Womanium AI+Climate Project

Matteo Selden, Pratham Gupta, Islam Mohammed Elgendy, Aratrika Gupta

August 9, 2024

Abstract

One of the most serious environmental problems is desertification, affecting millions of people globally. It is the procedure for land degradation in arid, semi-arid, and dry sub-humid areas. Timely and accurate detection of desertification is hence necessary for effective management and mitigation strategies over the land. In this paper, we investigated an approach for detecting desertification by using satellite images and advanced machine learning techniques which harness the use of image classification. We have presented a deep learning model based on convolutional neural networks for the automatic identification of patterns and features indicative of desertification processes, such as vegetation loss, soil degradation, and changes in land cover. In this respect, our study is on a holistic dataset of high-resolution satellite images from multiple sources— different geographical areas subject to desertification. The results have shown significant improvements with respect to our model which has been trained for 90% accuracy, hence making it an effective and accessible solution for the monitoring of desertification around the globe.

Keywords: desertification; image classification; convolutional neural networks (CNN)

1 Introduction

It is among the most critical environmental problems, posing a severe threat to ecosystems, biodiversity, and human livelihoods, more specifically in arid and semi-arid regions of the world. Desertification may be defined as a process whereby productive agricultural land is rendered unusable desert. It encapsulates land degradation, vegetation loss, and soil fertility loss and affects approximately 25% of the Earth's terrestrial land surface, with over a billion people affected in various parts of the world. This essentially implies that desertification results from various factors that include climatic change, unsupportable land use practices, deforestation, and overgrazing. In view of this, strategies to combat desertification need to be based on a better understanding of the extent and process of desertification, as well as the implementation of effective monitoring and mitigation strategies.

Between 2015 and 2019, at least 100 million hectares of healthy and productive land were degraded every year, affecting food and water security globally. The loss is equivalent to twice the size of Greenland, impacting the lives of 1.3 billion people, who are estimated to be directly exposed to land degradation. [Source: UNCCD]

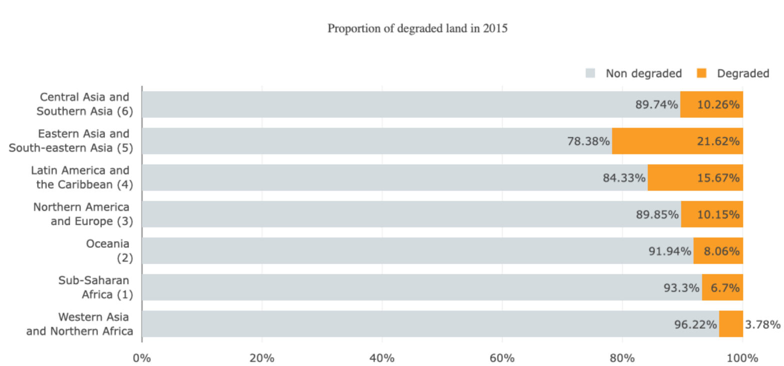

Figure 1: Image Source: https://data.unccd.int/worldwide-investment

Desertification and land degradation are very serious challenges. They lead to hunger and poverty, drive unemployment, forced migration and conflict, while increasing the risk of extreme weather related to climate change. But recent achievements in restoration and sustainable land management show that these problems are not insurmountable. Bold action and investments are needed to boost food security, improve livelihoods and help people adapt to a changing climate. [Source: FAO]

These calls to action motivated us to take on this problem with a positive mindset, knowing that via enhancing the scientific methods behind monitoring and prevention, there are the premises that with comprehensive policies and plans of action from the countries of the entire world, critical situations can be overcome.

We further explored the promising scientific breakthroughs of the last decade, with the goal to create an accessible tool for risk evaluation of desertification and take a step forward in preliminary monitoring of the problem globally. Our idea develops as follows: we wanted to architecture a model that could, given satellite images of an area, evaluate if there was a plausible risk of reduction of vegetation coverage. The actual realization of the project was planned to be divided into two sections: an image recognition model that could establish the presence of vegetation and the extension of desert areas; a second model that, given ulterior historical information about the climate in the region, could evaluate the evolution of the landscape in the future. We have concluded valuable insights, especially on the first intended subdivision of the project's development.

Desertification is often overlooked by the general public, but it's one of the most urgent problems currently afflicting the Earth, and not just humanity. Our choice to study this issue is a reflection of our commitment as a team to understanding and addressing one of the most significant environmental challenges of our time. We learned from the progress made so far by the scientific community, and went forward getting onboard with our enthusiasm and team collaboration. We hope that our work can contribute to a better understanding of how we can mitigate the impacts of climate change and build resilience in vulnerable regions.

2 Research background and proposed solution

2.1 Modelling and use of Machine Learning methods

Conventional methods of desertification detection and monitoring are usually based on ground-based observations and manual assessments, which are very time-consuming, labor-intensive, and narrow in scope. In this respect, satellite remote sensing becomes a powerful alternative because it offers an extended and continuing view of the Earth's surface. Recent advances in satellite technology have allowed the acquisition of high-resolution imagery that can monitor changes in land cover, vegetation health, and soil conditions over large areas.

In the recent past, machine learning has grown quite popular for processing and analyzing huge volumes of data acquired from satellites. Machine learning models could automate the detection process of patterns and features associated with desertification and, hence, provide timely and accurate insights into the conditions of such regions. Convolutional neural networks, a form of deep learning neural models, are exceptionally good at image analysis and therefore quite promising in extracting meaningful information from satellite images. [Men24]

Desertification has also been studied from a more classical standpoint as the transition of a biological system towards a catastrophic event such as the formation of a desert area. It's been shown that specific self-organized patterns in the vegetation are indicators of the ecosystems health, despite there being still a lot to uncover about how these evolve during time [RMG22]. The transition has been studied and simulated from a theoretical standpoint, starting from a study on the characteristics of various common dry-lands vegetation patterns. Specific sequences of stages seem to be in most cases a valid indicator of an occurring process of desertification [KG16]. Although these methods are valuable from a biological research standpoint, they aren't considered accurate enough yet, and would require area specific adaptations depending on on-site obtained data related to soil and vegetation.

Therefore, to contextualize properly, we'll give a general overview of the different techniques that have been used to study desertification in the last three decades.

Reaction-Diffusion Models: Reaction-diffusion models are widely used to simulate vegetation patterns in arid regions. These models describe the interaction between plant growth and resource diffusion, capturing the emergence of spatial patterns that precede desertification. [Kla99]

Spatial Pattern Analysis: Recent studies have highlighted the importance of spatial patterns in vegetation as early warning signals of desertification. Changes in spatial pattern characteristics, such as patch size and distribution, can indicate proximity to critical thresholds in ecosystem transitions. [She13]

Machine Learning Algorithms: Machine learning algorithms, such as decision trees, random forests, and support vector machines, are used to predict desertification risk based on remote sensing data and environmental factors. [Wu18] applied random forest models to classify land degradation areas in China, achieving high accuracy in identifying regions at risk of desertification.

Predictive Analytics: AI models are increasingly used for predictive analytics, allowing researchers to forecast future desertification trends and evaluate the effectiveness of intervention strategies. By leveraging large datasets, AI models can identify patterns and correlations that may be missed by traditional methods. [Xu20]

Our model was intended to combine both image recognition (using a Convolutional Neural Network) and predictive analytics, that grant great performance and accuracy without requiring a complete understanding of the biologic aspects of the phenomenon.

2.2 Modelling and use of Machine Learning methods

Lansat images are the most used satellite data for remote sensing desertification studies, mostly due to the access to a long time frame that allows easier change detection analysis. [DRM22]

In our particular case, we followed the example of many papers [DRM22, PL23] and considered data from the Landsat 8-9 OLI/TIRS C2 L1 directly downloaded from Earth Explorer. This is advantageous because being the Landsat 8-9 satellites sun-synchronous, with a near-polar orbit, the images are taken always around the same solar local time, allowing for consistent lighting conditions.

2.3 Towards preliminary global analysis

There is a marked geographical variation in the study areas considered historically in research papers, with a predominant focus on Asia and China in particular. In general, studies seem to concentrate in analysing desertification in highly populated countries, where there's a bigger socio-economic risk involved with future cathastrofic events. [DRM22]

Since there is a significant lack of studies that take on central Africa, we decided to focus on Zimbabwe for our case study, as further motivated afterwards. This was to further ensure that our model could be considered as a valid global resource, to make a step forward in fast preliminary analysis of the problem, rather than immediate high level research on a restrictive area.

3 Methods and Materials

3.1 Introduction and quantum possibilities

Our software application uses a Convolutional Neural Network (CNN) to classify satellite images into categories like "Desert," "Green Area," and others. The code handles data preprocessing, model training, and image classification. The model helps assess the percentage of land at risk of desertification by analyzing patterns in satellite imagery. Although we did not use quantum specifically but our project can easily be converted to something that uses Quantum Neural Networks (QNNs) which could process complex datasets, such as high-resolution satellite images, more efficiently than classical methods. We have considered Quantum Image Processing (QIP) as an option, but the low number of pixels this method currently supports made us discard it. If and when the number of qubits increases considerably, QIP might become a viable option to research on for this precise application.

The model could likely achieve high accuracy in classifying land cover types, which is essential for reliable desertification risk prediction with more data as well as time. The way the code works is by telling the user the distribution of vegetation cover based on a satellite images. This was intended to be applied to a dataset of images ranging from the past decade (from 2013 to 2023) to then evaluate the trends of land degradation if no mitigation is done. As this requires more research and dealing with large datasets, it was difficult to achieve in the period of time given to us the predictive functionality of the model, but this model lays a foundation for future projects working towards the same goal.

3.2 Training datasets

Accurate detection and monitoring of desertification call for high-quality data to capture the complexity of environmental conditions across affected regions. In this paper, we use the RSICB256 dataset, which is a comprehensive and well-curated collection of remote-sensing images designed for land-cover classification and change-detection tasks. This dataset will be instrumental in training and testing our machine-learning model, enabling it to effectively identify patterns of desertification from satellite images. The RSICB256 is a dataset of high-resolution remote-sensing images, each of 256 by 256 pixels. It has been compiled with a robust and representative collection of almost all land covers, including areas affected by desertification. Due to the fact that it is public and mostly applied by the remote sensing community in algorithm development and testing in image classification tasks, it is published as open data.

The dataset we referred to is from Kaggle Datasets.

The dataset includes multiple land cover classes relevant to desertification, such as barren land, grassland, shrubs, forest, and urban areas. It also covers a wide range of geographical locations, capturing varying climatic conditions and land-forms. Each image in the dataset has a resolution of 256 x 256 pixels, providing detailed spatial information crucial for detecting subtle changes in land cover associated with desertification. The dataset comes with labeled ground truth data, offering precise annotations for each image. It also includes temporal information that can be leveraged for analyzing changes over time.

The dataset was divided into training, validation, and test sets, ensuring that the model's performance could be rigorously evaluated. The training set was used for model training, the validation set for hyper-parameter tuning, and the test set for final performance assessment.

3.3 Methods

How to use our Model: As the code contains the way the model was made including the training and validation, there are two ways it can be used:

Making Your Own Model And Testing

- For making your own model, you need to install everything needed which can be found in the Jupyter notebooks.
- The next thing is either unzip archive.zip file which gives you data folder which contains the images needed by the model to be trained on. You can also just download the data folder from the GitHub repository.
- Then you put the data folder in a known place on your local machine and change the path in Data Pre-processing in the following [code snippet]:

```
labels = {'/content/data/cloudy' : 'Cloudy',
          '/content/data/desert' : 'Desert',
```

```
'/content/data/green_area' : 'Green_Area',
'/content/data/water' : 'Water',
}
```

Here the path should be wherever the data folder is.

- Then you follow the notebook which should run without any errors till you reach the part where we check the URL images. Then after that there is a code block where you can upload an image and get the prediction on.

Using Existing Model And Testing

- For this you just need to download the model
- Then you start from loading the model in the notebook and essentially follow along. The only thing that needs to be done is having the image named Forest_1768jpg from the data folder in the repository in your machine so you can input the path where it asked.
- As you follow along you reach the part where we check the URL images. Then after that there is a code block where you can upload an image and get the prediction on.

4 Case study

4.1 Area of study specifics

Anywhere on the Earth, human actions interact strongly with the environment. Since the relationship between human actions, climate change and direct and measurable consequences are complexly intertwined, and considering that each region has their own biophysical, social, economical and political environments, land degradation is a phenomenon that is considered to be best modelled locally and not globally.

The WAD3 report, which we have used to seek guidance in the choice of a target area to focus on, provides an interactive map that identifies regional areas of concern, where land degradation process might be underway. As expressed in the report, local studies, with awareness of the specifics of the area considered, are needed to offer guidance to prevention and concrete intervention.

We considered the following criteria, in order to orient our work properly. We searched for an area scarcely mentioned in the scientific literature and with little direct human intervention (low water stress, irrigation, productivity, population density) in order to help approaching the problem by reducing such high interference less documented deciding factors. Our case study landed on the choice of the furthest south-eastern region of Zimbabwe, specifically around Gonarezhou National Park. Approximately 5,053 square kilometers, it is one of Zimbabwe's largest national parks. In particular we searched for satellite images approximately inside the perimeter shown in the figures, which is defined by: (Lat: 20° 33' S, Lon: 032° 40' E), (Lat: 21° 24' S, Lon: 032° 40' E), (Lat: 21° 24' S, Lon: 031° 56' E), (Lat: 20° 33' S, Lon: 031° 56' E).

Figure 2:

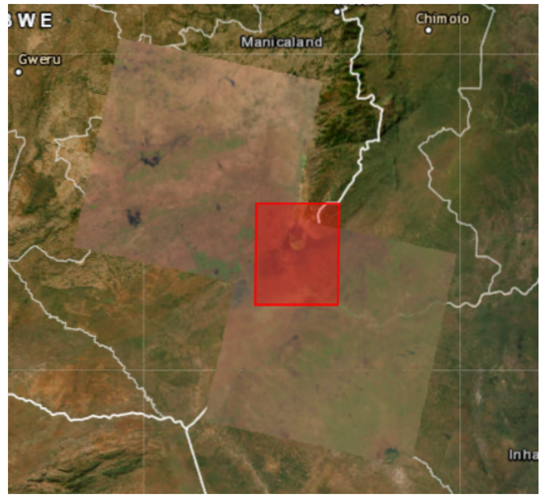

Figure 3:

Vegetation:
The park hosts a variety of habitats, including mopane woodlands, acacia grasslands, and dense river forests along the Save, Runde, and Mwenezi rivers. The region's diverse plant life includes both drought-resistant species and those more susceptible to arid conditions, making it a valuable area for assessing the resilience of various plant communities to environmental stress.

Interactions:
The interactions between humans, wildlife, and the environment create complex dynamics that are crucial for understanding land use patterns and their impact on desertification. The region is also affected by land-use practices such as agriculture and grazing, which can exacerbate soil degradation and contribute to desertification processes.

Climate:
The climate in the south-eastern region of Zimbabwe is predominantly semi-arid, with distinct wet and dry seasons. This climate is typical of areas prone to desertification, where changes in rainfall patterns and temperature can significantly impact vegetation health and ecosystem stability. The area receives most of its rainfall between November and March, with a prolonged dry season from April to October. The region experiences high temperatures, especially during the dry season.

Human interactions with the environment are a major driver of landscape changes, particularly in regions susceptible to desertification. Activities such as deforestation, overgrazing, and unsustainable agricultural practices significantly contribute to land degradation and the transformation of fertile areas into deserts. [WMN20]

An area at risk of desertification is more sensible to the catastrophic shift during the dry seasons. In Zimbabwe, average precipitations are drastically lower between May and September, which also correspond to the lowest surface air temperatures experienced in the area. The Pacific—the El Niño-Southern Oscillation, or ENSO for short, modifies the normal climatic cycles with frequently catastrophic effect for land productivity. The pattern shifts back and forth irregularly every two to seven years, causing variations in ocean surface temperature and disrupting the wind and rainfall patterns across the tropics. For instance, ENSO-induced droughts have been well-documented, including the severe droughts of 1991-1992 and 2015-2016, which led to widespread agricultural failure, food insecurity, and economic hardship [Mas14] [Bau17]. Understanding the interplay between ENSO and regional climate dynamics is crucial for developing adaptive strategies to mitigate the impacts of desertification and enhance the resilience of affected communities.

We choose images from the month of September, in order to monitor the area at the end of the period of drought, right before precipitation and temperatures rise. This is also a period in which, seen the already climatically drastic conditions that are unfavourable for agriculture and vegetation in general, the effects of ENSO are less impactful, while when the climatic patterns are disturbed at the beginning of the new vegetation cycle, the effects are much worse.

4.2 Testing and final evaluations

The testing provided results that conclude that a lot of work needs to be done to make the model usable. We'll provide some insight on our difficulties, knowing this is a starting point that needs further development.

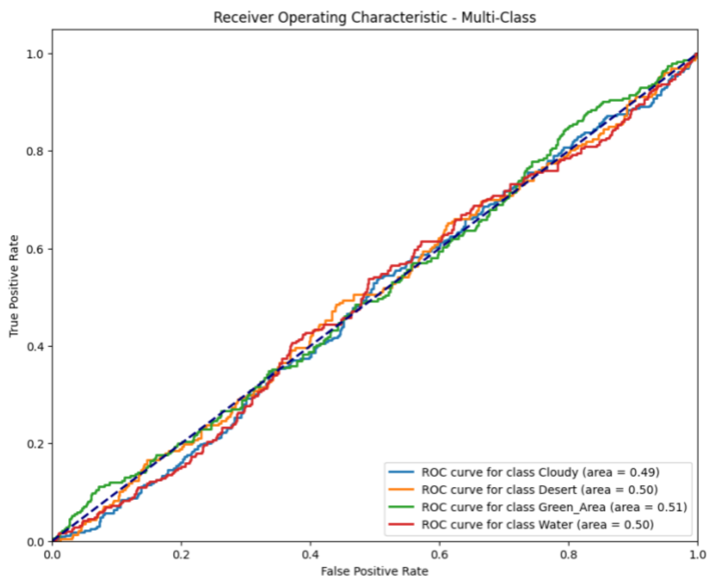

Figure 4:

ROC curves for each class (Cloudy, Desert, Green Area, and Water) are quite close to the diagonal line (AUC 0.50) which indicates that the model is not performing well for any of these classes. This suggests the model's predictions are close to random for each class.

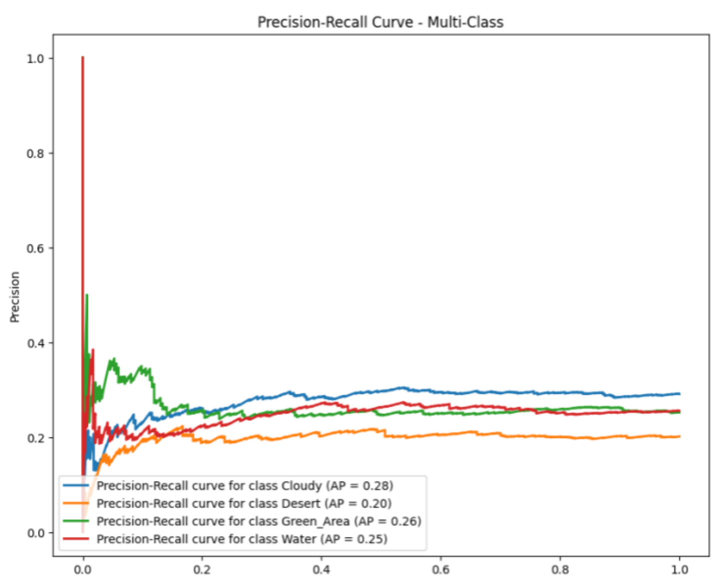

Figure 5:

The precision recall curves show that precision drops significantly as recall increases especially for classes like Desert and Water which indicates that the model has difficulty balancing precision and recall. Low AP scores further confirm the model's struggle with correctly identifying positives for each class.

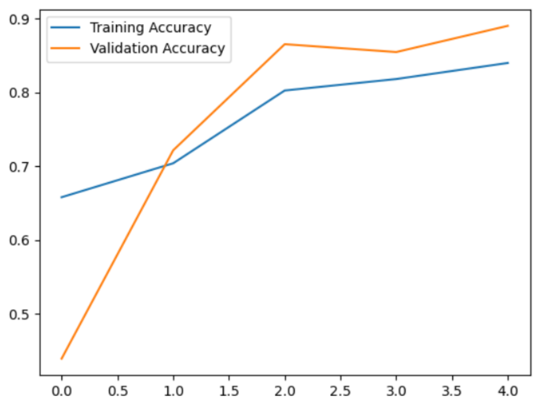

Figure 6:

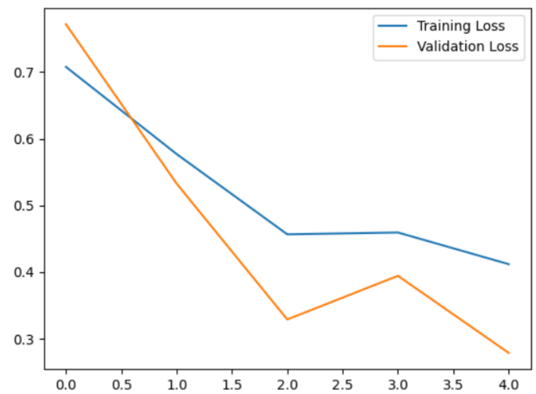

Figure 7:

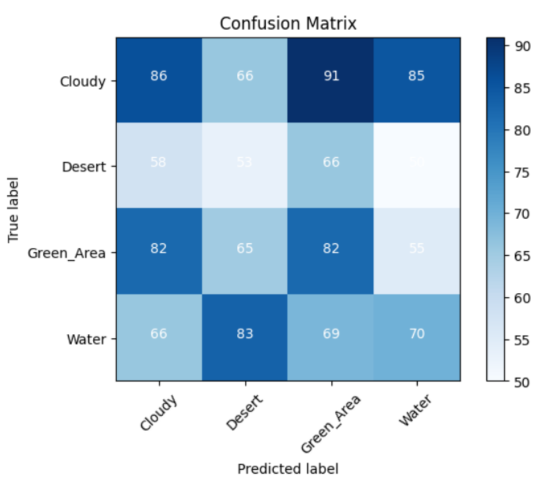

Figure 8:

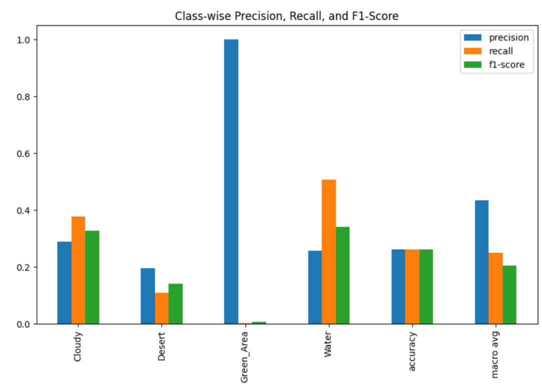

Figure 9:

For the Cloudy, Desert, Green Area, and Water classes, the precision, recall, and F1-scores are generally low which indicates poor model performance where the Green Area class has a notably high precision but low recall which means the model predicts Green Area very conservatively but does so correctly. The overall low scores across metrics indicate that the model needs improvement which is possibly in terms of better handling the class distribution , improving feature extraction, or adjusting the model architecture.

References

[Bau17] Vogel C. Nortje K. Naik M. Baudoin, M. A. Living with drought in South Africa: Lessons learnt from the recent El Niño drought period. *International Journal of Disaster Risk Reduction, 23, 128-137.*, 2017.

[DRM22] Booker Ogutu Daniela Rivera-Marin, Jadunandan Dash. The use of remote sensing for desertification studies: A review. *Journal of Arid Environments*, 2022.

[KG16] Sarah Iams Mary Silber Karna Gowda, Yuxin Chen. Assessing the robustness of spatial pattern sequences in a dryland vegetation model. 2016.

[Kla99] C. A. Klausmeier. Regular and irregular patterns in semiarid vegetation. *Science, 284(5421)*, 1999.

[Mas14] Maskey S. Mussa F. E. F. Trambauer-P. Masih, I. A review of droughts on the African continent: A geospatial and long-term perspective. *Hydrology and Earth System Sciences, 18(9), 3635-3649*, 2014.

[Men24] Li S. Akhmadi K. He P. Dong-G. (2024) Meng, X. Trends, turning points, and driving forces of desertification in global arid land based on the segmental trend method and shap model. *GIScience Remote Sensing, 61(1)*, 2024.

[PL23] Olivier Debeir Polina Lemenkova. Multispectral satellite image analysis for computing vegetation indices by r in the khartoum region of sudan, northeast africa. *Journal of Imaging*, 2023.

[RMG22] Justin M. Calabrese Emilio Hernandez-Garca Corina E. Tarnita Cristobal Lopez Juan A. Bonachela Ricardo Martinez-Garcia, Ciro Cabal. Integrating theory and experiments to link local mechanisms and ecosystem-level consequences of vegetation patterns in drylands. 2022.

[She13] Bastiaansen R. Holmgren M. Rietkerk M. Sheffer, E. The role of microtopography in determining vegetation pattern formation in dryland ecosystems. *Journal of Ecology, 101(2), 324-333*, 2013.

[WMN20] Ephraim Mpofu Walter Musakwa and Nesisa Analisa Nyathi. Local community perceptions on landscape change, ecosystem services, climate change, and livelihoods in Gonarezhou National Park, Zimbabwe. 2020.

[Wu18] Zhang C. Li J. Cheng Y. Wu, X. Land degradation risk mapping based on random forest models: A case study in yanchi county, china. *Ecological Indicators, 91, 102-110.*, 2018.

[Xu20] Li Y. Zhang L Xu, C. Predictive analytics for desertification risk assessment using ai-based models. *Environmental Research Letters, 15(4), 044014.*, 2020.