



**SACAIR2021**

Southern African Conference  
for Artificial Intelligence Research

AI for Science, Technology and Society

Online (Virtual)

6 - 10 December 2021

# SACAIR 2021

## Proceedings of the Second Southern African Conference for Artificial Intelligence Research

ISBN: 978-0-620-94410-6

### Editors:

***Edgar Jembere,  
Aurona Gerber,  
Anban Pillay, &  
Serestina Viriri***

**SACAIR**  
Southern African Conference for Artificial Intelligence Research



## Deep Learning of Drone Images for Soil Erosion Detection Using Geological and Topographic Landcover Features

Dollman G.J.<sup>1[0000-0003-1073-5150]</sup>, Kotzé E<sup>1[000-0002-5572-4319]</sup>, Heckert A.B.<sup>2[000-0001-9240-6267]</sup>, Choiniere J<sup>3[000-0002-1008-0687]</sup>

<sup>1</sup> University of the Free State, 205 Nelson Mandela Drive, Bloemfontein 9301, RSA

<sup>2</sup> Appalachian State University, 572 Rivers St, Boone, NC 28608, USA

<sup>3</sup> University of the Witwatersrand, 1 Jan Smuts Ave, Braamfontien, Johannesburg, 2000, RSA  
dollmangj@ufs.ac.za

**Abstract.** Soil erosion is a major threat to arable land and wildlife habitat, particularly in southern Africa, where a semiarid climate is paired with intensive stock grazing. The South African term “donga” is used for erosional features in semiarid areas that manifest as dry gullies formed by the action of ephemeral running water. These dongas are a major source of habitat degradation, but also represent areas that actively expose fossils, the primary source of information about Earth’s deep-time history. Classifying erosional features like dongas is a major challenge for remote sensing, with spectral classifiers being insufficiently accurate enough. However, automated means of classifying these features could greatly improve the efficiency of conservation efforts as well as palaeontological inquiry. To this end, we developed a spectral and topographic based convolutional neural network (CNN) deep learning model to classify land cover in parts of the Eastern Cape and Free State of South Africa. Aerial photography was used to create an orthomosaic drone map which was fused with a Digital Elevation Map (DEM) and then classified into five classes: donga, dried mudflats, sandstone, grass vegetation and shrub vegetation. The multi-layer level CNN model was trained with a dataset of 20320 patches with each class having approximately +4000 patches for each class. The model was hyper tuned on the attributes dropout rate, epoch size, batch size, momentum and learning rates and reported an average accuracy of 0.910 +0.108 and a F1 score of 0.910 + 0.111. These results are within range of current state-of-the-art land classifiers and this resulting land classifier model will make an excellent tool for determining erosional areas within a drone-based image map.

**Keywords:** Deep learning • soil erosion • landscape classification • convolutional neural network • remote sensing • fused data

### 1 Introduction

Soil erosion is one of the major causes of land degradation particularly in arid and semi-arid areas [1]. Gully erosion, known in southern Africa as “donga” formation is the most severe type of soil erosion within arid and semiarid landscapes [2]. The

ability to detect dongas accurately, remotely, and automatically would be beneficial for determining high risk areas in the arid and semi-arid southern African landscape, as well as determining priority areas for palaeontological fieldwork. A land classification approach is needed to support a predictive model that is being developed to locate prospectable sites for fossils within a meter square patch using drone map imagery.

Within the last decade, the use of drones to collect image data from remote areas has become an invaluable research tool [3, 4]. Drone surveys have many applications including geographical information systems (GIS), agricultural mapping, urban planning, land use and environmental monitoring [5–7]. Within supervised machine learning classification tasks, the high-resolution images available from drone imagery are ideal for training datasets [8]. To improve the accuracy of land cover classification, a combination of multi-source or multimodal data is recommended [9, 10].

Based on previous research success [9, 11–13], this paper considers utilising feature fusion for the land cover classification task with the drone image survey data. Considering the needs of the land classifier, that is the ability to identify high erosion areas like dongas, the RGB map data was fused with a digital elevation map (DEM), with a map representing the slope of the area and finally a map representing the aspect of the survey area. Fused datasets are state-of-the-art approaches to urban land classification tasks. Our study uses a fused dataset with a deep learning model to determine the height that is necessary for identification of dongas in a rural setting. Our study employs a Convolutional Neural Network (CNN), which is a deep learning technique that automatically: learns representation from training data; regulates and shares the weights with respect to the training data; and generalizes, optimizes, and reduces the parameters with the higher ability to discriminate and extract features [14]. Since this is a multi-class classification problem, the labels were classified into five (5) land cover type labels: sandstone, donga, dried mudflat, shrub vegetation and grass vegetation. This manuscript discusses the related works and motivation for the study within section 2, followed by the materials and methods section 3 where the study area is outlined (3.1), the dataset is discussed in detail (3.2), limitations for this study are outlined (3.3) and the deep learning model used is explained (3.4). Section 4 reports the results, Section 5 discusses the results. An error analysis is performed within section 6 and conclusions reached are outlined within Section 7.

## 2 Related Work

### 2.1 Machine and Deep Learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. machine learning approaches are generally categorised as utilising either unsupervised, semi-supervised or semi-supervised learning [15]. Every

instance within a dataset is represented by a feature that can be continuous, categorical, or binary.

Supervised learning learns a dataset with many features, and each example is associated with a label, whereas in unsupervised learning the instances are unlabelled [16]. Labelling data refers to the process of adding tags to the data, such as describing a blue pixel as a river from a satellite map. With semi-supervised learning the algorithm has limited labelled data and, from this, must infer patterns for the unlabelled data.

There are many approaches available within machine learning that encompass unsupervised, supervised, and semi-supervised learning. Supervised techniques include SVM, RF, SAM, Fuzzy ARTMAP, MD, radial basis function (RBF), decision tree (DT), multilayer perceptron (MLP), naive bayes (NB), maximum likelihood classifier (MLC) and fuzzy logic [17, 18]. The unsupervised classification approaches include Affinity Propagation (AP) cluster algorithm, fuzzy c-means algorithms, K-means algorithm, ISODATA and similar clustering techniques [19, 20].

Deep learning based techniques, a subset of machine learning, allow for features to be learned automatically from training data without human assistance [14, 21]. Deep learning is an application of Artificial Neural Networks (ANNs), using modern hardware allowing for the development, training and use of neural networks. The three major architecture types are CNNs, MLPs and Recurrent Neural Networks (RNN). CNNs have proven to be good classifiers for image classification while RNN appear to be better suited for natural language processing (NLP). MLPs are a basic type of deep neural network and suited to tasks where available processing power is limited.

## 2.2 Land Classification

The need to classify land is an essential function and can be seen in early work such as Rowe [22] and is often used to support urban planning, conservation and many more tasks.. Within the 1980s classifying a landscape required much time and painstaking work by experts. Modern land classifier tasks that apply machine learning algorithms to remotely sensed imagery have attracted a great deal of research [19, 23].

One recent analysis determined that ANN, SVM and RF generally produce better accuracy results in classifying land [24]. This result is supported by Talukdar [25], and narrows down the best approaches to RF or an ANN. Several studies have revealed that LULC classification using low resolution images introduces spectral and spatial limitations that affect accuracy, which supports the argument that high-resolution imagery should result in better accuracy [26].

However, it is worth noting that the classification tasks discussed here are normally based on spectral data only and usually from image sources with a low spatial density.

**Soil Erosion.** Soil erosion is the loss of topsoil (upper layer of soil) through erosion by natural processes or otherwise. Soil erosion is one of the major causes of land degradation particularly in arid and semi-arid areas [1]. The Middle Atlas region within Morocco is vulnerable to soil erosion from human activities and severe

weather. Using satellite images, the Revised Universal Soil Loss Equation (RUSLE) was used within a GIS environment to quantify the soil loss and then map the erosion risk as an indicator of landslides and vulnerable landscapes [27]. RUSLE is limited as an indicator as it only accounts for soil loss through sheet and rill erosion ignoring the effects of gully erosion and dispersive soils. Gullies or dongas are particularly indicative of high erosion rates and in this study we sought to classify them remotely. It appears that the automatic detection of donga's is a recent topic of interest to this research so efforts to classify dongas remotely was investigated. A study was undertaken to produce gully erosion susceptibility maps (GESMs) and quadratic discriminant analysis (QDA) models within the Shahroud Watershed, Semnan Province, Iran[28]. Results showed that under the curve (AUC) values for the LDA and QDA models are 0.875 and 0.862 respectively [28]. An ensemble machine learning model was developed to determine the susceptibility of an area within Iran to forming gullies from satellite imagery. A spatial model was developed to determine gullies using drone data with four ensemble classifiers within the Golestan Province, Iran. Bagging, random subspace (RS) [29] and rotation forests (RF) were the techniques used. Four machine-learning techniques were used for the study, BFTree, bag-BFTree, RS-BFTree and RF-BFTree. An accuracy reported of 0.936 for the BFTree and 0.735, respectively [30]. Arabameri [31] used RS with an evidential belief function (EBF), an LDA and a logistic model tree (LMT). The Cohen's Kappa was reported as 0.676 for the RS-FLSA, 0.702 for the RS-LMT, and 0.748 for RS-NBTree [31]. A similar study by Saha [32] within Eastern India with random forest (RF), Gradient Boosted Regression Tree (GBRT), naive Bayes Tree (NBT) and a Tree Ensemble (TE) method reported accuracy of 0.87 for RF. A deep learning approach was undertaken by Gafurov [33] where an automatic gully detection approach was attempted on the European part of Russia using ultra-high resolution satellite images using a CNN. The network was trained for 1000 epochs and resulted in a F1-score of 0.7. The authors reported that the lack of training data hampered model performance resulting in poor classification performance.

**Fused Data.** Recently, the use of fused data has been investigated for classification tasks which are a combination of multi-source or multimodal data [9,11–13,36,37]. [9] investigated improving the accuracy of land cover classification using CNN via drones using digital surface models (DSMs) and achieved an overall accuracy of 0.97. [36] demonstrated an approach to determine the nitrogen content in citrus-trees using the RF algorithm and spectral vegetation indices reporting a R<sup>2</sup> of 0.90. Using a drone imagery based orthomosaic model fused with a digital surface model (DSM) urban landscapes were classified using three machine learning (ML) models [10] and RF was the best scorer with a kappa coefficient of 0.938. In each case, utilising fused datasets achieved better performance from the machine and deep learning approaches.

For this study RGB drone orthomosaic map data has been fused with a digital elevation map (DEM), a map representing the slope of the area and a map representing the aspect of the drone survey area to improve model performance.

### 2.3 Convolutional Neural Networks (CNNs)

CNN is a highly effective technique used commonly with computer vision applications [14]. A CNN consists of three neural layers: convolutional layers, pooling layers, and fully connected layers. Various CNN methods use different combinations of these layers to achieve distinct results [14]. The convolutional layer creates a two-dimensional representation of the image, known as an activation map. After each convolutional layer an activation function is called. Either Rectified Linear Unit (*ReLU*) or *Tanh* can be utilised as an activation function depending on the context. This network type utilises a mathematical operation called convolution, which is a specialised kind of linear operation. Convolution is used in place of a general matrix multiplication. Pooling layers take the output from a convolutional layer and derive a summary statistic to reduce the computational requirements. Fully connected layers help map the representation between inputs and outputs. There are several techniques that can be applied to a layer to help prevent overfitting via regularisation, underfitting and explosion of gradients. Information from an image can be extracted at different levels, including pixel-level, object-level, and patch-level. Additionally, other structures can be used to improve CNNs generalization such as nonlinear activations, regularization (dropout and batch normalization), stochastic pooling, and so on.

Dropout involves the algorithm randomly omitting a percentage decided by the model designer of all feature detectors to prevent co-adaptations in the training data and enhance the generalisation ability [34–40].

Batch normalization as regularization is applied to each feature map (the output of previous activation function) to normalize the map within a batch to have zero mean and unit standard deviation [41]. Batch normalization uses Stochastic Gradient Descent (SGD) to improve training and minimizing loss before every convolutional layer [42].

## 3 Materials and Methods

### 3.1 Study Area

The location of the study is geographically positioned near an area called Qhemegha, Joe Qgabi District, Eastern Cape, South Africa. The site is known as the Qhemegha Bone Bed and is a rural area consisting of shales, red or purple mudstones and red to white sandstone. The area falls within the Lower Elliot Formation [43] and arguably contains fossils from the Late Triassic time period [44]. The RGB images were acquired July 31st, 2019 using a DJI Mavic Pro 2 drone (<https://www.dji.com/mavic-2>) from an altitude of 80 meters above ground surface. The ground resolution of the resulting orthomosaic map is 2.86 cm/pix with a resolution of 52634 x 32383. The DEM extracted from the model has a ground resolution of 1.15m/pix and a resolution of 1858 x 1210. The DEM and orthomosaic model were calculated from within Agisoft PhotoScan Professional (<https://www.agisoft.com/>) version 1.4.4. Data pre-processing was accomplished via scripting within Python using the Rasterio

(<https://rasterio.readthedocs.io/en/latest/>), and Keras [45] packages. For general GIS work QGIS Desktop 3.10.3 was utilised (<https://www.qgis.org/en/site/>).

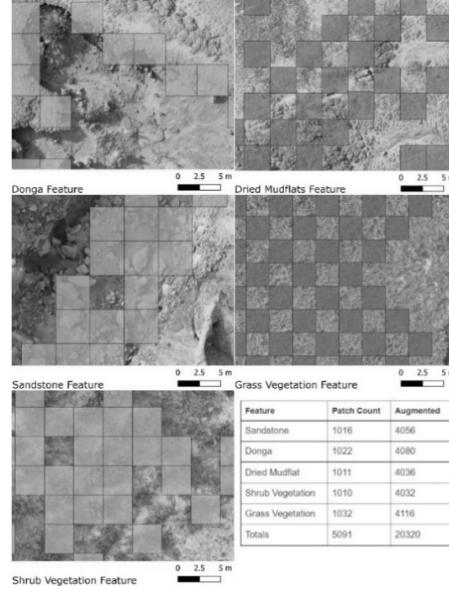
### 3.2 Dataset

For this study, a total of 5080 ground truth (GT) patches were sampled from the drone orthomosaic map. Each patch is an 83x83 pix image that matches one of the land covers classes. The five land cover classes considered for this work were donga (steep-sided gully created by soil erosion), dried mudflat, grass vegetation, shrub vegetation and sandstone (Fig. 1). The patch size of 83 x 83 was chosen as it matches a area of one meter squared. This is the size of the area of interest for the fossil prospecting model that this LULC was intended to be utilised for.

Data augmentation was applied on the dataset because the available dataset is limited and insufficient for training a CNN. The suitability of a transformation depends on the dataset being utilised. For this study each patch was rotated 90 degrees allowing for four times the amount of data to be used for training. As each patch consisted of multiple layers including slope and height data, colour augmentation was not utilised. This resulted in a balanced dataset of 20320 patches which was available for training, validation, and testing (Table 1).

**Table 1.** Complete breakdown of dataset used for training, validation and testing

| Feature          | Patch Count | Augmented    | Training     | Validation  | Test        |
|------------------|-------------|--------------|--------------|-------------|-------------|
| Sandstone        | 1016        | 4064         | 3252         | 406         | 406         |
| Donga            | 1022        | 4088         | 3266         | 409         | 409         |
| Dried Mudflat    | 1011        | 4044         | 3236         | 404         | 404         |
| Shrub Vegetation | 1010        | 4040         | 3236         | 404         | 404         |
| Grass Vegetation | 1021        | 4084         | 3266         | 409         | 409         |
| <b>Totals</b>    | <b>5080</b> | <b>20320</b> | <b>16256</b> | <b>2032</b> | <b>2032</b> |



**Fig. 1.** An example of the patches taken and a feature count.

**Training, Validation, and Testing Dataset.** From the dataset 80% was allocated for training, 10% for testing and 10% for the validation dataset. The training dataset is used to train the CNN to make predictions. The validation dataset is utilised during the training process to validate the results. The test dataset is unseen by the model and is used to give a more accurate indication of the model's performance. Within supervised learning, the training data set contains labelled instances known to the classifier. During each training iteration (known as an epoch) the model classifies all instances of the input based on the initial weights of the network. After the epoch, a loss value is produced which represents the error between the CNN's predictions and the actual class labels. This loss value is accumulated across the batch samples with the weights updated after each batch. The final goal of the training is to produce a model that minimizes the loss error.

Hyperparameter tuning is the process where settings are manipulated to ensure the best possible output [46]. Typical hyperparameters are learning rate, drop out settings, number of epochs, number of hidden layers, activation functions and batch size. The base settings are 50 epochs, 0.25 for the drop out setting, a batch size of 25 and a learning rate of 0.00001, a momentum of 0.9, an optimiser using mini-batch SGD and the loss function categorical crossentropy. The hyperparameter tuning was performed in batches according to the settings shown within Table 2.

A final configuration of a batch size of 25, 50 epochs, a learning rate of 0.001, a momentum of 0.6, the two hidden layers dropouts with 0.25 then 0.5 and SGD as the optimizer was chosen. During training the validation loss and validation accuracy are closely monitored to determine the effect of hyperparameter tuning.

**Table 2.** All settings trained during hyperparameter tuning, best settings are denoted by \*.

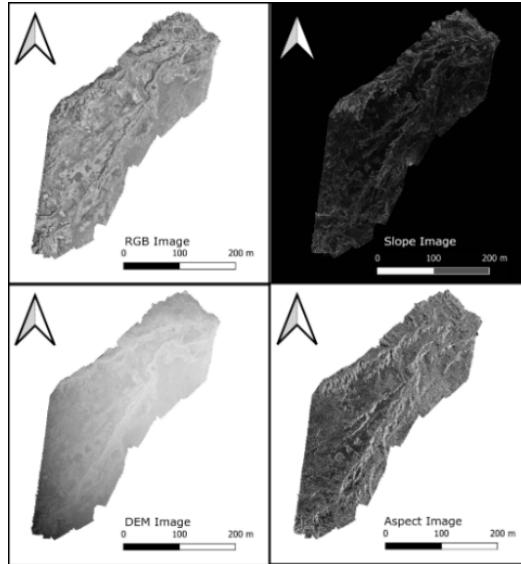
|              |        |       |      |     |      |     |     |
|--------------|--------|-------|------|-----|------|-----|-----|
| Batch Size   | 12     | 25*   | 35   | 40  | 60   | 80  | 100 |
| Epochs       | 10     | 50*   | 100  | 150 | 200  |     |     |
| Dropout Rate | 0.1    | 0.25* | 0.3  | 0.4 | 0.5* | 0.6 | 0.7 |
| Learn Rate   | 0.001* | 0.01  | 0.1  | 0.2 | 0.3  |     |     |
| Momentum     | 0.2    | 0.4   | 0.6* | 0.8 | 0.9  |     |     |

**Image Pre-Processing.** Image pre-processing is used to prepare the image data for the CNN training process. The orthomosaic model was created using Agisoft PhotoScan Professional and the DEM was generated from this resulting drone orthomosaic map. Within QGIS, the orthomosaic map and DEM were used to generate the aspect map and the slope map using the raster-based functions GDAL Aspect and GDAL Slope (Fig. 2). These disparate rasters were all projected to CRS EPSG:32635 and then combined using the merge function. This resulted in a single raster containing 6 layers (Table 2) with a dimension of 29954 x 39238.

Patch level processing is often used with CNNs to overcome challenges posted by speckle noise and segmentation optimization problems associated with pixel-level and object-level feature extraction [47]. With patch-level analysis (also known as tile-based analysis) images are divided into a grid of tiles of m x m and then each patch is analysed. To accomplish this in the model presented here, a grid was overlaid using the QGIS Grid function with a size of 1 meter square. Each labelled class is a patch with a size of 83x83 pix that is mapped to one of the land cover classes (Fig. 1). Before each patch is loaded into the CNN for training, the data is normalised. The real valued numeric attributes are rescaled into a range value between 0 and 1 without distorting the value. The ranges of values contained within each layer is different, so normalization is required (Table 3).

**Table 3.** All layers(bands) used within the training, validation, and testing dataset.

| Layers | Description                                |
|--------|--|
| Red    | Spectral value for red (0 - 255)           |
| Green  | Spectral value for green (0 - 255)         |
| Blue   | Spectral value for blue (0 - 255)          |
| DEM    | Height value in meters (1461.53 - 1485.75) |
| Aspect | Aspect value in degrees (0 - 360)          |
| Slope  | Slope value in degrees (0 - 90)            |



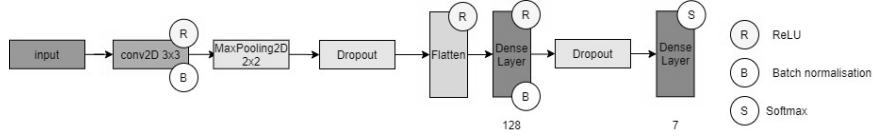
**Fig. 2.** The six layers are viewed from within QGIS, RGB layers combined top left, slope top right, DEM bottom left and aspect bottom right.

### 3.3 Limitations

All the imagery used for training the neural network was taken from July, a winter month. The spectral values during other seasons (especially summer) will differ due to the spectral differences of the patches and could reduce classification accuracy. Additionally, this study is limited to the Lower Elliot Formation within South Africa.

### 3.4 Deep Learning Model

For this study, the CNN network (Fig. 3) consisted of one 2D convolution layer that learned 64 kernels with a filter size of  $3 \times 3$ . A batch normalization layer is then run. The convolution results then went through a rectified linear unit (ReLU) layer before being down sampled via maxpooling with a filter size of  $2 \times 2$ . A dense layer with 128 neurons then followed with a batch normalization and another ReLU activation, followed by a final SoftMax dense layer with seven neurons indicating the number of classes. Note that dropout was implemented to avoid overfitting. A dropout of 0.25 was performed after the maxpooling layer, and another dropout of 0.25 was performed after the dense ReLU layer (Fig. 5). This simplistic single layer model was chosen because of the increased performance compared to CNN models with five convolutional layers and less.

**Fig. 3.** CNN model visualised

## 4 Results

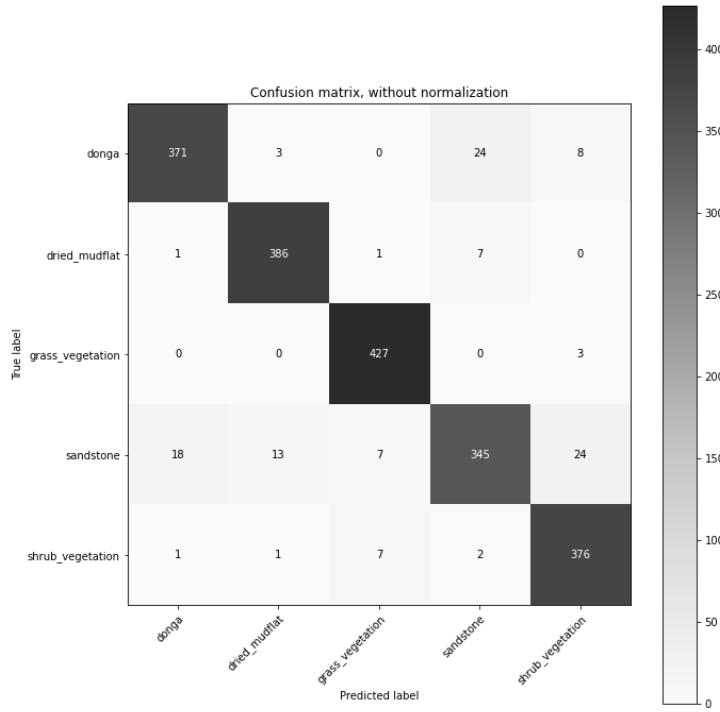
This study aimed to determine the optimal CNN to delineate and extract rural LULC classes that share similar spectral and shape characteristics while allowing for variance in structure, size, and surrounding contrast. Macro-averaging was used on all result evaluation to ensure the reliability of the metrics (Table 3) because the classifier is a multi-class classifier. Macro-averaging computes all metrics independently for each class treating each class equally. Micro-averaging is preferred when class imbalance is suspected, which is not the case for this dataset [50].

**Table 2.** Detailed test results available from the best CNN model.

|                         | <i>Precision</i> | <i>Recall</i> | <i>F1-Score</i> | <i>Patch Count</i> |
|-------------------------|------------------|---------------|-----------------|--------------------|
| <i>donga</i>            | 0.949            | 0.914         | 0.931           | 406                |
| <i>dried_mudflat</i>    | 0.958            | 0.977         | 0.967           | 395                |
| <i>grass_vegetation</i> | 0.966            | 0.993         | 0.979           | 430                |
| <i>sandstone</i>        | 0.913            | 0.848         | 0.879           | 407                |
| <i>shrub_vegetation</i> | 0.915            | 0.918         | 0.942           | 387                |
| <i>Accuracy</i>         | 0.941            | 0.941         | 0.941           | 0.941              |
| <i>Macro Avg</i>        | 0.940            | 0.941         | 0.940           | 2025               |

\*Macro averaging was utilised

Considering the confusion matrix for the model (Fig. 5); the classifier has the greatest trouble classifying between a donga and a sandstone. This is because spectrally, the donga and sandstone are the same. A donga is a formation of eroded sandstone. Without the height and slope data, a CNN cannot effectively differentiate between a donga and sandstone (Fig. 1). For the 42 miss-classifications it is likely that most of the patch was sandstone with a small erosion area, resulting in an incorrect label. However, these areas could be indicative of future erosion. The next major misclassification is between sandstone and dried mudflats, again it is likely the misclassifications are based on similar spectral attributes between the two classes (Fig. 1). The recall for sandstone is only 0.848 (Table 3) which supports the assertion that the spectral values of sandstone are problematic for classification.



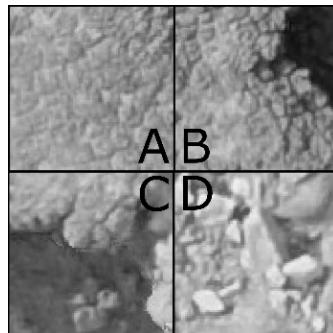
**Fig. 5.** Confusion matrix of best CNN model.

## 5 Discussion

The classification of a landscape with spectrally similar features is a challenge for any classifier. As can be seen from the confusion matrix (Fig. 5) it is evident that there are still challenges between classifying the similar land features. The original attempts of classifying land cover based only on spectral values resulted in poor performance (+- 70% F1-score) and struggled particularly to identify dongas. The proposed multi-layer CNN approach to LULC demonstrates great potential for classifying land of interest to palaeontologists to prospect for fossils. The ability to accurately identify dongas has applications within the impact on land areas lost to erosion. Identifying where heavy erosion has impacts on sustainability and food security. Traditional techniques to document the formation and existence of dongas are resource intensive and have difficulty reaching rural areas. The ability to automatically document the development of erosional features within arid areas using remote sensing platforms like a drone allow for a much better ability to monitor the problem. Finally, a high erosion zone such as a donga are desirable for palaeontologists to locate fossil sites and is a useful tool for virtual prospecting.

## 6 Error Analysis

It is clear from Fig. 5. that the classifier struggles to classify between dongas, mudflats and sandstone. In Fig. 6, the patch marked A on the left is classified as mudflats while the patch B on the top-right has been classified as mudflats but should have been labelled as donga because of the dip in the top right of the right image. However, since such a large proportion of the patch is a mudflat the classifier decided that that was the best fitting label. Within the same figure, between the bottom left patch C and the bottom right patch D, both the patches were labelled as a donga by the classifier where patch D should have been sandstone. Spectrally, patch C and D are very similar. There are several sandstone rocks each with a height difference, the number of large rocks may have resulted in the classifier to label this patch as a donga. The use of pixel-level rather than patch level image features would likely improve the ability to for the classifier to differentiate between the classes. A pixel level classifier could delineate the edge of a donga as the donga formation and the rest of the image as mudflats or sandstone. Another approach would be a larger dataset, more training data should allow the classifier to better determine between these edge cases.



**Fig. 6.** Panel A: correctly labelled mudflats; Panel B: donga incorrectly labelled as mudflats; Patch C: correctly labelled donga; Patch D: sandstone incorrectly labelled donga.

## 7 Conclusions

The aim of this study was to perform a rural LULC classification of land within the Lower Elliot Formation, South Africa that had been surveyed by a low flying drone. The ability of the classifier to determine dongas has a wider application within the field of soil erosion where drone imagery could be analysed and donga's identified automatically and accurately.

Accurate and reliable rural LULC information of areas being considered for palaeontological prospecting is critical to determine whether (or how much of) a particular area is prospectable or not. Drones allow for the acquisition of remote sensed data with a resolution as high as 1cm per pixel in a relatively inexpensive and

flexible approach. However, LULC on that level of detail to this author's knowledge does not exist and needed to be developed. The spectral similarities between the areas of interest and the inability of a spectral based CNN to determine depth (to be able to determine eroded formations) required the fusion of the datasets and then classification using a patch driven approach. The resulting classifier has an average F1-Score of 0.941 +-0.108 and an Accuracy of 0.910 +- 0.109. The applicability of this approach on the proposed dataset showed an excellent fit and is accurate and is reliable enough to be used as a feature in the prospectable fossil area predictive classifier. Future work could be to apply these techniques to hyper detailed satellite imagery and determine if the model's accuracy is maintained in arid areas such as Iran, Morocco, and other parts of South Africa. ResNet50 [50] and VGG19 [51] will be investigated as alternative LULC classifiers for the dataset.

## References

1. Morgan RP (1979) Soil erosion. Longman
2. Aber JS, Marzolff I, Ries J, Aber SE (2019) Small-format aerial photography and UAS imagery: Principles, techniques and geoscience applications. Academic Press
3. Bemis SP, Micklithwaite S, Turner D, et al (2014) Ground-based and UAV-Based photogrammetry: A multi-scale, high-resolution mapping tool for structural geology and paleoseismology. *Journal of Structural Geology* 69:163–178. <https://doi.org/10.1016/j.jsg.2014.10.007>
4. Floreano D, Wood RJ (2015) Science, technology and the future of small autonomous drones. *Nature* 521:460–466. <https://doi.org/10.1038/nature14542>
5. Murtha TM, Broadbent EN, Golden C, et al (2019) Drone-Mounted Lidar Survey of Maya Settlement and Landscape. *Latin American Antiquity* 30:630–636. <https://doi.org/10.1017/laq.2019.51>
6. Orengo HA, Garcia-Molsosa A (2019) A brave new world for archaeological survey: Automated machine learning-based potsherd detection using high-resolution drone imagery. *J Archaeol Sci* 112:105013. <https://doi.org/10.1016/j.jas.2019.105013>
7. Tezza D, Andujar M (2019) The State-of-the-Art of Human–Drone Interaction: A Survey. *IEEE Access* 7:167438–167454. <https://doi.org/10.1109/ACCESS.2019.2953900>
8. Ma L, Li M, Ma X, et al (2017) A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 130:277–293. <https://doi.org/10.1016/j.isprsjprs.2017.06.001>
9. Al-Najjar HAH, Kalantar B, Pradhan B, et al (2019) Land Cover Classification from fused DSM and UAV Images Using Convolutional Neural Networks. *Remote Sens (Basel)* 11:1461. <https://doi.org/10.3390/rs11121461>
10. Gibril MBA, Kalantar B, Al-Ruzouq R, et al (2020) Mapping Heterogeneous Urban Landscapes from the Fusion of Digital Surface Model and Unmanned Aerial Vehicle-Based Images Using Adaptive Multiscale Image Segmentation and Classification. *Remote Sens (Basel)* 12:1081. <https://doi.org/10.3390/rs12071081>
11. Irwin K, Beaulne D, Braun A, Fotopoulos G (2017) Fusion of SAR, optical imagery and airborne lidar for surface water detection. *Remote Sens (Basel)* 9:890. <https://doi.org/10.3390/rs9090890>

12. Hartfield KA, Landau KI, Leeuwen WJD van (2011) Fusion of high resolution aerial multispectral and lidar data: land cover in the context of urban mosquito habitat. *Remote Sens (Basel)* 3:2364–2383. <https://doi.org/10.3390/rs3112364>
13. Liu T, Abd-Elrahman A, Zare A, et al (2018) A fully learnable context-driven object-based model for mapping land cover using multi-view data from unmanned aircraft systems. *Remote Sensing of Environment* 216:328–344. <https://doi.org/10.1016/j.rse.2018.06.031>
14. Guo Y, Liu Y, Oerlemans A, et al (2016) Deep learning for visual understanding: A review. *Neurocomputing* 187:27–48. <https://doi.org/10.1016/j.neucom.2015.09.116>
15. Raschka S (2015) Python machine learning
16. Kotsiantis SB, Zaharakis ID, Pintelas PE (2007) Supervised Machine Learning : A Review of Classification Techniques General Issues of Supervised Learning Algorithms. [books.google.com](http://books.google.com) 31:501–520
17. Ma L, Liu Y, Zhang X, et al (2019) Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing* 152:166–177. <https://doi.org/10.1016/j.isprsjprs.2019.04.015>
18. Shih H, Stow DA, Tsai YH (2019) Guidance on and comparison of machine learning classifiers for Landsat-based land cover and land use mapping. *Int J Remote Sens* 40:1248–1274. <https://doi.org/10.1080/01431161.2018.1524179>
19. Maxwell AE, Warner TA, Fang F (2018) Implementation of machine-learning classification in remote sensing: an applied review. *Int J Remote Sens* 39:2784–2817. <https://doi.org/10.1080/01431161.2018.1433343>
20. Camps-Valls G, Benediktsson JA, Bruzzone L, Chanussot J (2011) Introduction to the issue on advances in remote sensing image processing. *IEEE J Sel Top Signal Process* 5:365–369. <https://doi.org/10.1109/JSTSP.2011.2142490>
21. Hossain MDZ, Sohel F, Shiratuddin MF, Laga H (2019) A comprehensive survey of deep learning for image captioning. *ACM Comput Surv* 51:1–36. <https://doi.org/10.1145/3295748>
22. Rowe JS, Sheard JW (1981) Ecological land classification: A survey approach. *Environ Manage* 5:451–464. <https://doi.org/10.1007/BF01866822>
23. Adam E, Mutanga O, Odindi J, Abdel-Rahman EM (2014) Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. *Int J Remote Sens* 35:3440–3458. <https://doi.org/10.1080/01431161.2014.903435>
24. Carranza-García M, García-Gutiérrez J, Riquelme J (2019) A framework for evaluating land use and land cover classification using convolutional neural networks. *Remote Sens (Basel)* 11:274. <https://doi.org/10.3390/rs11030274>
25. Talukdar S, Singha P, Mahato S, et al (2020) Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. *Remote Sens (Basel)* 12:1135. <https://doi.org/10.3390/rs12071135>
26. Manandhar R, Odeh I, Ancev T (2009) Improving the Accuracy of Land Use and Land Cover Classification of Landsat Data Using Post-Classification Enhancement. *Remote Sens (Basel)* 1:330–344. <https://doi.org/10.3390/rs1030330>
27. Senanayake S, Pradhan B, Huete A, Brennan J (2020) Assessing Soil Erosion Hazards Using Land-Use Change and Landslide Frequency Ratio Method: A Case Study of Sabaragamuwa Province, Sri Lanka. *Remote Sens (Basel)* 12:1483. <https://doi.org/10.3390/rs12091483>
28. Arabameri A, Pourghasemi HR (2019) Spatial modeling of gully erosion using linear and quadratic discriminant analyses in GIS and R. In: Spatial modeling in GIS and R for earth and environmental sciences. Elsevier, pp 299–321

29. Ho TK (1998) The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence* 20:832–844
30. Hosseinalizadeh M, Kariminejad N, Chen W, et al (2019) Spatial modelling of gully headcuts using UAV data and four best-first decision classifier ensembles (BFTree, Bag-BFTree, RS-BFTree, and RF-BFTree). *Geomorphology* 329:184–193. <https://doi.org/10.1016/j.geomorph.2019.01.006>
31. Arabameri A, Chen W, Lombardo L, et al (2020) Hybrid computational intelligence models for improvement gully erosion assessment. *Remote Sens (Basel)* 12:140. <https://doi.org/10.3390/rs12010140>
32. Saha S, Roy J, Arabameri A, et al (2020) Machine Learning-Based Gully Erosion Susceptibility Mapping: A Case Study of Eastern India. *Sensors* 20:. <https://doi.org/10.3390/s20051313>
33. Gafurov AM, Yermolayev OP (2020) Automatic gully detection: neural networks and computer vision. *Remote Sens (Basel)* 12:1743. <https://doi.org/10.3390/rs12111743>
34. Ba J, Frey B (2013) Adaptive dropout for training deep neural networks. In: Burges CJC, Bottou L, Welling M, et al (eds) *Advances in Neural Information Processing Systems 26*. Curran Associates, Inc., pp 3084–3092
35. Baldi P, Sadowski PJ (2013) Understanding Dropout. In: Burges CJC, Bottou L, Welling M, et al (eds) *Advances in Neural Information Processing Systems 26*. Curran Associates, Inc., pp 2814–2822
36. Hinton GE, Srivastava N, Krizhevsky A, et al (2012) Improving neural networks by preventing co-adaptation of feature detectors. eprint arXiv:12070580
37. McAllester D (2013) A PAC-Bayesian Tutorial with A Dropout Bound. arXiv:13072118 [cs]
38. Wager S, Wang S, Liang PS (2013) Dropout Training as Adaptive Regularization. In: Burges CJC, Bottou L, Welling M, et al (eds) *Advances in Neural Information Processing Systems 26*. Curran Associates, Inc., pp 351–359
39. Wang S, Manning C (2013) Fast dropout training. In: international conference on machine learning. pp 118–126
40. Warde-Farley D, Goodfellow IJ, Courville A, Bengio Y (2014) An empirical analysis of dropout in piecewise linear networks. arXiv:13126197 [cs, stat]
41. Marcos D, Volpi M, Kellenberger B, Tuia D (2018) Land cover mapping at very high resolution with rotation equivariant CNNs: Towards small yet accurate models. *ISPRS Journal of Photogrammetry and Remote Sensing*. <https://doi.org/10.1016/j.isprsjprs.2018.01.021>
42. Ioffe S, Szegedy C (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. International conference on machine learning 448
43. Bordy EM, Eriksson P (2015) Lithostratigraphy of the elliot formation (karoo supergroup), south africa. *South African Journal of Geology* 118:311–316. <https://doi.org/10.2113/gssaajg.118.3.311>
44. Bordy EM, Abrahams M, Sharman GR, et al (2020) A chronostratigraphic framework for the upper Stormberg Group: Implications for the Triassic-Jurassic boundary in southern Africa. *Earth-Science Reviews* 203:
45. Chollet F (2018) Keras: The python deep learning library. *Astrophysics Source Code Library*
46. Géron A (2019) *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media

47. Sameen MI, Pradhan B, Aziz OS (2018) Classification of Very High Resolution Aerial Photos Using Spectral-Spatial Convolutional Neural Networks. *Journal of Sensors* 2018:1–12. <https://doi.org/10.1155/2018/7195432>
48. Zerrouki N, Bouchaffra D (2014) Pixel-based or Object-based: Which approach is more appropriate for remote sensing image classification? In: 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, pp 864–869
49. Li W, Guo Q (2013) A new accuracy assessment method for one-class remote sensing classification. *IEEE transactions on geoscience and remote sensing* 52:4621–4632
50. Tran D, Mac H, Tong V, et al (2018) A LSTM based framework for handling multiclass imbalance in DGA botnet detection. *Neurocomputing* 275:2401–2413. <https://doi.org/10.1016/j.neucom.2017.11.018>
51. Xie, S., Girshick, R., Doll'ar, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on. pp. 5987–5995. IEEE (2017)
52. K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015