts TH, Kura Y,

- Lamoreux JF, Wettengel WW, Hedao P & Kassem KR 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience* 51, 11: 933–938. [online]. Available from: [https://doi.org/10.1641/0006-3568\(2001\)051\[0933:TEOTWA\]2.0.CO](https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO)
- Olthof I, Butson C & Fraser R 2005. Signature extension through space for northern landcover classification: A comparison of radiometric correction methods. *Remote Sensing of Environment* 95, 3: 290–302. [online]. Available from: <https://www.sciencedirect.com/science/article/pii/S0034425705000131>
- Ottosen T-B, Petch G, Hanson M & Skjøth CA 2020. Tree cover mapping based on Sentinel-2 images demonstrate high thematic accuracy in Europe. *International Journal of Applied Earth Observation and Geoinformation* 84, June 2019: 101947.
- Özbay E & Özbay FA 2021. A CNN Framework For Classification Of Melanoma And Benign Lesions On Dermatoscopic Skin Images. *International Journal of Advanced Networking and Applications* 13, 02: 4874–4883.
- Ozturk O, Sariturk B & Seker DZ 2020. Comparison of Fully Convolutional Networks (FCN) and U-Net for Road Segmentation from High Resolution Imageries. *International Journal of Environment and Geoinformatics* 7, 3: 272–279.
- Pal KK & Sudeep KS 2016. Preprocessing for Image Classification by Convolutional Neural Networks. *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*: 1778–1781.
- Pan Z, Xu J, Guo Y, Hu Y & Wang G 2020. Deep Learning Segmentation and Classification for Urban Village Using a Worldview Satellite Image Based on U-Net. *Remote Sensing* 12, 10. [online]. Available from: <https://www.mdpi.com/2072-4292/12/10/1574>
- Pax-Lenney M, Woodcock C, Macomber S, Gopal S & Song C 2001. Forest mapping with a generalized classifier and Landsat TM data. *Remote Sensing of Environment*, 77, 241–250. *Remote Sensing of Environment* 77: 241–250.
- Pekel J-F, Cottam A, Gorelick N & Belward AS 2016. High-resolution mapping of global surface water and its long-term changes. *Nature* 540, 7633: 418–422. [online]. Available from: <https://doi.org/10.1038/nature20584>
- Perez F bio, Vasconcelos C, Avila S & Valle E 2018. Data Augmentation for Skin Lesion Analysis. In *Lecture Notes in Computer Science*, 303–311. Springer International Publishing. [online]. Available from: [https://doi.org/10.1007%2F978-3-030-01201-4\\_33](https://doi.org/10.1007%2F978-3-030-01201-4_33)



- Phiri D, Simwanda M, Salekin S, Nyirenda VR, Murayama Y & Ranagalage M 2020. Sentinel-2 data for land cover/use mapping: A review. *Remote Sensing* 12, 14.
- Pitaloka D, Wulandari A, Basaruddin T & Yanti Liliana D 2017. ScienceDirect ScienceDirect Peer-review under responsibility of the scientific Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition ScienceDirect Peer-review under responsibility of the scientific Enhancing CNN with Preprocessing Stage. *Procedia Computer Science* 116: 0–000. [online]. Available from: [www.sciencedirect.com](http://www.sciencedirect.com)[www.sciencedirect.com](http://www.sciencedirect.com)[www.elsevier.com/locate/procedia1877-0509](http://www.elsevier.com/locate/procedia1877-0509)[www.sciencedirect.com](http://www.sciencedirect.com)[www.elsevier.com/locate/procedia1877-0509](http://www.elsevier.com/locate/procedia1877-0509) [Accessed 31 March 2020].
- Pontalba JT, Gwynne-Timothy T, David E, Jakate K, Androutsos D & Khademi A 2019. Assessing the Impact of Color Normalization in Convolutional Neural Network-Based Nuclei Segmentation Frameworks. *Frontiers in Bioengineering and Biotechnology* 7.
- Powers DMW 2007. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies* 2, December: 37–63. [online]. Available from: <http://arxiv.org/abs/2010.16061>
- Premaladha J & Ravichandran & KS 2016. SYSTEMS-LEVEL QUALITY IMPROVEMENT Novel Approaches for Diagnosing Melanoma Skin Lesions Through Supervised and Deep Learning Algorithms. [online]. Available from: <http://www.dermnet.org.nz/lesions/img/melanoma/> [Accessed 31 March 2020].
- Radoux J, Brockmann C, Consult B & Defourny P 2014. Automated Training Sample Extraction for Global Land. , May.
- Ramezan CA, Warner TA & Maxwell AE 2019. Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. *Remote Sensing* 11, 2.
- Robinson C, Hou L, Malkin K, Soobitsky R, Czawlytko J, Dilkina B & Jojic N 2019. Large scale high-resolution land cover mapping with multi-resolution data. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2019-June: 12718–12727.
- Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M & Rigol-Sanchez JP 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 67, 1: 93–104. [online]. Available from: <http://dx.doi.org/10.1016/j.isprsjprs.2011.11.002>



- Ronneberger O, Fischer P & Brox T 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *In International Conference on Medical Image Computing and Computer-Assisted Intervention*: 234–241.
- Roy DP, Li J, Zhang HK, Yan L, Huang H & Li Z 2017. Remote Sensing of Environment Examination of Sentinel-2A multi-spectral instrument ( MSI ) re f l e c t a n c e anisotropy and the suitability of a general method to normalize MSI re f l e c t a n c e to nadir BRDF adjusted re f l e c t a n c e. *Remote Sensing of Environment* 199: 25–38. [online]. Available from: <http://dx.doi.org/10.1016/j.rse.2017.06.019>
- Sarker I 2021. Deep Cybersecurity: A Comprehensive Overview from Neural Network and Deep Learning Perspective. *SN Computer Science* 2.
- Schulze, Katharina; Malek, Ziga; Verberg P 2019. Towards better mapping of forest management patterns: A global allocation approach. *Forest Ecology and Management*. [online]. Available from: [https://www.researchgate.net/publication/328282381\\_Towards\\_better\\_mapping\\_of\\_forest\\_management\\_patterns\\_A\\_global\\_allocation\\_approach](https://www.researchgate.net/publication/328282381_Towards_better_mapping_of_forest_management_patterns_A_global_allocation_approach) [Accessed 16 May 2021].
- Scott GJ, England MR, Starns WA, Marcum RA & Davis CH 2017. Training Deep Convolutional Neural Networks for Land-Cover Classification of High-Resolution Imagery. *IEEE Geoscience and Remote Sensing Letters* 14, 4: 549–553.
- Sekertekin A, Marangoz AM & Akcin H 2017. Pixel-based classification analysis of land use land cover using Sentinel-2 and Landsat-8 data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 42, 4W6: 91–93.
- Sharma RC, Hara K & Tateishi R 2017. High-resolution vegetation mapping in Japan by combining sentinel-2 and Landsat 8 based multi-temporal datasets through machine learning and cross-validation approach. *Land* 6, 3.
- Shen X, Hertzmann A, Jia J, Paris S, Price B, Shechtman E & Sachs I 2016. Automatic portrait segmentation for image stylization. *Computer Graphics Forum* 35, 2: 93–102.
- Shendryk Y, Rist Y, Ticehurst C & Thorburn P 2019. ISPRS Journal of Photogrammetry and Remote Sensing Deep learning for multi-modal classification of cloud , shadow and land cover scenes in PlanetScope and Sentinel-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 157, June 2018: 124–136. [online]. Available from: <https://doi.org/10.1016/j.isprsjprs.2019.08.018>

- Shimada M, Itoh T, Motooka T, Watanabe M, Shiraishi T, Thapa R & Lucas R 2014. New global forest/non-forest maps from ALOS PALSAR data (2007–2010). *Remote Sensing of Environment* 155: 13–31. [online]. Available from: <https://www.sciencedirect.com/science/article/pii/S0034425714001527>
- Shlien S & Smith A 1975. A rapid method to generate Spectral theme classification of LANDSAT imagery. *Remote Sensing of Environment* 4, C: 67–77.
- Shorten C & Khoshgoftaar TM 2019. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data* 6, 1. [online]. Available from: <https://doi.org/10.1186/s40537-019-0197-0>
- Showkat N & Parveen N 2017. (PDF) Non-Probability and Probability Sampling. In *Atshala*, 1–8. [online]. Available from: [https://www.researchgate.net/publication/319066480\\_Non-Probability\\_and\\_Probability\\_Sampling](https://www.researchgate.net/publication/319066480_Non-Probability_and_Probability_Sampling)
- Shyam L, Singh S, Ahlawat AK, Singh KM & Singh TR 2017. *A Review on Image Enhancement Methods on Different Domains*. 1. *International Journal of Engineering Inventions* 6 [online]. Available from: [www.ijejournal.com](http://www.ijejournal.com) [Accessed 3 January 2020].
- Silva-Flores R, Pérez-Verdín G & Wehenkel C 2014. Patterns of tree species diversity in relation to climatic factors on the Sierra Madre Occidental, Mexico. *PLoS ONE* 9, 8: 105034. [online]. Available from: [www.plosone.org](http://www.plosone.org) [Accessed 30 May 2021].
- Sjöqvist H, Långkvist M & Javed F 2020. An Analysis of Fast Learning Methods for Classifying Forest Cover Types. *Applied Artificial Intelligence* 34, 10: 691–709. [online]. Available from: <https://www.tandfonline.com/action/journalInformation?journalCode=uaai20> [Accessed 15 May 2021].
- Solórzano J V., Mas JF, Gao Y & Gallardo-Cruz JA 2021. Land use land cover classification with U-net: Advantages of combining sentinel-1 and sentinel-2 imagery. *Remote Sensing* 13, 18.
- Souza CM, Shimbo JZ, Rosa MR, Parente LL, Alencar AA, Rudorff BFT, Hasenack H, Matsumoto M, Ferreira LG, Souza-Filho PWM, de Oliveira SW, Rocha WF, Fonseca A V., Marques CB, Diniz CG, Costa D, Monteiro D, Rosa ER, Vélez-Martin E, Weber EJ, Lenti FEB, Paternost FF, Pareyn FGC, Siqueira J V., Viera JL, Neto LCF, Saraiva MM, Sales MH, Salgado MPG, Vasconcelos R, Galano S, Mesquita V V. & Azevedo T 2020. Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. *Remote Sensing* 12, 17.

- Stivaktakis R, Tsagkatakis G & Tsakalides P 2019. Deep Learning for Multilabel Land Cover Scene Categorization Using Data Augmentation. *IEEE Geoscience and Remote Sensing Letters* 16, 7: 1031–1035.
- Swaine M, Smit C, Tripodi S, Fonteix G, Tarabalka Y, Laureore L & Hyland J 2020. Operational pipeline for a global cloud-free mosaic and classification of sentinel-2 images. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 43, B3: 195–200.
- Taylor L & Nitschke G 2017. Improving Deep Learning using Generic Data Augmentation. , August. [online]. Available from: <http://arxiv.org/abs/1708.06020>
- Timothy R.H. Pearson Sandra L. Brown Richard A. Birdsey 2007. Measurement Guidelines for the Sequestration of Forest Carbon, northern research station: General Technical Report NRS-18. *United States Department of Agriculture, forest service*: 5.
- Tong X-Y, Lu Q, Xia G-S & Zhang L 2018. Large-scale Land Cover Classification in GaoFen-2 Satellite Imagery. [online]. Available from: <http://arxiv.org/abs/1806.00901> [Accessed 7 January 2020].
- Townshend JR, Masek JG, Huang C, Vermote EF, Gao F, Channan S, Sexton JO, Feng M, Narasimhan R, Kim D, Song K, Song D, Song X-P, Noojipady P, Tan B, Hansen MC, Li M & Wolfe RE 2012. Global characterization and monitoring of forest cover using Landsat data: opportunities and challenges. *International Journal of Digital Earth* 5, 5: 373–397. [online]. Available from: <https://www.tandfonline.com/action/journalInformation?journalCode=tjde20> [Accessed 18 April 2022].
- Tuia D & Camps-Valls G 2009. Recent advances in remote sensing image processing. *Proceedings - International Conference on Image Processing, ICIP* , December: 3705–3708.
- Umri BK, Wafa Akhyari M & Kusri K 2020. *Detection of COVID-19 in Chest X-ray Image using CLAHE and Convolutional Neural Network*. 2020 2nd International Conference on Cybernetics and Intelligent System, ICORIS 2020. Institute of Electrical and Electronics Engineers Inc.
- Venter ZS, Barton DN, Chakraborty T, Simensen T & Singh G 2022. Global 10 m Land Use Land Cover Datasets: A Comparison of Dynamic World, World Cover and Esri Land Cover. *Remote Sensing* 14, 16.

- Verhulp J & Van Niekerk A 2016. Effect of inter-image spectral variation on land cover separability in heterogeneous areas. *International Journal of Remote Sensing* 37, 7: 1639–1657.
- Weih RC & Riggan ND 2008. Object-based classification vs Pixel-based classification: Comparative importance of multi-resolution imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII*.
- Weinstein BG, Marconi S, Aubry-Kientz M, Vincent G, Senyondo H & White E 2020. DeepForest: A python package for RGB deep learning tree crown delineation. *bioRxiv*: 2020.07.07.191551. [online]. Available from: <https://doi.org/10.1101/2020.07.07.191551> [Accessed 15 May 2021].
- Wessel M, Brandmeier M & Tiede D 2018. Evaluation of Different Machine Learning Algorithms for Scalable Classification of Tree Types and Tree Species Based on Sentinel-2 Data. *Remote Sensing 2018, Vol. 10, Page 1419* 10, 9: 1419. [online]. Available from: <https://www.mdpi.com/2072-4292/10/9/1419/htm> [Accessed 4 November 2022].
- Wittke S, Yu X, Karjalainen M, Hyypä J & Puttonen E 2019. Comparison of two-dimensional multitemporal Sentinel-2 data with three-dimensional remote sensing data sources for forest inventory parameter estimation over a boreal forest. *International Journal of Applied Earth Observation and Geoinformation* 76: 167–178.
- Woodcock CE, Macomber SA, Pax-Lenney M & Cohen WB 2001. Monitoring large areas for forest change using Landsat: Generalization across space, time and Landsat sensors. *Remote Sensing of Environment* 78, 1–2: 194–203.
- Xiao H, Rasul K & Vollgraf R 2017. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*.
- Xie S, Girshick R, Dollár P, Tu Z & He K 2017. *Aggregated Residual Transformations for Deep Neural Networks*. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR): 5987–5995.
- Xu Y 2018. Building Extraction in Very High Resolution Remote Sensing Imagery Using Deep Learning. *Remote Sensing*.
- Yang C, Rottensteiner F & Heipke C 2018. CLASSIFICATION OF LAND COVER AND LAND USE BASED ON CONVOLUTIONAL NEURAL NETWORKS. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives IV*: 7–10.

- Yang Y & Newsam S 2010. *Bag-Of-Visual-Words and Spatial Extensions for Land-Use Classification*. ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. San Jose. Association for Computing Machinery.
- Yu X, Wu X, Luo C & Ren P 2017a. Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework. *GIScience and Remote Sensing* 54, 5: 741–758. [online]. Available from: <https://doi.org/10.1080/15481603.2017.1323377>
- Yu X, Wu X, Luo C & Ren P 2017b. GIScience & Remote Sensing Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural ne. [online]. Available from: <https://doi.org/10.1080/15481603.2017.1323377>
- Zanaga D, Van De Kerchove R, De Keersmaecker W, Souverijns N, Brockmann C, Quast R, Wevers J, Grosu A, Paccini A, Vergnaud S, Cartus O, Santoro M, Fritz S, Georgieva I, Lesiv M, Carter S, Herold M, Li L, Tsendbazar N-E, Ramoino F & Arino O 2021. ESA WorldCover 10 m 2020 v100. [online]. Available from: <https://doi.org/10.5281/zenodo.5571936>
- Zhang C, Jiang W, Zhang Y, Wang W, Zhao Q & Wang C 2022. Transformer and CNN Hybrid Deep Neural Network for Semantic Segmentation of Very-High-Resolution Remote Sensing Imagery. *IEEE Transactions on Geoscience and Remote Sensing* 60.
- Zhang Liangpei, Zhang Lefei & Du BO 2016. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geoscience and Remote Sensing Magazine* 4.
- Zheng Q, Yang M, Yang J, Zhang Q & Zhang X 2018. Improvement of Generalization Ability of Deep CNN via Implicit Regularization in Two-Stage Training Process. *IEEE Access* 6: 15844–15869.