Investigating the efficiency and capabilities of UAVs and Convolutional Neural Networks in the field of remote sensing as a land classification tool

Cameron Wesson[1], Dr Wilma Britz[1], Dr Robbert Duker[2]

[1]Nelson Mandela University, Geosciences Department
[2] Nelson Mandela University, Botany Department

DOI: http://dx.doi.org/10.4314/sajg.v12i.2.5

Abstract

The study aimed to determine the efficacy and capabilities of using high-resolution aerial imagery and a convolutional neural network (CNN) to identify plant species and monitor land cover and land change in the context of remote sensing. The full capabilities of a CNN were examined, including testing whether the platform could be used for land cover and the evaluation of land change over time. An unmanned aerial vehicle (UAV) was used to collect the aerial data of the study area. The CNN was encoded and operated in RStudio, while digitised data from the input imagery were used by the programme as training and validation data. The object in this respect was to learn about the relevant features of the landscape, and thereafter to classify the Opuntia invasive plant species. Accuracy assessments were carried out on the results to test the efficacy of the aerial imagery in terms of its accuracy and reliability. The classification achieved an overall accuracy of 93%, while the kappa coefficient score was 0.86. CNN was also able to predict the land coverage area of Opuntia to be within four percent (4%) of the ground truthing data. A change in land cover over time was detected by the programme after the manual clearing of the plant had been undertaken. This research has determined that the use of a CNN in remote sensing is a very powerful tool for supervised image classifications. It can be used for monitoring land cover in that it is able to accurately estimate the spatial distribution of plant species and to monitor the growth or decline in the species over time. As such, it is an efficient methodology and its use in remote sensing could be extended.

Keywords: *Convolutional Neural Network, UAVs, Remote Sensing, Plant Species Identification, Deep Learning, Invasive Species, Albany Thicket, Earth Observation, Land Cover/Change, Nature Conservation.*

1. Introduction

Aerial, non-invasive spectral imagery, captured by UAVs, aircraft, or satellites, is an ideal method for collecting data across large areas. Multispectral imagery is readily available and

captures data from 350-2500nm within the electromagnetic spectrum by making use of three to 10 broad spectral bands, the number of which is dependent on the number of detectors within a spectral sensor (Fischer & Kakoulli, 2006). However, owing to the limited bands available in multispectral imaging, it is not possible to detect vegetation species. Therefore, to remedy the situation, a hyperspectral dataset has been developed from a combination of imaging and point spectroscopy (Liu, Bruning, Garnett, & Berger, 2020; Lu & Park, 2008). The major difference between hyperspectral and multispectral images is that the former contains hundreds of narrow (10-20nm) spectral bands (Adam, Mutanga, & Rugege, 2010), which show a two-dimensional distribution of the reflected spectral signatures of objects (Liu, 2020), whereas the latter captures data within specific wavelength ranges across the electromagnetic spectrum.

Previous research has made use of multispectral and hyperspectral imaging systems for the distribution of vegetation biomes and the identification of plant species, respectively. However, although hyperspectral imagery has a high spectral resolution and is capable of identifying different plant species, and even distinguishing diseases in plants (Zhang, Yang, Pan, Yang, Chen, & Zhao, 2020), multispectral imagery cannot, on account of its low spectral resolution, detect and identify vegetation species. As such, an alternative method is needed for hyperspectral imaging, as it is a very complex system to use. Likewise, hiring aerial vehicles with a hyperspectral sensor on board is challenging, as there are few available and they are expensive to hire in South Africa. Therefore, the use of CNNs to classify images and perform object detection has been increasingly adopted over the last few years in the fields of remote sensing and GIS.

Although the use of CNNs as an object recognition tool has increased recently, they have been primarily used for detecting objects within images, as was done in the study by Kembuan et al. (2020), which created an Indonesian sign language dataset consisting of 2659 images and 26 six-letter categories. Hou et al. (2019) conducted a study whereby they used a CNN to detect objects in a video and then tracked those objects as they moved across the field of view. The results revealed an efficient way in which to identify objects, with an accuracy result of 89.23%. Ji et al. (2019) undertook a study in which they were able to use a CNN algorithm to classify electrocardiogram (ECG) data into five different categories, with an average accuracy of 99.21%, which contributed to the diagnosis of heart disease.

Even though CNNs have been extensively and successfully applied to object detection and recognition, as briefly mentioned above, they have only recently been used as a tool for image classification in the field of remote sensing. However, as camera systems have been developed and improved, the amount of data that can be collected has increased. This allows for the collection of very high spatial resolution aerial images and thus the identification and recognition of fine-grained features (Hoeser, Bachofer, & Kuenzer, 2020). Kattenborn et al. (2020) undertook a study in New Zealand which captured high spatial resolution imagery of a

glacier foreland with a complex vegetation biome that consisted of herbs and shrubs. Four vegetation species were classified using a CNN implemented in RStudio, and the results showed high accuracy and a reliable classification of the plant species. In a similar study conducted by Kattenborn et al., (2019) invasive woody plant species in Chile were classified using similar methods to those in the Kattenborn (2020) study mentioned above. In the Chile study, a CNN was able to classify the invasive woody species in an aerial image captured at the canopy level. Schiefer et al. (2020) used a CNN to map forest tree species at the canopy level in Germany, thereby identifying and classifying 14 different species from the aerial images, with 89% accuracy.

Convolutional neural networks were developed from artificial neural networks (ANNs), which embrace a layered structure of algorithms on which deep learning, a subset of machine learning, relies (O'Shea & Nash, 2015). CNNs were designed specifically to analyse spatial patterns and are, therefore, the most effective algorithms for classifying high spatial resolution aerial data (Schiefer, 2020). Traditional classifiers used in the field of remote sensing are pixel-based, while CNNs use textures to detect objects, such as the contextual signal of many neighbouring pixels, their corners, edges, shapes, colours and sizes (Kattenborn, 2020; Schiefer, 2020). CNNs are also self-learning algorithms, which allow for a more accurate classification of an image. Moreover, the basic architecture of a CNN consists of an input layer, a convolution layer that applies the filters obtained from the training data, a pooling layer that reduces data dimensionality, and a fully connected layer that arranges the output image and results (Kembuan, 2020; Lee & Song, 2019; O'Shea, 2015).

The Albany thicket biome, which is found predominately in the Eastern Cape, South Africa, supports a large array of different fauna species and is known to be a highly effective carbon sink (Marais, Cowling, Powell, & Mills, 2009). The vegetation within this biome generally comprises medium-sized shrubland to low forestland, and owing to the dense shrubland, is renowned for being impenetrable (Hoare et al., 2006). However, encroaching alien plants threaten the indigenous Albany thicket plant species, capture their resources and destroy the natural habitat/ecosystem (Hoare, 2006). In particular, the prickly pear (*Opuntia*), which is a widespread invader of the Albany thicket, thrives in the Eastern Cape and is a fast-spreading species. The prickly pear forms dense stands, which can be as much as 10 metres wide and one to five metres high (Everitt, Escobar, Alaniz, & Davis, 1991). It is a succulent shrub with large rounded thick leaves and thorns as protection.

Invasive plant species (IPS) have negative effects on the Albany thicket biome and cause events, such as natural resource capture, encroachment, and colonisation, which in their turn lead to the destruction of the natural vegetation (Kohli, Dogra, Batish, & Singh, 2009). The Albany thicket biome is one of seven natural vegetation biomes found in South Africa and plays a very important role as the backbone of multiple ecosystems for a range of fauna species

(Kota & Shackleton, 2015). Owing to the density of the bush, mapping the spatial distribution of plant species in this biome is challenging (Hoare, 2006). However, because a knowledge of the distribution of invasive species is vital for conservation efforts, high-resolution aerial imagery has been successfully paired with convolutional neural networks (CNNs) to detect and identify plant species at the canopy level. It was in this current study that this technique was applied for the first time to the Albany thicket to find and map the spatial distribution of the prickly pear (*Opuntia*) plant species.

2. Methodology

2.1. Study Area

The data collection site was on the Sibuya Game Reserve, located on the southern coast of the Eastern Cape, near Kenton-on-Sea (figure 1). The indigenous flora in the study region is made up primarily of Albany thicket vegetation species. The Kowie thicket species (Hoare, 2006) is widespread across the reserve and thrives in clay-rich soils (Hoare, 2006), which is a common soil type in the area. *Opuntia* is currently the most widespread of the invaders on the reserve; it thrives in the prevailing climatic conditions and can easily start growing from a segment of itself when dropped or cut (Everitt, 1991). Large fauna species (elephants) are also responsible for its widespread distribution across the reserve. As seen in figure 2, three species of *Opuntia* occur on the reserve. However, this study will not attempt to detect the three species separately but rather to classify them collectively as *Opuntia*.

Data were collected across three plots that range in size from 16 to 22 hectares. The plots were chosen on account of the high concentration of the prickly pear invader on them and the natural occurrence of thicket, both of which make it difficult to detect alien plant species (AIPs). The three plots seen in figure 1 were further broken down into 45 subplots, measuring 60 by 60 metres in size. This was done to reduce the data dimensionality and to focus on the areas of the subplots with high concentrations of *Opuntia*. Subplots 1 to 20 issued from plot 1, subplots 21 to 31 from plot 2, and subplots 32 to 45 from plot 3.

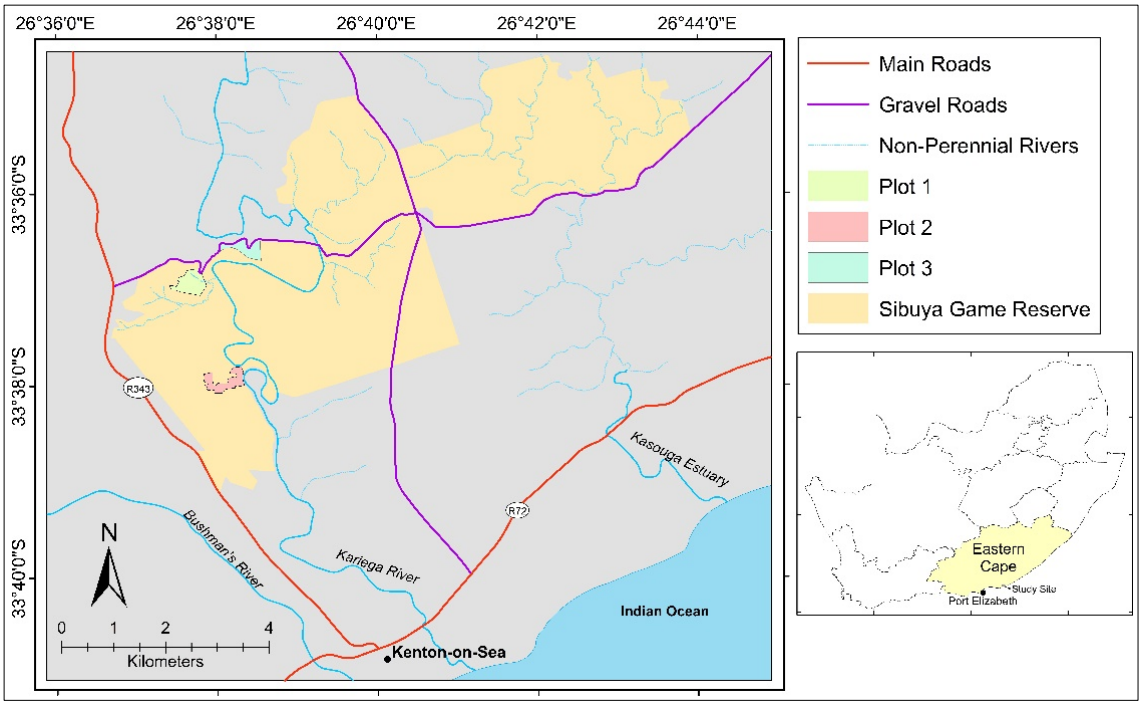

Figure 1: Location of the study site.

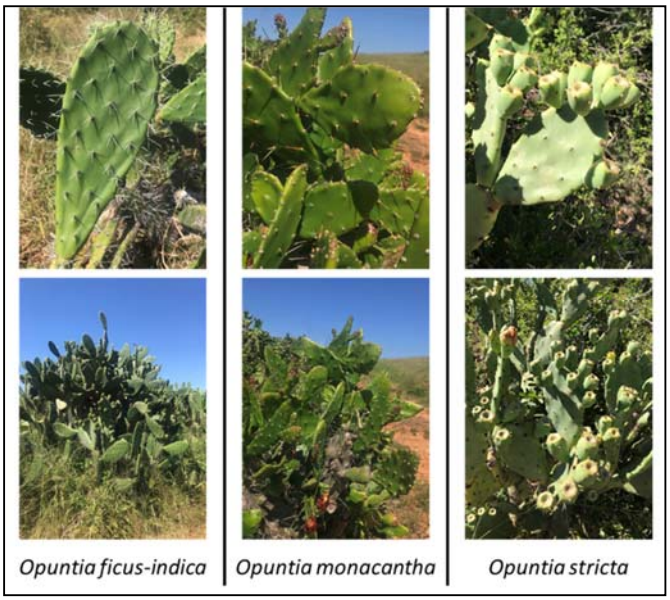

Figure 2: Three species of *Opuntia* found in the study area.

2.2. Data Collection

High-resolution aerial imagery will be the main data source for the data analysis. The images were collected utilizing Unmanned Aerial Vehicles (UAVs). UAVs were selected as the platform for collecting the images over satellite imagery and manned aerial vehicles, with

camera sensors attached such as planes or helicopters due to the flexibility they allow during the data collection process and the high resolution of imagery that can easily be captured.

The UAV used in this study was the DJI Mavic 2 Pro, which had a Hasselblad 20MP (megapixel) one-inch complementary metal-oxide-semiconductor (CMOS) RGB sensor attached to it with a 3-axis gimbal. In addition, the sensor allowed the drone to be flown at a high altitude while still permitting it to capture high resolution images. The DJIFlightPlanner was used to create precise flight plans for the UAV over each plot. This allowed for a forward and side overlap of 70% between images and a flight speed of 17.7km/h to be maintained. Furthermore, it permitted a flight height of 70 metres to secure an image resolution between 1.94 cm/pixel and 2.35 cm/pixel. The flight plans were then exported to a .csv file to be used in Litchi, a software that connected to the drone and controlled it during the data collection process, which was based on the flight files that had been imported into the software.

2.3. Data processing

The processing involved removing any distortions or bad images from the aerial photos and then merging the individual images of each plot to create a single orthomosaic of the plots. Agisoft Metashape Professional (version 1.8.1 build 13915 (64bit)) software was used to merge the images and generate the orthomosaics. When a drone captures images, a geolocation is associated with each image, and Agisoft uses this geolocation to align the images and initiate the merging process. After the orthomosaic of each plot had been created, the footprints of each subplot were cropped from the respective orthomosaic to produce the 60 by 60-metre image that would later be classified.

The classification of the aerial data was performed using a CNN, which is a powerful form of artificial machine learning that was specifically developed for the identification and recognition of objects in imagery and videos. Moreover, a CNN is a powerful deep-learning algorithm, which can detect objects in RGB images, whereas traditional remote sensing classifiers require multispectral or hyperspectral imagery for land cover classifications. To execute the CNN, RStudio version 1.4.1106 was used to encode and run the classifier.

Figure 3 presents an overview of a basic CNN architecture and indicates the filter size, which is used to extract a portion of the image to be used in the first convolution layer for processing. After this, the data is reduced in a pooling layer, which is a single epoch, and yet another convolution layer and another pooling layer is then created to improve on the results. This process is repeated for the entire image and thus multiple epochs can be set by the user. Once the feature extraction has been completed, the data enter the fully connected layer where an entire dataset is compiled, with a dense layer thus being formed, which is then used to produce the output data.

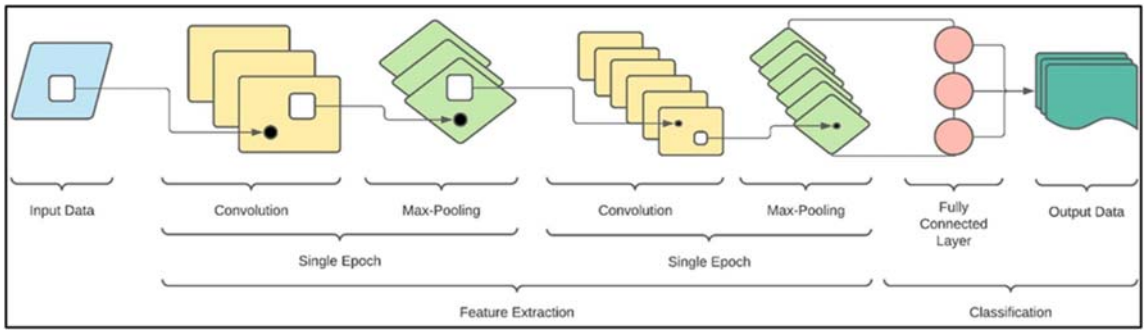

Figure 3: The basic architecture of a CNN.

The U-Net CNN architecture was used for image processing in the study. Developed in 2015 by Olaf Ronneberger, Philipp Fisher, and Thomas Brox, it was originally developed for biomedical image segmentation but has since been adapted slightly and used in other machine-learning fields. The U-Net architecture has been built in such a way that it has a contracting path that captures the context. The architecture also has an expanding path that is symmetrical to the contracting path and maps the contextual information to match the original input image resolution (Schiefer, 2020).

In the layout of the U-Net structure used, the contracting path and the expanding path were each made up of four blocks, with one block in the centre. Each block in the contracting path consisted of two 3x3 convolutions, with each convolution being followed by a batch normalisation and a gaussian error linear units (gelu) activation function. After the execution of the convolutions and activation functions, a max pooling of 2x2 with a striding of two, was performed, which completed each block and reduced the data dimensionality of the feature maps by half.

The central block had the same structure as the contracting path blocks. However, it had only two convolutions and activation functions, and not a max-pooling operation. Within the expanding path, the four blocks each consisted of an up-sampling of the feature maps by a 2x2 up-convolution, which reduced the feature maps from the previous block by half. The feature maps were then concatenated with feature maps from the equivalent contracting path. Once this was done, a 3x3 convolution was performed, followed by a batch normalisation and a gelu activation function. This process was repeated three times in each up-block. Through the expanding path, the feature map value was halved in each block, and the spatial dimension was multiplied by two. Lastly, the pixel-wise classification was executed, and a 1x1 convolutional layer with the sigmoid activation function was applied.

2.4. Reference data and classification

The CNN needs reference data to learn the features and attributes of the target class. A 25 by 25-metre area was drawn up in each subplot. *Opuntia* from the aerial view was then digitized manually and the shapefiles were fed into the CNN programme. A high standard was upheld with the execution of the manual digitization process as the CNN learns the plant's shapes and characteristics from the input data. Poor quality reference data could negatively affect the CNN classification results.

The programme broke the reference data shapefiles down into tiles of 128 by 128 pixels. This was done because the graphics card on the machine used for processing performs best with images that size. The digitised, tiled data were then split by the programme into training data and validation data. Most importantly, the digitised data from the three plots were used as input data for the validation and training data programme.

This meant, however, that the classifier would have an extremely large dataset to contend with and different areas from which to learn, as the training data for Plot 1 would be used for classifying plots 2 and 3, and *vice versa* in the case of plots 2 and 3. Nevertheless, as the images of the plots captured slightly different vegetation densities and were taken at different times, inputting all data from the three plots enabled a wide range of variables and features to be learnt by the classifier, and should have theoretically provided a higher-quality result.

Once the classifier had learnt the training data, the subplots were inputted for classification. The classifier broke each image down into a smaller 128 by 128-pixel tile, classified it, moved on to the next tile, and then rebuilt all the tiles to match the size of the original input image. Validation was then performed on the output data by the programme.

2.5. Accuracy Assessment

Two major accuracy assessments were carried out to test the full efficacy of the CNN. The first accuracy assessment was conducted to determine how well the CNN had learnt the features to be classified. This was encoded in RStudio and a portion of the reference data was used to determine the cross entropy and dice coefficient loss functions. These functions are run in each epoch[1] during the learning phase of a CNN. Cross entropy looks at each pixel and compares the class prediction (neural network prediction) with the target vector, or the masks that were inputted. The closer the output value of the cross entropy is to zero, the better the performance of the neural network. The dice coefficient loss function measures the overlap between the

[1] An epoch is a process where the programme extracts information from the training data and learns features. It is repeated to increase the accuracy of the learned features.

class prediction and the target vector. The output of the dice coefficient is rated 0 to 1, where 1 represents perfect overlap and 0 represents no overlap.

The second accuracy assessment was to determine the trustworthiness and reliability of the classification of the orthomosaics by calculating the user's and producer's accuracy and by computing the kappa coefficients, which would rate the classifications. Control points were placed on two classes, the *Opuntia* species and other, the latter included indigenous vegetation, bare ground, and any other relevant features in the plots that weren't *Opuntia*. The other class was selected to test whether the classifier had overclassified or misclassified *Opuntia*. Functions were run in ArcMap to create error matrix tables for each subplot and from this, the respective user's and producer's accuracy results were calculated alongside the kappa coefficient.

2.6. Land cover and change detection

In the study, the land cover analysis determined the land coverage of *Opuntia* per plot. Using the CNN predictions for each subplot, the area occupied by *Opuntia* per subplot could be calculated in ArcGIS. For each prediction attribute table, a new field called "size" was added and the field calculator was used to determine the area of all the predictions in the respective subplots. To ensure the accuracy of the predictions and the land cover analysis, a small 10 by 10-metre area was drawn up in six subplots. Within this area, the *Opuntia* was manually digitised to an extremely high level. Following the steps above, the area of the manually digitised *Opuntia* was then calculated. The CNN predictions were then added, and again the steps above were followed to calculate the area of coverage in the small region. The manually digitised results could then be compared to the CNN prediction results and the percentage difference between them was calculated.

The change detection aimed to determine whether the CNN could be used to monitor the change in vegetation densities and population numbers over time. Unfortunately, to wait for more *Opuntia* to grow was not an option; therefore, a subplot was selected, and the AIPs (the *Opuntia* plants) were manually cleared from the area. However, not all the *Opuntia* plants were removed from the subplot[2]. Before clearing started, the UAV was flown over the plot and again after it had been cleared. The aerial data was then classified, and the land cover quantities could be compared.

[2] This was done on purpose so that the plot would still have training data and it would be possible to determine whether the CNN could still detect the remaining plants.

3. Results and Discussion

3.1. Classification training data

For the classification, a total of 16 701 jpeg image tiles and their respective png masks were inputted into the classifier, which split the input data into training and validation data at an 80:20 ratio (80% to training data). Other studies in plant identification using CNNs have not published the amount of data that was inputted into the CNN for training and validation. However, the results of the classification (see below) indicate that the number of images that was inputted into the programme was more than sufficient for the CNN to adequately learn all the features and variables for the study.

3.2. CNN model training results

The two loss functions that were run during the learning process reached a minimum after 30 epochs had run over the training data. The loss functions used the validation data to test the learning rates and determine the results. In figure 4, the cross entropy and dice coefficient loss results can both be seen: by epoch 30, the cross entropy had reached a value of 0,1005602, which could result in acceptable classification probabilities. The study conducted by Schiefer et al. (2020) achieved a cross entropy between 0.02 and 0.03, which is an excellent result. However, the plants and canopy structures investigated by Schiefer et al. (2020) were completely different from the ones in the current study, which means that comparing the results is not straightforward. In figure 4, the dice coefficient loss attained a value of 0,976899 by epoch 30, and with the curve levelling out from the 14th epoch onwards, it could be concluded that the dice coefficient loss graph has an ideal shape. This suggests that the programme could identify features and achieve a high overlap percentage early in the learning process.

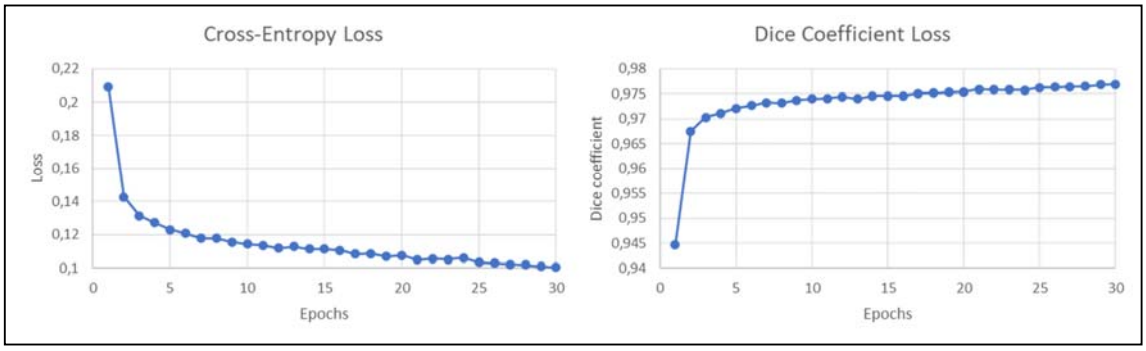

Figure 4: The cross entropy and dice coefficient loss function results.

3.3. Classification results

Once the features had been learned, the CNN was able to classify all 45 subplot images. The outputs of the CNN predictions were in tiff files and were transformed using polygonization in

QGIS to structure the predictions in vector polygon format. In figure 5, an example of the CNN classification can be seen: the image on the left is the subplot before classification and the image on the right is the subplot with the CNN *Opuntia* classification overlaid in pink.

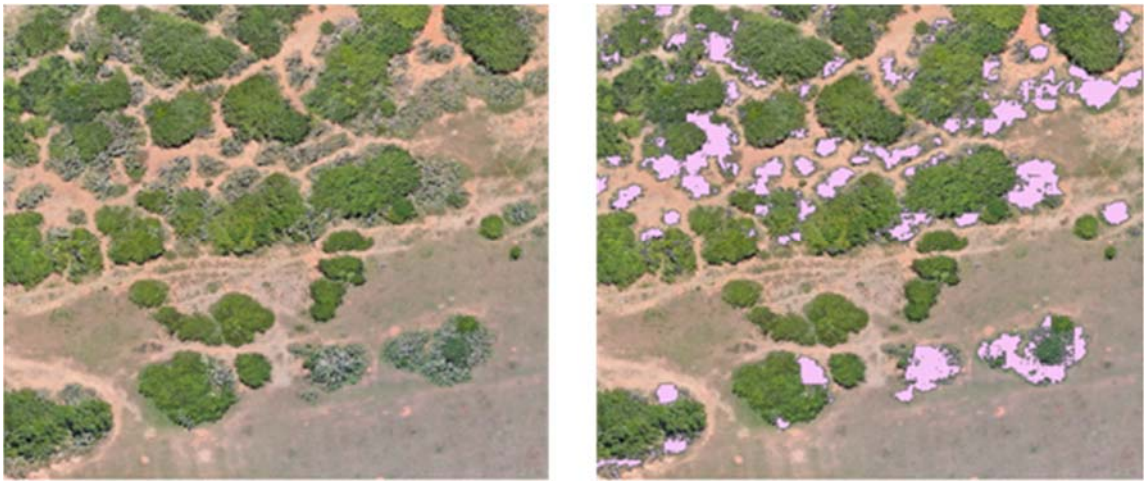

Figure 5: An example of the classification of *Opuntia* in one of the subplots. The indigenous vegetation has not been classified: only the invasive species, which are represented by the pink polygons.

3.4. Classification accuracies

The user's accuracy test is a measurement of the reliability of the classification. The producer's accuracy test is a measurement of how well a given area (class) has been classified. The overall accuracy test gives a single percentage of how well a map has been classified. The kappa coefficient is a measure of the complete agreement of a matrix, and the output is from 0 to 1, with 0 being poor, 0.4-0.6 being moderate, and 1 being perfect. These accuracy tests were selected as they have been well-developed and can be considered benchmark accuracy tests in the field of remote sensing. The user's and producer's accuracy tests were performed on the two classes, namely, the *Opuntia* class and the other class, which consists of all other features in the scene such as the indigenous vegetation and bare ground.

In table 1, the accuracy results applicable to the 45 subplots are displayed as an average value. These results were all very high and do indeed correspond with the learning results shown above. For the accuracy results from the two classes, the results were very high and showed that the *Opuntia* had been classified accurately and that the results are trustworthy. The overall accuracy of the test was also very good at 93%, thus showing that a high level of classification had been achieved. Based on the estimation in the study by Rwanga and Ndambuki (2017), the kappa coefficient score average was rated almost perfect.

Table 1: The average accuracy results from the 45 subplots.

Opuntia class		Other class		Overall accuracy	Kappa coefficient
User's Accuracy	Producer's Accuracy	User's Accuracy	Producer's Accuracy		
94%	93%	92%	94%	93%	0.86

The study by Kattenborn et al. (2019) achieved an overall classification accuracy that ranged from 84% to 90% for four different plant classes, while the study by Oldeland et al. (2021) achieved an overall classification accuracy of 95% and a kappa score of 0.937. These studies, alongside the current study, show how well neural networks are able to learn features and classify images. However, in the current study, only one plant was observed (one target class) and that class, the *Opuntia*, has a distinct shape compared with the surrounding bush (apart from the euphorbia). This allowed the CNN to perform well and easily identify and classify the plant, which would have boosted the classification accuracy results. Other studies, which investigated multiple species that were similar to plants in the surrounding bush, achieved an overall classification accuracy between 84% and 90% (Kattenborn, 2020; Schiefer, 2020). However, the classification results show that the CNN is a powerful and efficient tool for land classification and further research should be done on different target classes.

3.5. Weaknesses of the classification

Although the classification proved effective in identifying and classifying *Opuntia* in the subplots, there were weaknesses. Three commonly occurring issues were noted in the 45 subplots: the misclassification of dead wood and river euphorbia, and the over-classification of *Opuntia*. The misclassification of dead wood occurred only when it had a white-silver reflectance, while that of the river euphorbia occurred as a result of its similarity in terms of its canopy level, shape, and colour to these particular features of the *Opuntia*. The programme never misclassified the entire stands of dead wood or of euphorbia, however. It classified only small portions or pieces of the vegetation. The last shortcoming of CNN was that it occasionally over-classified the *Opuntia*. This occurred when *Opuntia* was growing alongside the natural vegetation such that the programme infrequently classified the natural bush as *Opuntia*. An example of this slight overclassification is evident in figure 6, where it can be seen how the classifier made straight lines rather than identifying the shape of the plant.

3.6. Land cover analysis

Comparisons were made between the manual digitisation and the CNN digitisation of six subplots. In table 2, the results from the coverage tests, whereby the square metres covered by *Opuntia* in the six subplots, as well as the average percentage of the land that the *Opuntia*

covered, are shown. Compared with the manual classification, the programme calculated a greater square metre area of *Opuntia* for five of the subplots and less of an area for the other subplot. However, the average difference in square metres between the CNN predictions and the manual classification was 0.6064 m^2, which works out to be a four percent (4%) difference between the two (in favour of the CNN classifications). This is an acceptable level of variation and shows that the CNN could be used to accurately determine the amount of *Opuntia* within the study area. Land cover studies are very important, and these results show that a CNN methodology could be used in future remote sensing studies to look at the land cover coverage of objects, especially the finer detailed items, thanks to the high-resolution imagery used.

Table 2: Comparison of manual and CNN digitising of land cover in the same areas.

Subplots	Size of area (m^2)	*Opuntia* from manual digitising		*Opuntia* from CNN digitising		Difference	
		Area (m^2)	Percent coverage	Area (m^2)	Percent coverage	Area (m^2)	Percentage
1	100	9,3500	9,3%	10,4294	10,4%	1,0794	11,5%
13	100	22,9232	22,9%	23,9875	24,0%	1,0643	4,6%
21	100	26,5376	26,5%	27,3776	27,4%	0,8400	3,2%
27	100	15,7569	15,8%	15,1264	15,1%	-0,6305	-4,0%
34	100	9,0982	9,1%	9,9034	9,9%	0,8052	8,8%
43	100	8,1749	8,2%	8,6549	8,7%	0,4800	5,9%
Average	100	15,3068	15,3%	15,9132	15,9%	0,6064	4,0%

3.7. Land cover change

The land cover change assessment was done to test the potential of the CNN programme in detecting land cover change over time. Table 3 shows the results of the clearing of Subplot 28. The table indicates the area in square metres covered by *Opuntia* before and after clearing and the percentage cover. According to the coverage results, only 62.082 m^2 of the area covered by *Opuntia* was cleared, which worked out to be 11.1% of the *Opuntia* in the subplot at the canopy level. A total percentage coverage change of 1.72% of the total plot coverage was determined. While this result was very low, it was expected, as only a small proportion of the area of the total plot coverage occupied by the plant had been cleared from the research site Nevertheless, the results prove that the CNN was able to efficiently detect a difference in the before and after images.

Table 3: CNN classification of *Opuntia* before and after clearing.

	Area occupied by *Opuntia* (m^2)	**Percent coverage**
Before Clearing	555,784	15,44%
After Clearing	493,702	13,71%

In figure 6, a visual representation of the clearing can be seen. The top left image is an image from before clearing while the bottom left is the same area as the top left image, except that it shows the area after clearing. As stated above, while nearly all the *Opuntia* was cleared, some plants were left behind, as indicated in the lower portion of the bottom left image. The two images on the right are the CNN classifications of the *Opuntia* for the respective images on the left. It can be seen how the entire stand of *Opuntia* was classified in the upper image, but in the lower image, only the remaining *Opuntia* stand in the lower section was classified.

The CNN did not become confused by the two images and could still accurately classify the *Opuntia*, thereby indicating the change between the images taken before and after the clearing of the plant. This confirms that the CNN programme can be used not only to determine the spatial distribution of a plant but also to monitor the change in its distribution over time, thereby allowing researchers and landowners to determine whether an IPS is spreading or declining and how much of the plant species has been or needs to be removed. Again, this shows how the CNN methodology has many applications and much potential in the remote sensing field.

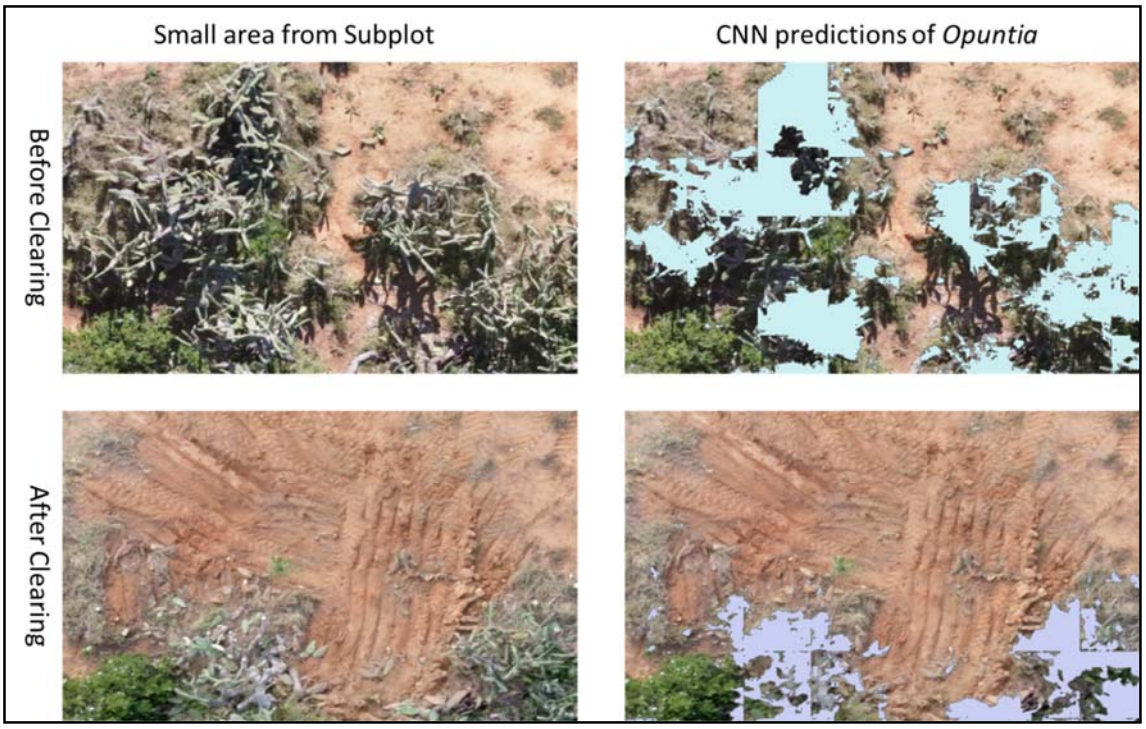

Figure 6: Visual representation of CNN classification before and after clearing.

3.8. CNN overall performance

The results of the classification revealed that the CNN applied in this study performed at a high level. However, even though CNN did perform beyond expectations, the *Opuntia* has a

very distinct shape compared to the surrounding indigenous Kowie thicket species. This would have undeniably been a contributor to the highly accurate results for the CNN's classifications.

The CNN's ability to process data and classify plants accurately depends on how well it has been trained and the quantity of the input data. However, the specific CNN programme used in the current study was the U-Net CNN, which is designed to deal with images as the input and is able, therefore, to operate with less input data than other neural networks. Another factor that would affect the capabilities of a neural network is the quality of the input data, which refers not only to how well the target class has been digitised, but also to the variety in the digitising method. Moreover, digitising needs to be precise. If features are not part of the target class or if the drawn polygons do not represent the true shape of the target, then the CNN would not know what shapes to look for.

The digitised training data need to include a variety of scenarios focusing on the target class. For example, there needs to be training data in respect of the different light and climatic conditions (e.g., for sunny and cloudy days); different elevations (e.g. for slightly different pixel sizes according to changes in altitude); different plant properties/conditions (e.g., for small, large, and even dying plants); and different geomorphological areas (e.g., for flat land, hills, or slopes). The greater the variety in the training data (while still maintaining quality and quantity), the greater the number of features that the CNN can learn about the target class, and the more efficient the performance of the neural network.

4. Significance of Research

The investigations around this research study relied on the application of a pre-existing tool, namely, CNN machine learning in the field of remote sensing to distinguish between the various types of land use and to identify objects. Moreover, the study looked at applying a new methodology to further extend its capabilities of land classification by looking at the identification of small-scale features. Most other classifiers require large multispectral or hyperspectral datasets to accurately detect and identify objects; however, in their classification of objects, CNNs are powerful enough to require only RGB imagery at a high resolution level. This allows new types of land classifications to be performed that contribute to and extend the land classification field of remote sensing and GIS. This study shows the capabilities of some new techniques and how they could further extend the remote sensing field.

5. Conclusion

The aim of this research study was to determine the potential of using UAVs and neural networks to classify invasive plants within the Albany thicket biome, as well as to assess the beneficial aspects of using this method. Neural networks have only recently been used to

classify plant species from high-resolution aerial imagery, and this is the first time that a study of this capacity has been performed in respect of the Albany thicket in South Africa. The results from the study have been promising, in that high accuracy levels were achieved when the classification results of this study were aligned with those of other studies that had used similar methodologies. This shows that the method of classification has the potential for further exploration in the Earth observation field. The results from the land clearing analysis revealed that the neural network is able to detect a change between two images of the same area where land clearing has taken place, thereby proving that the CNN and high spatial resolution levels in the imagery could help landowners and researchers to monitor the growth or decline of plant species over time. The study achieved the research aim, namely, to determine the efficacy of CNNs in the classification of the high spatial resolution aerial imagery of a select collection of Albany thicket plots. Moreover, the study determined the reliability, accuracy, suitability, and capabilities of using CNNs and commercially available UAVs for this task.

6. Acknowledgments

My thanks are due to the co-authors for their assistance in preparing this article, in helping with the collection of the data, and for all the support that they provided. A special thank you is directed at Dr A Potts from the NMU Botany Department for his assistance with problem-solving errors in the code used for the analysis of the data and for allowing us the use of his research computer for the processing of all the data used in this study. We thank the South African National Space Agency (SANSA) for providing the funding for this study. Our thanks are also due to the owner of the Sibuya Game Reserve for allowing us access to the land and to conduct our data collection trips.

7. References

Adam, E., Mutanga, O., & Rugege, D. (2010). Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: A review. *Wetlands Ecology and Management*, *18*(3), 281–296. https://doi.org/10.1007/s11273-009-9169-z

Everitt, J. H., Escobar, M. A., Alaniz, M. A., & Davis, M. R. (1991). Light reflectance characteristics of pricklypear. *Journal Of Range Management*, *44*(6), 587–592.

Fischer, C., & Kakoulli, I. (2006). Multispectral and hyperspectral imaging technologies in conservation: current research and potential applications. *Studies in Conservation*, *51*(sup1), 3–16. https://doi.org/10.1179/sic.2006.51.supplement-1.3

Hoare, D. B., Mucina, L., Rutherford, M. C., Volk, J. H. J., Euston-Brown, D. I. W., Palmer, A. R., Powrie, L. W., Lechmere-Oertel, R. G., Proches, S. M., Dold, A. P., & Ward, R. A. (2006). Albany Thicket Biome. *The Vegetation of South Africa, Lesotho and Swaziland*, *Strelitzia 19*, 541–567.

Hoeser, T., Bachofer, F., & Kuenzer, C. (2020). Object detection and image segmentation with deep learning on earth observation data: A review-part II: Applications. *Remote Sensing*, *12*(18). https://doi.org/10.3390/RS12183053

Hou, B., Li, J., Zhang, X., Wang, S., & Jiao, L. (2019). Object Detection and Trcacking Based on Convolutional Neural Networks for High-Resolution Optical Remote Sensing Video. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 5433–5436. https://doi.org/10.1109/IGARSS.2019.8898173

Ji, Y., Zhang, S., & Xiao, W. (2019). Electrocardiogram classification based on faster regions with convolutional neural network. *Sensors (Switzerland)*, *19*(11). https://doi.org/10.3390/s19112558

Kattenborn, T., Eichel, J., & Fassnacht, F. E. (2019). Convolutional Neural Networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery. *Scientific Reports*, *9*(1), 1–10. https://doi.org/10.1038/s41598-019-53797-9

Kattenborn, T., Eichel, J., Wiser, S., Burrows, L., Fassnacht, F. E., & Schmidtlein, S. (2020). Convolutional Neural Networks accurately predict cover fractions of plant species and communities in Unmanned Aerial Vehicle imagery. *Remote Sensing in Ecology and Conservation*, *6*(4), 472–486. https://doi.org/10.1002/rse2.146

Kembuan, O., Caren Rorimpandey, G., & Milian Tompunu Tengker, S. (2020). Convolutional Neural Network (CNN) for Image Classification of Indonesia Sign Language Using Tensorflow. *2020 2nd International Conference on Cybernetics and Intelligent System, ICORIS 2020*, *26*. https://doi.org/10.1109/ICORIS50180.2020.9320810

Kohli, R., Dogra, K., Batish, D., & Singh, H. P. (2009). Impact of Invasive Plants on the Structure and Composition of Natural Vegetation of Northwestern Indian Himalayas1. *Weed Technology*, *18*, 1296–1300. https://doi.org/10.1614/0890-037X(2004)018[1296:IOIPOT]2.0.CO;2

Kota, Z., & Shackleton, S. E. (2015). Harnessing local ecological knowledge to identify priority plant species for restoration of the Albany Thicket, South Africa. *Forests, Trees and Livelihoods*, *24*(1), 43–58. https://doi.org/10.1080/14728028.2014.943305

Lee, H., & Song, J. (2019). Introduction to convolutional neural network using Keras; An understanding from a statistician. *Communications for Statistical Applications and Methods*, *26*(6), 591–610. https://doi.org/10.29220/CSAM.2019.26.6.591

Liu, H., Bruning, B., Garnett, T., & Berger, B. (2020). Hyperspectral imaging and 3D technologies for plant phenotyping: From satellite to close-range sensing. *Computers and Electronics in Agriculture*, *175*(June), 105621. https://doi.org/10.1016/j.compag.2020.105621

Lu, R., & Park, B. (2008). Hyperspectral and multispectral imaging for food quality and safety. *Sensing and Instrumentation for Food Quality and Safety*, *2*(3), 131–132. https://doi.org/10.1007/s11694-008-9060-2

Marais, C., Cowling, R., Powell, M., & Mills, A. (2009). Establishing the platform for a carbon sequestration market in South Africa: The Working for Woodlands Subtropical Thicket Restoration Programme. *XIII World Forestry Congress, Buenos Aires …*, *October*, 18–23. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Establishing+the+platform+for+a+carbon+sequestration+market+in+South+Africa+:+The+Working+for+Woodlands+Subtropical+Thicket+Restoration+Programme#0

O'Shea, K., & Nash, R. (2015). An Introduction to Convolutional Neural Networks. *Cornell University*, 1–11. http://arxiv.org/abs/1511.08458

Oldeland, J., Revermann, R., Luther-Mosebach, J., Buttschardt, T., & Lehmann, J. R. K. (2021). New tools for old problems — comparing drone- and field-based assessments of a problematic plant species. *Environmental Monitoring and Assessment*, *193*(2). https://doi.org/10.1007/s10661-021-08852-2

Rwanga, S., & Ndambuki, J. (2017). Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS. *International Journal of Geosciences*, *8*, 611–622.

Schiefer, F., Kattenborn, T., Frick, A., Frey, J., Schall, P., Koch, B., & Schmidtlein, S. (2020). Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, *170* (November), 205–215. https://doi.org/10.1016/j.isprsjprs.2020.10.015

Zhang, N., Yang, G., Pan, Y., Yang, X., Chen, L., & Zhao, C. (2020). A review of advanced technologies and development for hyperspectral-based plant disease detection in the past three decades. *Remote Sensing*, *12*(19), 1–34. https://doi.org/10.3390/rs12193188