



1949

**SUPERVISED MACHINE LEARNING FOR GULLY  
MAPPING AND MODELING USING LOW-COST, HIGH-  
RESOLUTION SENSORS AND OPEN-SOURCE  
GEOSPATIAL DATA IN A SEMI-ARID ENVIRONMENT**

Thesis for the Degree of Doctor of Philosophy (Ph.D.)

Kwanele Phinzi

Supervisor: Prof. Szilárd Szabó DSc

UNIVERSITY OF DEBRECEN

Doctoral Council of Natural Sciences and Information Technology

Doctoral School of Earth Sciences

Debrecen, 2023



*I hereby declare that I prepared this thesis within the Doctoral Council of Natural Sciences and Information Technology, Doctoral School of Earth Sciences, University of Debrecen, to obtain a PhD Degree in Natural Sciences at Debrecen University.*

*The results published in the thesis are not reported in any other PhD thesis.*

*Debrecen, 22 May 2023*

.....  
*Signature of the candidate*

*I hereby confirm that the candidate, Kwanele Phinzi, conducted his studies under my supervision in the Department of Physical Geography and Geoinformatics, Doctoral School of Earth Sciences, between 2019 and 2023. The candidate's independent studies and research work significantly contributed to the results published in the thesis.*

*I also declare that the results published in the thesis are not reported in any other thesis.*

*I support the acceptance of the thesis.*

*Debrecen, 22 May 2023*

.....  
*Signature of the supervisor*



**SUPERVISED MACHINE LEARNING FOR GULLY MAPPING AND  
MODELING USING LOW-COST, HIGH-RESOLUTION SENSORS  
AND OPEN-SOURCE GEOSPATIAL DATA IN A SEMI-ARID  
ENVIRONMENT**

Dissertation submitted in partial fulfilment of the requirements for the doctoral (PhD) degree  
in Earth Sciences

Written by Kwanele Phinzi certified Geographer (MSc. Environmental Science)

Prepared in the framework of the doctoral school of the University of Debrecen  
(Natural and Anthropogenic Processes of the Lithosphere and the Hydrosphere programme)

Dissertation advisor: Prof. Szabó Szilárd DSc

The official opponents of the dissertation:

Dr. ....  
Dr. ....

The evaluation committee:

chairperson: Dr. ....  
members: Dr. ....  
Dr. ....  
Dr. ....  
Dr. ....

The date of the dissertation defence: ..... 20...



## **Dedication**

*To my late mother, Thandiwe Phinzi*



## TABLE OF CONTENTS

<b>1. INTRODUCTION.....</b>	<b>11</b>
<b>2. SIGNIFICANCE AND RATIONALE OF THE STUDY .....</b>	<b>13</b>
<b>3. LITERATURE REVIEW .....</b>	<b>16</b>
<b>3.1. Gully definition.....</b>	<b>16</b>
<b>3.2. Factors controlling gully erosion .....</b>	<b>18</b>
<b>3.3. An overview of machine learning (ML) .....</b>	<b>20</b>
<b>3.4. Gully mapping using remote sensing .....</b>	<b>25</b>
<b>3.5. Gully susceptibility modeling.....</b>	<b>27</b>
<b>3.6. Resampling techniques .....</b>	<b>29</b>
<b>3.7. Evaluation metrics .....</b>	<b>32</b>
<b>4. METHODS AND MATERIALS .....</b>	<b>35</b>
<b>4.1. Study area description.....</b>	<b>35</b>
<b>4.1.1. Study Area #1 .....</b>	<b>39</b>
<b>4.1.2. Study Area #2 .....</b>	<b>40</b>
<b>4.1.3. Study Area #3 .....</b>	<b>41</b>
<b>4.1.4. Study Area #4 .....</b>	<b>42</b>
<b>4.2. Data .....</b>	<b>43</b>
<b>4.3. Geo-environmental variables.....</b>	<b>44</b>
<b>4.4. Variable selection .....</b>	<b>49</b>
<b>4.5. Reference data collection for gully mapping and susceptibility modeling     .....</b>	<b>50</b>
<b>4.6. Machine learning (ML) and hyperparameter tuning.....</b>	<b>54</b>
<b>4.6.1. Random forest (RF) .....</b>	<b>55</b>
<b>4.6.2. Support vector machines (SVM).....</b>	<b>55</b>
<b>4.6.3. Regularized discriminant analysis (RDA) .....</b>	<b>56</b>
<b>4.6.4. Linear discriminant analysis (LDA).....</b>	<b>56</b>
<b>4.6.5. Maximum likelihood classifier (MLC) .....</b>	<b>57</b>
<b>4.6.6. K-nearest neighbor (K-NN) .....</b>	<b>57</b>
<b>4.6.7. Minimum distance (MD).....</b>	<b>58</b>

4.6.8. <i>Artificial neural network (ANN)</i> .....	58
4.6.9. <i>Partial least squares (PLS)</i> .....	59
4.6.10. <i>Stochastic gradient boosting (SGB)</i> .....	59
4.7. Model performance evaluation .....	60
4.8. Statistical analysis .....	62
<b>5. RESULTS AND DISCUSSION</b> .....	<b>63</b>
5.1. Low-cost, high-resolution sensors for gully mapping .....	64
5.1.1. <i>SPOT-7 multispectral image</i> .....	64
5.1.2. <i>PlanetScope multispectral image</i> .....	69
5.1.3. <i>Visual range SPOT-7 image</i> .....	72
5.2. Quantifying gully classification accuracy across different study areas .	73
5.2.1. <i>Study Area #1: accuracy and factors biasing model performance at class-level</i> .....	73
5.2.2. <i>Study Area #2: accuracy as a function of an algorithm, resampling technique, and season</i> .....	79
5.2.3. <i>Study Area #3: efficacy of algorithms based on limited spectral information</i> .....	84
5.2.4. <i>Visual analysis of gully classification across different study areas</i> ....	86
5.3. Gully characteristics' influence on the precise mapping of gullies and their density .....	93
5.4. Feature selection and multicollinearity analysis .....	95
5.5. Analyzing algorithms' performance when using feature sets of varying sizes.....	98
5.6. Gully susceptibility and key controlling geo-environmental variables	103
<b>6. SUMMARY AND CONCLUSIONS</b> .....	<b>107</b>
<b>7. ACKNOWLEDGEMENT</b> .....	<b>114</b>
<b>8. APPENDICES</b> .....	<b>116</b>
<b>9. REFERENCES</b> .....	<b>119</b>

## 1. INTRODUCTION

Despite ongoing global research and soil conservation efforts, gully erosion remains the most significant environmental challenge affecting many countries worldwide. It contributes significantly (50-80%) to the sediment losses in semi-arid regions (Poesen *et al.*, 2002; Marzloff and Poesen, 2009), mainly due to sparse vegetation cover and prolonged dry periods with occasionally high-intensity rainfall regimes. The significant extent of gully erosion in semi-arid regions is a matter of great apprehension because these regions account for nearly 50% of the earth's landmass, supporting a significant fraction of global livestock and food production (Bailey, 1998; Safriel *et al.*, 2006; Reynolds *et al.*, 2007; Anadón *et al.*, 2014). Due to gully erosion reducing the availability and quality of productive lands, many semi-arid countries with a firm reliance on agriculture are likely to suffer food shortages. For example, Africa, where almost half of the population (48%) depends on agriculture (NEPAD, 2013), records nearly 9% of annual average yield losses from past erosion, leading to severe food shortages (Ashiagbor *et al.*, 2013). A semi-arid country, South Africa is among the most erosion-affected countries on the continent.

Approximately 85% of the land in South Africa is at risk of erosion (Rensburg, 2010; Parwada and Van Tol, 2017), while 70% is already considered eroded (Garland, Hoffman and Todd, 2000). Moreover, nearly 6% of the population in South Africa relies on agriculture for their sustenance (Department of Agriculture, 2007). Consequently, soil erosion in general and specifically gully erosion pose a significant hazard to food security in numerous rural regions that depend on subsistence farming in the country (Phinzi, 2018). Besides, gullies damage infrastructure (buildings and roads), increase local flood risks, and considerably reduce dam water quality and quantity due to siltation and sedimentation. From the economic standpoint,

gully erosion has far-reaching implications, costing the country about R12 billion (\$836 million) a year to purify silted dam water (Hoffman and Ashwell, 2001). Unless the severity and spatial extent of gully erosion are known, it is challenging to implement targeted conservation measures. Therefore, accurate gully mapping, though not straightforward, is the most practical approach for detecting gullies in remote locations and may aid active gully rehabilitation and soil management efforts.

Although gully erosion research spans over a century (Castillo and Gómez, 2016; Liu *et al.*, 2021), gully modeling is still in its infancy compared to rill and sheet erosion, with well-established and widely adopted models worldwide, including the universal soil loss equation (USLE) (Wischmeier and Smith, 1978). In addition, due to their geographic specificity, intricate physical mechanisms, and difficulties in acquiring observational data (Roberts *et al.*, 2022), existing gully models have not been widely used globally. Although traditional field measurements and surveys are commonly used, these methods are often expensive and limited to accessible locations, hampering a precise understanding and quantification of gullies, especially for inaccessible locations (Vrieling, Sterk and de Jong, 2010). Another well-established approach is the manual digitisation and interpretation of gullies based on aerial or satellite images, but it is also time-consuming and subjective, which greatly impedes repeatability. An alternative approach involves the utilization of machine learning (ML)-based techniques to automatically extract and model gullies from satellite images and related open-source geospatial data.

The utilization of these data has witnessed a significant surge in popularity in recent years (Magliulo *et al.* 2020; Žížala *et al.* 2019; Vrieling *et al.* 2007), driven by advancements in computer processing capabilities and the augmented availability of such data free of cost. The emergence of high-

resolution ( $\leq 3\text{m}$ ) imagery has opened new avenues for extracting narrow geomorphic features, such as gullies, in fine spatial detail. In particular, low-cost multispectral sensors, such as Systeme Pour l'Observation de la Terre (SPOT-7) and PlanetScope, offer a reasonably high spatial resolution (1.5-3m) for mapping gullies with sufficient detail. Despite these sensors' low cost or free availability, their application in gully mapping utilizing ML has received limited attention in severely gullied semi-arid regions such as South Africa.

Besides gully mapping, in recent years, ML algorithms have been increasingly utilized in predicting gully-susceptible areas at the catchment level. Several studies (Dewitte *et al.*, 2015; Pourghasemi *et al.*, 2017; Arabameri, Chen, *et al.*, 2019; Huang *et al.*, 2022; Lana, Castro and Lana, 2022; Kulimushi *et al.*, 2023) have applied this approach, which involves the identification of areas with varying degrees of gully vulnerability. Rather than extracting individual gully features, these studies aim to map areas susceptible to gully formation through modeling and analyzing the relationship between gully occurrence and multiple geo-environmental covariates. This approach identifies spatial variability in gully susceptibility, providing a more comprehensive understanding of gully formation processes (Conoscenti and Rotigliano, 2020).

## **2. SIGNIFICANCE AND RATIONALE OF THE STUDY**

The United Nations (UN) has made a worldwide call to tackle the problems of land degradation (Goal No. 15), promote sustainable agriculture, and reduce food insecurity (Goal No. 02). As gullies significantly contribute to land degradation, which undermines sustainable agriculture and leads to food insecurity, it is essential to comprehend their spatial distribution and related factors to attain these UN sustainable development goals. Furthermore, the Department of Agriculture, Forestry, and Fisheries (DAFF) in South

Africa has recognized the necessity to map gully erosion and its severity at a national level (Mararakanye and Le Roux 2012), and several scholars in the country have responded to the call (Le Roux *et al.*, 2008; Mararakanye and Le Roux, 2012; Phinzi and Ngetar, 2017; Ebhuoma *et al.*, 2022).

Recently, researchers have utilized Machine Learning (ML)-based models to identify medium-sized gully features (Makaya *et al.*, 2019), generate gully susceptibility maps (Conoscenti *et al.* 2014; Dewitte *et al.* 2015; Huang *et al.* 2022; Kulimushi *et al.* 2023), and identify the factors that trigger gully erosion (Amiri *et al.* 2019; Bernini *et al.* 2021; Lana *et al.* 2022; Rahmati *et al.* 2022). Although these studies have demonstrated successful ML applications in gully erosion analysis, there is still a vast potential for further development, with many possibilities yet to be explored. In addition, the procedures are intricate, and various limitations must be investigated from multiple perspectives. Currently, it is unclear which combination of factors, such as resampling techniques, seasons, class number (binary or multiclass), and algorithms, provide the best accuracy, particularly in the context of low-cost, high-resolution sensors such as SPOT-7 and PlanetScope, which are not yet well researched. To date, no study has systematically evaluated the impact of different combinations of these factors on gully classification accuracy using these sensors. This dissertation proposes the utilization of these sensors to map gullies and examine the different combination of factors, including the impact of gully characteristics on classification accuracy.

Furthermore, concerning gully susceptibility modeling, previous studies have relied on a fixed set of predictor variables. However, this practice limits our comprehension of how various ML algorithms perform regarding the accuracy and processing speed when utilizing feature sets of varying sizes. Therefore, a comprehensive study on ML applications to gully mapping and modeling is crucial, considering various factors influencing accuracy. More

importantly, this dissertation presents cost-effective ML methods centered on open-source geospatial data for gully mapping and susceptibility modeling. These techniques have the potential to be used in different regions of the country to support targeted gully rehabilitation and soil management endeavors. Specifically, the objectives of this dissertation are to:

1. determine if low-cost, high-resolution sensors improve gully mapping in semi-arid regions,
2. quantify gully classification accuracy and analyze factors biasing model performance on a class level,
3. examine how different gully morphological characteristics affect the precise mapping of gullies using high-resolution satellite data,
4. select geo-environmental variables with the greatest predictive power to model gully susceptibility, and
5. analyze algorithms' performance when using input feature sets of varying sizes.

This dissertation encompasses four interrelated studies conducted in various parts of Eastern South Africa. The first study presents a gully mapping approach using a pan-sharpened multispectral SPOT-7 image, which was based on three algorithms: random forest (RF), support vector machines (SVM), and linear discriminant analysis (LDA). The study evaluated two different class numbers (binary and multi-class) and six combinations of study areas used for training and testing data (Phinzi *et al.*, 2020). The second study used a SPOT-7 visual range satellite image and various machine learning algorithms, including k-Nearest Neighbor (k-NN), minimum distance (MD), maximum likelihood classifier (MLC), and RF, to map gullies based on their morphological features, such as shape/appearance, size, depth, and length (Phinzi, Holb and Szabó, 2021). The third study applied SVM and RF to

examine the impact of bootstrapping and k-fold cross-validation (CV) on the accuracy of gully mapping using PlanetScope data in both the dry and wet seasons (Phinzi, Abriha and Szabó, 2021). The fourth and last study focuses on gully susceptibility and identifies geo-environmental variables with the most predictive power for modeling gully susceptibility. Unlike the preceding three studies, the findings of the fourth study are not yet published as it is currently undergoing a review process for publication.

### **3. LITERATURE REVIEW**

This section defines a gully, focusing on mechanisms leading to its formation, main characteristics, and types. Next, factors contributing to gully erosion are reviewed. Furthermore, the section reviews the role of supervised ML in gully mapping and susceptibility modeling, highlighting some of the most widely used ML algorithms, resampling techniques, and satellite products. Finally, model validation, focusing on various matrices commonly used, is reviewed.

#### **3.1. Gully definition**

Gully is a typical form of surface erosion induced by water. Generally, it is defined as a steep-sided (U- or V-shaped) channel resulting from erosion due to sporadic water flow, often during and after intensive rainfall (Poesen, Vandaele and Wesemael, 1998). Gullies can be either temporary (called ephemeral gullies) or permanent (also called classical gullies), depending on their morphology, location in the landscape, and dominant erosion mechanisms causing them (Poesen, Vandaele and Wesemael, 1998; Poesen *et al.*, 2003). Kirkby and Bracken (2009) present a detailed overview of these gullies, classifying them based on their internal morphology using two ratios. Figure 1 illustrates this internal morphology-based gully classification. The first ratio is the side-slope to channel slope, which is high for classical gullies

and lower for ephemeral gullies. The second ratio is the gradient of the gully channel relative to the surface gradient it dissects. For a gully to extend headward, there must be a significant portion of the gully where this ratio is less than unity, allowing the depth of the gully to increase upstream (Poesen *et al.*, 2002; Kirkby and Bracken, 2009). The ratio of the gully side-slope gradient to the initial surface gradient, which defines the channel way, is obtained by multiplying these two ratios and must be greater than unity (Kirkby and Bracken, 2009). Classical gullies are in the lower right corner, while typical valley formation happens with higher valley slopes (Figure 1). Badlands can form when valleys are incised into surfaces with slopes nearly as steep as the valley itself (Kirkby and Bracken, 2009). Such badlands are characterized by regions with a high density of gully systems, typically separated by short, steep, and devoid-of-vegetation slopes (Boardman *et al.*, 2003).

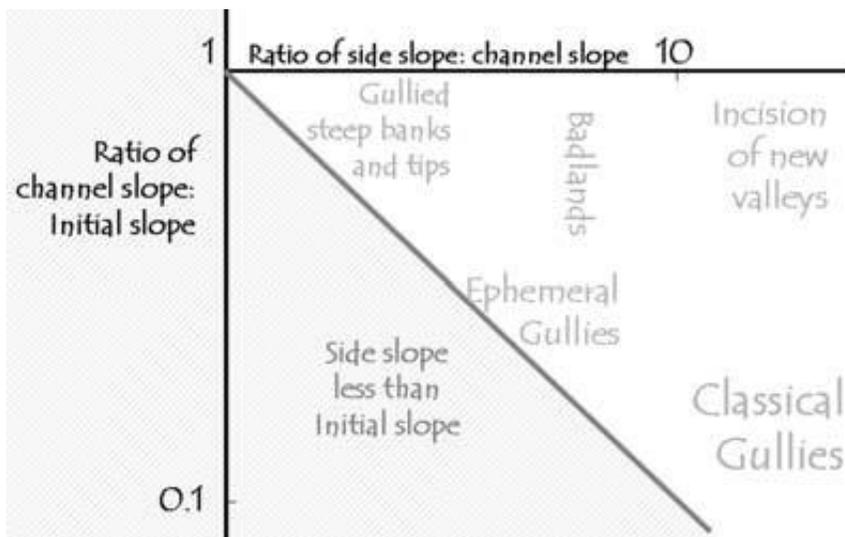


Figure 1. Classification of gullies based on their internal morphology (Kirkby and Bracken, 2009).

Gullies can also be distinguished based on their depth, duration of occurrence, and location. Small concentrated flow channels of 0.5 – 50 cm deep that occur over a short period in agricultural fields result in ephemeral gullies (Momm *et al.*, 2012). Normal tillage operations can quickly fill ephemeral gullies, although they can reform again in the exact location by subsequent runoff events (Foster, 2005; Zhang *et al.*, 2007; Bennett and Wells, 2019). However, when left unabated, ephemeral gullies can become permanent and too large to be obliterated by normal tillage operations (Liu *et al.*, 2021).

On the contrary, permanent gullies are long-duration erosional features with relatively deep (>50 cm) channels that interfere with farming operations (Poesen *et al.*, 2003; Luffman, Nandi and Spiegel, 2015; Bennett and Wells, 2019). Permanent gully systems often exhibit an inclination to expand and become increasingly challenging and expensive to remove (Kirkby and Bracken, 2009). Although heavy rainfall and concentrated surface water flow are critical in forming and developing gullies, subsurface flow through soil pipes or macropores is another crucial mechanism driving gully formation (Faulkner, 2006; Wilson, 2009). Details of these factors are presented in the following section.

### **3.2. Factors controlling gully erosion**

Gully erosion is driven by various natural and anthropogenic factors, including rainfall, topography, soil type, land use, and vegetation (Valentin, Poesen and Li, 2005; Le Roux and Sumner, 2012). Thus, the success of gullying depends on the cumulative and synergistic effects of these natural and anthropogenic factors.

Rainfall is a prerequisite for any water-borne erosion (Phinzi, 2018; Phinzi and Njoya Silas Ngetar, 2019) and is responsible for detaching, transporting, and depositing soil particles. However, it is worth noting that not

all raindrops have the capacity to cause erosion. The erosive properties of rainfall, referred to as rainfall erosivity, are determined by factors such as the size and duration of raindrops. Larger raindrops tend to dislodge more soil particles than smaller raindrops due to their higher velocity and intensity, leading to greater rainfall erosivity (Cruse *et al.*, 2000; Parlak and Parlak, 2010). These rainfall properties play a crucial role in the initial stages of soil erosion, known as splash erosion, where the greatest detachment of soil particles occurs. However, short-duration rainfall has a limited impact on other forms of erosion, especially gully erosion, which requires sustained runoff.

Gullies are frequently found in mountainous or hilly regions with steep slopes (Valentin, Poesen and Li, 2005). Steep slopes are conducive to high runoff velocity, which leads to rills and gullies forming. However, under certain climatic conditions, these steep slopes may result in lower runoff volumes than gentle slopes where soil crusts mainly develop, causing higher runoff (Valentin, d'Herbès and Poesen, 1999). Several studies have reported gullies in low-elevation areas with gently to nearly flat slopes (Pham *et al.*, 2020; Chowdhuri *et al.*, 2021; Phinzi, Holb and Szabó, 2021; Huang *et al.*, 2022). Catchment drainage is pivotal in gully erosion hydrological processes (Ebhuoma *et al.*, 2022). According to Valentin *et al.* (2005), for a given slope (S), there must be a critical drainage area (A) in order to generate sufficient runoff that can concentrate and initiate gully formation.

The erodibility of soil, or its ability to be eroded, depends on soil type. The vulnerability of different soils to erosion varies based on their cohesiveness. Soils with high cohesion, fast infiltration rates, high organic matter content, and improved soil structure are more resistant to erosion. Additionally, the size of soil particles also plays a crucial role in determining soil erodibility, with smaller and non-cohesive particles being more susceptible to erosion than larger particles. Soils prone to crusting, as they

enhance runoff generation and concentration downslope, are generally subject to sheet and gully erosion (Valentin, Poesen and Li, 2005). Due to the limited vegetation in arid and semi-arid regions, including South Africa, the soils in these areas are prone to crusting, which leads to runoff production and the formation of gullies (Kakembo and Rowntree, 2003; Valentin, Poesen and Li, 2005). Subsurface water flow through soil pipes also gives rise to gullying. Soils that have dispersive characteristics and contain sodic layers are susceptible to forming pipes, which can progress into gullies when the roofs covering them collapse (Beckedahl and de Villiers, 2000; Faulkner *et al.*, 2004; Faulkner, 2006).

Human activities, specifically land use, play a significant role in affecting erosion. Therefore, the type of human activity determines the likelihood of soil erosion in a specific area. For example, certain agricultural practices increase the probability of gullying. Indeed, the natural processes of gully formation are accelerated by implementing more intensive farming systems (Valentin, Poesen and Li, 2005), such as overgrazing, mono-cropping, and other unsound farming methods. Usually, natural habitats such as grasslands and forests effectively reduce soil erosion due to their high levels of vegetative cover. However, according to Le Roux *et al.* (2008), grasslands in South Africa are highly susceptible to erosion.

### **3.3. An overview of machine learning (ML)**

Although introduced in the 1990s, the application of machine learning (ML) to gully erosion gained popularity only after 2010 (Svoray *et al.*, 2012; Lana, Castro and Lana, 2022). Since then, the usage of ML in gully erosion has continued to increase, particularly in the past three years. A search conducted on Scopus in May 2023 using keywords such as "gully" and "machine learning" revealed that 330 research papers had been published on the subject between 2012 and 2022, more than 50% of which were released in

the last three years (2020-2022). This remarkable surge in the adoption of ML for gully erosion can be attributed to the growing availability of open-source geospatial data, a critical component for developing predictive ML models.

A typical ML process entails training a model, also known as an algorithm, on a given data (training set) for which we know the outcome and applying this algorithm to make a prediction when we do not know the outcome (Figure 2). The learning algorithm essentially recognizes patterns from data without explicitly being programmed. It can learn the target object's (gully) characteristics from input data and identify them when presented with new data (testing set). The success of any ML algorithm lies in its ability to generalize well and make an accurate prediction or classification using new data (testing/validation data) for which the outcome is unknown.

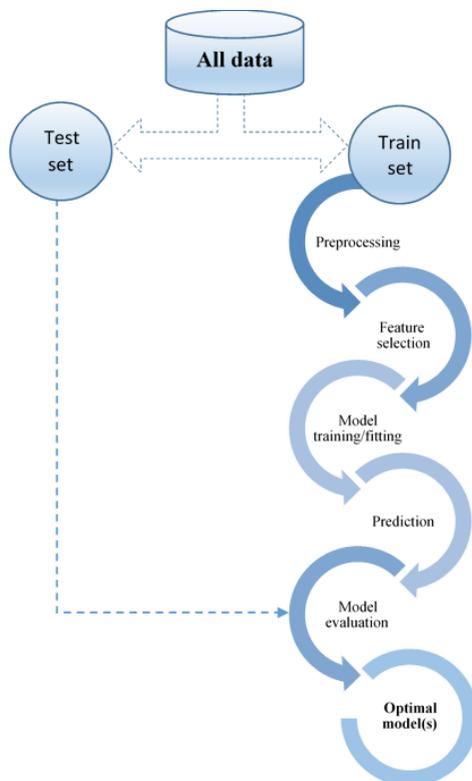


Figure 2. A typical ML process for a classification task.

ML algorithms can be broadly grouped into supervised and unsupervised depending on the amount and type of supervision required during the learning phase (Boehmke and Greenwell, 2019). The key difference between the two is that unsupervised ML, also known as cluster analysis or class discovery, does not require a training set or prior labeled examples/data except to specify the number of clusters in advance (Gentleman and Carey, 2008). Unsupervised ML has been successfully applied in many fields, including natural language processing, computer vision, networking (mobile and internet), and speech recognition (Usama *et al.*, 2019). A recent literature survey indicates that principal component analysis (PCA), k-means, partial least squares, and hierarchical clustering are among unsupervised learners most commonly used (Alloghani *et al.*, 2020), often in exploratory data analysis (Boehmke and Greenwell, 2019). This dissertation focuses on supervised ML. Table 1 summarizes some of the most commonly used algorithms in gully mapping. While conventional ML algorithms are predominantly used in gully mapping, the advent of deep learning (DL), although not a focus of this study, cannot be ignored.

Indeed, interest in DL has been shown in many fields, including remote sensing (Chen *et al.*, 2014; Kussul *et al.*, 2017; Vetrivel *et al.*, 2018). DL algorithms are considered advantageous over conventional ML algorithms mainly because of their high level of automation of feature extractors and their ability to adapt to new future challenges, solving highly complex problems (Ball, Anderson and Chan, 2017). It has significantly impacted precision medicine, speech recognition, cancer diagnosis, predictive forecasting, and self-driving cars (Shrestha and Mahmood, 2019). Convolutional neural network (CNN), another form of DL architecture, was successfully applied to detect and discriminate gullies from other linear types of surface erosion (Gafurov and Yermolayev, 2020). Similarly, Liu *et al.* (2022) successfully

employed DL to automatically recognize and detect ephemeral gullies from remotely sensed data. Although DL overcomes several limitations of earlier generation ML techniques, its predictive performance depends on highly optimized algorithms and requires massive training data, which introduces computational burdens (L. Zhang *et al.*, 2016; Chen *et al.*, 2018; Ma *et al.*, 2019).

For this reason, ML algorithms are still widely applied and preferred over DL in many remote sensing applications. Since knowing which algorithm will best perform is difficult, applying and evaluating different algorithms is common (Boehmke and Greenwell, 2019). Support vector machines (SVM), random forest (RF), artificial neural network (ANN), and to some extent, boosted regression trees (BRT) are the most frequently used algorithms in gully erosion. For example, Lana *et al.* (2022) employed RF, logistic regression (LR), naïve bays (NB), and ANN to model gully susceptibility and reported that RF achieved the highest performance. Garosi *et al.* (2019) assessed the performance of four algorithms: SVM, RF, NB, and generalized linear models (GLM), and reported the best predictive performance by RF, followed by SVM. Similarly, Huang *et al.* (2022) assessed gully erosion susceptibility using RF, SVM, ANN, and GLM and concluded that RF and SVM were better than other algorithms. Sahour *et al.* (2021) compared DL to BRT and multiple linear regression and found that BRT was better than DL in modeling soil erosion. This dissertation also employs RF and SVM together with other algorithms to map gully based on low-cost, high-resolution, remotely sensed data.

Table 1. A non-exhaustive list of supervised algorithms commonly used in gully erosion research.

Name	Abbreviation	Brief Description
k-nearest Neighbor	KNN	This non-parametric algorithm uses distance functions to classify features. The algorithm achieves this by finding the closest pixels to unknown pixels (Thanh Noi and Kappas, 2017; Phinzi, Holb and Szabó, 2021).
Minimum distance	MD	The Minimum Distance (MD) classifier, as described by Richards and Richards (1999), is a non-parametric method used to classify data, wherein each unidentified pixel is assigned to a class with the shortest minimum distance.
Maximum likelihood classifier	MLC	MLC is a parametric classifier that assumes a normal distribution of data. The algorithm assigns each pixel to a class with the highest probability of membership (Bolstad and Lillesand, 1991).
Random forest	RF	Introduced by (Breiman, 2001), RF is a popular ML method that addresses the shortcomings of decision trees by ensembling randomly numerous decision trees to improve predictive performance. RF accomplishes this by bootstrap or bagging aggregation, where multiple predictors are generated using classification trees.
Support vector machines	SVM	SVM is a binary classifier in statistical learning theory (Vapnik, 1999), capable of solving two-class and multi-class classification problems. SVM aims to find a hyperplane (a decision boundary) that best separates the two classes of data points (called support vectors) in a feature space using the so-called <i>kernel trick</i> (Unay and Gosselin, 2007).
Boosting	-	Traditionally developed for classification problems (Valiant 1984), boosting is a classification approach whereby several weak classifiers are combined to form a robust classifier (Kuhn and Johnson, 2013). There are many variants of boosting algorithms, including AdaBoost, stochastic gradient boosting (SGB), and extreme gradient boosting (Xgboost), boosted regression trees (BRT).

Discriminant analysis	DA	DA is a generative classification approach that uses Bayes' theorem to model the conditional distribution of the predictors $\mathbf{X}$ in the response classes $\mathbf{Y} = 1$ (i.e., <i>gully</i> ) and $\mathbf{Y} = 0$ (i.e., <i>non-gully</i> ) (Welch, 1939; Friedman, 1989; Gareth <i>et al.</i> , 2013). The algorithm uses a linear or quadratic function to find one or more linear combinations of predictors that best discriminate or separate the response classes (Alkarkhi and Alqaraghuli, 2018; Kabir, 2021).
Partial Least squares	PLS	PLS intends to form components that capture most of the information in the explanatory variables helpful in predicting the response variable (Garthwaite, 1994; Tang <i>et al.</i> , 2014). It achieves this by constructing linear combinations (components) of the original predictors from which a set of latent variables with the best predictive power is extracted (Abdi, 2003; Chen and Hoo, 2011), then regressing the response variable on these latent variables (Chung and Keles, 2010).
Artificial neural network	ANN	ANN consists of layers with artificial neurons that mimic the biological neurons' function in the human brain. An input layer of neurons (also called nodes or units), one or more hidden layers of neurons, and a final layer of output neurons define a typical architecture of an ANN. Neurons in each layer are connected to other neurons in the next layer, and each connection is associated with a specific weight (Wang, 2003). These weights determine the importance of each predictor in gully extraction, where predictors with larger weights contribute considerably to gully mapping.

### 3.4. Gully mapping using remote sensing

Remote sensing has been defined variously due to its vast applications. This dissertation defines it as the science, technology, or art of collecting data about material objects without physical contact (Campbell and Wynne, 2011; Lillesand, Kiefer and Chipman, 2015). This form of data acquisition is cost-effective and fast relative to traditional methods. Remotely sensed data are collected using various airborne and spaceborne sensors. The former include

high spatial resolution aerial photographs collected by aircraft or unmanned aerial vehicles (UAV). UAVs represent the most advanced means of data acquisition in remote sensing (Alijani, Hasanlou and Azizi, 2018). Armed with multispectral sensors, UAVs combined with structure from motion (SfM) technology are also capable of producing high spatial resolution multispectral images, and LiDAR (light detection and ranging) point clouds from which high-quality digital surface models (DSMs), with submeter spatial resolution, can be derived. LiDAR data have been exploited to characterize and quantify gullies (D'Oleire-Oltmanns *et al.*, 2012; Lannoeye *et al.*, 2016; Wang *et al.*, 2016; Gong *et al.*, 2019). Recently, Niculiță, Mărgărint, and Tarolli (2020) used UAV and LiDAR data for monitoring gully geomorphic changes.

Although accurate and detailed characterization of gullies can be achieved with UAV and LiDAR data, the cost of acquiring these data is relatively high, restricting their application to small areas. For this reason, broadband sensors onboard satellites are commonly used. The most widely used spaceborne sensors with global coverage include Landsat, Sentinel, and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). For example, Landsat data were used to quantify gullies' spatial and temporal extent (Dube *et al.*, 2017). Phinzi and Ngetar (2017) employed three Landsat-derived vegetation indices to map gully erosion, of which the soil-adjusted vegetation index (SAVI) achieved the highest overall classification accuracy (83%). However, they faced obstacles attributable to the relatively coarse spatial resolution (15m x 15m) of the Landsat-8 image. Consequently, they proposed the fusion and pan-sharpening of Landsat-8 imagery with higher-resolution counterparts to enhance the identification and analysis of erosion features. Orti *et al.* (2020) and Makaya *et al.* (2019) mapped gullies at the catchment scale using Sentinel data, yielding 60% and 77% accuracy for gully classification, respectively. Vrieling *et al.* (2007)

employed ASTER imagery to gully mapping and achieved a greater than 90% user's accuracy.

These sensors are free of charge, which explains their widespread applications in gully erosion and other geomorphic features. An essential benefit of using these sensors in gully erosion is their broad spatial coverage, which permits accessibility of extensive remote gullied locations, a current challenge to field surveys. With additional spatial detail, high-resolution (<5m) sensors like Ikonos, GeoEye, WorldView, and RapidEye permit accurate extraction of even small and narrow gully features at pixel and object levels. For instance, Utsumi et al. (2020) extracted gullies from the RapidEye image using object-based image analysis. Shruthi et al. (2015) also used object-based image analysis to quantify temporal changes in gullied areas from Ikonos-2 and GeoEye-1 images. Mararakanye and Nethengwe (2012) extracted gully features from SPOT-5 data. The choice of the sensor used mainly depends on the characteristics of the gullies to be mapped, the scale of the study area (large or small), and image acquisition costs, among other factors. High spatial resolution sensors are generally suited for small area mapping where narrow geomorphic gully features exist, while broadband images are suitable for detecting extensively gullied areas over broad spatial coverage. The 3-meter PlanetScope multispectral data and 1.5-meter pan-sharpened SPOT data, freely available for the study area, offer tremendous possibilities for mapping gullies but have yet to be fully exploited in previous investigations.

### **3.5. Gully susceptibility modeling**

In order to recommend and execute effective preventive measures for reducing the risk of new gullies and impeding or reversing the growth of existing ones (Kirkby and Bracken, 2009), it is imperative to have a

comprehensive understanding of the susceptibility of different areas to gully formation. Besides merely extracting gully features from satellite images, ML algorithms can also predict or simulate areas susceptible to gully erosion. This is accomplished by establishing a relationship between the factors that contribute to the formation of gullies and the actual occurrence of gullies, thereby enabling the forecasting or modeling of vulnerable regions.

Numerous scholars have effectively developed models to predict gully susceptibility at catchment scales in different regions across the globe, including China (Huang *et al.*, 2022), Iran (Arabameri, Chen, *et al.*, 2019; Hosseinalizadeh *et al.*, 2019a), Democratic Republic of the Congo (Kulimushi *et al.*, 2023), Algeria (Dewitte *et al.*, 2015), India (Roy and Saha, 2021), Brazil (Lana, Castro and Lana, 2022), South Africa (Bernini *et al.*, 2021), United States of America (Han, Guzman and Chu, 2022), Italy (Conoscenti *et al.*, 2014), and Morocco (Hitouri *et al.*, 2022). Researchers use various ML methods to assess their performance to identify the most accurate predictive model. For instance, Gayen *et al.* (2019) utilized four ML techniques, including multivariate additive regression spline (MARS), flexible discriminant analysis (FDA), RF, and SVM, to map gully susceptibility and found RF to yield the highest accuracy. Chowdhuri *et al.* (2021) used ANN, CNN, and deep NN (DNN) to predict gully susceptibility and discovered that DNN outperformed the other methods. Hosseinalizadeh *et al.* (2019a) employed FT, NBTree, and RF models to forecast gully headcut susceptibility and identified RF as the most effective method. Similarly, Huang *et al.* (2022) applied RF, SVM, ANN, and GLM to gully susceptibility, and RF was the most accurate model. However, many studies prioritize the accuracy of ML models for prediction and tend to overlook computation time, which is also a crucial aspect of performance evaluation. Previous research has suggested the importance of considering both aspects in performance evaluation (Gislason,

Benediktsson and Sveinsson, 2006; Belgiu and Drăgu, 2016). Indeed, it is desirable for an ML model to achieve high accuracy while minimizing computation time, especially in gully modeling, where several features are involved.

In addition to gully susceptibility modeling, ML-based methods enable the assessment of various factors and their interactions, presenting a significant potential for identifying the most crucial factors influencing gully erosion (Vanmaercke *et al.*, 2021). For instance, Rahmati *et al.* (2022) investigated the factors contributing to the initiation of gully erosion, encompassing physical factors such as topography, soil properties, and lithology, anthropogenic factors such as land use and distance from roads, and other relevant geo-technical and hydrological factors. The authors determined that distance from roads was the most influential factor in the onset of gully erosion. Similarly, Amiri *et al.* (2019) utilized the Boruta algorithm to evaluate the significance of effective factors in gully erosion and identified land use, distance from rivers, and clay percentage as the most critical factors for the occurrence of gully erosion in their study. Although progress has been made in gully susceptibility modeling, a significant research gap exists in our understanding of how various machine learning (ML) models might perform with different subsets of predictor variables in terms of accuracy and computation time. This gap arises from the common practice in gully susceptibility studies of employing a fixed set of predictor variables, which hinders our ability to gain insights into the relative performance of ML models when using feature sets of varying sizes.

### **3.6. Resampling techniques**

It is of utmost importance for the model to generalize well on unseen data and one way of ensuring this is to split the data into train-test sets where the training set is used to build the model while the testing set is used to

evaluate the model's performance. However, evaluating the model on a single test (or *hold-out*) set tends to be highly variable and unreliable, especially when using a small data set, i.e.,  $n < 1000$  (Hawkins, Basak and Mills, 2003; Molinaro, Simon and Pfeiffer, 2005). Two commonly used resampling techniques, including the  $k$ -fold cross-validation and bootstrapping, overcome this limitation by repeatedly re-fitting a model to parts of the training set and evaluating its performance on other parts (Boehmke and Greenwell, 2019).

$K$ -fold cross-validation (CV) randomly partitions the training data into small  $k$  groups or folds of nearly equal sizes in which the first fold is used in the first iteration as a testing set while the remaining folds are treated as the training data (Gareth *et al.*, 2013). Figure 3 illustrates this procedure. In subsequent iterations, the training data is reshuffled, a different subset is used for model testing, and the rest is considered the training set. This procedure is repeated  $k$  times with a different group of observations treated as a test set each time (Gareth *et al.*, 2013). The generalization error is recorded for all iterations, resulting in a  $k$  estimate of the generalization error. Thus, the  $k$ -fold cross-validation is calculated by averaging the  $k$ -test errors, approximating the expected error when the model is applied to unseen data (Boehmke and Greenwell, 2019). Although there is no formal rule regarding the size of  $k$ , 5-fold or 10-fold are typically used in literature (Abdi, 2020; Csatáriné Szabó *et al.*, 2020; Phinzi, Abriha and Szabó, 2021; Varga *et al.*, 2021; Saha *et al.*, 2022). These  $k$  values are preferred because they have proven empirically to yield reasonable generalization error estimates that do not typically suffer from excessively high bias and variance (Gareth *et al.*, 2013). Furthermore, a repeated 10-fold CV is often suggested for smaller data sets ( $n < 10000$ ) as it improves the accuracy of the estimated generalization error and provides an estimate of its variability (Kim, 2009; Boehmke and Greenwell, 2019).

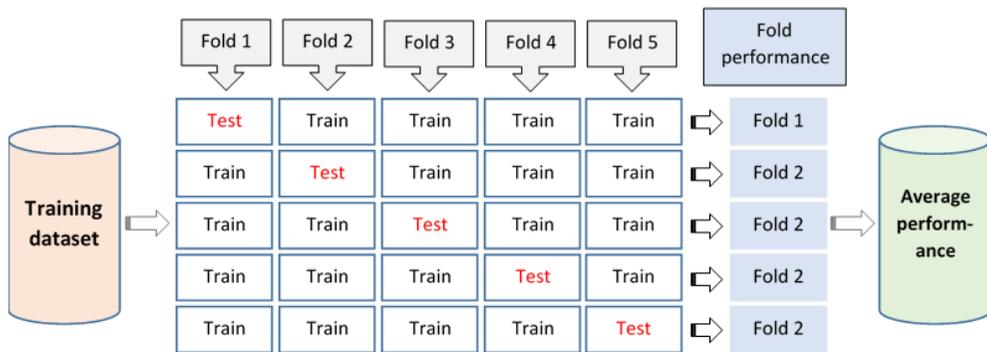


Figure 3. A typical  $k$ -fold cross-validation procedure (adapted from Boehmke and Greenwell 2019).

Bootstrapping is a widely applied resampling technique that can quantify an algorithm’s uncertainty (Gareth *et al.*, 2013) while improving its stability and avoiding overfitting. It involves sampling (with replacement) a set of observations from the original data randomly, resulting in a bootstrap sample of the same size as the original data (Figure 4).

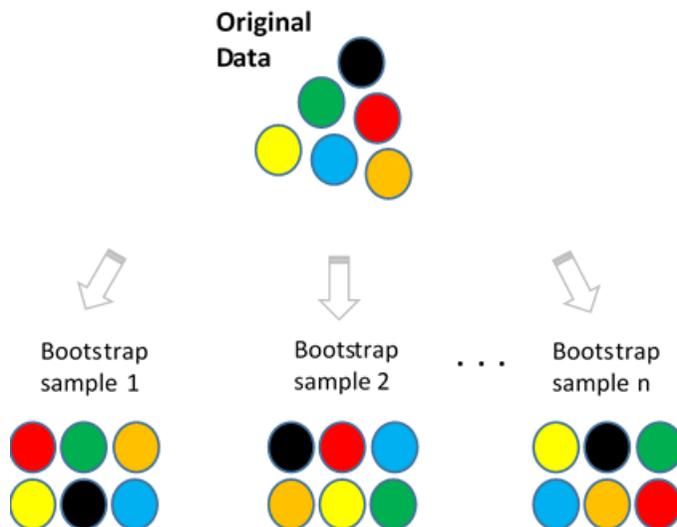


Figure 4. A typical bootstrapping procedure (adapted from Boehmke and Greenwell 2019).

Observations in the bootstrap sample are used to build a model, whereas the original observations outside a specific bootstrap sample, referred to as *out-of-bag* (OOB), validate the model.

Given that the samples are selected with replacement, each bootstrap sample may contain duplicate values, potentially leading to lower variance in the error estimate (Efron, 1983; Boehmke and Greenwell, 2019). It is highly probable that similar predicted values will result, even if a different set of bootstrapped samples are used to construct multiple models (Kuhn and Johnson, 2013). While this issue can pose a challenge when utilizing smaller datasets, it is frequently insignificant for average to large datasets (Boehmke and Greenwell, 2019).

### **3.7. Evaluation metrics**

The effectiveness of an algorithm is ultimately determined by its ability to perform well on unseen data in the real world. There is no singular, precise method for evaluating the performance of a model, but rather a variety of accuracy metrics that can be derived from the contingency table, also known as the confusion or error matrix (Congalton, 1991). This matrix summarizes each possible combination of predicted and actual values resulting in four possible outcomes, including true positive (TP), false negative (FN), false positive (FP), and true negatives (TN). For example, in the context of an algorithm for predicting gully pixels, TP occurs when the algorithm correctly identifies a gully pixel. Conversely, TN occurs when the algorithm correctly identifies a non-gully pixel, while FP occurs when it mistakenly predicts a non-gully pixel as a gully. On the other hand, FN occurs when the algorithm fails to predict the presence of a gully when one exists. Several evaluation metrics can be computed, including overall accuracy (OA), sensitivity or producer's accuracy (PA), specificity or user's accuracy (UA), the area under the curve (AUC), kappa coefficient, and F1-score. Table 2 summarizes these

accuracy metrics. These accuracy metrics highlight distinct quality aspects and may necessitate careful analysis (Foody, 2020). A comprehensive explanation of each accuracy metric and appropriate examples can be found in Boehmke and Greenwell (2019) and Irizarry (2019).

Table 2. Accuracy metrics commonly used in gully erosion (TP=true positive, FN=false negative, FP=false positive, and FN=false negative, Po=observed agreement, and Pe=agreement expected by chance).

Accuracy metric	Brief description	Formula
Overall accuracy (OA)	It measures the proportion of correct predictions made by the model out of all the predictions.	$\frac{TP + TN}{TP + TN + FN}$
Sensitivity (recall)	Also known as producer's accuracy (PA), this metric measures the proportion of positive instances that the model correctly identifies.	$\frac{TP}{TP + FN}$
Specificity	Also called user's accuracy (UA), it measures the proportion of negative instances that the model correctly identifies	$\frac{TN}{TN + FP}$
Precision	It calculates the proportion of true positive predictions made by the model among all the positive predictions.	$\frac{TP}{TP + FP}$
Kappa coefficient	It is the ratio of the observed agreement between the model and ground truth to the agreement that would be expected by chance.	$\frac{Po - Pe}{1 - Pe}$
F1-score	The F1-score is the harmonic mean of precision and recall, providing a trade-off between the two metrics.	$2 \times \frac{precision \times recall}{precision + recall}$

The Kappa coefficient and OA have been frequently utilized for evaluating the overall classification performance of an algorithm, as noted by

Congalton (1991). One primary concern regarding using OA is that it can be highly susceptible to fluctuations in class prevalence, meaning that its value can be influenced by the relative abundance of different classes (Foody, 2020).

Consequently, the Kappa coefficient is viewed as a reliable measure of agreement, considering the algorithm's predictions of TP, TN, FP, and FN (Congalton, Oderwald and Mead, 1983). In addition, it offers a more nuanced understanding of agreement than basic metrics, such as OA, and is particularly useful when the distribution of positive and negative instances in the data is unbalanced. However, in recent years, the use of Kappa in remote sensing classification accuracy has decreased in popularity (Thanh Noi and Kappas, 2017; Heydari and Mountrakis, 2018). This is because several researchers, including Pontius Jr and Millones (2011), Flight and Julious (2015), and Delgado and Tibau (2019), have cautioned against relying on Kappa due to its limitations. One significant limitation of Kappa is its high sensitivity to the distribution of marginal totals, which can result in unreliable outcomes (Flight and Julious, 2015).

In light of the issues associated with OA and the Kappa coefficient, scholars have been advised to present both class-level and overall metrics in their reports (Olofsson *et al.*, 2014; Congalton and Green, 2019). An ideal classifier should possess high sensitivity (true positive rate) and precision, meaning it should be able to make accurate predictions regarding the presence and absence of an event while minimizing false positives and negatives (Boehmke and Greenwell, 2019). The F1-score, a metric derived from the harmonic mean of precision and recall, is a reliable indicator of a model's performance per class. However, the performance of individual classes may be more effectively measured by the AUC, which represents the area under the ROC curve. The ROC curve plots the true positive rate against the false positive rate for various classification thresholds. A higher AUC value

indicates superior classification performance and better discrimination between positive and negative instances. Alternatively, scatter plots depicting sensitivity (PA) versus specificity (UA) are frequently utilized, along with a predefined accuracy threshold, to assess the performance of a classification model at the class level (Phinzi *et al.*, 2020; Likó *et al.*, 2022).

## **4. METHODS AND MATERIALS**

### **4.1. Study area description**

This research was conducted in four rural study areas (#1-#4) in Eastern South Africa that suffer from severe soil erosion in the form of gullying (Figure 5). Despite differences in physical and geographical features (Figures 6 and 7), all four areas share a semi-arid climate with cold, dry winters and hot, rainy summers, with the highest rainfall occurring between November and January. The study areas experience an average annual rainfall of 511-671 mm (Figure 7c), with temperatures ranging from 7 to 30 °C. The study areas span an elevation range of 837-1604 m (Figure 6a), characterized by excessively steep slopes exceeding 60% (Figure 6b). The geological composition underlying these study areas comprises Clarens, Elliot, Molteno (CEM), Suurberg, Drakensberg, Lebombo (SDL), Tarkastad, and Adelaide rock formations (Figure 6c). The predominant soil types are Luvisols, with a lesser presence of Ferralsols and Planosols (Figure 6d). The Transitional and Temperate Forest Scrub Type (TTFST) emerges as the prevailing vegetation type throughout all the study areas (Figure 7a). The prevailing climatic conditions in the study area facilitate agricultural (i.e., cultivation) pursuits that primarily define the land use patterns observed (Figure 7b). The subsequent sections provide descriptions of the distinctive characteristics defining each study area.

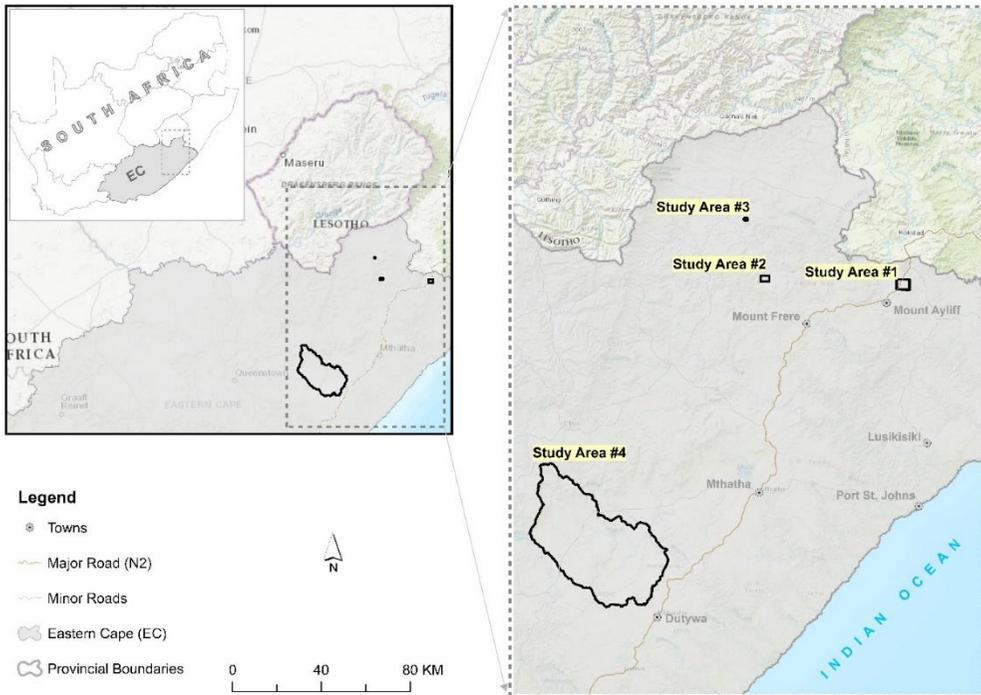


Figure 5. Location of four study areas in the Eastern Cape (EC) Province of South Africa.

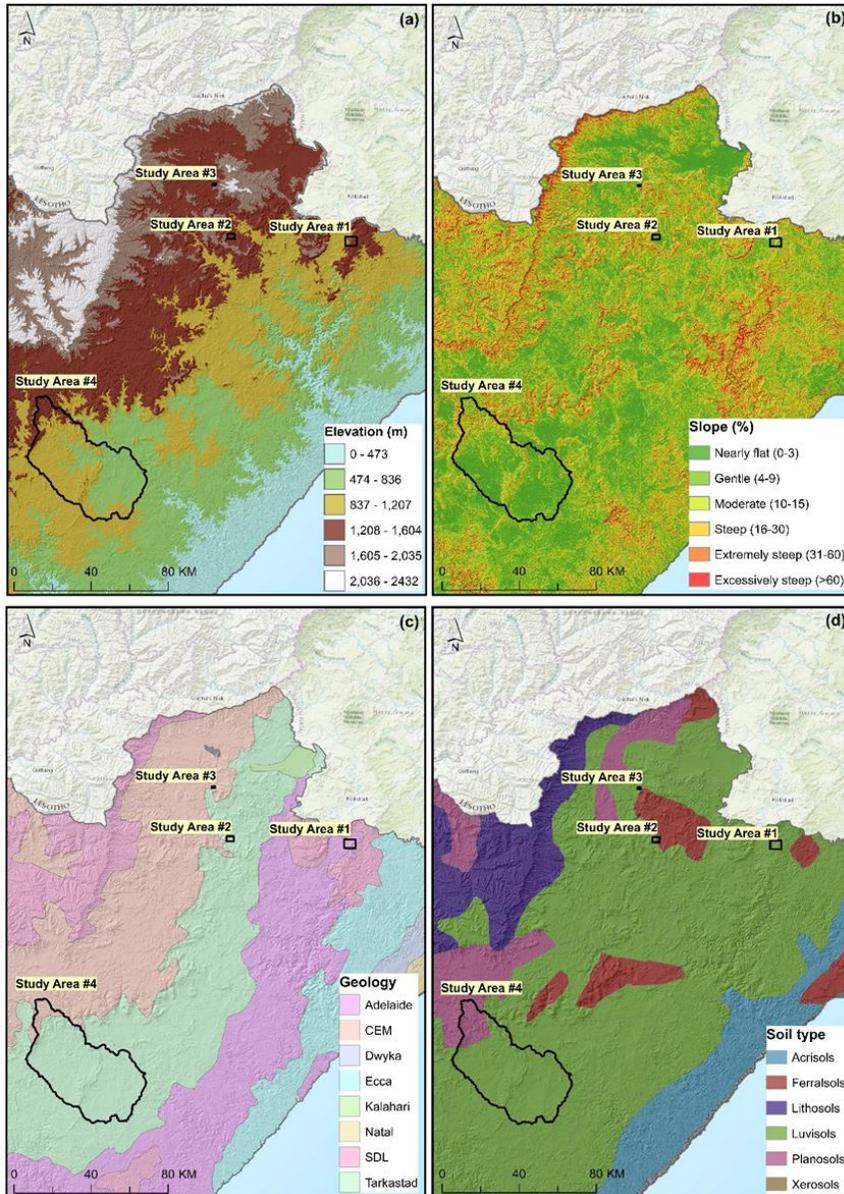


Figure 6. Topographic and geo-environmental characteristics of the study areas: (a) elevation, (b) slope, (c) geology, and (d) soil type. Note: CEM (Clarens, Elliot, Molteno), SDL (Suurberg, Drakensberg, Lebombo).

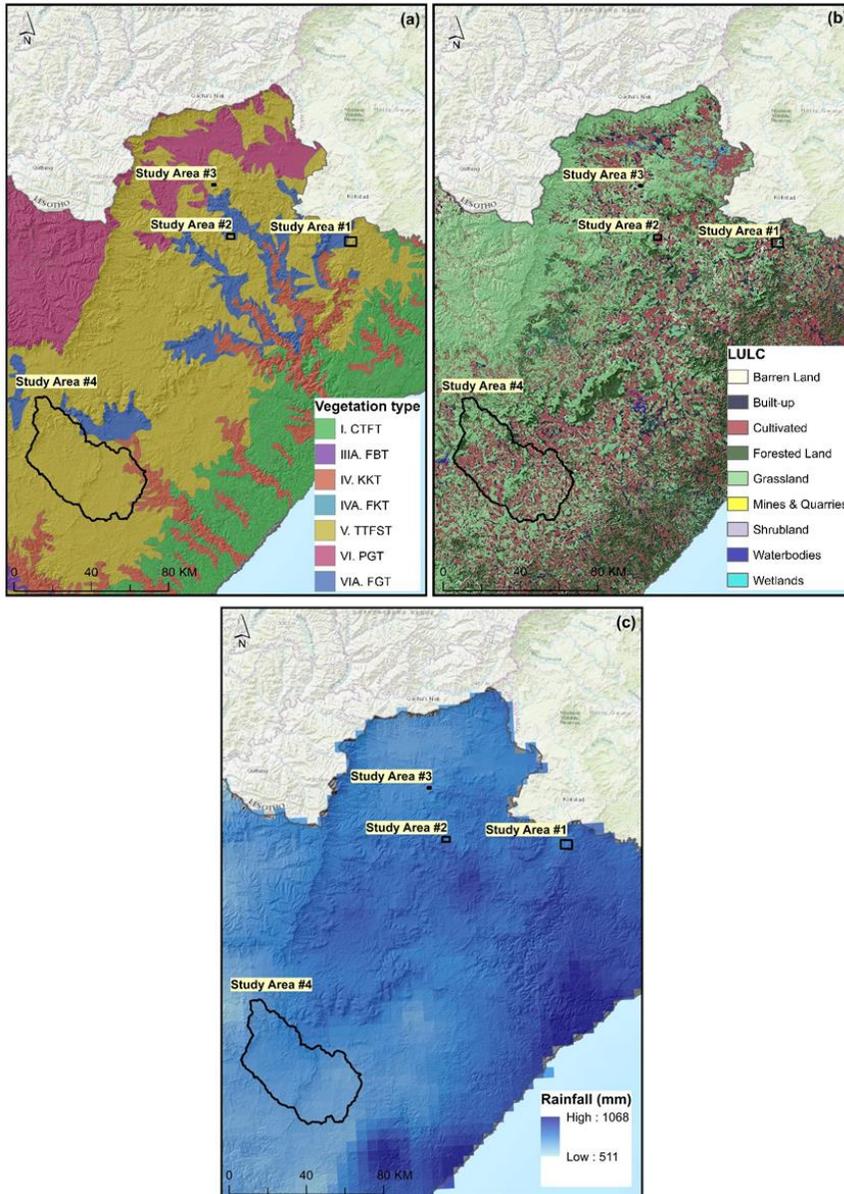


Figure 7. Physical and climatological characteristics of the study areas: (a) vegetation types, (b) land use/land cover (LULC), and (c) rainfall. Note: CTFT (Coastal Tropical Forest Types), FBT (False Bushveld Types), KKT (Karoo and Karroid Types), FKT (False Karoo Types), TTFST (Temperate and Transitional Forest and Scrub Types), PGT (Pure Grassveld Types), FGT (False Grassveld Types).

#### 4.1.1. Study Area #1

The study area comprises three small sites (1A - 1C), each covering 1.26 km<sup>2</sup>, as shown in Figure 8. The area is characterized by extensive erosion features, particularly gullies on gentle slopes and rills on steep slopes. Land use is mixed, including rural settlements, unpaved roads, and agriculture. The topography varies, with elevations ranging from about 1098 m in central areas to over 1500 m in hilly northern and eastern regions. The geological setting of the area is composed of Adelaide rock formations overlain by Luvisols as a dominant soil type. The vegetation in the area is primarily grassland, located in elevated and mountainous areas. On the other hand, low-lying areas, where human activities like farming and settlement occur, have minimal vegetation cover. A pan-sharpened multispectral SPOT-7 image was utilized to map gullies in sites 1A - 1C.

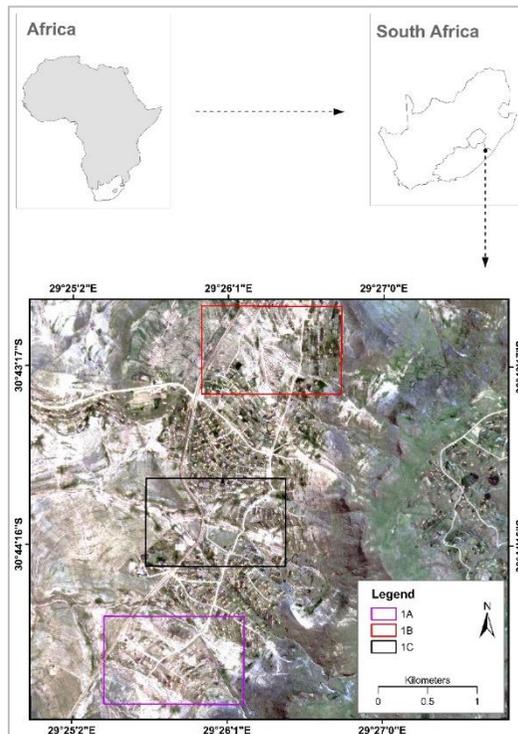


Figure 8. The geographical location of the study area showing three gully sites (1A, 1B, and 1C).

#### 4.1.2. Study Area #2

The area covers approximately 10 km<sup>2</sup> (Figure 9). The primary land use types are subsistence agriculture (crop farming and livestock rearing) and settlement. The most common vegetation type is grassland, with some forest patches in the northwestern part of the study area. The topography ranged from 1213 m – 1658 m, with the northwestern and southwestern parts being steeper than other sections. The area's geomorphology is characterized by steep mountain slopes with gently undulating foot slopes (Le Roux and Sumner 2012).

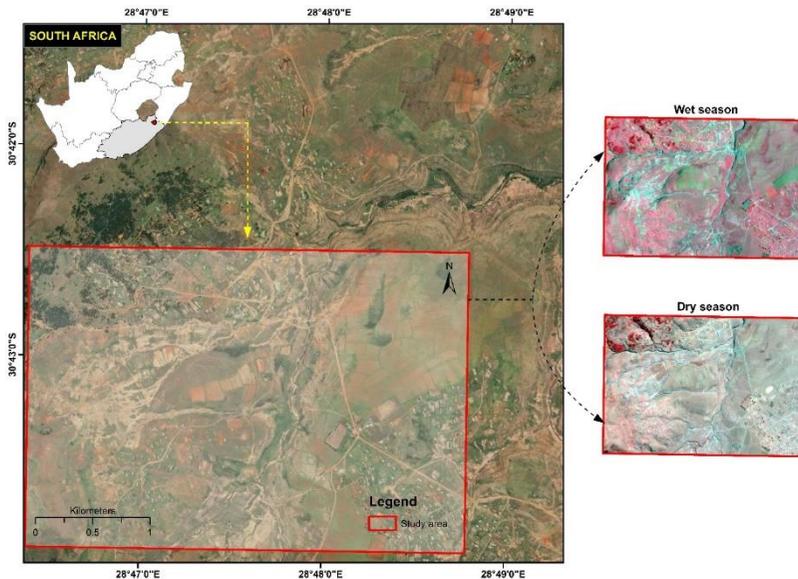


Figure 9. Location of the study area (PlanetScope false-color images).

The area is underlain by mudstone and sandstone of the Beaufort Group (Hilbich *et al.*, 2007), which includes the lower Adelaide Subgroup . Soils derived from these rocks are inherently erodible. Indeed, the extensive erosion in the area is potentially due to the prevalence of duplex and dispersive soils (van Breda Weaver, 1991; Beckedahl and de Villiers, 2000). The area comprises continuous and discontinuous gully networks with various

characteristics, including narrow, wide, vegetated, shallow, deep with shadows, and so on (Le Roux and Sumner 2012; Phinzi et al. 2021b). Gully mapping was conducted using PlaneScope data obtained during the dry and wet seasons.

#### 4.1.3. Study Area #3

The study area has a surface area of 1.47 km<sup>2</sup> and is characterized by permanent gullies of varying length, depth, and width (Figure 10). These gully characteristics are observable at four locations (sites 3A – 3D) within the study area. Gullies in sites 3A and 3D have a similar pattern but differ in depth, whereas gullies in sites 3B and 3C are primarily linear and elongated. The elevation ranges from 1445 m to 1584 m, with the highest values in the western parts and the lowest in the far southeastern parts.

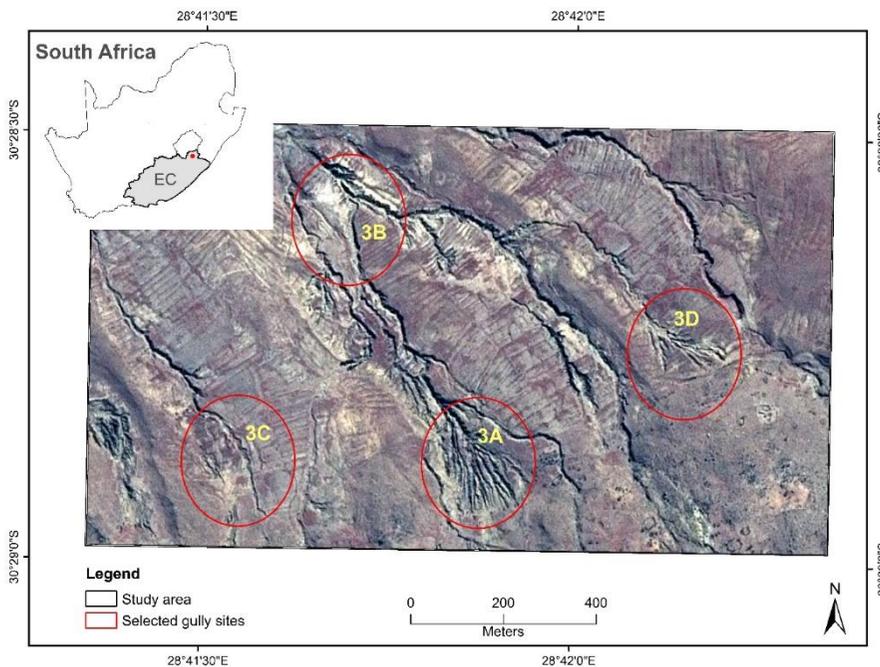


Figure 10. Location of the study area and selected gully sites (3A - 3D).

Highland Sourveld Grassland (Acocks, 1988) is the primary vegetation, and the area is predominantly used for livestock grazing, making it susceptible to gully erosion. In addition, the predominant geological composition of the area consists of mudstone and sandstone derived from the Beaufort Group (Hilbich et al. 2007), encompassing the lower Adelaide Subgroup and the upper Tarkastad Subgroup. The entire area is characterized by the presence of Luvisols throughout. The visual range SPOT-7 image was employed for gully mapping.

#### **4.1.4. Study Area #4**

The study was conducted in a tertiary catchment, covering an area of approximately 2145 km<sup>2</sup> with an elevation range of 534-1772 m above sea level (Figure 11). The area is drained by the Mgwali River, a tributary of the Mbashe River system. Grassland is the dominant natural vegetation type, with scattered forest patches. Agriculture, mainly consisting of subsistence crops and livestock farming, is the main land use activity, although there are a few commercial farms. The underlying geological rock types are Mudstones and sandstones of the Tarkastad Formation and Molteno Formation, which are prone to erosion due to their easily weathered parent material (DWA, 2010). Soils in the area are predominantly Chromic Luvisols, which have a high silt content (ISRIC, 2002), covering about 87% of the catchment, while Eutric Planosols and Solodic Planosols, commonly found in hilly areas, cover the remaining 13%. Planosols are highly unstable soils, making them susceptible to gully erosion (Du Plessis *et al.*, 2020).

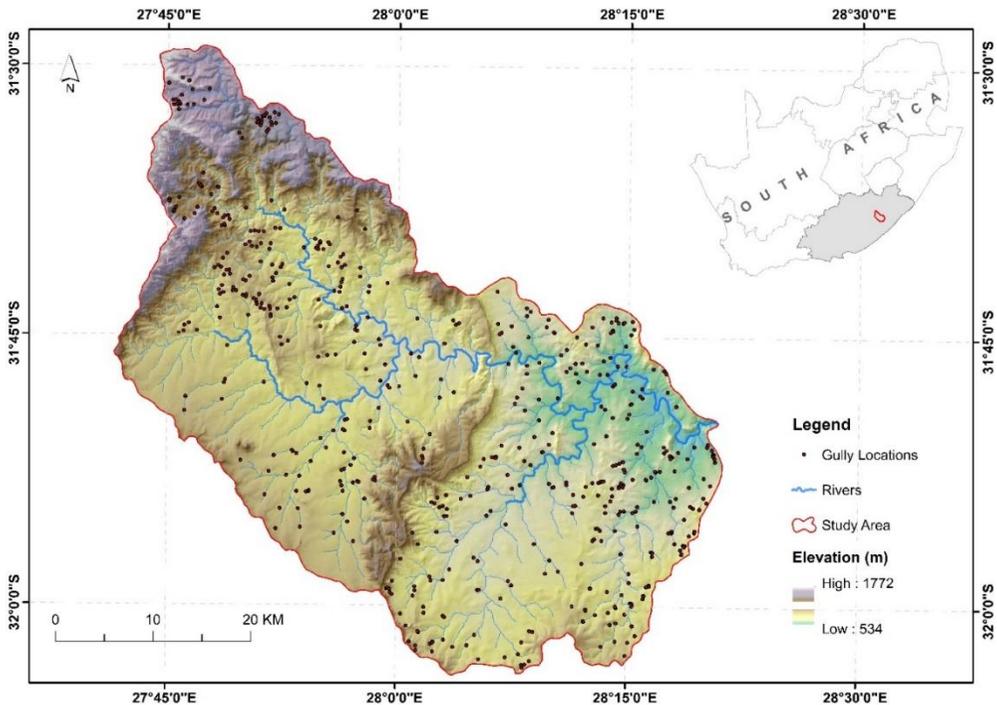


Figure 11. Study area map showing the distribution of gully locations (data source: shuttle radar topography mission – SRTM DEM).

#### 4.2. Data

The remote sensing data used in this study, comprising cloud-free satellite imagery and elevation data, were acquired at no expense. The images were obtained from the South African National Space Agency (SANSA), including multispectral and visual range SPOT-7 scenes from June 2017. The visual range image had a resolution of 1.3 m and consisted of red, green, and blue (RGB) bands, while the multispectral image had a resolution of 5.5 m and included RGB and near-infrared (NIR) bands, as well as a high-resolution panchromatic band of 1.5 m. To correct for atmospheric reflectance, the images were converted to top-of-atmosphere reflectance using the “atmospheric reflectance function” in ArcGIS 10.4. Additionally, the low

geometric resolution of the multispectral image was enhanced using the Gram-Schmidt pan-sharpening method in ENVI software (Maurer, 2013; Grochala and Kedzierski, 2017; Abriha *et al.*, 2018).

The study also utilized PlanetScope multispectral data, specifically Level 3B images for wet (January 2017) and dry (June 2017) seasons. The images were obtained from the Planet explorer website (<https://www.planet.com/explorer>, accessed on 30 July 2020), which provides orthorectified scenes that had already undergone necessary radiometric and geometric corrections and were projected to the Universal Traverse Mercator (UTM) projection system based on the world geodetic system (WGS84) datum. The PlanetScope image contained four spectral bands, including RGB and near-infrared bands, with a spatial resolution of 3m and a temporal resolution of 1 day. To obtain a digital elevation model (DEM), a 30 m Shuttle Radar Topography Mission (SRTM) was downloaded from the Earth Explorer website (<https://earthexplorer.usgs.gov/>) on 26 January 2021.

### **4.3. Geo-environmental variables**

Gully erosion is influenced by various topographic, environmental, hydrological, and socio-economic factors, collectively known as geo-environmental predictors, features, variables, or covariates. For this dissertation, 22 geo-environmental predictors were selected based on our research objectives, data availability, and relevant literature (Valentin, Poesen and Li, 2005; Le Roux and Sumner, 2012; Conoscenti *et al.*, 2014; Phinzi and Njoya Silas Ngetar, 2019; Ghaedi and Shojaian, 2020; Ebhuoma *et al.*, 2022) (see Table 3). Appendix 1 illustrates the spatial distribution of these geo-environmental predictors. Elevation data was derived from the void- SRTM DEM from which other topographic factors, including slope, aspect, plan curvature, profile curvature, slope length and steepness (LS-factor), topographic wetness index (TWI), terrain ruggedness index (TRI), stream

power index (SPI), distance from rivers, and drainage density, were computed using ArcGIS 10.4 software (ESRI, 2022).

SPI plays a crucial role as a predictor variable in assessing gully susceptibility, as it not only quantifies the erosive capacity of surface runoff but also identifies areas with a high potential for gully formation. Calculation of the SPI was performed using the equation (1) proposed by Moore et al. (1991):

$$SPI_i = \ln(CA_i) \times \tan(G_i) \quad (1)$$

Where  $i$  is the grid cell,  $CA_i$  is the catchment area (m), and  $G_i$  is the slope gradient at the grid cell.

TWI is a significant topographic characteristic in evaluating gully susceptibility, as it defines regions within a catchment that accumulate high amounts of water. Calculation of this topographic attribute was performed using the following equation (2):

$$TWI = \ln\left(\frac{CA}{\tan\beta}\right) \quad (2)$$

Where  $CA$  is the catchment area, and  $\beta$  is the slope gradient.

The slope length and steepness, collectively referred to as the LS factor, account for the overall effect of topography on soil erosion. The LS factor was calculated based on the following empirical equation (Moore and Burch, 1986):

$$LS = \left(\frac{A}{22.13}\right)^m \times \left(\frac{\sin\beta}{0.0896}\right)^n \quad A = (\text{Flow accumulation} \times \text{Cell size}) \quad (3)$$

Where  $A$  is the upslope contributing area per unit cell (m),  $m$  (0.4) is a variable slope length exponent,  $\beta$  is the slope gradient, and  $n$  (1.3) is a slope steepness exponent.

Soil data consisting of soil physical (organic matter content, sand, silt, and clay content) and chemical properties such as cation exchange capacity (CEC), calcium carbonate (CaCO<sub>3</sub>), and pH were obtained from the digital soil map of the world (FAO, 2003). The soil erodibility, represented by the K-factor, was computed in ArcGIS 10.4 using the empirical relation of Williams (1995) based on the soil's physical properties (Equation 4):

$$K_{USLE} = f_{csand} * f_{cl-si} * f_{orgc} * f_{hisand} \quad (4)$$

$$f_{csand} = \left( 0.2 + 0.3 * \exp \left[ -0.256 * m_s * \left( 1 - \frac{m_{silt}}{100} \right) \right] \right)$$

$$f_{cl-si} = \left( \frac{m_{silt}}{m_c + m_{silt}} \right)^{0.3}$$

$$f_{orgc} = \left( 1 - \frac{0.0256 * orgC}{orgC + \exp[3.72 - 2.95 * orgC]} \right)$$

$$f_{hisand} = \left( 1 - \frac{0.7 * \left( 1 - \frac{m_s}{100} \right)}{\left( 1 - \frac{m_s}{100} \right) + \exp[-5.5 + 22.9 * \left( 1 - \frac{m_s}{100} \right)]} \right)$$

where the  $f_{csand}$  is a factor that gives low soil erodibility values for soils with high coarse sand content and high values for soils with little sand,  $f_{cl-si}$  is a factor that gives low soil erodibility for soils with high clay-to-silt ratios,  $f_{orgc}$  is a factor that reduces soil erodibility for soils with high organic carbon content,  $f_{hisand}$  is a factor that reduces soil erodibility for soils with extremely high sand content,  $m_s$  is the percent sand content (0.05 – 2.0 mm diameter particles),  $m_{silt}$  is the percent silt content (0.002 – 0.05 mm diameter particles),  $m_c$  is the percent clay content (<0.002 mm diameter), and  $orgC$  is the percent organic carbon content of the layer (%).

The normalized difference vegetation index (NDVI) was computed from a cloud-free Landsat-9 Operational Land Imager (OLI) acquired on 08

February 2022, which we downloaded from the USGS website. The land use/land cover (LULC) map for the study area was extracted from the South African National Land Cover (SANLC) dataset available on the Department of Forestry, Fisheries, and Environment website. The study area comprised eight LULC classes, including built-up land, barren land, cultivated land, forest, grassland, mines and quarries, water bodies, and wetlands.

The land type map was prepared from the South African Land Type Survey database (Land Type Survey Staff, 1972-2006). The map comprises nine broad land types grouped according to the prevailing climate, terrain, and dominant soil types found within the land type (Van Zijl, Le Roux and Turner, 2013; Du Plessis *et al.*, 2020). Distance from roads was computed from the roads network data available at <https://www.hotosm.org/>. Geology for this research consisting mainly of Mudstones and sandstones of the Tarkastad Formation and Molteno Formation was downloaded from the South African National Space Agency (SANSA). Long-term (1981-2021) annual gridded rainfall data from the Climate Hazards Group Infrared Precipitation (CHIRPS) (Funk *et al.*, 2015) were used. The CHIRPS product provides quasi-global rainfall estimates at a spatial resolution of 0.05° and was resampled to 30 m. Finally, the 2020 population density dataset was downloaded from the WorldPop database ([www.worldpop.org](http://www.worldpop.org)) in Geotiff format at a spatial resolution of 100 m.

Table 3. Geo-environmental predictors considered for gully susceptibility modeling (SRTM= Shuttle Radar Topography Mission, DEM=digital elevation model, TWI=topographic wetness index, LS=slope length and steepness, TRI=terrain ruggedness index, NDVI=normalized difference vegetation index, LULC=land use/land cover, CEC=cation exchange capacity, CaCO<sub>3</sub>=calcium carbonate, SPI=stream power index).

	<b>Predictors</b>	<b>Class range</b>	<b>Data source</b>
topographical	Slope	0 – 67.07°	SRTM DEM ( <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a> )
	Aspect	9 classes	SRTM DEM ( <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a> )
	Plan curvature	-9 – 8	SRTM DEM ( <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a> )
	Profile curvature	-10 – 11	SRTM DEM ( <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a> )
	LS-factor	0 – 3228.69	SRTM DEM ( <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a> )
	TWI	2.73 – 24.99	SRTM DEM ( <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a> )
	Elevation	534 – 1772 m	SRTM DEM ( <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a> )
	TRI	0.89 – 0.11	SRTM DEM ( <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a> )
Environmental	NDVI	-0.23 – 0.72	Landsat-9 ( <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a> )
	LULC	8 classes	Department of forestry, fisheries and environment ( <a href="https://egis.environment.gov.za/">https://egis.environment.gov.za/</a> )
	Land type	9 classes	Land Type Survey Staff (1972-2006) Roads shapefile
	Distance from roads	0 – 12647.2 m	( <a href="https://www.hotosm.org/">https://www.hotosm.org/</a> )
	K-factor	0.06 – 0.12	FAO world soil database
	Soil pH	6.2 – 6.9	FAO world soil database
	CEC	8.4 – 13.1 cmol/kg	FAO world soil database
	CaCO <sub>3</sub>	0 – 1	FAO world soil database
Geology	2 classes	South African National Space Agency ( <a href="http://atlas.sansa.org.za/atlas-geology.html">http://atlas.sansa.org.za/atlas-geology.html</a> )	
Hydrological	SPI	-13.82 – 13.35	SRTM DEM ( <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a> )
	Distance from rivers	0 – 3079.5 m	SRTM DEM ( <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a> )
	Drainage density	0 – 4.51 km/km <sup>2</sup>	SRTM DEM ( <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a> )
	Rainfall	651.94 – 766.19 mm	CHIRPS (Funk <i>et al.</i> , 2015)
Socio-economic	Population density	0 – 85 people/km <sup>2</sup>	WorldPop database ( <a href="http://www.worldpop.org">www.worldpop.org</a> )

#### 4.4. Variable selection

Choosing the most suitable predictors for the model is crucial to avoid overfitting and multicollinearity, which in turn improves the predictive performance of the model. The recursive feature elimination (RFE) algorithm executed in R software (R Core Team, 2021) was used to identify and remove uninformative predictors from the dataset. This process involves removing one predictor at a time through iteration until the optimal subset of predictors is achieved (Csatáriné Szabó *et al.*, 2020; Varga *et al.*, 2021). A random forest-based RFE was used with a 10-fold cross-validation, repeated five times. To check for multicollinearity, the correlation matrix, tolerance, and variance inflation factor (VIF) were computed. Collinearity was considered present if bivariate correlations among predictors exceeded 0.75, tolerance was less than 0.1, and VIF was greater than 10 (Mason and Perreault Jr, 1991; Gareth *et al.*, 2013; Kuhn and Johnson, 2013; Vatcheva *et al.*, 2016). If collinearity was detected, the variables concerned were removed, resulting in an optimal feature set of sixteen predictors. The relative importance of these predictors was then ranked using RF-based variable importance. From this optimal feature set, three feature sets were generated: a large set with 16 predictors, a medium set with 12 predictors, and a small set with 6 predictors. The small feature set comprised the top six most important predictors. The medium set consisted of the small set and an additional six important predictors, while the large set included all sixteen predictors. These feature sets were used to examine the performance of the ML models when the number of predictors varied. The terms "large," "medium," and "small" sets were used throughout the manuscript for convenience.

#### **4.5. Reference data collection for gully mapping and susceptibility modeling**

The reference data for all study areas were collected based on field surveys, satellite images, and Google Earth image interpretation. While the main objective of this dissertation was to map gullies, including other land cover classes in the classification process was crucial for effectively distinguishing gullies from the surrounding land cover. Areas where the land cover was identifiable in the field, and the images were delineated in each study area. For example, in the case of Study Area #1, seven land cover classes were distinguished on pan-sharpened SPOT-7 data, including stressed vegetation (SV), bare soil (BS), gully (G), mixed bare soil (MS) (exposed rocks, unpaved roads, and bare soils), dense vegetation (DV), roads (R) and, settlement (S). The classification process was conducted utilizing RF, SVM, and LDA algorithms within the R software. In addition, the two-class approach was also investigated, where all non-gully classes were reclassified into one class. Throughout this dissertation, the seven-class solution is called the ‘multi-class’ (m) approach, while the two-class solution is a ‘binary’ (b) approach. A stratified random sampling of the entire dataset was conducted, with 20795, 31784, and 22512 pixels collected for sites 1A, 1B, and 1C, respectively. A subset of 1000-1000 cases was randomly selected for the binary approach, while 350 cases per category were chosen for the multi-class approach to avoid autocorrelation. A 10-fold CV with three repeats was applied within the R software environment, where the entire dataset (pixels) was randomly divided into ten subsamples, of which nine subsamples were used for model training while one subsample was used for testing. This process was repeated until all subsamples were tested against each other and required no separate training or testing database (Brownlee, 2014; Heckel *et al.*, 2020).

In Study area #2, the identification and classification of gullies from the PlanetScope image were carried out using random RF and SVM algorithms implemented in the Python software. Prior to this procedure, the area was classified into seven different land cover categories for each of the two seasons: forest, agriculture, built-up areas, bare soil, gullies, and mixed bare soil that consisted of exposed rocks, unpaved roads, and exposed soil, primarily in plowed fields. The reflectance values of gullies during two different seasons were compared by computing the Normalized Difference Vegetation Index (NDVI). A total of 17,757 pixels for the wet season and 30,597 pixels for the dry season were extracted from the PlanetScope image. The 5-fold CV was repeated 20 times, while bootstrapping was repeated 100 for each season. In the former, the reference data were split into train and test sets, with a 50:50 ratio, while in the latter, a sample of the same size as the original dataset was obtained by repeatedly drawing random samples (100 times), with replacement, from the original dataset.

In Study Area #3, four land cover classes were distinguished: grassland, stressed vegetation, gully, and bare soil. Training data for each land cover type was obtained by digitizing polygons. Collecting training pixels using polygons offers a significant advantage in terms of efficiency and time-saving. This method allows for the swift collection of pixels falling within a designated polygon, as opposed to the laborious process of manually collecting individual pixels using points. These training polygons were strategically distributed across the study area to ensure each land cover category was well represented. Stratified random sampling was employed to collect 200 ground truth points for the four land cover classes, with 78 points for grassland, 54 for stressed vegetation, 41 for gully, and 27 for bare soil. While the Normalized Green Red Difference Index (NGRDI) was not used as a feature in the classification process, it was computed separately to provide

insights into the distribution of vegetation in gullied areas. The NGRDI was calculated in ArcMap using the formula  $(G - R)/(G + R)$ , where G represents the Green and R represents the Red bands. Spectral profiles of the NGRDI for gullies were extracted from gully transects ranging from 20 m to 100 m in length using ENVI 5.3. The initial land cover map was reclassified into a binary map with two classes (gully and non-gully) to generate a gully map. The resulting binary map was then clipped into four specific gully sites (sites 3A-3C) to analyze further the gullies' morphological characteristics, including their shape, size, length, width, and depth. For each of the four selected sites, gullies were manually digitized based on high-resolution (0.5 m) aerial photography, and digitized polygons were created in ArcMap at a scale of 1:500. These polygons were then converted into a raster format. The manually digitized gullies were then used to assess the accuracy of the automatically extracted gullies by comparing the extent of each type. For instance, the proportion of the area affected by gullies as identified by an algorithm was compared to the proportion manually delineated in each site. Furthermore, the accuracy of the automated classification was evaluated through the utilization of a confusion matrix (Congalton, 1991; Congalton and Green, 2019). Finally, the gully density map of Study Area #3 was created in ArcGIS by applying the Line Density tool, which allowed for the visualization of the severity or intensity of gullies across the entire area. In addition, the lengths and density of gullies per unit area were calculated, with meters selected as the unit of measurement for both length and area (gully length in meters per square meter). The input data for computing the density map consisted of the gully map generated using an algorithm with the highest overall and class-level accuracies.

The reference data for Study Area #4 consisted of 592 gully points that were collected from Google Earth. The gullies in this study area are extensive

and intricate, with narrow linear and dendritic patterns, and a dense network of smaller gullies. Due to the complexity of the gullies, the locations were mainly collected at gully heads or in the middle part of long narrow linear gullies with limited branching or visible gully heads. Most gullies in the area are permanent and display diverse morphological characteristics such as lengths ranging from 30-274 meters, depths from 1.22-6.90 meters, and widths from 4.66-15 meters. The gullies primarily exhibit V-shaped and U-shaped cross-sections (Phinzi, Holb and Szabó, 2021). Despite their complexity, the relatively large size and lack of vegetation cover made it easy to locate gullies both in the field and on Google Earth. Likewise, a total of 592 non-gully points were collected from locations where the land cover was different from that of gullies. The resulting gully and non-gully points were then transformed into a vector shapefile and exported to R software to extract pixel values of the 16 geo-environmental predictors. The reference data contained 1184 observations with the response variable being binary ( $g$  = gully and  $ng$  = non-gully) and 16 predictors (large set). Note that the reference data for the small and medium feature sets contained six and twelve predictors, respectively. After checking for missing values, the reference data were randomly split into training (70%) and testing (30%) datasets. This partitioning procedure was also applied to the small and medium feature sets.

A two-step process was followed for gully susceptibility modeling using six ML techniques (ANN, PLS, RDA, RF, SGB, and SVM) and a bivariate statistical method (frequency ratio). First, the caret package (Kuhn, 2008) available in R software was utilized to train all ML methods with 10-fold cross-validation repeated five times, resulting in 50 candidate models for each ML method. Following this procedure, a final model with the highest accuracy was selected and applied to predict gully erosion at a pixel level, resulting in a binary output ( $g$  and  $ng$  pixels) for each ML method. Second, the

frequency ratio (FR) was then used to assign weights to each geo-environmental predictor and reclassify the binary output of each ML method into four gully susceptibility classes (e.g., low, moderate, high, and very high). FR is a bivariate statistical technique successfully applied to various natural hazards such as flooding (Rahmati, Pourghasemi and Zeinivand, 2016; Shafapour Tehrani *et al.*, 2019), landslides (Lee and Sambath, 2006; Anbalagan *et al.*, 2015), and, more recently, gully erosion (Roy and Saha, 2019; Amare *et al.*, 2021; Azedou *et al.*, 2021; Lana, Castro and Lana, 2022). It expresses the ratio between the occurrence and non-occurrence of a natural hazard (in this case, gully erosion) based on its spatial relationship with associated influencing factors (i.e., gully predictors) (Lee and Pradhan, 2007). Much like conditional probability, an FR ratio of  $>1$  represents a strong relationship, while a ratio of  $<1$  represents a weak relationship between gullies and predictor classes (Anbalagan *et al.*, 2015).

#### **4.6. Machine learning (ML) and hyperparameter tuning**

The accuracy of gully erosion classification depends on numerous factors, which encompass the choice of algorithms as well as the specific characteristics of the study area and the remote sensing data employed. Thus, various algorithms including RF, SVM, PLS, MLC, K-NN, RDA, RDA, ANN, SGB, and MD were utilized to identify gully features and model gully susceptibility. These algorithms were chosen due to their widespread application in gully-related research, rendering them suitable for examining various previously untested factors that impact accuracy. In Study Area #1, the pan-sharpened SPOT-7 product was classified using RF, SVM, and LDA algorithms, while in Study Area #2, only SVM and RF algorithms were applied to the PlanetScope data. Python programming language was used to execute the algorithms in Study Area #1 and #2. In Study Area #3, RF, ML, K-NN, and MD algorithms were used to extract gullies from a visual range

SPOT-7 image, and these were run in the Sentinel Application Platform (SNAP), an image-processing software developed by the European Space Agency (ESA). In Study Area #4, six algorithms (RF, PLS, RDA, SVM, ANN, and SGB) were executed in R programming language (R Core Team, 2021). Appendix 2 presents the process of hyperparameter tuning and the chosen optimal values for the six algorithms across three feature sets of different sizes. The following subsections describe each algorithm, including associated hyperparameters.

#### ***4.6.1. Random forest (RF)***

RF is an ensemble learning method that combines multiple decision trees to create a more accurate and stable model (Breiman, 2001). It constructs a forest of decision trees, each trained on a distinct and randomly selected subset of the training data and features (Camps-Valls and Bruzzone, 2009). The trees are developed independently, and their predictions are aggregated to produce a final prediction, which reduces the model's variance. The algorithm has two hyperparameters: a *mtry* parameter, which specifies the number of features randomly selected at each split of the decision tree, and an *ntree* parameter, which defines the number of decision trees included in a forest. A grid search was applied to find a *mtry* value that maximizes accuracy.

#### ***4.6.2. Support vector machines (SVM)***

SVM works by finding the hyperplane that maximizes the margin between the classes. The margin is the distance between the hyperplane and the closest data points from each class. This hyperplane is the maximum margin hyperplane, the optimal decision boundary for the classification problem. In cases where the data points are not linearly separable, the SVM uses a kernel function to transform them into a higher-dimensional feature space where they can be linearly separable (Lantz, 2015; Boehmke and

Greenwell, 2019). This study applied SVM using a radial kernel function. The algorithm has two crucial parameters that affect its performance. The first parameter, cost (C), penalizes the misclassification of training data. The second parameter,  $\gamma$ , controls the influence of a single training data point on the decision boundary. The process of hyperparameter tuning was conducted using the grid search method.

#### **4.6.3. Regularized discriminant analysis (RDA)**

Discriminant analysis (DA) is a machine learning method that models the conditional distribution of predictors X in predefined response classes  $Y = 1$  (i.e., gully) and  $Y = 0$  (i.e., non-gully) using Bayes' theorem (Welch, 1939; Friedman, 1989; Gareth *et al.*, 2013). DA identifies linear combinations of predictors that best discriminate the response classes and makes predictions for new observations using linear or quadratic discriminant functions (Alkarkhi and Alqaraghuli, 2018). Regularized discriminant analysis (RDA) is a regularization method based on linear DA (LDA) and quadratic DA (QDA) that incorporates tuning parameters to improve classification performance. RDA is advantageous over its predecessors as it has tuning parameters that make it a robust classifier. In this dissertation, I fine-tuned the RDA parameters using a tune length of 5. The "rda" (Friedman, 1989) method in the caret package was employed. Although all discriminant analysis methods rely on the assumption of multivariate normal data distribution, RDA is expected to perform better than LDA and QDA due to its incorporation of tuning parameters (Wu *et al.*, 1996).

#### **4.6.4. Linear discriminant analysis (LDA)**

LDA seeks to find a linear combination of features that maximizes the separation between classes in a dataset. LDA aims to project the original data into a lower-dimensional space while preserving class-discriminatory

information. It does so by modeling the data distribution in each class and computing the mean and covariance matrix for each class. Once the mean and covariance matrix are computed for each class, LDA uses these parameters to calculate a linear discriminant function to project new data onto a lower-dimensional space. The linear discriminant function is essentially a weighted sum of the input features, where the weights are learned during training to optimize the separation between classes. LDA does not have any hyperparameters.

#### ***4.6.5. Maximum likelihood classifier (MLC)***

MLC is a probabilistic algorithm that models the probability distribution of each class in the input data and uses Bayes' theorem to calculate the probability of a new observation belonging to each class (Bolstad and Lillesand, 1991). The algorithm first estimates the probability distribution of each class in the training data by calculating the mean and variance of the data for each class. Once the probability distributions are estimated, the algorithm can classify new observations by calculating the likelihood of the observation belonging to each class. The class with the highest likelihood is then chosen as the predicted class for the observation. Minimum and maximum power set sizes are two important hyperparameters. After experimentation with different power values and observing the resulting classification outcomes, values of two (minimum power) and seven (maximum power), which are default values were used.

#### ***4.6.6. K-nearest neighbor (K-NN)***

The k-NN algorithm works by finding the k nearest data points in the training set to a new, unseen data point and using the labels of those neighbors to predict the label or value of the new data point (Thanh Noi and Kappas, 2017). The "nearest" data points are determined based on the Euclidean

distance or other distance metrics between the new data point and the training set. In the case of classification, the k-NN algorithm selects the most common class label among the k-nearest neighbors and assigns that label to the new data point. The value of the k hyperparameter, which represents the number of neighbors, can significantly affect the performance of the k-NN algorithm. A k value of five was used after experimentation with several values.

#### ***4.6.7. Minimum distance (MD)***

MD is a simple algorithm that calculates the distance between a new, unseen data point and the class centroids (center of each class distribution in the feature space) in the training set (Richards and Richards, 1999). The algorithm then assigns the class label of the closest centroid to the new data point. It assumes that the data points in each class are normally distributed and have the same covariance matrix, which allows calculating the Euclidean distance between the new data point and the class centroids. Like the MLC, the MD classifier involves two hyperparameters: the minimum and maximum power set size. A procedure similar to MLC was also followed to determine the optimal values for these hyperparameters. Specifically, a value of two was selected for the minimum power set size, and a value of seven was chosen for the maximum power set size parameter.

#### ***4.6.8. Artificial neural network (ANN)***

Artificial neural networks (ANN) are composed of layers of artificial neurons that imitate the function of biological neurons in the human brain. A typical ANN architecture includes an input layer of neurons, one or more hidden layers of neurons, and a final layer of output neurons. Neurons in each layer are linked to other neurons in the next layer through connections with specific weights (Wang, 2003). These weights determine the relative importance of each geo-environmental predictor in predicting gully

occurrence. Predictors with larger weights contribute more significantly to gully prediction. ANN was executed with the "nnet" (Venables and Ripley, 2013) method in the caret package (Kuhn, 2008). To find the best combination of hyperparameters that maximizes predictive performance, we conducted a grid search of two hyperparameters: the number of neurons in the hidden layer (size = 5, 10, 15) and a regularization parameter to prevent over-fitting (decay = 0.001, 0.01, 0.1).

#### ***4.6.9. Partial least squares (PLS)***

Initially designed for dimension reduction (Wold, 1966), PLS is becoming popular for solving classification problems, including gully susceptibility (Pham *et al.*, 2020; Pourghasemi *et al.*, 2020). PLS intends to form components that capture most of the information in the explanatory variables helpful in predicting the response variable (Garthwaite, 1994). It achieves this by constructing linear combinations (components) of the original predictors from which a set of latent variables with the best predictive power is extracted (Abdi, 2003), then regressing the response variable on these latent variables (Chung and Keles, 2010). PLS only has one hyper-parameter, the number of components. A tune length of 15 was selected to find the number of components with the highest accuracy. PLS was executed with the "pls" (Wehrens and Mevik, 2007) method in the caret package.

#### ***4.6.10. Stochastic gradient boosting (SGB)***

Boosting was traditionally developed for classification problems (Valiant, 1984) and concerned with combining several weak classifiers to form a robust classifier (Kuhn and Johnson, 2013). There are many variants of boosting algorithms, and SGB (Friedman, 2002), also known as gradient boosting machines (GBM), is among the most recent and popular algorithms. Much like RF, SGB uses bagging, a technique where a set of random decision

trees are generated, and each tree is trained on a random subset of the training data. The main difference is that SGB generates an ensemble of several shallow trees sequentially, where each tree learns and improves on the previous one, while RF generates an ensemble of deep independent trees (Boehmke and Greenwell, 2019). SGB uses a sequential ensemble approach in which boosting starts with a weak model and sequentially boosts its performance by building a new tree at each iteration from a random subsample of the training set, improving the model's prediction accuracy (Moisen *et al.*, 2006; Boehmke and Greenwell, 2019). SGB was implemented through the caret package using the "gbm" (Ridgeway, 2007) method. Hyper-parameters include the depth of each tree (interaction depth = 1, 5, 9), the number of trees (n trees = 1500), the learning rate of the algorithm (shrinkage = 0.1), and the minimum number of observations in the for the trees terminal node (n minobsinode = 20). A grid search method was used to find a combination of these parameters with the highest accuracy.

#### **4.7. Model performance evaluation**

The overall performance of RF, SVM, and LDA in Study Area #1 was evaluated with a repeated 10-fold CV procedure. This procedure was repeated thrice, generating 30 models and their associated OA measures. Study Area #1 comprised three sites (1A-1C), and the model was trained using data (pixels) from one site and tested using data (pixels) from another independent site. This process was repeated for all possible combinations of the three sites, resulting in 36 candidate models. These models were developed using three different algorithms (RF, SVM, and LDA), two approaches to class numbers (binary and multi-class), and six combinations of study sites as train and test sets (1A→1B, 1B→1A, 1B→1C, 1C→1B, 1A→1C, 1C→1A). The performance of these models was evaluated by calculating the medians and quartiles of their OA. Despite the reliability of CV as a tool, it does not offer

insights into the accuracy of specific classes. As a result, the confusion matrix, which computes class-level metrics such as PA and UA, was utilized to supplement the analysis.

For Study Area #2, the performance of RF and SVM algorithms was assessed using OA, computed using a 5-fold cross-validation and bootstrapping. The 5-fold CV was repeated 20 times to increase the reliability of the models, resulting in 100 candidate models from which final accuracies were computed. The dataset was randomly shuffled in each repetition, creating new folds. In the bootstrapping procedure, the models were validated on samples outside of the bootstrap sample, and this process was repeated 100 times, generating 100 models for evaluation. Additional performance metrics, including PA, UA, F1-score, commission, and omission errors, were calculated. The omission error is the difference between 100% accuracy and the PA, which occurs when a pixel is not classified into the appropriate category. In contrast, the commission error is the difference between 100% accuracy and UA, which arises when a pixel is incorrectly assigned to a category to which it does not belong. Moreover, unbiased areal coverages (ha) of gullies, along with their standard errors (ha) and associated  $\pm$  95% confidence intervals (ha), were computed following the “good practice” recommendations for accuracy assessment of Olofsson et al. (2014). Six models were developed based on a combination of classifiers (SVM and RF), seasons (dry and wet), and resampling methods (bootstrapping and cross-validation), namely rf-d-b, rf-d-cv, rf-w-cv, rf-w-b, svm-d-cv, svm-d-b, svm-w-cv, and svm-w-b.

The performance evaluation of RF, ML, K-NN, and MD in Study Area #3 was based on accuracy metrics computed from the confusion matrix, such as overall accuracy (OA), kappa coefficient, producer’s accuracy (PA), and user’s accuracy (UA). The “compute confusion matrix” function in ArcGIS

10.4 created a confusion matrix. Similarly, in Study Area #4, the confusion matrix computed in R software evaluated the performance of six ML algorithms (ANN, PLS, RDA, RF, SGB, and SVM) based on testing data.

#### **4.8. Statistical analysis**

Statistical analyses were conducted in R 3.6.2 software (R Core Team, 2020) with WRS2 (Mair and Wilcox, 2019), jamovi 1.2., and GAMLj module (Gallucci, 2019). For Study Area #1, the normality assumption of the reference data of land cover classes was checked using the Shapiro-Wilk test. The Shapiro-Wilk test is a statistical hypothesis test utilized on a sample, where the null hypothesis assumes that the sample is drawn from a normal distribution (Mohd Razali and Bee Wah, 2011). In the case of this dissertation, the null hypothesis was rejected, indicating that the reflectance values of land cover classes do not follow a normal distribution. While the reflectance values were skewed, the classification accuracy measures (OA, PA, and UA) followed a normal distribution. Then, hypothesis testing was applied to determine if the land cover classes had identical medians or if they differed from one another concerning reflectance values. The Yuen test with a 0.2 trim value and 599 times bootstrapping was used for the binary approach, while robust ANOVA was used with a post hoc test based on trimmed means for the multiclass approach. Effect sizes were calculated for the Yuen test, while Dunnett's test (Tallarida and Murray, 1987) was used to compare gullies with other land cover classes based on confidence intervals of differences instead of effect sizes. Thus, the number of comparisons and degree of freedom were limited, which avoided a complete factorial comparison. This step was necessary, given that the objective was to map gullies. General Linear Modelling (GLM) was used, with three sites, the number of classes (binary or multiclass), and algorithms considered as factors in different combinations to evaluate model performance on classification accuracy measures, where the

UA and PA were the dependent variables in this analysis. The effect size measure used was  $\omega^2$ , which is less biased by low sample size, limited to the results of 36 models (Levine and Hullett, 2002).

In Study Area #2, the NDVI values of the images were compared for gullies in the wet and dry seasons using the robust Mann-Whitney test, with Monte Carlo p (pMC) and 9999 permutations. The General Linear Model (GLM) was used to analyze the effects of spectral bands (four bands: RGB + NIR), seasons (wet and dry), and land cover classes (seven classes). The statistical interactions were also examined to determine if factorial variables had a common effect, such as whether the effects of spectral bands differed by land cover classes or differed in the dry or wet seasons. In addition, effect size ( $\omega^2$ ) was calculated as a standardized measure of the variables' contribution to the model, where higher values indicate an enormous contribution. An effect size of  $\omega^2 > 0.14$  was considered a significant effect, as suggested by Field (2013). Finally, the Dunnett test (Lee and Lee, 2018) was used to determine whether gullies differed significantly from other land cover types. This test is designed to perform multiple comparisons of the factor groups' means against one control group's mean (i.e., gullies), which can reveal subtle reflectance differences while identifying all possible overlaps in reflectance with the gullies. This test limited the number of comparisons to six instead of 21 (complete factorial approach). This approach was taken to identify all possible differences in reflectance with the gullies.

## **5. RESULTS AND DISCUSSION**

This section presents and discusses the research findings in light of existing literature. Subsection 5.1 examines the capabilities of low-cost, high-resolution satellite imagery for gully mapping. Subsection 5.2 presents the accuracy results for gully classification, particularly considering factors that

may influence model performance at the class level. In Subsection 5.3, the focus shifts towards examining the morphological characteristics of gullies and how they influence the accuracy of their mapping and density measurement. Finally, Subsections 5.4-5.6 investigate gully susceptibility modeling using feature sets of varying sizes and analyze key geo-environmental variables that explain the distribution of gullies in the study area.

## **5.1. Low-cost, high-resolution sensors for gully mapping**

### ***5.1.1. SPOT-7 multispectral image***

Results demonstrated that in a binary approach, reflectance values were significantly different across all SPOT-7 bands, except for the blue band in 1A (Figure 12). However, the effect sizes for blue and green bands were small, medium for the red band, and large for the NIR band (Table 4), indicating that the NIR band was the most effective in differentiating gullies. Subsequently, a multi-class approach was utilized to analyze seven categories (Figure 13), and the robust ANOVA test confirmed the significance of the models for each combination of bands and study sites (Table 5).

After the ANOVA analysis, a post hoc test was conducted to identify the differences among categories, as the latter only confirms that there is at least one significantly different category. Given many possible combinations with seven categories, the analysis focused only on comparing gullies with the other categories, which was the dissertation's primary objective. Gullies did not significantly differ from stressed vegetation and roads in the red, green, and blue bands but significantly differed from other categories (Table 6). The mean differences of land cover categories were also reported, with confidence intervals consistent with the post hoc test, except for 1B, where the robust post hoc test revealed significant differences (Figure 14). The confidence intervals

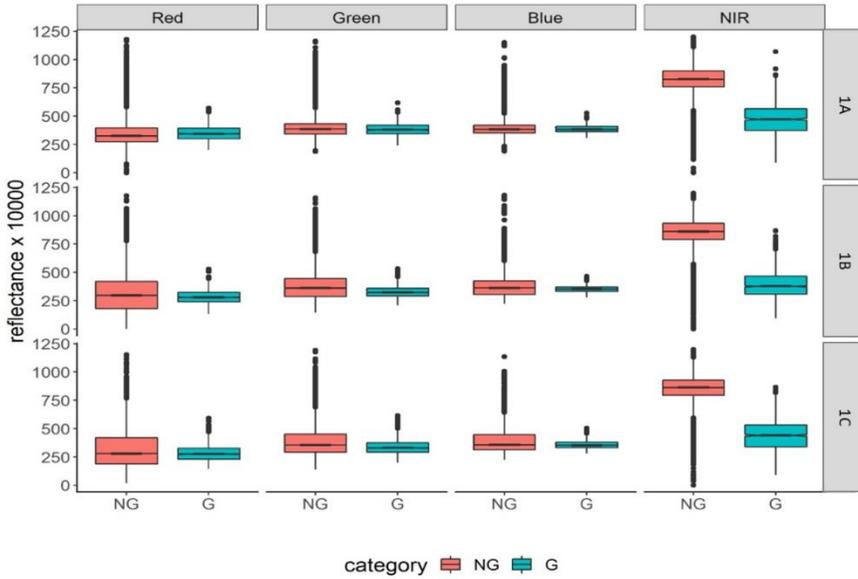


Figure 12. Distribution of reflectance values of SPOT-7 image by bands (red, green, blue (RGB) and near-infrared (NIR)), study areas (1A–1C), and classification categories (NG: non-gully, G: gully).

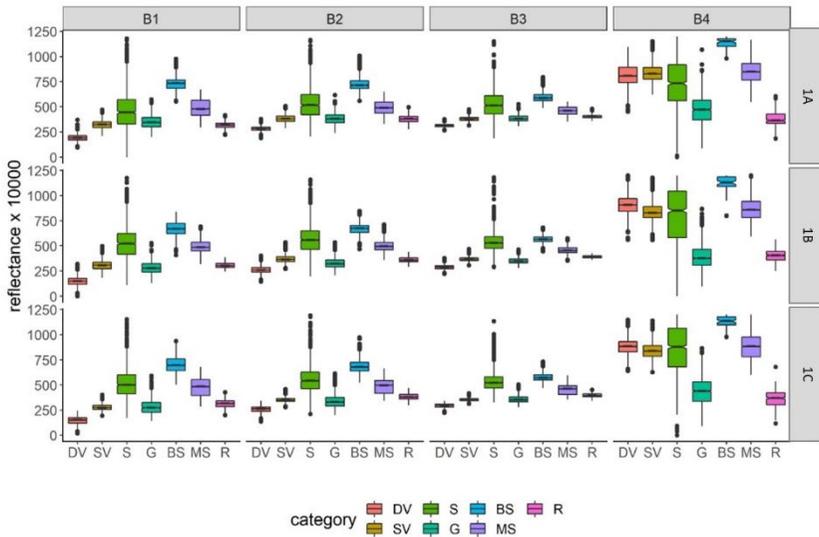


Figure 13. Distribution of reflectance values of SPOT-7 image by bands (RGB+NIR), study sites (1A–1C), and classification categories (DV: dense vegetation, SV: stressed vegetation, S: settlement, G: gully, BS: bare soil, MS: mixed bare soil, R: road, B1: red, B2: green, B3: blue, B4: NIR).

Table 4. Results of robust independent samples t-test performed on SPOT-7 bands using the binary (gully—non-gully) approach (study areas: 1A–1C, t: the value of t-statistic, p: significance,  $\xi$ : effect size).

Bands	1A			1B			1C		
	t	p	$\xi$	t	p	$\xi$	t	p	$\xi$
Red	8.3	<0.001	0.159	9.55	<0.001	0.14	6.74	<0.001	0.11
Green	2.99	0.003	0.062	28.21	<0.001	0.38	15.98	<0.001	0.28
Blue	1.07	0.286	0.02	14.42	<0.001	0.2	9.56	<0.001	0.17
NIR	86.1	<0.001	0.964	171.7	<0.001	0.98	113.3	<0.001	0.98

Table 5. Results of robust ANOVA performed on SPOT-7 bands using the multiclass (7-classes) approach (study areas: 1A–1C, F: the value of F-statistic, p: significance).

Bands	1A		1B		1C	
	F	p	F	p	F	p
Red	10105	<0.001	25193	<0.001	12660	<0.001
Green	9309	<0.001	21188	<0.001	10571	<0.001
Blue	9718	<0.001	25694	<0.001	13036	<0.001
NIR	3590	<0.001	10905	<0.001	4317	<0.001

Table 6. Results of robust ANOVA performed on SPOT-7 bands using the multiclass (7-classes) approach (LC: land cover, study areas: 1A–1C, G: gully, DV: dense vegetation, SV: stressed vegetation, S: settlement, BS: bare soil, MS: mixed soil, R: road).

LC	1A				1B				1C			
	Red	Green	Blue	NIR	Red	Green	Blue	NIR	Red	Green	Blue	NIR
G-DV	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
G-SV	<0.001	0.275	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.281	0.275	0.883	<0.001
G-S	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
G-BS	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
G-MS	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
G-R	<0.001	0.528	<0.001	<0.001	0.614	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

(based on Dunnett's test statistics) indicated non-significance between gullies and roads (Figure 14).

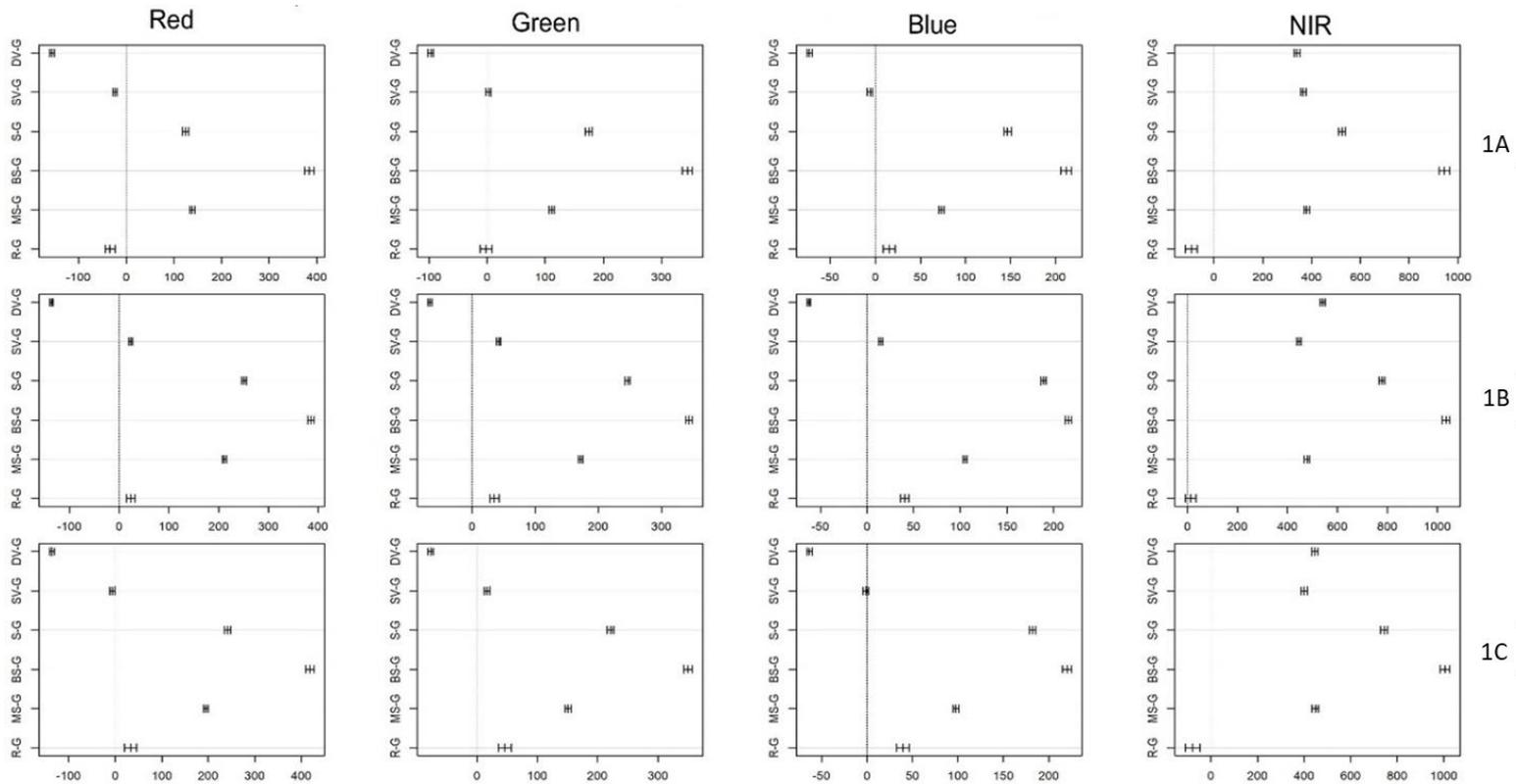


Figure 14. Mean differences between gullies (G) and other land cover categories (mean  $\pm$  95% confidence intervals; 95% confidence intervals coinciding with 0 are not significant differences,  $p > 0.05$ ; DV: dense vegetation, SV: stressed vegetation, S: settlement, BS: bare soil, MS: mixed bare soil, R: roads) by SPOT 7 bands (columns) and study areas (rows).

Although gullies exhibit spectral heterogeneity, remote sensing data can still be used for gully mapping. SPOT-7 multispectral bands were found to have spectral differences that were reasonably discernible, despite significant variability. While the RGB bands showed only small to moderate effect sizes, the NIR band showed a large effect size. While p-values are used to determine the significance of differences, effect sizes provide a measure of the magnitude of differences, and standardized measures demonstrate the relevance of differences with land cover categories. Even though slight differences (with effect sizes  $<0.3$ ) may still be significant, they can lead to misclassifications. These findings are consistent with previous research (Szabó et al. 2016).

While effect sizes were not calculated for the land cover pairs in the multiclass approach, valuable information on differences was provided by confidence intervals (Figure 14). In general, all pairs had significant differences, and the confidence intervals were mainly within a small range, except for specific land cover categories, such as gullies, stressed vegetation, and roads, which showed non-significant differences in RGB bands (primarily in red and green bands), but always had significant differences in the NIR band. However, the NIR band successfully discriminates between roads and gullies, as indicated by the confidence intervals close to the "zero" line, indicating non-significance. This was not reflected in the p-values (all of which were  $p < 0.001$ ). However, if confidence intervals and their distance from zero as an effect size are used, these cases indicate low values with lower efficiency in distinguishing between categories. Despite criticisms of the statistical evaluation, considering the results, it can be argued that the reference dataset contained reliable data about land cover categories and could presumably be used in classification models.

### ***5.1.2. PlanetScope multispectral image***

Results showed that NDVI values for the wet season ranged from -0.36 to 0.81 and were generally higher than those for the dry season, which ranged from -0.41 to 0.59 (Figure 15). The distribution of NDVI values for the dry season was bimodal, representing non-vegetation pixels in the first mode and vegetation pixels in the second mode. Conversely, the wet season had a multimodal distribution, with the first mode representing non-vegetation pixels and the last two modes indicating vegetated areas, namely, vegetation and forest pixels.

The study also analyzed the determinants of reflectance by examining spectral bands, land cover classes, and seasons. Statistical analysis revealed that these factors and their interactions significantly influenced reflectance ( $p < 0.001$ ), explaining 92.3% of the variance (Table 7). The difference between dry and wet seasons had the most significant effect on reflectance (0.868), while the impact of bands and land cover classes was similar but slightly lower (~0.6), indicating a substantial effect. Results further showed that reflectance varied across bands, land cover classes, and seasons, and the contribution of their interactions was significant. The interaction between seasons and land cover classes had a smaller effect size than interactions with bands (0.141), but it was still substantial. Finally, the interaction of all factors had a significant effect, although with a smaller value (0.185).

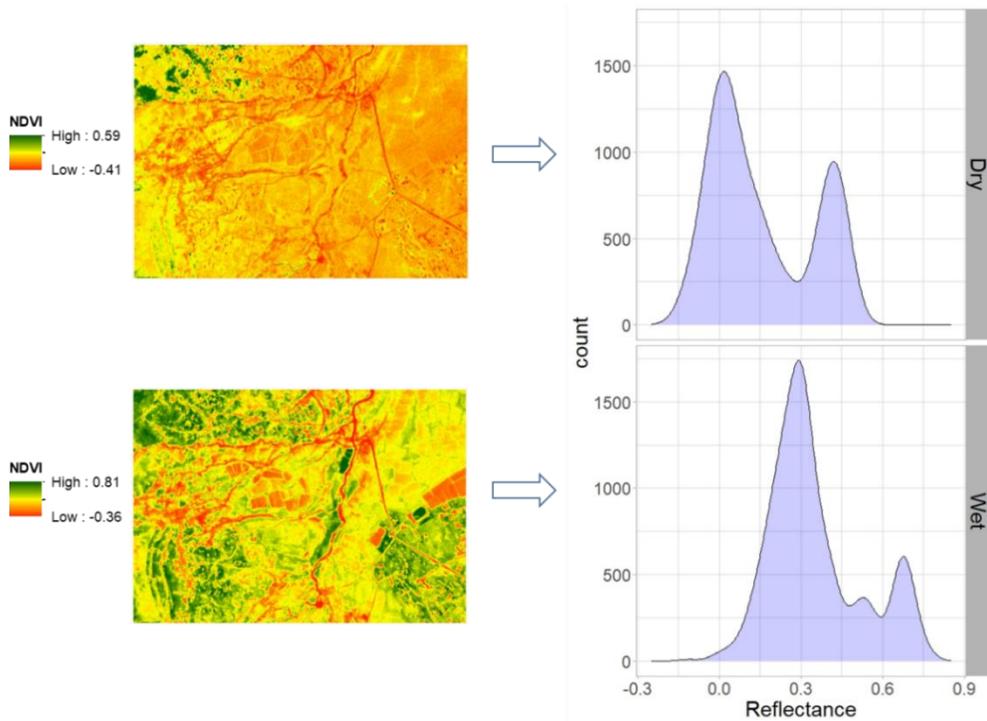


Figure 15. Distribution of NDVI reflectance values in the dry and wet seasons.

Table 7. Results of General Linear Modelling (GLM) performed with reflectance as an independent variable (SS: Sum of Squares, df: degree of freedom, F: F-statistic, p: significance,  $\omega^2p$ : effect size;  $p < 0.05$ : significance level).

Variables	SS	df	F	p	$\omega^2p$
Model	6.99e0+9	55	860.4	< .001	0.923
Bands	1.00e0+9	3	2256.1	< .001	0.633
Season	3.80e0+9	1	25715.0	< .001	0.868
Class	9.79e0+8	6	1104.2	< .001	0.629
Bands $\times$ Season	4.48e0+8	3	1010.0	< .001	0.436
Bands $\times$ Class	5.30e0+8	18	199.3	< .001	0.477
Season $\times$ Class	9.62e0+7	6	108.5	< .001	0.141
Bands $\times$ Season $\times$ Class	1.34e0+8	18	50.3	< .001	0.185
Residuals	5.70e0+8	3860			
Total	2.96e+10	3916			

The post hoc test conducted with the Dunnett test showed significant differences ( $p < 0.001$ ) between gullies and other land cover classes in the dry season but not in the wet season, except for a few specific cases (Figure 16). Furthermore, a variable importance analysis using a RF algorithm in the R software was conducted to assess the significance of the original bands in discriminating gullies from the surrounding areas. The results revealed that NIR and red bands exhibited the highest level of influence in both seasons, while the blue band had minimal impact (Table 8). Moreover, the dissertation also evaluated the performance of NDVI as a spectral index and found that it was less effective in differentiating gullies from other land cover classes than the original bands. The NDVI-based gully values did not differ significantly from mixed bare soil and vegetation values in the dry season and performed better in the wet season but did not differ significantly from the built-up class values.

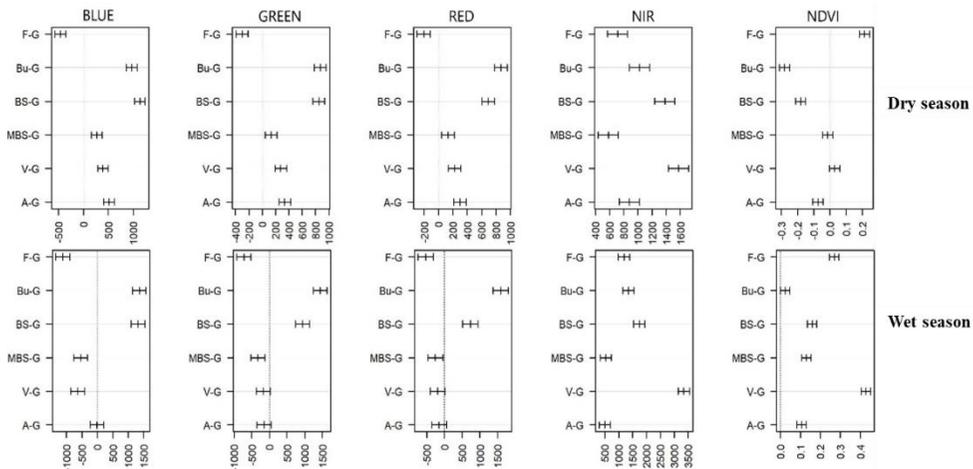


Figure 16. Differences of gullies and other land cover types' reflectance by bands and seasons (G: gully; F: forest; Bu: built-up; BS: bare soil; MBS: mixed bare soil; V: vegetation; A: agriculture; mean  $\pm$  95% confidence intervals; the difference was not significant if confidence range intersects the dashed line).

Table 8. PlanetScope bands ranking in discriminating gullies against the surrounding land cover.

<b>Dry Season</b>		<b>Wet Season</b>	
<b>Band</b>	<b>Importance (%)</b>	<b>Band</b>	<b>Importance (%)</b>
NIR	31	NIR	35
Red	26	Red	32
Green	25	Green	21
Blue	17	Blue	12

### ***5.1.3. Visual range SPOT-7 image***

Despite having limited spectral information (i.e., only RGB bands), the SPOT-7 image effectively discriminated gullies from other land cover types. This was due to the high spatial resolution of the image (1.3 m), which enabled the detection of most gullies, even without the near-infrared (NIR) band. The Normalized Green-Red Difference Index (NGRDI) was used to discriminate gullies from vegetation and other surrounding land cover classes, as demonstrated by the NGRDI map and the corresponding spectral profiles of gullies (Figure 17). In addition, the study sites' spectral profiles revealed the gullies' morphology, including the V-shaped gully in site 3C and the dendritic network of gullies in sites 3A and 3D. Gullies with these morphological characteristics were the most detectable in the study.

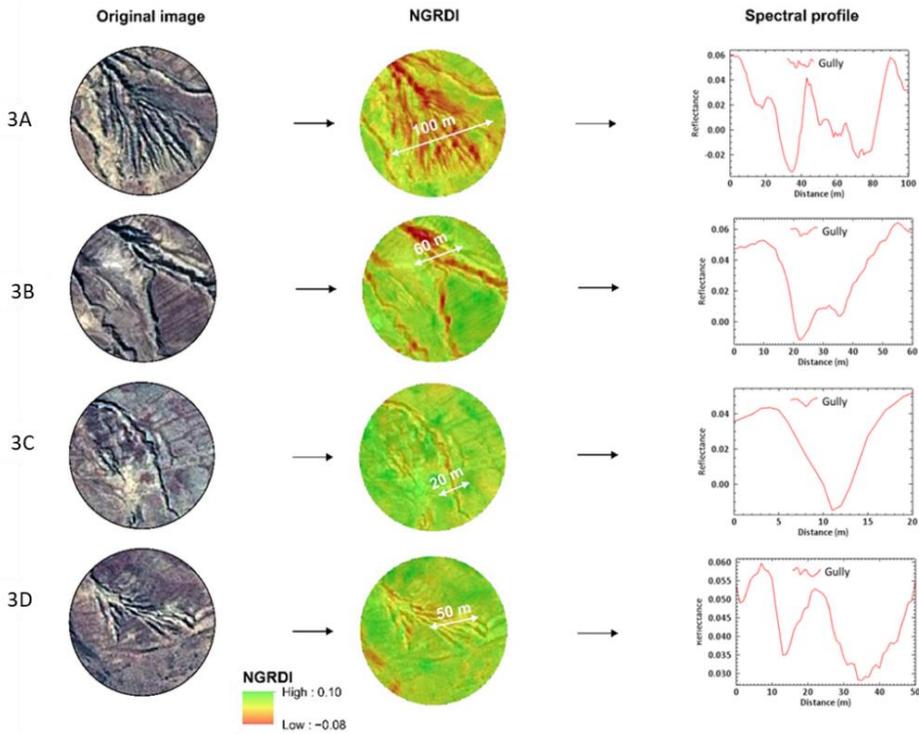


Figure 17. Gully transects (20–100 m) with Normalized Green Red Difference Index (NGRDI)-based spectral profiles of selected gullies.

## 5.2. Quantifying gully classification accuracy across different study areas

### 5.2.1. Study Area #1: accuracy and factors biasing model performance at class-level

The first study area (Study Area #1), which utilized the SPOT-7 multispectral image, demonstrated that accurate identification of gullies could be achieved using either RF or SVM classification algorithms. Both approaches produced similar levels of accuracy, with median values ranging from 92% to 96% (Figure 18). While the binary approach resulted in better OA values, the multiclass (m) classification approach was only marginally less accurate, with a difference of 2%. In all study sites (1A-1C), the top three

performing models were consistently SVMb, followed by RFb and LDAb in the binary approach. However, the multiclass classification approach exhibited some variation in fourth place between SVMm and RFm, while LDAm consistently underperformed. Although SVMb had a slight advantage over RFb in the top two places, RFb had higher lower quartile values, suggesting that it may be more reliable overall. Moreover, the multiclass approach was generally less effective than the binary approach, with even the lowest quartile values for LDAb being higher than the upper quartile values for the best-performing multiclass solution, resulting in differences ranging from 2% to 8%.

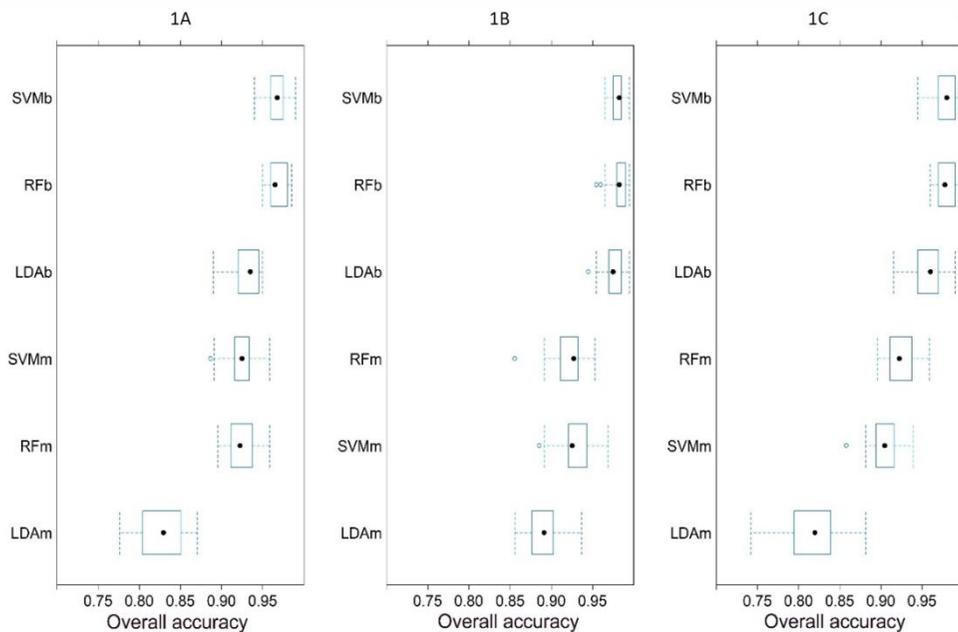


Figure 18. Classification results of the applied algorithms ranked by overall accuracies of 30 models (10-fold cross-validation with three repetitions) by study areas (LDA: Linear Discriminant Analysis, RF: Random Forest, SVM: Support Vector Machine; b: binary, m: multiclass).

The evaluation of classifiers on a class-level in Study Area #1 was based on only gullies' accuracy metrics. The results revealed that most

classifiers failed to achieve PA and UA values above 80%, as shown in Figure 19. Notably, the LDA algorithm performed poorly, with UA values below 30% for multiclass classification, resulting in many commission errors where many pixels were classified as gullies that belonged to other categories. However, some successful LDA models, such as L-m-1B-1A, were comparable to particular RF and SVM models. Out of the best 80-80% quarter, seven successful models comprised five RF and two SVM models, including two binary and five multiclass types. However, R-b-1C-1B, despite having a PA of less than 80%, achieved the highest UA at 86%. On the other hand, the highest PA was obtained by the LDA model, L-b-1B-1C, which reached 99.5% but had a low UA of only 18.5%. According to the results depicted in Figure 20, LDA exhibited better performance at a class level when using the binary approach, specifically concerning PA.

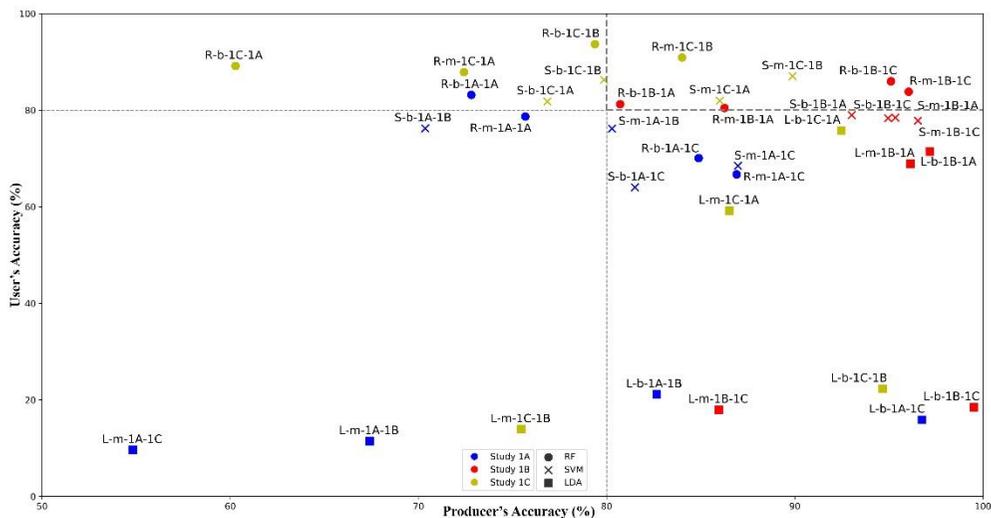


Figure 19. Class level accuracy metrics of different classifications of gullies by algorithms, number of categories, and study areas (S: SVM, R: RF, L: LDA; b: binary, m: multiclass; first number: number of the area where the models were applied, second number: number of the area where the model was trained; dashes line sections (upper right) indicate >80% accuracy quarter).

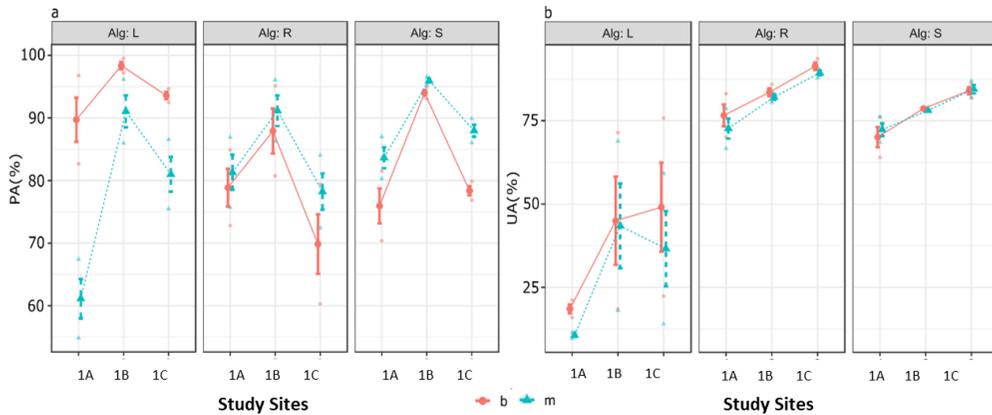


Figure 20. PA (a) and UA (b) values (median  $\pm$  quartiles) of gullies by classification algorithms (Alg: algorithm; L: LDA, R: RF, S: SVM), number of classes (b: binary; m: multiclass) and study areas (1A-1C).

In addition, a statistical analysis was carried out to identify factors that might have influenced the performance of the classification models at a class level. GLM indicated that both the study sites and the algorithms used could account for the differences in efficiency observed among the classification models. The study site had a significant impact on both the PA and UA, while the applied algorithm significantly affected only the UA (Tables 9-10). The interaction between the algorithms and study sites did not have a significant effect on the PA. The effect size ( $\omega^2$ ) revealed a substantial influence of the study sites on PA and the algorithms on UA. Moreover, there was a significant interaction between the applied algorithms and the type of classification approach (binary or multiclass), indicating that these algorithms exhibit distinct performance based on the number of categories involved.

Table 9. Summary of General Linear Modelling (GLM) performed with PA as an independent variable (Alg: algorithm, type: binary or multiclass approach, stud: Study site; SS: Sum of Squares, df: degree of freedom, F: F-statistic, p: significance,  $\omega^2$ : effect size;  $p < 0.05$  is highlighted with bold).

Source	SS	df	F	p	$\omega^2$
Model	3200.5	17	3.632	<b>0.005</b>	0.554
Alg	174.9	2	1.687	0.213	0.017
type	25.5	1	0.493	0.492	0.006
stud	1431.9	2	13.811	<b>&lt; 0.001</b>	0.317
Alg $\times$ type	946.7	2	9.132	<b>0.002</b>	0.201
Alg $\times$ stud	323	4	1.558	0.228	0.028
type $\times$ stud	99.1	2	0.956	0.403	0.001
Alg $\times$ type $\times$ stud	199.4	4	0.962	0.452	0.002
Residuals	933.1	18			
Total	4133.6	35			

Table 10. Summary of GLM performed with UA as an independent variable (Alg: algorithm, type: binary or multiclass approach, stud: Study site; SS: Sum of Squares, df: degree of freedom, F: F-statistic, p: significance,  $\omega^2$ : effect size;  $p < 0.05$  is highlighted with bold).

Source	SS	df	F	p	$\omega^2$
Model	20720.3	17	4.0014	<b>0.003</b>	0.586
Alg	17390.9	2	28.547	<b>&lt; 0.001</b>	0.633
type	82.6	1	0.2711	0.609	0.008
stud	2413.4	2	3.9615	<b>0.038</b>	0.068
Alg $\times$ type	98.6	2	0.1618	0.852	0.019
Alg $\times$ stud	666.8	4	0.5473	0.703	0.021
type $\times$ stud	18.7	2	0.0307	0.97	0.022
Alg $\times$ type $\times$ stud	49.4	4	0.0405	0.997	0.044
Residuals	5482.8	18			
Total	26203.1	35			

Concerning classification accuracies, RF and SVM algorithms were robust and not affected by data distribution, making them less biased by outliers, whereas LDA assumes multivariate normality and a balanced number of elements in categories (Belgiu and Drăgu, 2016; Xiong *et al.*, 2020), leading

to lower classification accuracy. Consequently, both RF and SVM outperformed LDA in our study, but LDA<sub>b</sub> performed better than RF<sub>m</sub> and SVM<sub>m</sub> in binary classification. Multiclass classification results showed lower OA than binary classification, and it may be more appropriate to use categories where possible, such as gully and non-gully categories in this case. While the binary approach generally performed better than the multiclass approach (Beygelzimer, Langford and Zadrozny, 2004), the multiclass approach was more efficient for gully identification. These results are consistent with previous studies that showed the multiclass approach to have better performance in identifying gullies (Allwein, Schapire and Singer, 2000). Only two binary approach models using RF and SVM classifiers were in the best 80% quadrant, while LDA's performance varied and provided ambiguous results due to its susceptibility to outliers. These results suggest that RF and SVM classifiers are more suitable for gully identification in the study areas with these reference datasets, as they are less biased by outliers, while outliers severely biased LDA models.

The applied GLM showed the multivariate effects of classification algorithms on the classification accuracy of gullies at the class level. The number of classes (binary or multiclass) did not significantly affect the performance, and the study demonstrated that algorithms applied to space-borne images like SPOT-7 could detect gullies automatically. Although gullies are challenging to detect automatically, this study showed that the spectral bands of SPOT-7 and pan-sharpening can contribute to the successful extraction of gullies. Visual interpretation of high-resolution images has been the preferred method for monitoring gullies over large areas. However, there is a need for more reliable methods that can help to automatically identify areas affected by gullies for consistent monitoring over time. Appropriate algorithms and satellite images seem adequate for these surveys, but digital

elevation models (DEMs) should be incorporated where available to discern gullies accurately from other land cover categories. The study demonstrated that the methodology applied could be adopted for larger areas, although a significant limitation was the inability to discriminate unpaved roads from bare soil and exposed rocks. To address this issue, the classes were consolidated into a single class, and this decision did not impact the identification of gullies.

**5.2.2. Study Area #2: accuracy as a function of an algorithm, resampling technique, and season**

In Study Area #2, the performance of RF and SVM algorithms were examined using two different resampling methods, namely bootstrapping and CV, for both the wet and dry season scenes of PlanetScope data. Results showed that the CV resampling method consistently produced better OA than bootstrapping for both seasons (Figure 21).

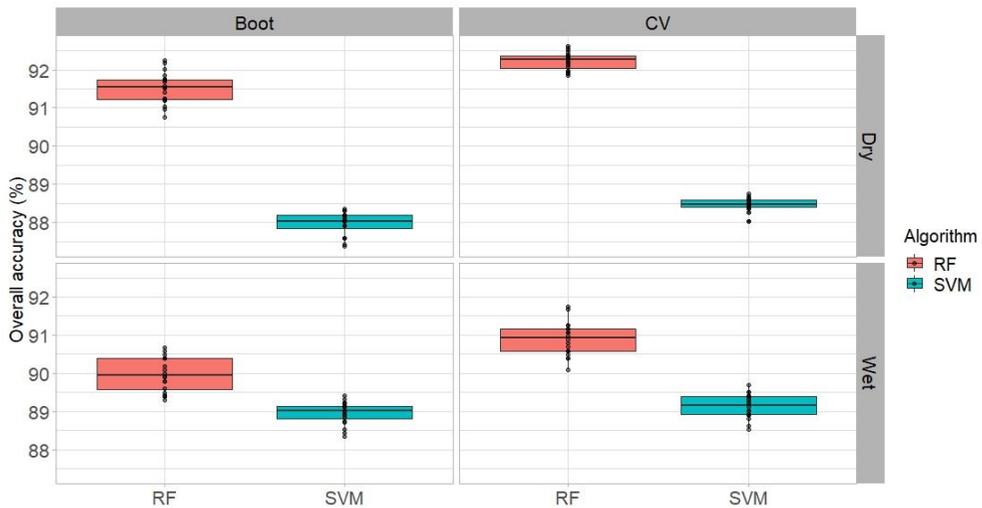


Figure 21. Accuracy assessment based on overall accuracy (OA) by the classification algorithm (RF: random forest, SVM: support vector machine), resampling method (boot: bootstrapping, CV: cross-validation), and season (wet and dry).

Additionally, two key trends emerged from these findings: firstly, RF consistently outperformed SVM regardless of the season or resampling method used; and secondly, the dry season generally yielded higher OA than the wet season, although this did not translate to higher accuracy in gully classification at the class level.

The class-level performance of RF and SVM was evaluated based on unbiased UA and PA. All models performed well, with UA values above 70% (Figure 22). Among the models, svm-d-b exhibited the best performance with a UA of 93.4%, while the worst performance was recorded by the rf-w-b model (77%). However, PA was generally lower than UA for most models, with only half recording a PA greater than 70%. The svm-w-cv model achieved the best PA (89.2%), with the rest falling below 70%, and the svm-d-b model had the lowest PA (32.5%).

An unbiased area estimate of gullies (in hectares) was also provided (Table 11), where svm-w-cv achieved the most accurate gully areal coverage of 57.2 ha with the highest PA (89.2%) and lowest standard error (3.7 ha). On the other hand, the rf-w-b model had the highest standard error (11.5 ha) and recorded a gully area of  $55.2 \pm 25$  ha. In terms of F1-score ranking, rf-d-b, and rf-d-cv algorithms achieved the best results ( $>0.90$ ), but RF algorithms belonging to the wet season had relatively low scores (0.82). However, all SVM algorithms (svm-d-cv, svm-d-b, svm-w-cv, and svm-w-b) recorded lower F1-scores ranging from 0.85 to 0.88. Both resampling techniques had the same omission error (14.9%) but slightly different commission errors, with bootstrapping having a 40.8% error of commission compared to a 37.8% error for CV (Table 12).

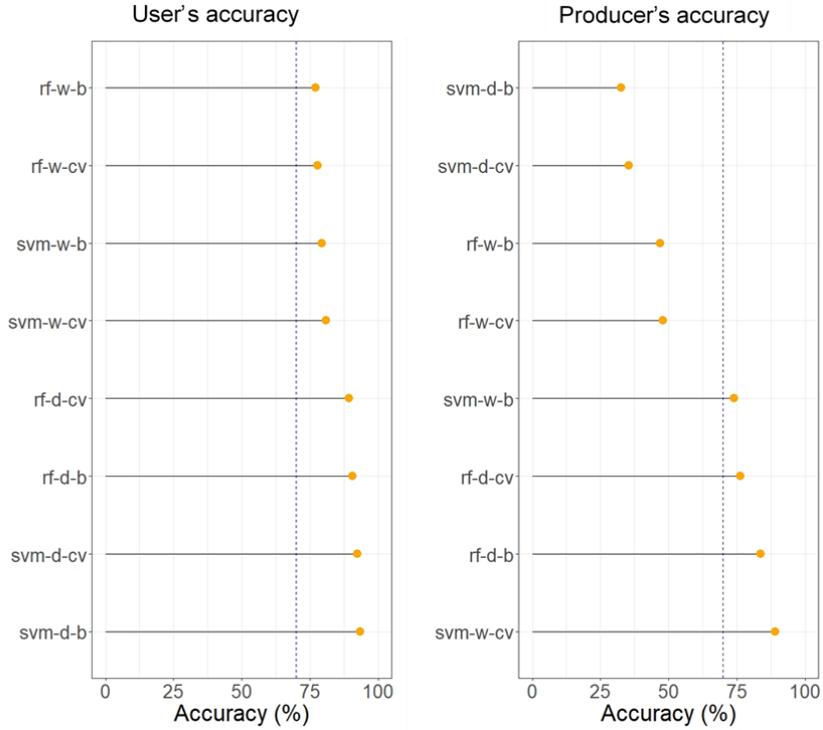


Figure 22. Unbiased user's and producer's accuracy (rf: random forest, svm: support vector machine, w: wet season, d: dry season, cv: cross-validation, b: bootstrapping, blue dashed line is 70% accuracy benchmark).

Table 11. Estimated gully area (ha) with associated standard error (ha) at  $\pm$  95% CI (ha) for each algorithm (rf: random forest, svm: support vector machine, d: dry, w: wet, b: bootstrapping, cv: cross-validation, CI: confidence interval).

Algorithm	Area (ha)	Standard Error (ha)	$\pm$ 95% CI (ha)	PA (%)	UA (%)	F1-score
rf-d-b	88	6.1	14.4	83.6	90.6	0.92
rf-d-cv	91.3	7.6	17.1	76.3	89.3	0.91
rf-w-cv	54.6	11.3	24.3	47.9	77.9	0.82
rf-w-b	55.2	11.5	25.0	46.8	77	0.82
svm-d-cv	32.6	10.1	21.1	35.4	92.3	0.86
svm-d-b	31.1	10.5	21.8	32.5	93.4	0.85
svm-w-cv	57.2	3.7	18.8	89.2	81	0.88
svm-w-b	57.4	6.4	19.3	74.1	79.4	0.86

Table 12. Summary of average error for resampling techniques, classifier, and season (RF: random forest, SVM: support vector machine, CV: cross-validation).

<b>Error</b>	<b>Resampling Technique</b>		<b>Classifier</b>		<b>Season</b>	
	<b>Bootstrap</b>	<b>k-fold CV</b>	<b>RF</b>	<b>SVM</b>	<b>Dry</b>	<b>Wet</b>
Commission (%)	40.8	37.8	36.4	42.2	43.1	35.5
Omission (%)	14.9	14.9	16.3	13.5	8.6	21.2
Standard error (ha)	8.6	8.2	9.1	7.7	8.6	8.2

Incorporating remotely sensed data products can introduce errors, so assessing and identifying these errors is critical for data assimilation (Povey and Grainger, 2015). In addition, different resampling methods can also impact classification accuracy and the final model selection. This dissertation explored the influence of bootstrapping and k-fold CV techniques on gully classification for different seasons (dry and wet) and classifiers (SVM and RF). Results showed that k-fold CV outperformed bootstrapping in terms of commission error. Previous research (Kohavi, 1995; Kim, 2009) also found k-fold CV superior to bootstrapping for accuracy estimation and model selection. Furthermore, bootstrapping was found to have bias issues for both large and small samples despite its low variance, which implies better performance for small samples with k-fold CV.

While this study generally agrees with previous studies, it is essential to note that bootstrapping and k-fold CV performance varied considerably with algorithm and season at the class level. Although most studies using these resampling techniques do not often use class accuracy metrics, it is crucial to consider different accuracy metrics at the class level to increase the reliability of the results. However, the different accuracy metrics used in this study, such as UA, PA, standard error, and F1-score, sometimes disagreed. For example, some algorithms had high PA values but low corresponding UA values or vice versa. Based on the F1-score, the best algorithms belonged to RF, but the study relied on the standard error as a reliable measure of accuracy due to the

disagreement among various accuracy metrics. Therefore, to make sound conclusions regarding the performance of the algorithms being studied, it is advised to consider various accuracy metrics (Chicco and Jurman, 2020), which should be selected based on the specific aims of the research.

Concerning seasons, the algorithms were found to be more effective in identifying gullies on bare soil surfaces during the wet season. These findings are consistent with previous studies that found wet season images to be more effective (Vrieling *et al.*, 2007). However, the success of gully identification depends on various factors such as gully morphological characteristics (Phinzi, Holb and Szabó, 2021), sensor type and resolution, and classification algorithms (Lu and Weng, 2007). For instance, Sentinel and Landsat images have been found to perform better in dry than wet seasons (Sepuru and Dube, 2018). Despite the successful identification of gullies in the wet season, there were instances where gullies were filled with vegetation, which made automatic classification impossible. In such cases, high-resolution aerial photographs and/or dry-season PlanetScope images were used for visual interpretation.

The appearance of gullies was also an essential factor in the success of gully classification, with linear-shaped, continuous gullies being more straightforward to detect than areas with high gully density and transitional zones to non-gully (Orti *et al.*, 2020). The SVM algorithm combined with CV (svm-w-cv) performed the best in the wet season, with the lowest standard error (3.7 ha) and the highest PA (89.2%). The RF model (rf-d-b) had a slightly different standard error (6.1 ha) and PA (83.6%). However, half of the models had a PA below 70%. Despite this variability, the estimated gully areas (in hectares) based on area-weighted metrics were unbiased and reliable.

### 5.2.3. Study Area #3: efficacy of algorithms based on limited spectral information

Four classifiers - namely RF, K-NN, MD, and MLC - were evaluated based on their performance in extracting gullies from a visual range SPOT-7 image. All four classifiers generally achieved high accuracy values above 0.80, with RF recording the highest OA (0.94) and kappa (0.89). K-NN followed closely with OA and kappa values of 0.92 and 0.86, respectively (Figure 23). Finally, MD obtained an OA of 0.86 and a kappa of 0.76, while MLC had the lowest OA (0.83) and kappa (0.72).

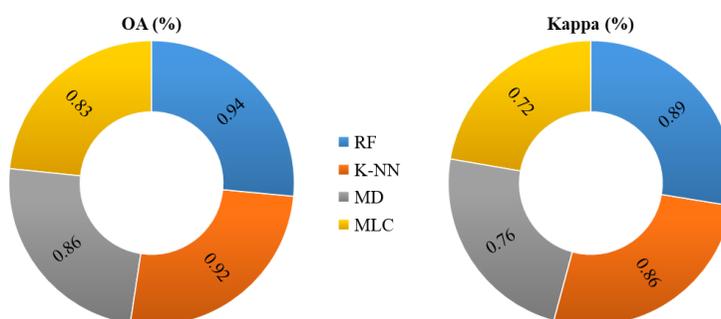


Figure 23. Overall accuracy (OA) and kappa coefficients (RF: random forest, K-NN: K-nearest neighbor, MD: minimum distance, MLC: maximum likelihood classifier).

Results showed significant variations in UA and PA across different classifiers, as presented in Figure 24. K-NN, MD, and RF performed excellently in identifying the BS class, each achieving 1.00 (100%) UA. Conversely, the MLC classifier exhibited a relatively low UA of 0.60 but had the highest PA (100%) compared to the other classifiers. For the gully (G) classification, MD was found to be the most effective classifier with 100%

UA, although the corresponding PA was only 0.80. RF achieved the highest OA (0.94) but had an 87% UA for gully classification.

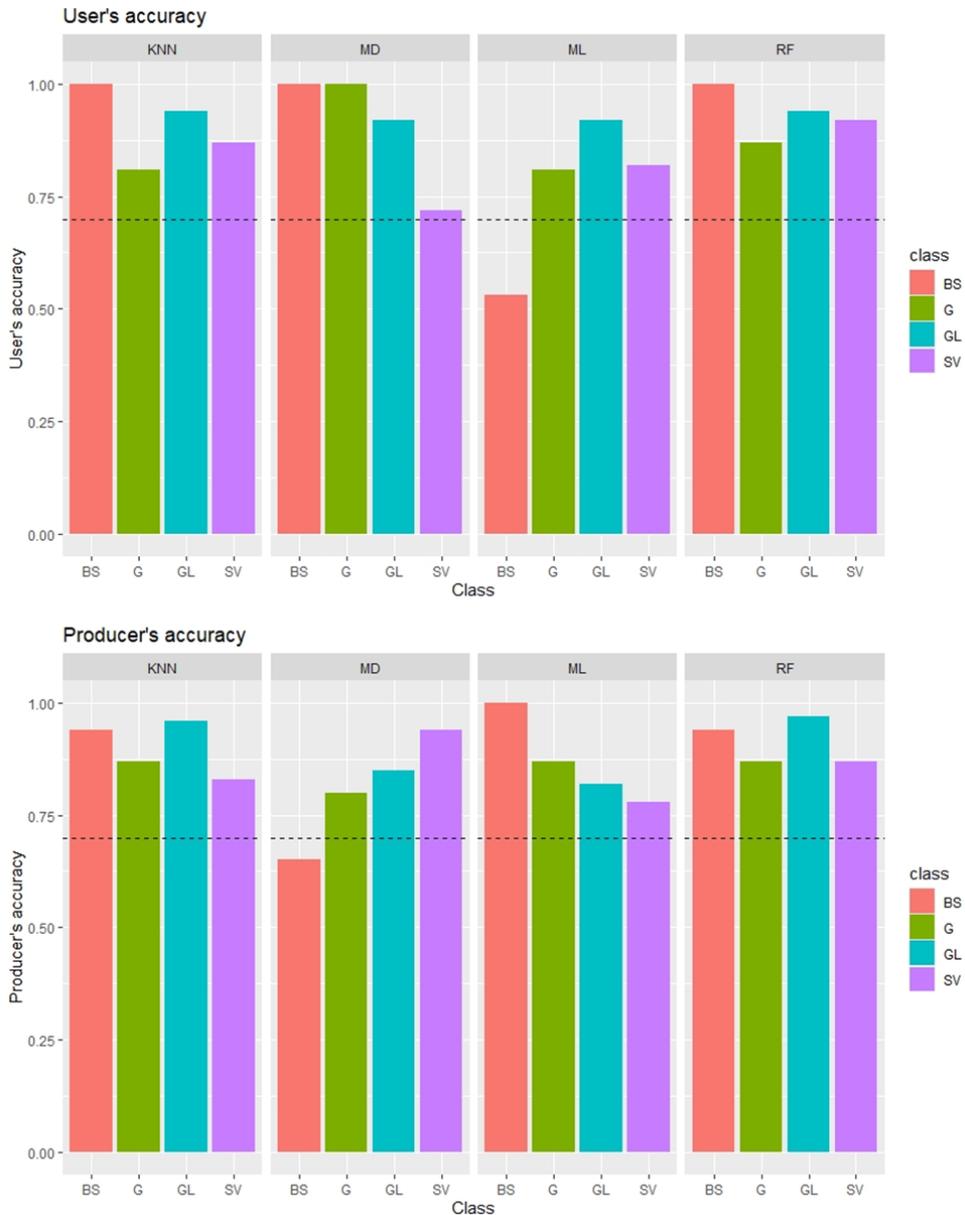


Figure 24. User's and producer's accuracy (dashed black line indicates class accuracy benchmark of 0.70; GL: grassland, SV: stressed vegetation, G: gully, BS: bare soil).

According to Everitt et al. (2008), it is recommended to achieve an overall accuracy (OA) rate of at least 0.85 and class-specific accuracies, namely producer's accuracy (PA) and user's accuracy (UA), of 0.70 as a benchmark for operational purposes. All the applied classifiers except for the MLC classifier surpassed the 0.85 benchmarks for OA, with RF (0.94) and KNN (0.92) achieving above 0.90 and MD obtaining an OA of 0.86. RF's superior performance compared to other methods is unsurprising, as similar results have been reported (Khatami, Mountrakis and Stehman, 2016). The superior performance of the RF classifier can be attributed to its ability to generate improved classification results as an ensemble of classifiers compared to its performance as an individual classifier (Rodriguez-Galiano *et al.*, 2012).

Although OA is a widely used and valuable metric with a straightforward interpretation (Heydari and Mountrakis, 2018), it has the limitation of concealing class-specific performance (He and Garcia, 2009). The algorithms could classify gullies with PAs and UAs above the 0.70 benchmarks. It is also important to note that a relatively large gully area does not necessarily result in higher accuracies, particularly concerning the MD classifier, which had the smallest proportion of gully area among all the classifiers evaluated in all four sites, but achieved no commission error (i.e., 100% UA) and had 20% omission error. KNN and ML had identical commission (19%) and omission (13%) errors, while RF recorded 13% in both omission and commission errors. Overall, these errors were relatively low compared to previous studies conducted in South Africa (Mararakanye and Nethengwe, 2012; Phinzi and Ngetar, 2017; Makaya *et al.*, 2019).

#### ***5.2.4. Visual analysis of gully classification across different study areas***

Gully classification results for Study Area #1 using multiclass and binary approaches are presented in Figure 25 and Figure 26, respectively. The

performance of SVM and RF were found to be similar, while LDA showed different results, particularly in site 1 (1A→1B, 1A→1C), site 1B (1B→1C), and site 1C (1C→1B). However, in some instances, such as site 1B (1B→1A) and site 1C (1C→1A), LDA produced comparable outcomes to SVM and RF. In addition, the multiclass and binary approaches had comparable gully extraction outcomes, though their accuracy results varied slightly.

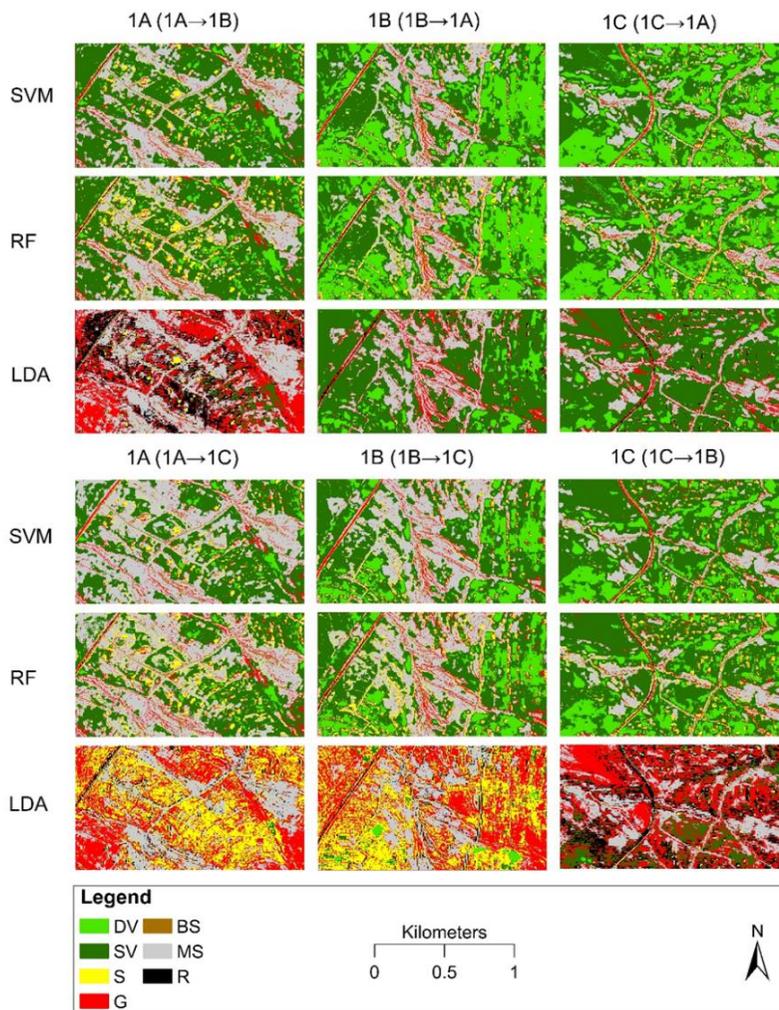


Figure 25. Classification results of the multiclass approach (DV: dense vegetation, SV: stressed vegetation, S: settlement, G: gully, BS: bare soil, MS: mixed bare soil, R: road).

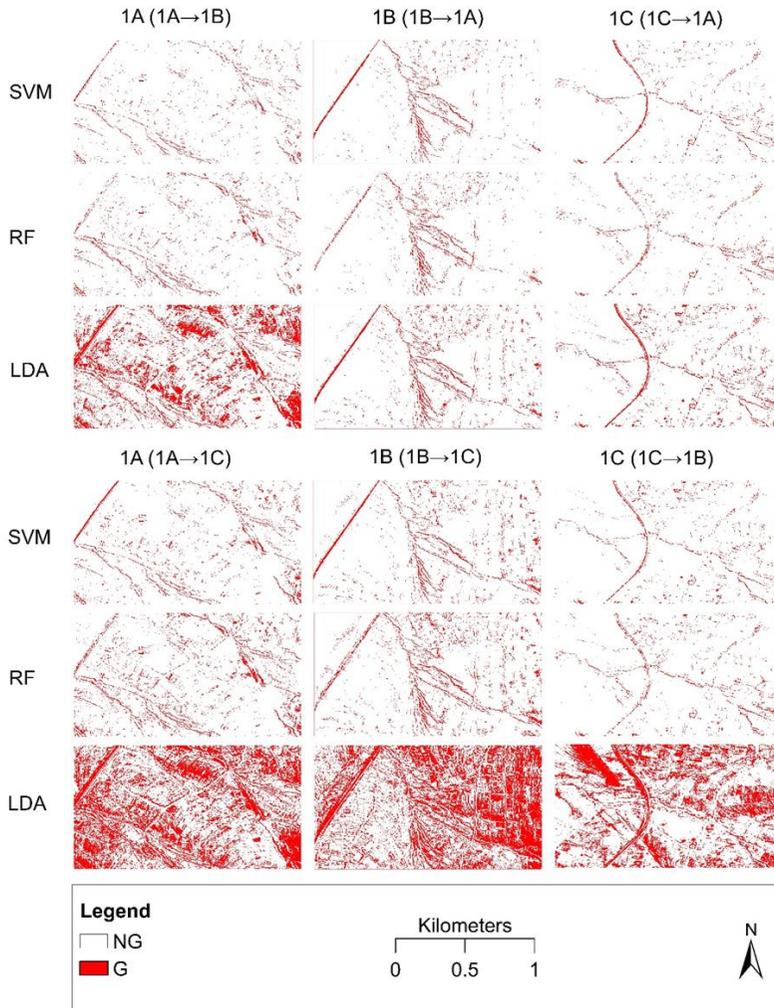


Figure 26. Classification results of the binary approach (G: gully, NG: non-gully).

In Study Area #2, gullies were effectively distinguished from other land cover classes during the dry and wet seasons; however, there were noticeable differences in the distribution of extracted gullies between the two seasons (Figure 27). The wet season exhibited a higher abundance of gullies than the dry season. This contrast was more apparent in Figure 27a, corresponding to rf-d-b, and Figure 27b, representing the svm-w-cv model. In addition, variations in gully reflectance between the two seasons impacted

gully classification. The wet season had more vegetation, leading to greater spectral differences (Figure 28). Conversely, in the dry season, most gullies shared spectral characteristics with the bare surfaces they intersected, making it harder to extract gullies on bare soil surfaces. This result may explain the high commission error (43.1%) and standard error (8.6 ha) observed in the dry season (Section 4.2.2).

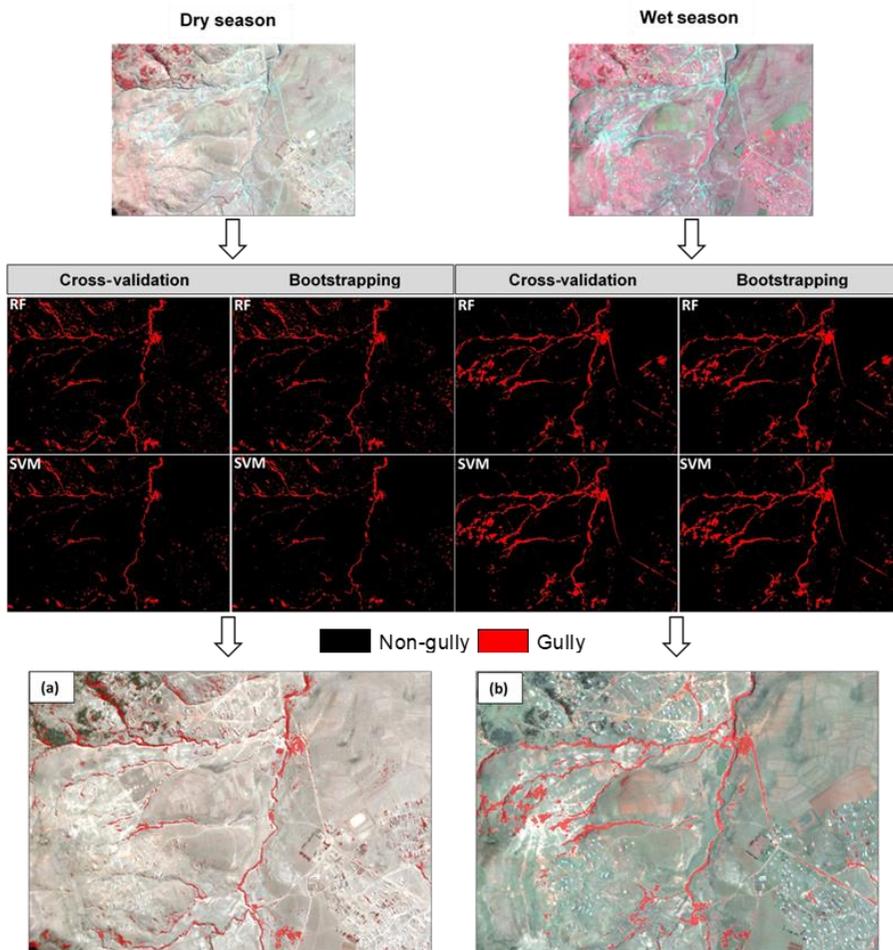


Figure 27. Spatial distribution of gullies: (a) rf-d-b and (b) svm-w-cv correspond to the best models for gully mapping in the dry and wet seasons, respectively (rf: random forest, svm: support vector machine, w: wet season, d: dry season, cv: cross-validation, b: bootstrapping).

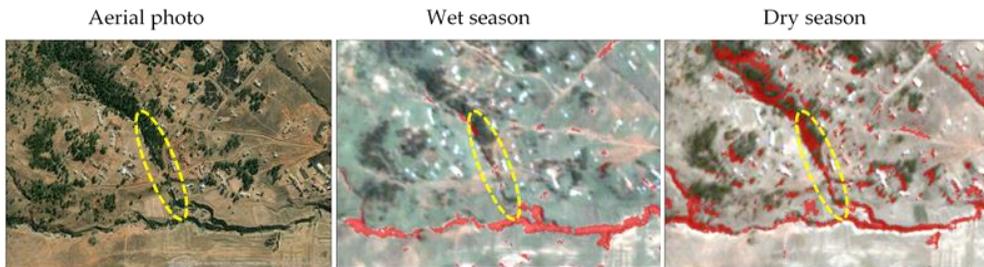


Figure 28. An example of a vegetated gully (dashed yellow ellipse) in the dry and wet seasons.

The first step in Study Area #3 to distinguish gullies from surrounding areas involved performing a general land cover classification, which produced different results with various algorithms (Figure 29). Grassland (GL) was the most common land cover in all classifiers, ranging from 47-57%, followed by the stressed vegetation (SV) class, which accounted for 27-31%. The gully (G) class was the smallest, ranging from 5-8%, while bare soil (BS) represented 5% of the MD classifier results. The MLC algorithm had the highest proportions of BS (19%) and G (8%), while RF and KNN had the largest percentages of GL (57% and 53%, respectively). In contrast, the MD classifier had the highest percentage of SV (31%), but the lowest percentage (5%) of G and BS.

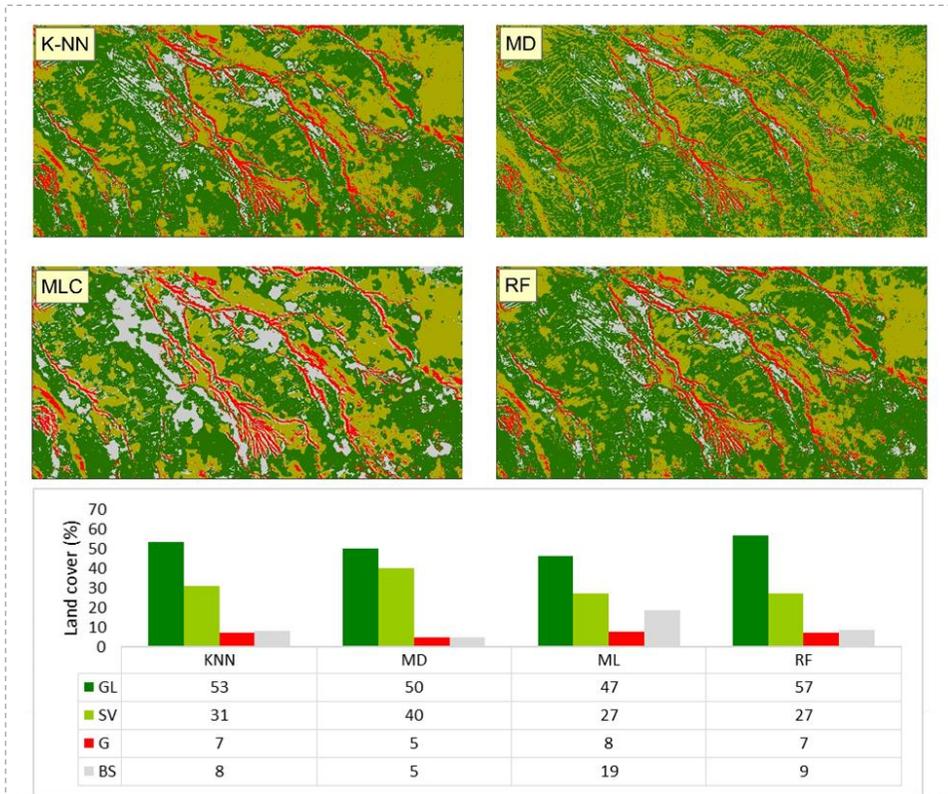


Figure 29. Land cover classification (GL: grassland, SV: stressed vegetation, G: gully, BS: bare soil) based on different algorithms (KNN: k-nearest neighbor, MD: minimum distance, ML: maximum likelihood, RF: random forest).

The second step involved mapping gullies in different sites within Study Area #3. Site 3A generally had the largest proportion of gullied area, ranging from 12-19% (9321-10344 m<sup>2</sup>; Figure 30, Table 13) among different classifiers. Sites 3C and 3D had the least gullied areas of 1668 m<sup>2</sup> to 3147 m<sup>2</sup> (3-6%) for the latter and 1039 m<sup>2</sup> to 1774 m<sup>2</sup> (2-3%) for the former. Across all selected gully sites, the MD classifier consistently recorded the smallest gully area. In four selected sites, other classifiers produced the same results, except for site 3A, where the ML classifier had the highest areal extent of gully

erosion (19%), followed by RF and KNN, recording 17%, and then MD with 12%.

The algorithms produced spatial distribution patterns similar to the actual (digitized) gullies. However, there was a noticeable disparity in the proportion of areas classified as gullies by different algorithms. Across all study sites, the algorithms failed to detect a considerable amount of gullies, with the most substantial difference observed at site 3C. For example, KNN, ML, and RF algorithms classified 1774 m<sup>2</sup> as gullied area, while the MD algorithm only classified 1039 m<sup>2</sup>, considerably less than 8064 m<sup>2</sup> (actual gully area).

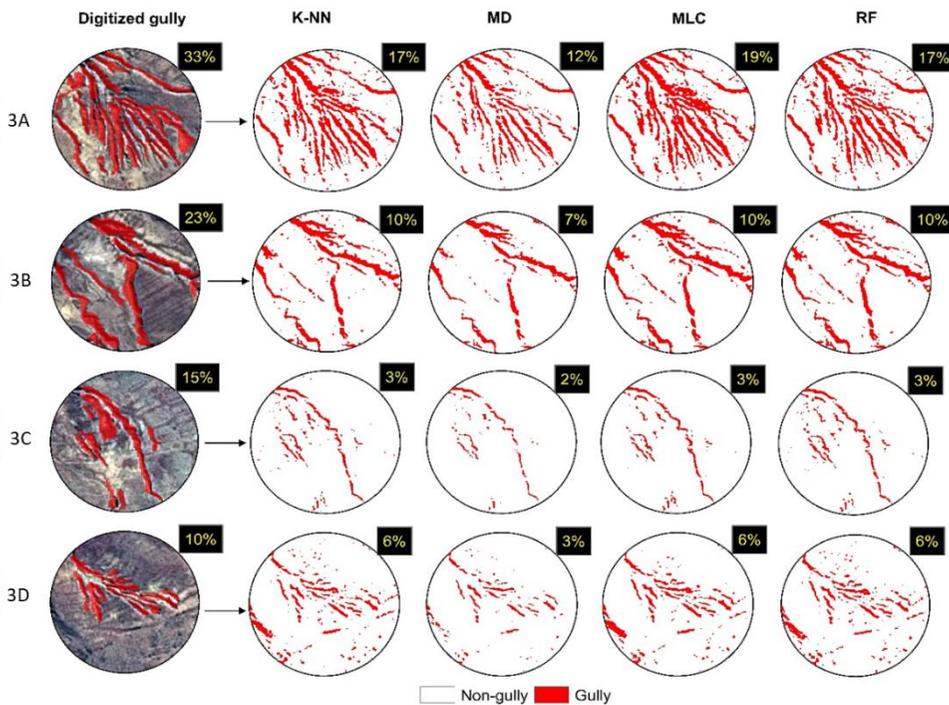


Figure 30. Selected gully sites (3A–3D) showing the spatial distribution of the classified gullies by different algorithms (KNN: k-nearest neighbor, MD: minimum distance, ML: maximum likelihood, RF: random forest) and actual (digitized) gullies.

Table 13. Aerial extent (m<sup>2</sup>) of actual gullies (digitized) and classified gullies based on different algorithms (KNN: k-nearest neighbor, MD: minimum distance, ML: maximum likelihood, RF: random forest).

Method	Site 3A		Site 3B		Site 3C		Site 3D	
	Pixels	Area (m <sup>2</sup> )						
KNN	911	9321	547	5601	173	1774	306	3129
MD	632	6463	393	4026	102	1039	163	1668
ML	1011	10,344	558	5706	157	1610	302	3088
RF	898	9193	548	5607	171	1747	308	3147
Digitized	1750	17,908	1210	12,382	788	8064	556	5690

### 5.3. Gully characteristics' influence on the precise mapping of gullies and their density

Gullies' characteristics strongly influenced their classification. The applied algorithms were most efficient in areas where gullies exhibited a dendritic pattern, such as site 3A. However, in site 3D, where gullies also exhibited a dendritic network, classifiers were less effective in detecting gullies, possibly due to their shallower depth than those at site 3A. The most comprehensive and deepest gullies were found at site 3B, where the depth was more important than the width in discriminating gullies. Shallow and wide gullies were more challenging to detect, particularly on bare soil. Gully length was not a significant factor in detection efficiency. Site 3C's longest gully had shallow walls, resulting in lower classifier efficiency. There was a discrepancy between the actual digitized gullies and those derived by the algorithms in all sites, possibly due to differences in resolution between the aerial photograph used for digitization (0.5 m) and the SPOT imagery used for classification (1.3 m) or because the classifiers mainly detected gullies with steep-sided walls or with shadows. In contrast, digitized gullies were captured in their correct

shape with exact boundaries. Similar issues were reported in a previous study where the pixel resolution of the input data was not fine enough to capture the flat parts of the gullies, resulting in the under-classification of gullies (d'Oleire-Oltmanns *et al.*, 2014). Nevertheless, it is worth acknowledging that the disparity between digitized and classified gullies is not a significant focus of concern in this dissertation. The digitized gullies were employed solely for the purpose of validating the accuracy of classified gullies in relation to their spatial extents.

The gully density serves as an indicator of the quantity of gullies, considering both their width and length, within a particular area (Li, Xiong and Tang, 2019). Results showed that the gully density varied between 0.12 m/m<sup>2</sup> and 0.61 m/m<sup>2</sup> (Figure 31) and was influenced by the appearance and pattern of the gullies. Site 3A, characterized by a dendritic network of gullies and a lack of vegetation cover, exhibited the highest gully density. Linear and V-shaped gullies with steep-sided walls in site 3B also had high densities, which varied across the study area. Previous research has demonstrated that slope steepness plays a significant role in determining gully density. For example, Zhang *et al.* (2016b) reported a positive correlation between slope gradient and gully density on hillslopes, while Muñoz-Robles *et al.* (2010) found that most areas with gullies in their study had steep slopes. In contrast, the studied gullies in this dissertation were distributed on gently sloping agricultural land, indicating that slope was not the primary factor influencing gully density, despite hilly areas. These findings align with the research conducted by Mararakanye and Le Roux (2012), which also concluded that gullies are more prevalent on agriculturally suitable land compared to land categorized unsuitable. Kakembo *et al.* (2009) similarly discovered that gullies were more common in areas characterized by gentle slopes.

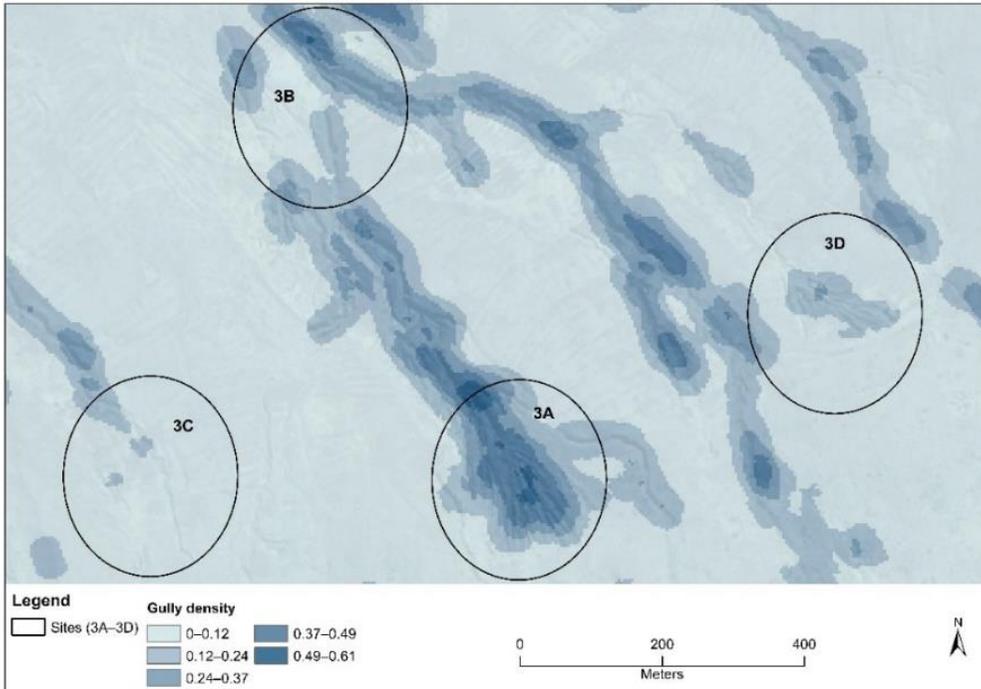


Figure 31. Gully density ( $\text{m}/\text{m}^2$ ) map of the study area.

#### 5.4. Feature selection and multicollinearity analysis

Variables were first ranked in their relative importance in the modeling process. This ranking was carried out using RF variable importance analysis in the R software. Results indicated that NDVI (importance=100%), elevation (26%), TWI (25%), population density (23%), SPI (22%), and LULC (19%) had considerable predictive power for gully susceptibility modeling (Figure 32a). The least essential variables, including aspect (0%), geology (0.20%), TRI (0.27%), and distance from roads (3.83%), were removed by the RFE algorithm, retaining 18 variables (out of 22). The applied algorithm showed no improvement in accuracy after the 18<sup>th</sup> variable, denoted by a blue-dashed line (Figure 32b). Although calcium carbonate ( $\text{CaCO}_3$ ) and pH were part of the retained predictors, they had a high correlation ( $r^2 > 0.75$ ) with other predictors and hence were also removed, resulting in 16 predictors (Figure

32c). These predictors were grouped into three feature sets (small, medium, and large sets) based on their relative importance.

Multicollinearity diagnostic tools, including tolerance and VIF, were used to check if multicollinearity exists among predictors in each feature set. Almost all predictors had relatively high tolerance ( $>0.22$ ) and low VIF ( $<4.5$ ) values which indicate the non-existence of multicollinearity (Table 14). Such findings are comparable to previous studies (Arabameri, Chen, *et al.*, 2019; Pham *et al.*, 2020; Hitouri *et al.*, 2022; Jaafari *et al.*, 2022), where VIF values of  $\leq 5$  were reported. On the contrary, elevation yielded the lowest tolerance (0.18) and highest VIF (5.65) in a larger set. Nevertheless, these values also fall within the acceptable VIF threshold, considering a long-standing rule of thumb (i.e.,  $VIF < 10$ ) for the non-existence of multicollinearity (Gareth *et al.*, 2013; Vatcheva *et al.*, 2016). Thus, all geo-environmental predictors in each feature set met these criteria and were used in the modeling process.

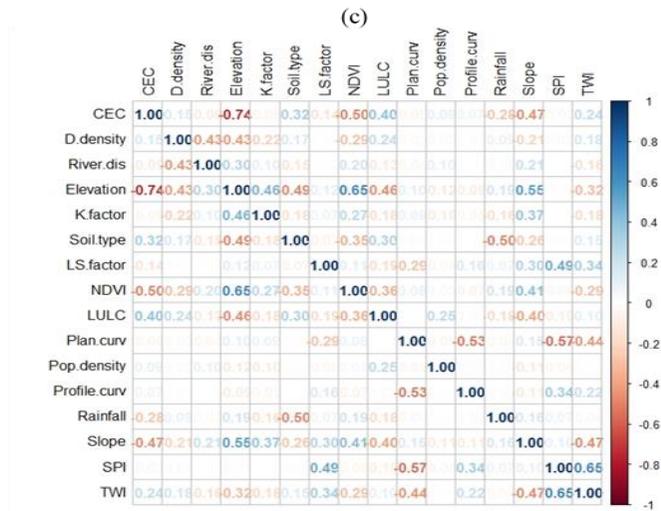
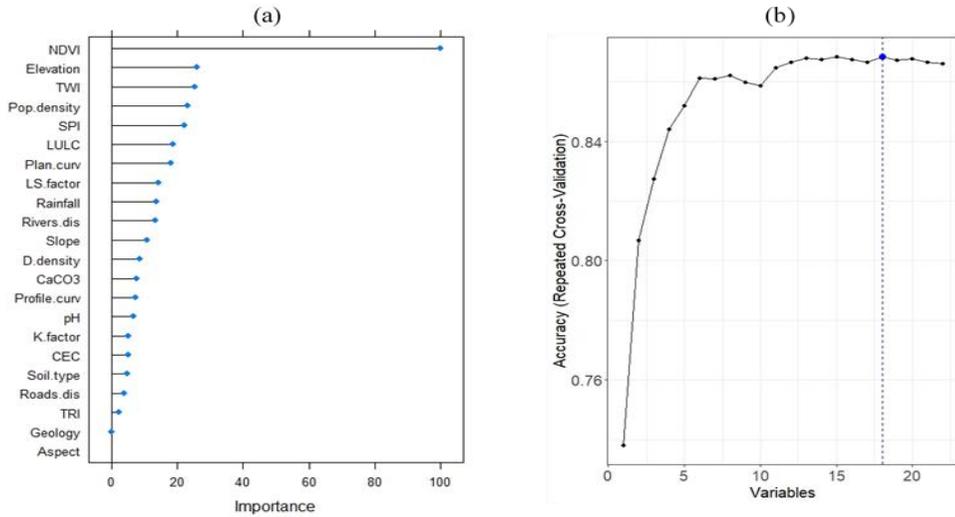


Figure 32. Ranking and selection of important predictors based on: (a) variable importance, (b) recursive feature elimination (the blue-dashed line marks the eighteenth variable), and (c) correlation matrix. Note: CEC: cation exchange capacity, D.density: drainage density, River.dis: distance from rivers, Plan.curv: plan curvature, Pop.density: population density, Profile.curv: profile curvature, SPI: stream power index, TWI: topographic wetness index.

Table 14. Tolerance (Tol) and variance inflation factor (VIF) for geo-environmental predictors in three feature sets (large, medium, and small sets) (TWI: topographic wetness index, LS: slope length, and steepness, NDVI: normalized difference vegetation index, LULC: land use/land cover, CEC: cation exchange capacity, SPI: stream power index).

Large set			Medium set			Small set		
Variables	Tol	VIF	Variables	Tol	VIF	Variables	Tol	VIF
Population density	0.87	1.16	Rainfall	0.90	1.2	Population density	0.92	1.09
Distance from rivers	0.75	1.33	Population density	0.88	1.13	LULC	0.73	1.37
Profile curvature	0.70	1.42	Distance from rivers	0.79	1.27	NDVI	0.55	1.80
LULC	0.67	1.49	LULC	0.68	1.47	SPI	0.52	1.92
LS factor	0.63	1.59	Drainage density	0.68	1.47	Elevation	0.48	2.07
Rainfall	0.63	1.59	LS factor	0.66	1.52	TWI	0.46	2.17
Drainage density	0.62	1.60	Plan curvature	0.59	1.69			
Land type	0.57	1.76	NDVI	0.52	1.92			
K factor	0.54	1.86	Elevation	0.42	2.37			
NDVI	0.51	1.94	Slope	0.31	3.21			
Plan curvature	0.50	1.98	SPI	0.29	3.50			
CEC	0.32	3.14	TWI	0.25	3.96			
Slope	0.30	3.39						
SPI	0.26	3.86						
TWI	0.23	4.38						
Elevation	0.18	5.65						

### 5.5. Analyzing algorithms' performance when using feature sets of varying sizes

The performance of six different algorithms was evaluated regarding the accuracy and processing time, using three different feature subsets (small, medium, and large), as shown in Figure 33. The results showed that SVM was the most efficient algorithm with medium and small feature sets, achieving the

highest F1-score (0.897), OA (0.898), and specificity (0.908) with the medium set and the least computation time (<2 minutes). SGB followed SVM closely in most evaluation metrics but was more efficient with larger feature sets and took more time (>5 minutes) to compute. RF was the most computationally expensive algorithm and performed poorly with small and medium sets but consistently yielded high accuracies with a large feature set. PLS and RDA were the fastest algorithms, with PLS taking less than a minute to compute, and both produced relatively high accuracies with a medium feature set. ANN was also sensitive to the number of input features, taking several minutes of computation with a larger feature set but consistently achieved relatively high accuracies with the small feature set, outperforming PLS and RDA (except in sensitivity) across most evaluation metrics. Overall, the performance of each algorithm varied depending on the feature set size and the evaluation metric used.

All six algorithms produced similar outputs regarding the susceptibility maps and portrayed reasonable gully susceptibility, except RDA, which exhibited a different pattern (Figure 34). Class-wise metrics, particularly specificity (<0.85), suggest that RDA highly misclassified the absence of gully erosion. This misclassification pattern can also be observed with a small and medium feature set where RDA produced the worst specificity values ( $\leq 0.83$ ). Better performance of SVM can be observed in its gully susceptibility map. Given the two-class (gully and non-gully) response variable, it is not surprising that SVM, a binary classifier, yielded impressive results. An additional advantage of SVM over other algorithms is using various kernel functions to find an optimal hyperplane that separates data points (support vectors) of two classes (gully and non-gully). Specifically, the applied radial kernel function ensured perfect separation of the gully and non-gully support vectors in the feature space. Rahmati et al. (2017) also found

SVM with the radial kernel function to be more accurate for predicting gully susceptibility than other kernels. Despite these advantages, RF outperformed SVM in most gully susceptibility studies (Garosi *et al.*, 2019; Gayen *et al.*, 2019; Pourghasemi *et al.*, 2020; Hitouri *et al.*, 2022; Huang *et al.*, 2022). However, in our case, the advantage of RF was experienced only with the larger predictor set, implying more computational time. Similarly, SGB required more predictors to reach its maximum accuracy but at the expense of computation time, albeit shorter than RF.

On the contrary, despite using the smallest feature set and hence shorter computation time, ANN was efficient and produced high-quality maps comparable to those of other algorithms with the best predictive performance. Likewise, PLS delivered an impressive gully susceptibility map with the shortest possible computation time (e.g., <1 minute). These results demonstrate that a fixed set of features for comparison would not provide an accurate assessment since specific algorithms, such as ensemble-based ones like RF and SGB, require larger input feature sets to achieve their full predictive potential. On the other hand, SVM and ANN performed well with fewer input features and less computational time.

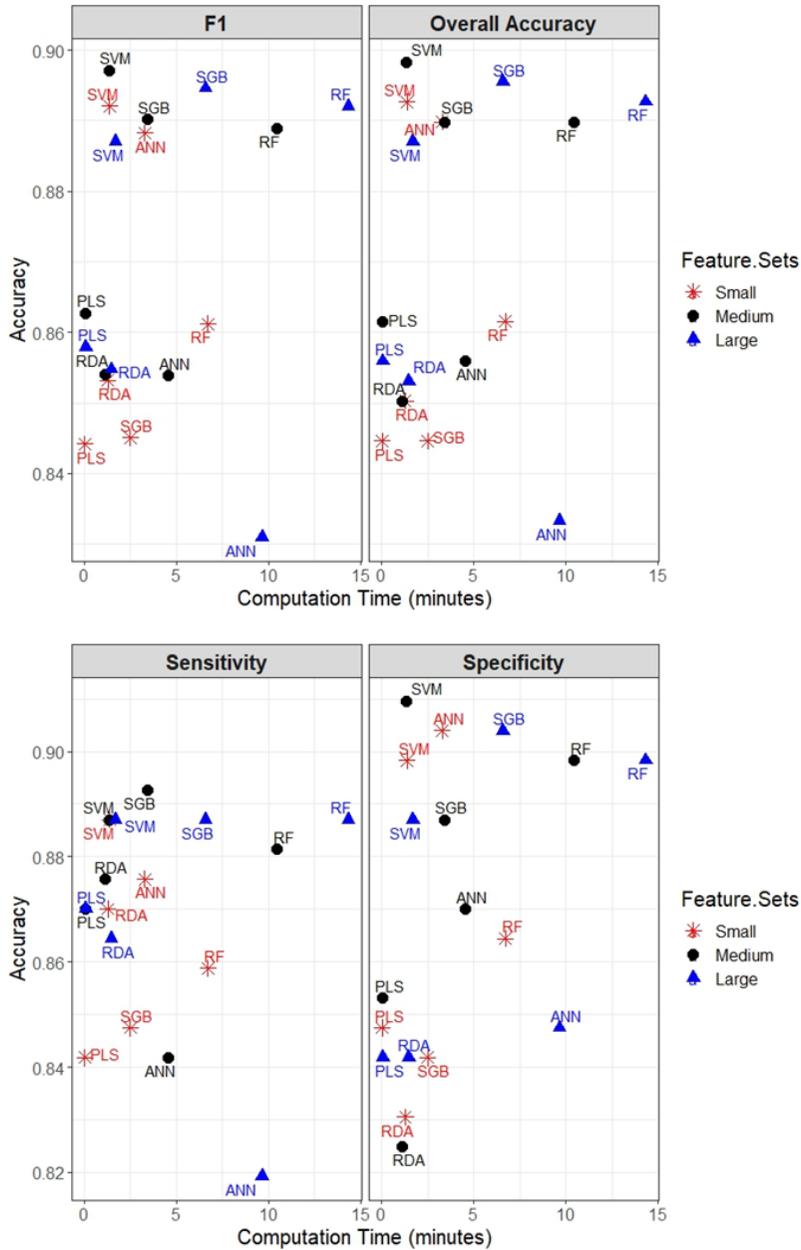


Figure 33. Predictive performance of ML models (ANN: artificial neural network, PLS: partial least squares, RDA: regularized discriminant analysis, RF: random forest, SGB: stochastic gradient boosting, SVM: support vector machines) using smaller, medium, and larger feature sets.

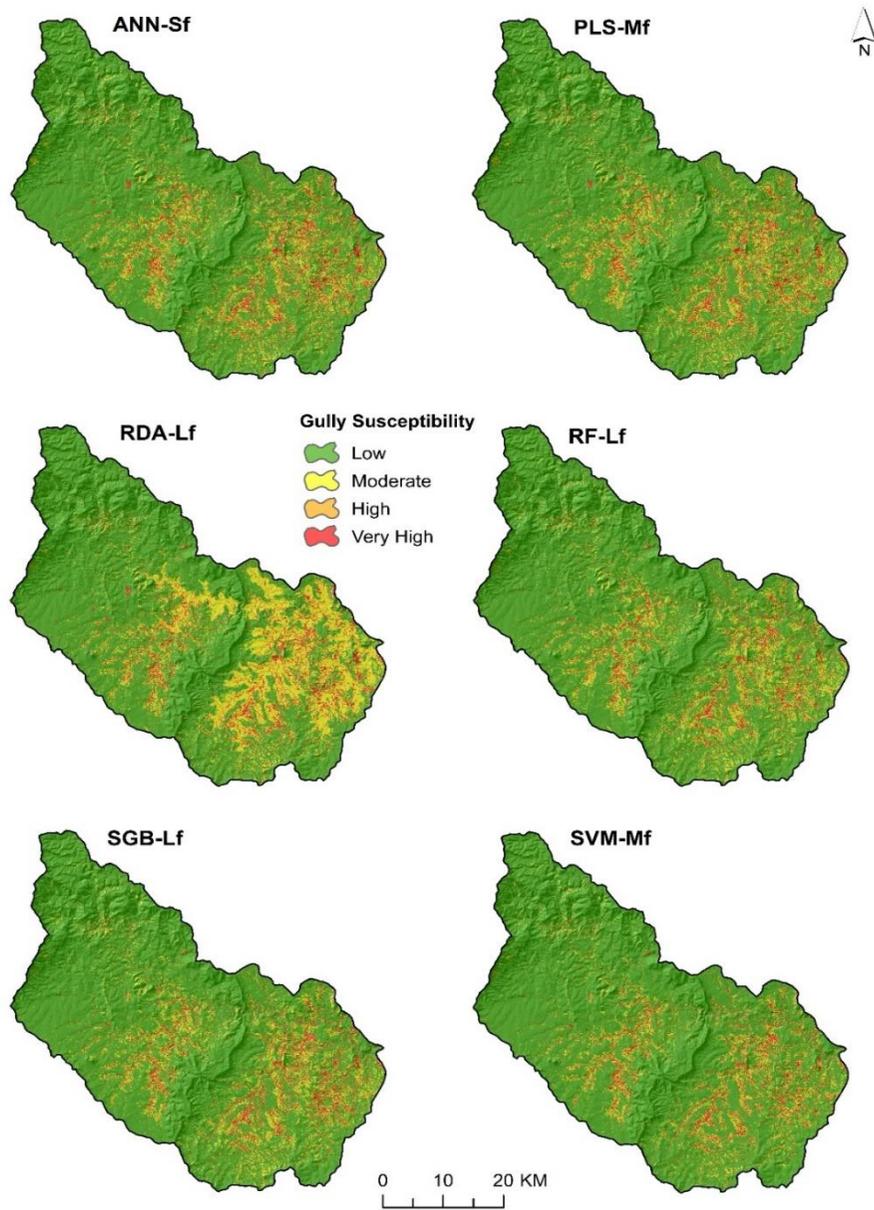


Figure 34. Gully susceptibility maps for each algorithm computed with smaller (Sf), medium (Mf) or larger (Lf) feature sets (ANN: artificial neural network, PLS: partial least squares, RDA: regularized discriminant analysis, RF: random forest, SGB: stochastic gradient boosting, SVM: support vector machines).

## 5.6. Gully susceptibility and key controlling geo-environmental variables

All six algorithms indicated that a considerable portion (71-84%) of the catchment is of low gully susceptibility, whereas only 0.2-2.6% is under high gully susceptibility (Figure 35). Furthermore, ranging from 4.5-6.6%, the proportion of very high gully susceptibility was almost the same across all algorithms. A very high susceptibility class corresponds to severely gullied areas, suggesting these algorithms can detect individual gullies. In particular, the SVM-derived map is a good example illustrating areas with varying degrees of gully susceptibility (Figure 36). SVM was used to produce this final map due to its superior predictive performance.

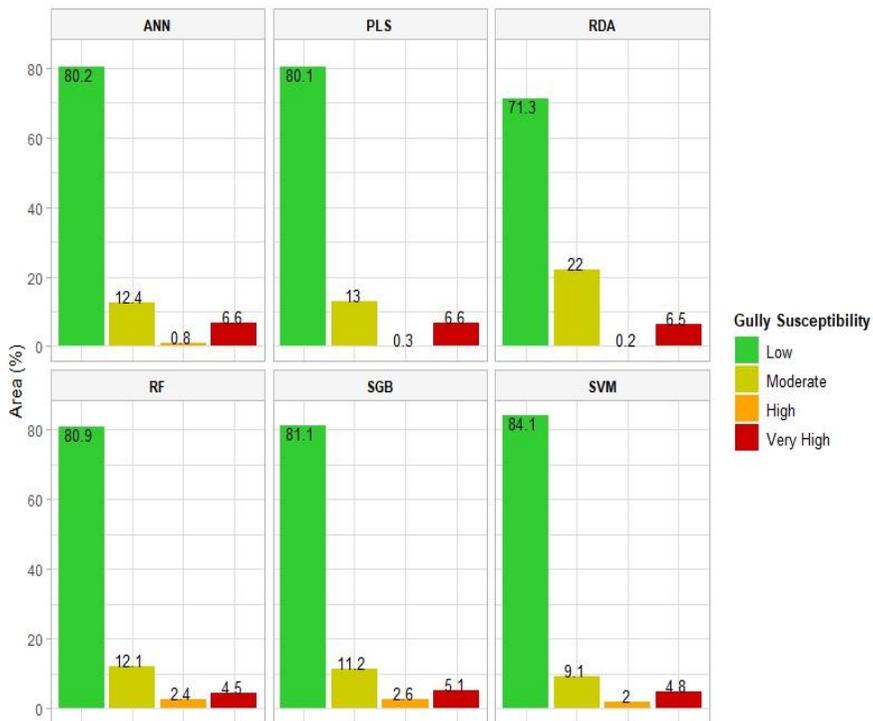


Figure 35. The proportion of area under different levels of gully susceptibility (ANN: artificial neural network, PLS: partial least squares, RDA: regularized discriminant analysis, RF: random forest, SGB: stochastic gradient boosting, SVM: support vector machines).

Medium to very high gully susceptibility is primarily confined to low-lying areas with gentle to flat slopes throughout the catchment (Figure 36c-f), although some elevated and hilly parts (i.e., Figure 36a) fall under these gully susceptibility classes. Extensive gully systems are remarkable in the central (Figure 36c), eastern (Figure 36e), and southeastern (Figure 36f) parts of the catchment. On the contrary, the western section of the catchment, particularly those areas with commercial farming activity, is predominantly less susceptible to gully erosion (Figure 36b).

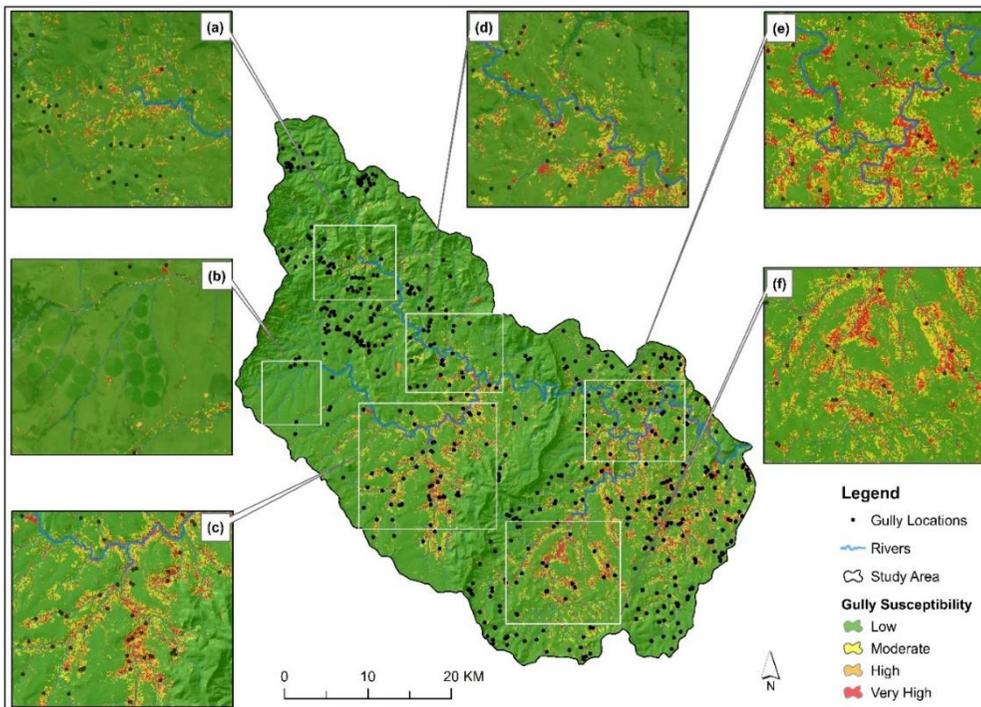


Figure 36. Distribution of gully locations across different parts (a-f) of the catchment with varying degrees of gully susceptibility

Among geo-environmental variables, NDVI (importance=100%), followed by elevation (importance=26%), TWI (importance=25), population density (importance=23%), SPI (importance=22%), and LULC

(importance=19%), had the greatest predictive power for gully susceptibility modeling. This finding is consistent with previous studies where one or more variables were the most critical in predicting gully susceptibility. For example, NDVI (Barakat *et al.*, 2022; Jaafari *et al.*, 2022), elevation (Pham *et al.*, 2020; Chowdhuri *et al.*, 2021; Huang *et al.*, 2022), TWI (Jiang *et al.*, 2021; Han, Guzman and Chu, 2022), SPI (Barakat *et al.*, 2022), and LULC (Roy and Saha, 2022) were among the top five most important factors contributing to gully erosion. It is not surprising that NDVI ranked first in importance (100%) in this study because of the unique spectral reflectance of gullies (lower values) against vegetation (higher values). In addition, it has been proven helpful for extracting gullies and other erosional features (Vaidyanathan *et al.*, 2002; Phinzi and Ngetar, 2017). For example, Bernini *et al.* (2021) found NDVI to be the most critical factor influencing gully erosion in the Mkhomanzi catchment in South Africa. The authors reported that areas with an NDVI range of 0-0.30 were susceptible to gully erosion, while another study by Phinzi and Ngetar (2017) reported 0.15-0.25 values. In this study, >96% of gully pixels occurred within an NDVI range of -0.23-0.35, whereas <4% of gully pixels were within 0.35-0.72 NDVI values (Table 15), suggesting that areas with little or no vegetative cover are the most susceptible to gully erosion.

Similarly, areas with low elevation (538-765m) were more vulnerable to gully erosion than highly elevated areas (1039-1772m). Due to greater surface runoff with higher erosive power, low-lying areas with poor vegetation cover are usually predisposed to gully erosion. These low-lying areas had varying degrees of water accumulation, as indicated by TWI values ranging from <6.13-7.53, with about 80% of gully pixels falling within this range. Low TWI values indicate less wetness or water accumulation, and high values indicate areas where surface water flow is more likely to accumulate (Sørensen, Zinko and Seibert, 2006), making the area vulnerable to gully erosion given higher

erosive power (high SPI values). Areas with SPI values of <3 were found to be the most susceptible to gully erosion, as over 93% of gully pixels occurred within this SPI range. These results are in agreement with related studies

Table 15. Relationship between the six most critical geo-environmental predictors and gully erosion.

<b>Gully factor</b>	<b>Class</b>	<b>Gully pixels</b>	<b>Gully pixels (%)</b>
NDVI	-0.23-0.18	132725	18.068
	0.18-0.24	196895	26.803
	0.24-0.29	167825	22.846
	0.29-0.35	133145	18.125
	0.35-0.72	24277	3.305
Elevation (m)	538-765	276441	37.625
	765-893	342962	46.679
	893-1039	24540	3.340
	1039-1255	10286	1.400
	1255-1772	748	0.102
TWI	<6.13	287932	39.189
	6.13-7.53	295262	40.187
	7.53-9.36	62294	8.479
	9.36-11.89	9437	1.284
	11.89-24.99	52	0.007
Population density (people/km <sup>2</sup> )	1	654515	89.143
	1-85	0	0.000
SPI	<-6.25	156181	21.257
	-6.25- -1.56	117122	15.941
	-1.56-0.36	166513	22.663
	0.36-3.23	169113	23.017
	3.23-13.35	46048	6.267
LULC	Forested		
	Land	8126	1.106
	Grassland	318769	43.389
	Waterbodies	2703	0.368
	Wetlands	1073	0.146
	Barren Land	58494	7.962
	Cultivated	239029	32.535
	Built-up Mines & Quarries	26123	3.556
	606	0.082	

where a considerable proportion (i.e., >60%) of gully pixels fell within TWI of <10 and SPI<9 (Arabameri, Cerda, *et al.*, 2019; Gayen *et al.*, 2019; Hosseinalizadeh *et al.*, 2019b).

Concerning LULC, grassland (gully pixels=43.39%) and cultivated lands (gully pixels=32.54%) were the most susceptible LULC classes to gully erosion. According to Le Roux *et al.* (2008), grasslands in South Africa are highly susceptible to erosion, which is not surprising given the predominantly rural character of the catchment and the widespread involvement of the local population in subsistence agriculture, including both crop and livestock farming. This view is supported by numerous studies in South Africa, which have indicated that human activities, including unsound agricultural practices and overgrazing, have contributed to soil erosion in rural areas (Beckedahl and de Villiers, 2000; Kakembo and Rowntree, 2003; Mhangara, Kakembo and Lim, 2012; Phinzi and Njoya S. Ngetar, 2019). For example, a recent study showed that 40% of erosion in communal areas is due to overgrazing and cattle tracks, while only 8.4% is attributed to population pressure (Olivier, Van De Wiel and De Clercq, 2022). The present study revealed that areas with lower population densities are more prone to gully erosion than those with higher densities, which is consistent with the rural nature of the catchment, which has sparse human settlements and extensive grasslands where gullies commonly occur.

## **6. SUMMARY AND CONCLUSIONS**

Gully erosion poses a significant challenge to sustainable agriculture, particularly in semi-arid regions such as South Africa, where subsistence agriculture is vital. In order to address this issue effectively, it is imperative to accurately identify gullies through satellite imagery and model areas susceptible to gully erosion with a high degree of precision.

This practical approach is essential for successful gully rehabilitation efforts in semi-arid environments. Therefore, the objectives of this dissertation were to:

- 1) *Determine if low-cost, high-resolution sensors improve gully mapping in semi-arid regions.*

Three low-cost, high-resolution satellite imagery were used, comprising two SPOT-7 images and a PlanetScope image. The study found that SPOT-7 multispectral bands showed discernible spectral differences, with the NIR band having the most significant effect size. While p-values indicate significance, effect sizes provide insight into the magnitude of differences, and standardized measures demonstrate relevance to land cover categories. Specific categories, such as gullies and stressed vegetation, showed non-significant differences in RGB bands but significant differences in the NIR band.

However, in Study Area #3, the visual range SPOT-7 image without the NIR band effectively discriminated against gullies due to its high spatial resolution (1.3 m). NGRDI successfully discriminated gullies from vegetation, as seen in their spectral profiles. However, vegetation identification during the training phase was challenging due to the absence of the NIR band and the dry season. High-resolution Google Earth images helped identify vegetation cover during training.

PlanetScope spectral bands, land cover classes, and seasons significantly influenced reflectance, and their interactions explained 92.3% of the variance. The difference between dry and wet seasons had the most significant effect on reflectance, while the impact of bands and land cover classes was slightly lower but still substantial. NIR and red bands were the most influential in discriminating gullies

from surrounding areas, while the blue band had the most negligible impact. NDVI was less effective in differentiating gullies from other land cover classes than the original bands. Gullies significantly differed from other land cover classes in the dry season but not in the wet season, except for a few specific cases.

2) *Quantify gully classification accuracy and analyze factors biasing model performance on a class level.*

The study shows that accurate identification of gullies can be achieved using RF or SVM classification algorithms, with similar levels of accuracy ranging from 92% to 96%. The binary approach resulted in better overall performance, with RF being more reliable than SVM. GLM analysis showed that study sites and algorithms significantly influenced classification models' efficiency, explaining 59.3% of the variance in UA and 56.1% in PA. The type of classification approach (binary or multiclass) also had a significant impact on algorithm performance. The multiclass classification approach resulted in lower overall accuracy than the binary approach but was more efficient for gully identification.

The study evaluated the performance of RF and SVM algorithms for gully identification in Study Area #2 using two resampling methods (CV and bootstrapping) for both wet and dry season scenes of PlanetScope data. RF consistently outperformed SVM, and the dry season yielded higher OA than the wet season. UA values for all models were above 70%, and PA was generally lower than UA for most models. SVM-w-cv achieved the most accurate gully areal coverage of 57.2 ha, while the rf-w-b model had the highest standard error (11.5 ha). The F1-score ranking showed that RF

algorithms achieved better results than SVM algorithms. Both resampling techniques had the same omission error (85.1%) but slightly different commission errors, with bootstrapping having a 40.8% error of commission compared to a 37.8% error for CV.

Four classifiers (RF, K-NN, MD, and MLC) were evaluated for extracting gullies from a SPOT-7 image in Study Area #3. All classifiers achieved high accuracy values above 0.80, with RF recording the highest OA (0.94) and kappa (0.89), followed closely by K-NN with OA and kappa values of 0.92 and 0.86, respectively. Results showed significant variations in UA and PA across different classifiers. K-NN, MD, and RF performed excellently in identifying the BS class, each achieving 1.00 UA. MD was found to be the most effective classifier with 100% UA for gully classification, although the corresponding PA was only 0.80, while RF achieved the highest OA (0.94) but had an 87% UA for gully classification.

- 3) *Examine how different gully morphological characteristics affect the precise mapping of gullies using high-resolution satellite data.*

Gully's characteristics strongly influenced their classification in Study Area #3. Algorithms were most efficient in areas with dendritic pattern gullies but less effective in detecting shallow and wide gullies, particularly on bare soil. Gully length was not a significant factor in detection efficiency. There was a discrepancy between the actual digitized gullies and those derived by the algorithms in all sites, possibly due to resolution differences or the classifiers' limitations in detecting gullies with shallow walls. The gully density varied between 0.12 m/m<sup>2</sup> and 0.61 m/m<sup>2</sup>, influenced by the appearance and pattern of the gullies where gullies of dendritic

form, linear and V-shaped gullies, had the highest gully density. The slope was not the primary factor influencing gully density, as gullies commonly occurred in gently sloping areas.

- 4) *Select geo-environmental variables with the greatest predictive power to model gully susceptibility.*

The study found that the most influential geo-environmental variables for predicting gully susceptibility were NDVI, elevation, TWI, population density, SPI, and LULC. NDVI was the most significant factor due to the spectral differences between gullies and vegetation. Areas with little or no vegetation cover were the most susceptible to gully erosion, and low-lying areas with poor vegetation were particularly vulnerable due to higher surface runoff and erosive power. The study also found that areas with low elevation, low TWI values indicating less water accumulation, high SPI values indicating areas of high surface water flow, and specific land use classes such as grasslands and cultivated lands were more susceptible to gully erosion. Lower population densities were also associated with higher susceptibility to gully erosion, consistent with the rural nature of the catchment area with sparse human settlements and extensive grasslands.

- 5) *Analyze algorithms' performance when using input feature sets of varying sizes.*

This study predicted gully susceptibility and compared six algorithms based on accuracy and processing time using small, medium, and large feature sets. Results demonstrated that the algorithm performance varied with feature set size. SVM was the most efficient with medium and small sets, while SGB closely followed

SVM, it required a large feature set to reach maximum accuracy. RF was computationally expensive but performed well with large sets. PLS and RDA were the fastest and achieved acceptable accuracies with a medium set. ANN was sensitive to input features but performed well with a small set.

Overall, this dissertation draws the following conclusions and recommendations. Low-cost, high-resolution sensors such as SPOT-7 and PlanetScope, which were obtained at no cost for the test areas, can map gullies in semi-arid regions despite their limited spectral information. The successful mapping of gullies depends on various factors such as the algorithm, season of image acquisition, gully characteristics, resampling technique, study site, and class number approach, all of which were investigated. The multiclass approach performed exceptionally well at the class level and is strongly suggested for accurate gully identification on SPOT-7 images. RF with bootstrapping resampling technique is recommended for mapping gullies in the dry season using the PlanetScope image, while SVM with k-fold CV is recommended for gully mapping in the wet season.

Concerning gully susceptibility modeling, NDVI had considerable predictive power, followed by elevation, TWI, population density, SPI, and LULC. Therefore, it is recommended that these geo-environmental variables, available at no cost, be involved in gully susceptibility prediction in semi-arid environments with similar environmental conditions as the study area. Different algorithms performed differently across feature sets of varying sizes, and a fixed set of features would not give an accurate assessment. Therefore, ensemble-based algorithms like RF and SGB need larger feature sets for better performance. Although this can be computationally expensive, adding high-predictive features can enhance accuracy. Conversely, SVM and ANN

perform well with fewer input features and less computational time, so using these algorithms, especially SVM, can improve accuracy and processing speed.

## **7. ACKNOWLEDGEMENT**

I am immensely grateful to all those who have played a role in the success of this research. In particular, I would like to extend my deepest gratitude to the individuals and institutions listed below:

I sincerely thank the Tempus Public Foundation, through the Stipendium Hungaricum Scholarship Programme, for providing the funding necessary for this research. I am also indebted to the Department of Higher Education and Training (DHET) of South Africa for the additional support, including the research grant received.

My supervisor, Prof. Szilárd Szabó, deserves special recognition for his timely advice, nurturing, and diligent guidance, which helped me complete this research. Besides, the collaboration we shared was immensely gratifying. You have fostered a spirit of self-reliance, essential to becoming an independent researcher, for which I will forever remain appreciative.

I wish to express my gratitude to Dr. Mester Tamás and the Soil Lab staff at UniDeb, including Mrs. Krisztina Sósne Mező and Ms. Csilla Tóth, for their invaluable expertise and support in facilitating the processing and measurement of soil samples.

I sincerely thank Dr. Loránd Szabó for his valuable assistance in acquiring PlanetScope images, without which the specific objectives of this dissertation would have been achieved. Similarly, I am very grateful to the South African National Space Agency (SANSA) for providing SPOT-7 images without any associated costs.

I also want to express my appreciation to my siblings and friends for their encouragement and support throughout my studies. My special appreciation

goes to my brothers, Gedli, Gwala, and Shinga, for their generous assistance in data collection.

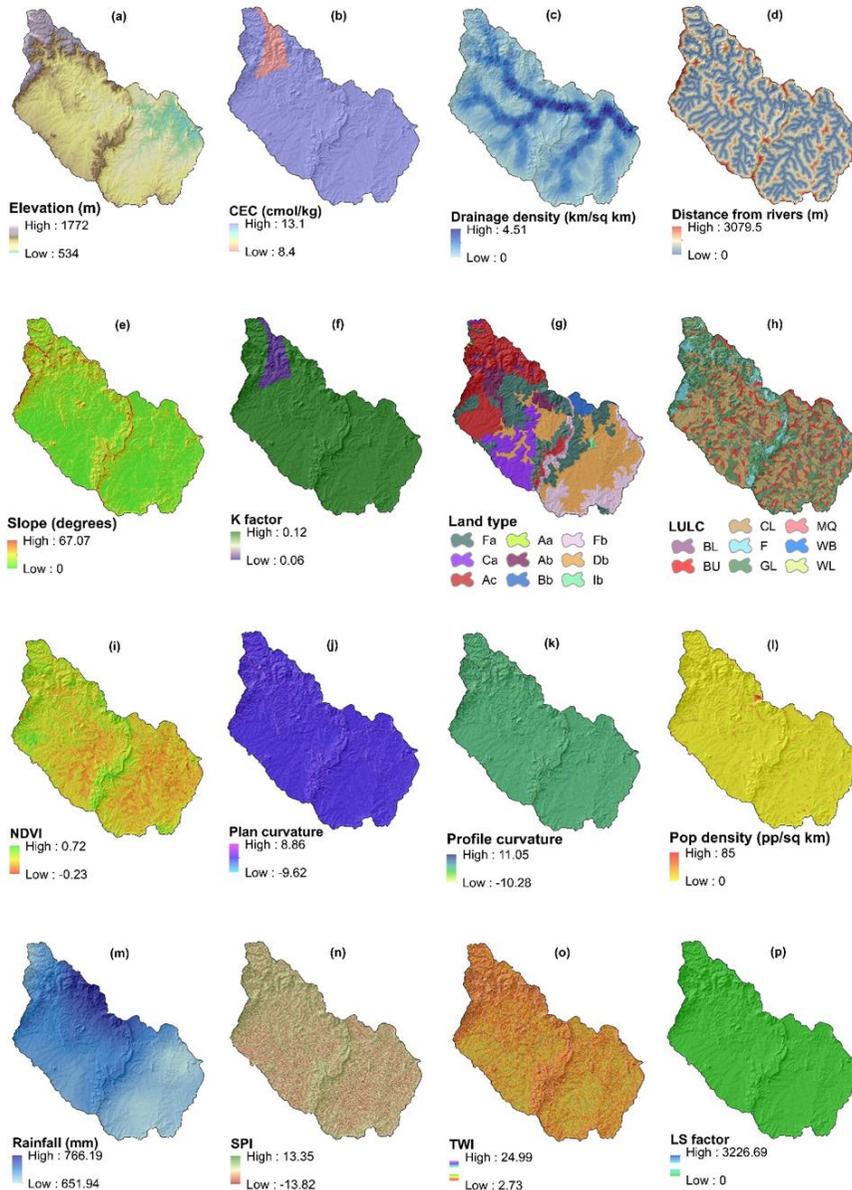
I am profoundly grateful to my dearest wife, Lelo Phinzi, who has always been an inexhaustible source of emotional support. I sincerely appreciate your understanding and consistent support during my prolonged absence in South Africa. I can confidently declare that my extended period of absence from home has now finally come to an end!

I am deeply indebted to my late mother, Thandiwe Phinzi, for everything she provided, not least her firm belief in me, her boundless affection and unwavering support, and the numerous sacrifices she made to ensure my education, which ultimately led me to the culmination of earning this Ph.D. Indeed, I owe you beyond what I could ever compensate.

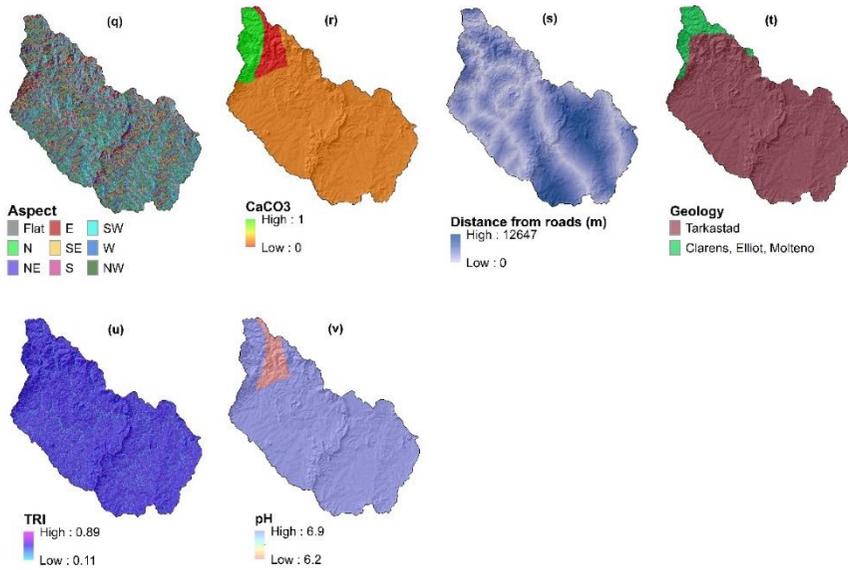
Above all, glory be to the Almighty God, the Creator, *Thixo wokhokho bethu*, for all the countless extraordinary blessings bestowed upon me, for which words alone are inadequate to quantify.

## 8. APPENDICES

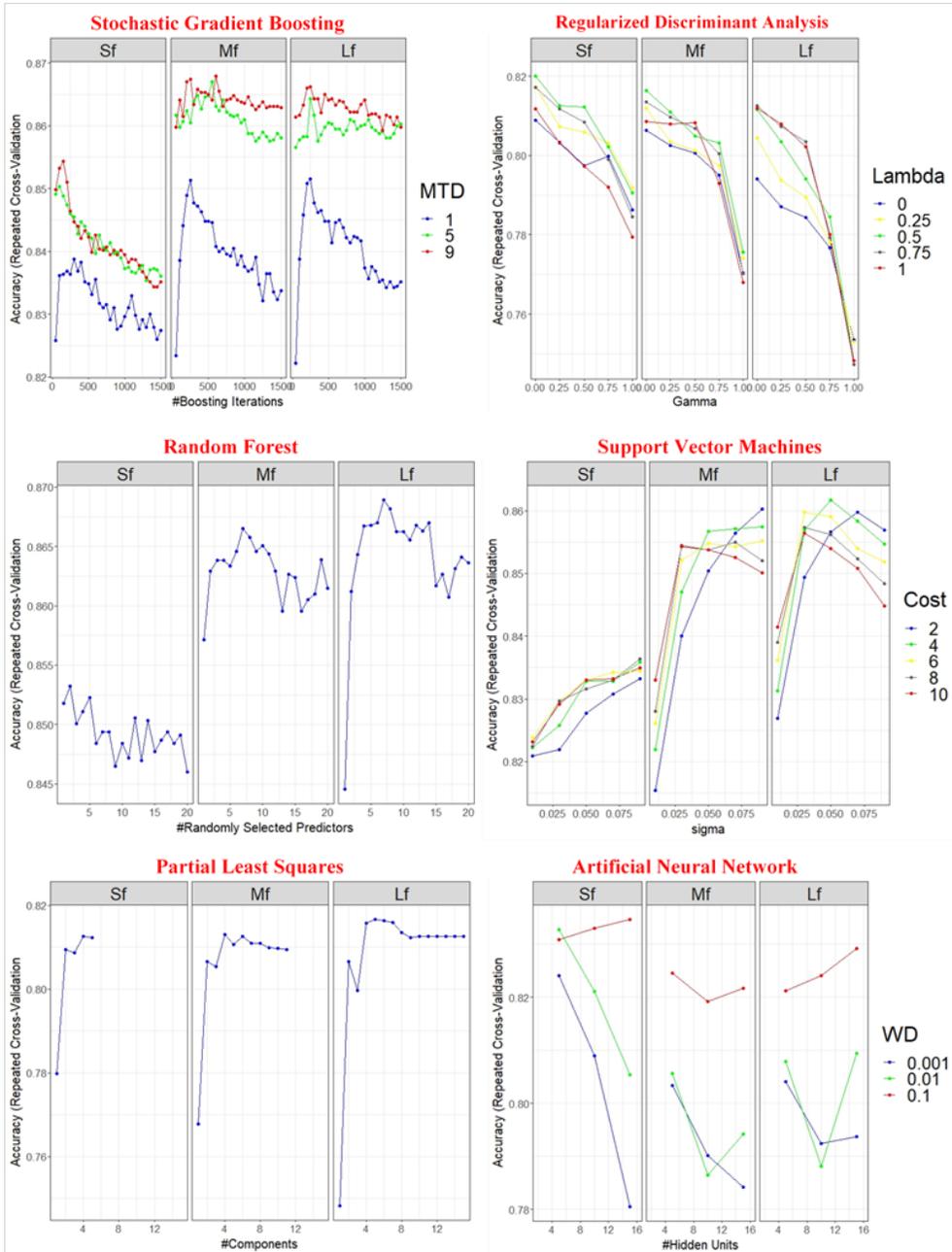
Appendix 1. Spatial distribution of geo-environmental covariates (TWI=topographic wetness index, LS=slope length and steepness, TRI=terrain ruggedness index, NDVI=normalized difference vegetation index, LULC=land use/land cover, CEC=cation exchange capacity, CaCO<sub>3</sub>=calcium carbonate, SPI=stream power index).



Appendix 1. Continued.



Appendix 2. Hyperparameter tuning and the optimal combination of values for the final models of six algorithms across three feature sets of different sizes (Sf: small feature, Mf: medium feature, and Lf: large feature sets).



## 9. REFERENCES

- Abdi, A. M. (2020) 'Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data', *GIScience and Remote Sensing*, 57(1), pp. 1–20. doi: 10.1080/15481603.2019.1650447.
- Abdi, H. (2003) 'Partial least square regression (PLS regression)', *Encyclopedia for research methods for the social sciences*, 6(4), pp. 792–795.
- Abriha, D. *et al.* (2018) 'Identification of roofing materials with discriminant function analysis and random forest classifiers on pan-sharpened worldview-2 imagery – A comparison', *Hungarian Geographical Bulletin*, 67(4), pp. 375–392. doi: 10.15201/hungeobull.67.4.6.
- Acocks, J. P. H. (1988) *Veld types of South Africa*.
- Alijani, G., Hasanlou, M. and Azizi, Z. (2018) 'Classifying UAVSAR Polarimetric Synthetic Aperture Radar (PolSAR) Imagery Using Target Decomposition Features', in *Proceedings*. MDPI, p. 333.
- Alkarkhi, A. F. M. and Alqaraghuli, W. A. A. (2018) 'Discriminant analysis and classification', in Alkarkhi, A. F. M. and Alqaraghuli, W. A. A. (eds) *Easy statistics for food science with R*. London: Academic Press, p. 213.
- Alloghani, M. *et al.* (2020) 'A systematic review on supervised and unsupervised machine learning algorithms for data science', *Supervised and unsupervised learning for data science*, pp. 3–21.
- Allwein, E. L., Schapire, R. E. and Singer, Y. (2000) 'Reducing multiclass to binary: A unifying approach for margin classifiers', *Journal of machine learning research*, 1(Dec), pp. 113–141.
- Amare, S. *et al.* (2021) 'Susceptibility to gully erosion: applying random forest (RF) and frequency ratio (FR) approaches to a small catchment in Ethiopia', *Water*, 13(2), p. 216.
- Amiri, M. *et al.* (2019) 'Assessment of the importance of gully erosion effective factors using Boruta algorithm and its spatial modeling and mapping using three machine learning algorithms', *Geoderma*, 340, pp. 55–69.
- Anadón, J. D. *et al.* (2014) 'Effect of woody-plant encroachment on livestock production in North and South America', *Proceedings of the National Academy of Sciences*, 111(35), pp. 12948–12953.

Anbalagan, R. *et al.* (2015) 'Landslide hazard zonation mapping using frequency ratio and fuzzy logic approach, a case study of Lachung Valley, Sikkim', *Geoenvironmental Disasters*, 2(1), pp. 1–17.

Arabameri, A., Chen, W., *et al.* (2019) 'Comparison of machine learning models for gully erosion susceptibility mapping', *Geoscience Frontiers*, (August). doi: 10.1016/j.gsf.2019.11.009.

Arabameri, A., Cerda, A., *et al.* (2019) 'Proposing a novel predictive technique for gully erosion susceptibility mapping in arid and semi-arid regions (Iran)', *Remote Sensing*, 11(21), p. 2577.

Ashiagbor, G. *et al.* (2013) 'Modeling soil erosion using RUSLE and GIS tools', *Int J Remote Sens Geosci*, 2(4), pp. 1–17.

Azedou, A. *et al.* (2021) 'A methodological comparison of three models for gully erosion susceptibility mapping in the rural municipality of El Faïd (Morocco)', *Sustainability*, 13(2), p. 682.

Bailey, R. G. (1998) 'Ecoregions', *New York*.

Ball, J. E., Anderson, D. T. and Chan, C. S. (2017) 'Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community', *Journal of Applied Remote Sensing*, 11(04), p. 1. doi: 10.1117/1.jrs.11.042609.

Barakat, A. *et al.* (2022) 'Mapping of water-induced soil erosion using machine learning models: a case study of Oum Er Rbia Basin (Morocco)', *Earth Systems and Environment*, pp. 1–20.

Beckedahl, H. R. and de Villiers, A. B. (2000) 'Accelerated erosion by piping in the Eastern Cape Province, South Africa', *South African Geographical Journal*, 82(3), pp. 157–162. doi: 10.1080/03736245.2000.9713709.

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

Bennett, S. J. and Wells, R. R. (2019) 'Gully erosion processes, disciplinary fragmentation, and technological innovation', *Earth surface processes and landforms*, 44(1), pp. 46–53.

Bernini, A. *et al.* (2021) 'Evaluation of Gully Erosion Susceptibility Using a Maximum Entropy Model in the Upper Mkhomazi River Basin in South Africa', *ISPRS International Journal of Geo-Information*, 10(11), p. 729.

Beygelzimer, A., Langford, J. and Zadrozny, B. (2004) ‘Weighted One-Against-All’, *American Association for Artificial Intelligence*, pp. 720–725.

Boardman, J. *et al.* (2003) ‘Development of badlands and gullies in the Sneeuberg, Great Karoo, South Africa’, *Catena*, 50(2–4), pp. 165–184.

Boehmke, B. and Greenwell, B. M. (2019) *Hands-on machine learning with R*. CRC Press.

Bolstad, P. V. and Lillesand, T. M. (1991) ‘Rapid maximum likelihood classification’, *Photogrammetric Engineering & Remote Sensing*, 57(1), pp. 67–74.

van Breda Weaver, A. (1991) ‘The distribution of soil erosion as a function of slope aspect and parent material in Ciskei, Southern Africa’, *GeoJournal*, 23(1), pp. 29–34.

Breiman, L. (2001) ‘Random forests’, *Machine learning*, 45(1), pp. 5–32.

Brownlee, J. (2014) *Machine learning mastery*, URL: <http://machinelearningmastery.com/discover-feature-engineering-howtoengineer-features-and-how-to-getgood-at-it>.

Campbell, J. B. and Wynne, R. H. (2011) *Introduction to remote sensing*. Guilford Press.

Camps-Valls, G. and Bruzzone, L. (2009) *Kernel methods for remote sensing data analysis*. John Wiley & Sons.

Castillo, C. and Gómez, J. A. (2016) ‘A century of gully erosion research: Urgency, complexity and study approaches’, *Earth-Science Reviews*, 160, pp. 300–319.

Chen, G. *et al.* (2018) ‘Symmetrical dense-shortcut deep fully convolutional networks for semantic segmentation of very-high-resolution remote sensing images’, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(5), pp. 1633–1644.

Chen, Y. *et al.* (2014) ‘Deep learning-based classification of hyperspectral data’, *IEEE Journal of Selected topics in applied earth observations and remote sensing*, 7(6), pp. 2094–2107.

Chen, Y. and Hoo, K. A. (2011) ‘Application of partial least square regression in uncertainty study area’, in *Proceedings of the 2011 American Control Conference*. IEEE, pp. 1958–1962.

Chicco, D. and Jurman, G. (2020) ‘The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation’, *BMC genomics*, 21(1), pp. 1–13.

Chowdhuri, I. *et al.* (2021) ‘Evaluation of different DEMs for gully erosion susceptibility mapping using in-situ field measurement and validation’, *Ecological Informatics*, 65, p. 101425.

Chung, D. and Keles, S. (2010) ‘Sparse partial least squares classification for high dimensional data’, *Statistical applications in genetics and molecular biology*, 9(1).

Congalton, R. G. (1991) ‘A review of assessing the accuracy of classifications of remotely sensed data’, *Remote Sensing of Environment*, 37(1), pp. 35–46. doi: 10.1016/0034-4257(91)90048-B.

Congalton, R. G. and Green, K. (2019) *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press.

Congalton, R. G., Oderwald, R. G. and Mead, R. A. (1983) ‘Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques’, *Photogrammetric engineering and remote sensing*, 49(12), pp. 1671–1678.

Conoscenti, C. *et al.* (2014) ‘Gully erosion susceptibility assessment by means of GIS-based logistic regression: A case of Sicily (Italy)’, *Geomorphology*, 204, pp. 399–411. doi: 10.1016/j.geomorph.2013.08.021.

Conoscenti, C. and Rotigliano, E. (2020) ‘Predicting gully occurrence at watershed scale: Comparing topographic indices and multivariate statistical models’, *Geomorphology*, 359, p. 107123.

Cruse, R. M. *et al.* (2000) ‘Water drop impact angle and soybean protein amendment effects on soil detachment’, *Soil Science Society of America Journal*, 64(4), pp. 1474–1478.

Csatáriné Szabó, Z. *et al.* (2020) ‘Uncertainty and overfitting in fluvial landform classification using laser scanned data and machine learning: A comparison of pixel and object-based approaches’, *Remote Sensing*, 12(21), p. 3652.

d’Oleire-Oltmanns, S. *et al.* (2014) ‘Detection of gully-affected areas by applying object-based image analysis (OBIA) in the region of Taroudannt, Morocco’, *Remote Sensing*, 6(9), pp. 8287–8309. doi: 10.3390/rs6098287.

D'Oleire-Oltmanns, S. *et al.* (2012) 'Unmanned aerial vehicle (UAV) for monitoring soil erosion in Morocco', *Remote Sensing*, 4(11), pp. 3390–3416. doi: 10.3390/rs4113390.

Delgado, R. and Tibau, X.-A. (2019) 'Why Cohen's Kappa should be avoided as performance measure in classification', *PloS one*, 14(9), p. e0222916.

Department of Agriculture, D. of (2007) *Strategic Plan Strategic for the Department of Agriculture*. doi: 10.1017/CBO9781107415324.004.

Dewitte, O. *et al.* (2015) 'Predicting the susceptibility to gully initiation in data-poor regions', *Geomorphology*, 228, pp. 101–115.

Dube, T. *et al.* (2017) 'Use of Landsat series data to analyse the spatial and temporal variations of land degradation in a dispersive soil environment: A case of King Sabata Dalindyebo local municipality in the Eastern Cape Province, South Africa', *Physics and Chemistry of the Earth*, 100, pp. 112–120. doi: 10.1016/j.pce.2017.01.023.

DWA (2010) *Mbashe River trends report (2007-2010) Department of Water Affairs river health programme Eastern Cape*.

Ebhuoma, O. *et al.* (2022) 'Soil Erosion Vulnerability Mapping in Selected Rural Communities of uThukela Catchment, South Africa, Using the Analytic Hierarchy Process', *Earth Systems and Environment*, pp. 1–14.

Efron, B. (1983) 'Estimating the error rate of a prediction rule: improvement on cross-validation', *Journal of the American statistical association*, 78(382), pp. 316–331.

ESRI (2022) 'ArcGIS Desktop (Version 10.4)'. ESRI Inc.

Everitt, J. H. *et al.* (2008) 'Comparison of QuickBird and SPOT 5 satellite imagery for mapping giant reed', *Journal of Aquatic Plant Management*, 46(1), pp. 77–82.

FAO (2003) *The digital soil map of the world, Land and Water Development Division, FAO*. Rome.

Faulkner, H. *et al.* (2004) 'Variations in soil dispersivity across a gully head displaying shallow sub-surface pipes, and the role of shallow pipes in rill initiation', *Earth Surface Processes and Landforms: the Journal of the British Geomorphological Research Group*, 29(9), pp. 1143–1160.

Faulkner, H. (2006) 'Piping hazard on collapsible and dispersive soils in

Europe’, *Soil erosion in Europe*, pp. 537–562.

Field, A. (2013) *Discovering statistics using IBM SPSS statistics*. sage.

Flight, L. and Julious, S. A. (2015) ‘The disagreeable behaviour of the kappa statistic’, *Pharmaceutical statistics*, 14(1), pp. 74–78.

Foody, G. M. (2020) ‘Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification’, *Remote Sensing of Environment*, 239, p. 111630.

Foster, G. R. (2005) ‘Modeling ephemeral gully erosion for conservation planning.’, *International Journal of Sediment Research*, 20(3), pp. 157–175.

Friedman, J. H. (1989) ‘Regularized discriminant analysis’, *Journal of the American statistical association*, 84(405), pp. 165–175.

Friedman, J. H. (2002) ‘Stochastic gradient boosting’, *Computational statistics & data analysis*, 38(4), pp. 367–378.

Funk, C. *et al.* (2015) ‘The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes’, *Scientific data*, 2(1), pp. 1–21.

Gafurov, A. M. and Yermolayev, O. P. (2020) ‘Automatic gully detection: Neural networks and computer vision’, *Remote Sensing*, 12(11), p. 1743.

Gallucci, M. (2019) ‘GMLj: General analysis for linear models’, *GAMLj: General analyses for linear models*.

Gareth, J. *et al.* (2013) *An introduction to statistical learning: with applications in R*. Springer.

Garland, G. G., Hoffman, M. T. and Todd, S. (2000) ‘Soil degradation’, *A National Review of Land Degradation in South Africa*. South African National Biodiversity Institute, Pretoria, South Africa, pp. 69–107.

Garosi, Y. *et al.* (2019) ‘Assessing the performance of GIS-based machine learning models with different accuracy measures for determining susceptibility to gully erosion’, *Science of the Total Environment*, 664, pp. 1117–1132.

Garthwaite, P. H. (1994) ‘An interpretation of partial least squares’, *Journal of the American Statistical Association*, 89(425), pp. 122–127.

Gayen, A. *et al.* (2019) ‘Gully erosion susceptibility assessment and

management of hazard-prone areas in India using different machine learning algorithms’, *Science of the total environment*, 668, pp. 124–138.

Gentleman, R. and Carey, V. J. (2008) ‘Unsupervised machine learning’, in *Bioconductor case studies*. Springer, pp. 137–157.

Ghaedi, S. and Shojaian, A. (2020) ‘Spatial and temporal variability of precipitation concentration in Iran’, *Geographica Pannonica*, 24(4).

Gislason, P. O., Benediktsson, J. A. and Sveinsson, J. R. (2006) ‘Random forests for land cover classification’, *Pattern recognition letters*, 27(4), pp. 294–300.

Gong, C. *et al.* (2019) ‘Analysis of the development of an erosion gully in an open-pit coal mine dump during a winter freeze-thaw cycle by using low-cost UAVs’, *Remote Sensing*, 11(11), p. 1356.

Grochala, A. and Kedzierski, M. (2017) ‘A method of panchromatic image modification for satellite imagery data fusion’, *Remote Sensing*, 9(6), pp. 11–13. doi: 10.3390/rs9060639.

Han, J., Guzman, J. A. and Chu, M. L. (2022) ‘Gully erosion susceptibility considering spatiotemporal environmental variables: Midwest US region’, *Journal of Hydrology: Regional Studies*, 43, p. 101196.

Hawkins, D. M., Basak, S. C. and Mills, D. (2003) ‘Assessing model fit by cross-validation’, *Journal of chemical information and computer sciences*, 43(2), pp. 579–586.

He, H. and Garcia, E. A. (2009) ‘Learning from imbalanced data’, *IEEE Transactions on knowledge and data engineering*, 21(9), pp. 1263–1284.

Heckel, K. *et al.* (2020) ‘Predicting Forest Cover in Distinct Ecosystems: The Potential of Multi-Source Sentinel-1 and -2 Data Fusion’, *Remote Sensing*, 12(2), p. 302. doi: 10.3390/rs12020302.

Heydari, S. S. and Mountrakis, G. (2018) ‘Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26 Landsat sites’, *Remote Sensing of Environment*, 204(February 2017), pp. 648–658. doi: 10.1016/j.rse.2017.09.035.

Hilbich, C. *et al.* (2007) ‘A landscape-based model to characterize the evolution and recent dynamics of wetlands in the Umzimvubu headwaters, Eastern Cape, South Africa’, *Wetlands: Monitoring, Modelling and*

*Management*, p. 61.

Hitouri, S. *et al.* (2022) ‘Hybrid Machine Learning Approach for Gully Erosion Mapping Susceptibility at a Watershed Scale’, *ISPRS International Journal of Geo-Information*, 11(7), p. 401.

Hoffman, T. and Ashwell, A. (2001) *Nature divided: land degradation in South Africa*. University of Cape Town Press.

Hosseinalizadeh, M. *et al.* (2019a) ‘Gully headcut susceptibility modeling using functional trees, naïve Bayes tree, and random forest models’, *Geoderma*, 342, pp. 1–11.

Hosseinalizadeh, M. *et al.* (2019b) ‘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, pp. 184–193.

Huang, D. *et al.* (2022) ‘Assessment of gully erosion susceptibility using different DEM-derived topographic factors in the black soil region of Northeast China’, *International Soil and Water Conservation Research*.

Irizarry, R. A. (2019) *Introduction to data science: Data analysis and prediction algorithms with R*. CRC Press.

ISRIC (2002) *Luvvisols (lv)*.

Jaafari, A. *et al.* (2022) ‘Understanding land degradation induced by gully erosion from the perspective of different geoenvironmental factors’, *Journal of Environmental Management*, 315(April), p. 115181. doi: 10.1016/j.jenvman.2022.115181.

Jiang, C. *et al.* (2021) ‘Spatial modeling of gully head erosion on the Loess Plateau using a certainty factor and random forest model’, *Science of the Total Environment*, 783, p. 147040.

Kabir, M. R. (2021) ‘Adopting Andersen’s behavior model to identify factors influencing maternal healthcare service utilization in Bangladesh’, *PloS one*, 16(11), p. e0260502.

Kakembo, V. and Rowntree, K. M. (2003) ‘The relationship between land use and soil erosion in the communal lands near Peddie town, Eastern Cape, South Africa’, *Land Degradation and Development*, 14(1), pp. 39–49. doi: 10.1002/ldr.509.

Kakembo, V., Xanga, W. W. and Rowntree, K. (2009) ‘Topographic

thresholds in gully development on the hillslopes of communal areas in Ngqushwa Local Municipality, Eastern Cape, South Africa’, *Geomorphology*, 110(3–4), pp. 188–194. doi: 10.1016/j.geomorph.2009.04.006.

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

Kim, J.-H. (2009) ‘Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap’, *Computational statistics & data analysis*, 53(11), pp. 3735–3745.

Kirkby, M. J. and Bracken, L. J. (2009) ‘Gully processes and gully dynamics’, *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group*, 34(14), pp. 1841–1851.

Kohavi, R. (1995) ‘A study of cross-validation and bootstrap for accuracy estimation and model selection’, in *Ijcai*. Montreal, Canada, pp. 1137–1145.

Kuhn, M. (2008) ‘Building predictive models in R using the caret package’, *Journal of statistical software*, 28, pp. 1–26.

Kuhn, M. and Johnson, K. (2013) *Applied predictive modeling*. Springer.

Kulimushi, L. C. *et al.* (2023) ‘Soil erosion susceptibility mapping using ensemble machine learning models: A case study of upper Congo river sub-basin’, *CATENA*, 222, p. 106858.

Kussul, N. *et al.* (2017) ‘Deep learning classification of land cover and crop types using remote sensing data’, *IEEE Geoscience and Remote Sensing Letters*, 14(5), pp. 778–782.

Lana, J. C., Castro, P. de T. A. and Lana, C. E. (2022) ‘Assessing gully erosion susceptibility and its conditioning factors in southeastern Brazil using machine learning algorithms and bivariate statistical methods: A regional approach’, *Geomorphology*, 402, p. 108159.

Land Type Survey Staff (no date) *Land Type Survey Database*. Pretoria: ARC–Institute for Soil, Climate and Water. Pretoria.

Lannoeye, W. *et al.* (2016) ‘The use of SfM-photogrammetry to quantify and understand gully degradation at the temporal scale of rainfall events: an example from the Ethiopian drylands’, *Physical Geography*, 37(6), pp. 430–451.

Lantz, B. (2015) *Machine Learning with R: Expert techniques for predictive modeling to solve all your data analysis problems*. Packt Publishing Ltd.

Lee, S. and Lee, D. K. (2018) ‘What is the proper way to apply the multiple comparison test?’, *Korean journal of anesthesiology*, 71(5), p. 353.

Lee, S. and Pradhan, B. (2007) ‘Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models’, *Landslides*, 4(1), pp. 33–41.

Lee, S. and Sambath, T. (2006) ‘Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models’, *Environmental Geology*, 50(6), pp. 847–855.

Levine, T. R. and Hullett, C. R. (2002) ‘Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research’, *Human Communication Research*, 28(4), pp. 612–625. doi: 10.1111/j.1468-2958.2002.tb00828.x.

Li, J., Xiong, L. and Tang, G. (2019) ‘Combined gully profiles for expressing surface morphology and evolution of gully landforms’, *Frontiers of Earth Science*, 13(3), pp. 551–562. doi: 10.1007/s11707-019-0752-1.

Likó, S. B. *et al.* (2022) ‘Tree species composition mapping with dimension reduction and post-classification using very high-resolution hyperspectral imaging’, *Scientific Reports*, 12(1), p. 20919.

Lillesand, T., Kiefer, R. W. and Chipman, J. (2015) *Remote sensing and image interpretation*. John Wiley & Sons.

Liu, B. *et al.* (2022) ‘Ephemeral gully recognition and accuracy evaluation using deep learning in the hilly and gully region of the Loess Plateau in China’, *International Soil and Water Conservation Research*, 10(3), pp. 371–381.

Liu, G. *et al.* (2021) ‘Three decades of ephemeral gully erosion studies’, *Soil and Tillage Research*, 212, p. 105046.

Lu, D. and Weng, Q. (2007) ‘A survey of image classification methods and techniques for improving classification performance’, *International journal of Remote sensing*, 28(5), pp. 823–870.

Luffman, I. E., Nandi, A. and Spiegel, T. (2015) ‘Gully morphology, hillslope erosion, and precipitation characteristics in the Appalachian Valley and Ridge province, southeastern USA’, *Catena*, 133, pp. 221–232.

- Ma, L. *et al.* (2019) ‘Deep learning in remote sensing applications: A meta-analysis and review’, *ISPRS journal of photogrammetry and remote sensing*, 152, pp. 166–177.
- Magliulo, P., Russo, F. and Lo Curzio, S. (2020) ‘Detection of permanently eroded landsurfaces through multitemporal analysis of Landsat data: a case study from an agricultural area in southern Italy’, *Environmental Earth Sciences*, 79(3). doi: 10.1007/s12665-020-8814-y.
- Mair, P. and Wilcox, R. (2019) ‘Robust statistical methods in R using the WRS2 package’, *Behavior Research Methods*. doi: 10.3758/s13428-019-01246-w.
- Makaya, N. P. *et al.* (2019) ‘Assessing the potential of Sentinel-2 MSI sensor in detecting and mapping the spatial distribution of gullies in a communal grazing landscape’, *Physics and Chemistry of the Earth*, 112(May 2018), pp. 66–74. doi: 10.1016/j.pce.2019.02.001.
- Mararakanye, N. and Nethengwe, N. S. (2012) ‘Gully Erosion Mapping Using Remote Sensing Techniques’, *South African Journal of Geomatics*, 1(2), pp. 109–118.
- Mararakanye, N. and Le Roux, J. J. (2012) ‘Gully location mapping at a national scale for South Africa’, *South African Geographical Journal*, 94(2), pp. 208–218. doi: 10.1080/03736245.2012.742786.
- Marzolff, I. and Poesen, J. (2009) ‘The potential of 3D gully monitoring with GIS using high-resolution aerial photography and a digital photogrammetry system’, *Geomorphology*, 111(1–2), pp. 48–60.
- Mason, C. H. and Perreault Jr, W. D. (1991) ‘Collinearity, power, and interpretation of multiple regression analysis’, *Journal of marketing research*, 28(3), pp. 268–280.
- Maurer, T. (2013) ‘HOW TO PAN-SHARPEN IMAGES USING THE GRAM-SCHMIDT PAN-SHARPEN METHOD &ndash; A RECIPE’, *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-1/W1(May), pp. 239–244. doi: 10.5194/isprsarchives-xl-1-w1-239-2013.
- Mhangara, P., Kakembo, V. and Lim, K. J. (2012) ‘Soil erosion risk assessment of the Keiskamma catchment, South Africa using GIS and remote sensing’, *Environmental Earth Sciences*, 65(7), pp. 2087–2102. doi: 10.1007/s12665-011-1190-x.

Mohd Razali, N. and Bee Wah, Y. (2011) 'Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests', *Journal of Statistical Modeling and Analytics*, 2(1), pp. 21–33.

Moisen, G. G. *et al.* (2006) 'Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods', *Ecological modelling*, 199(2), pp. 176–187.

Molinaro, A. M., Simon, R. and Pfeiffer, R. M. (2005) 'Prediction error estimation: a comparison of resampling methods', *Bioinformatics*, 21(15), pp. 3301–3307.

Momm, H. G. *et al.* (2012) 'AGNPS GIS-based tool for watershed-scale identification and mapping of cropland potential ephemeral gullies', *Applied engineering in agriculture*, 28(1), pp. 17–29.

Moore, I. D. and Burch, G. J. (1986) 'Physical basis of the length-slope factor in the universal soil loss equation', *Soil Science Society of America Journal*, 50(5), pp. 1294–1298.

Moore, I. D., Grayson, R. B. and Ladson, A. R. (1991) 'Digital terrain modelling: a review of hydrological, geomorphological, and biological applications', *Hydrological processes*, 5(1), pp. 3–30.

Muñoz-Robles, C. *et al.* (2010) 'Factors related to gully erosion in woody encroachment in south-eastern Australia', *Catena*, 83(2–3), pp. 148–157. doi: 10.1016/j.catena.2010.08.002.

NEPAD (2013) *Agriculture in Africa - Transformation and Outlook*.

Niculiță, M., Mărgărint, M. C. and Tarolli, P. (2020) 'Using UAV and LiDAR data for gully geomorphic changes monitoring', in *Developments in earth surface processes*. Elsevier, pp. 271–315.

Olivier, G., Van De Wiel, M. J. and De Clercq, W. P. (2022) 'Intersecting views of gully erosion in South Africa', *Earth Surface Processes and Landforms*.

Olofsson, P. *et al.* (2014) 'Good practices for estimating area and assessing accuracy of land change', *Remote Sensing of Environment*, 148, pp. 42–57.

Orti, M. V. *et al.* (2020) 'Use of TanDEM-X and Sentinel products to derive gully activity maps in Kunene Region (Namibia) based on automatic iterative Random Forest approach', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Parlak, M. and Parlak, A. O. (2010) ‘Measurement of splash erosion in different cover crops’, *Turkish Journal of Field Crops*, 15(2), pp. 169–173.

Parwada, C. and Van Tol, J. (2017) ‘Soil properties influencing erodibility of soils in the Ntabelanga area, Eastern Cape Province, South Africa’, *Acta Agriculturae Scandinavica, Section B—Soil & Plant Science*, 67(1), pp. 67–76.

Pham, Q. B. *et al.* (2020) ‘Head-cut gully erosion susceptibility modelling based on ensemble Random Forest with oblique decision trees in Fareghan watershed, Iran’, *Geomatics, Natural Hazards and Risk*, 11(1), pp. 2385–2410.

Phinzi, K. (2018) *Spatio-temporal appraisal of water-borne erosion using optical remote sensing and GIS in the Umzintlava catchment (T32E), Eastern Cape, South Africa*. University of KwaZulu-Natal.

Phinzi, K. *et al.* (2020) ‘Machine learning for gully feature extraction based on a pan-sharpened multispectral image: Multiclass vs. Binary approach’, *ISPRS International Journal of Geo-Information*, 9(4). doi: 10.3390/ijgi9040252.

Phinzi, K., Abriha, D. and Szabó, S. (2021) ‘Classification efficacy using k-fold cross-validation and bootstrapping resampling techniques on the example of mapping complex gully systems’, *Remote Sensing*, 13(15), p. 2980.

Phinzi, K., Holb, I. and Szabó, S. (2021) ‘Mapping Permanent Gullies in an Agricultural Area Using Satellite Images: Efficacy of Machine Learning Algorithms’, *Agronomy*, 11(2), p. 333.

Phinzi, K. and Ngetar, N. S. (2017) ‘Mapping soil erosion in a quaternary catchment in Eastern Cape using geographic information system and remote sensing’, *South African Journal of Geomatics*, 6(1), p. 11. doi: 10.4314/sajg.v6i1.2.

Phinzi, K. and Ngetar, Njoya S. (2019) ‘Land use/land cover dynamics and soil erosion in the Umzintlava catchment (T32E), Eastern Cape, South Africa’, *Transactions of the Royal Society of South Africa*, 74(3), pp. 223–237. doi: 10.1080/0035919X.2019.1634652.

Phinzi, K. and Ngetar, Njoya Silas (2019) ‘The assessment of water-borne erosion at catchment level using GIS-based RUSLE and remote sensing: A review’, *International Soil and Water Conservation Research*, 7(1), pp. 27–46. doi: 10.1016/j.iswcr.2018.12.002.

Du Plessis, C. *et al.* (2020) ‘Machine learning digital soil mapping to inform gully erosion mitigation measures in the Eastern Cape, South Africa’, *Geoderma*, 368, p. 114287.

Poesen, J. *et al.* (2002) ‘Gully erosion in dryland environments’, *Bull, LJ, Kirkby, MJ (Eds.), Dryland rivers: hydrology and geomorphology of semi-arid channels*. Wiley, Chichester, UK, pp. 229–262.

Poesen, J. *et al.* (2003) ‘Gully erosion and environmental change: Importance and research needs’, *Catena*, 50(2–4), pp. 91–133. doi: 10.1016/S0341-8162(02)00143-1.

Poesen, J., Vandaele, K. and Wesemael, B. van (1998) ‘Gully erosion: importance and model implications’, in *Modelling soil erosion by water*. Springer, pp. 285–311.

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

Pourghasemi, H. R. *et al.* (2017) ‘Performance assessment of individual and ensemble data-mining techniques for gully erosion modeling’, *Science of the Total Environment*, 609, pp. 764–775. doi: 10.1016/j.scitotenv.2017.07.198.

Pourghasemi, H. R. *et al.* (2020) ‘Gully erosion spatial modelling: Role of machine learning algorithms in selection of the best controlling factors and modelling process’, *Geoscience Frontiers*, 11(6), pp. 2207–2219.

Povey, A. C. and Grainger, R. G. (2015) ‘Known and unknown unknowns: uncertainty estimation in satellite remote sensing’, *Atmospheric Measurement Techniques*, 8(11), pp. 4699–4718.

R Core Team (2020) ‘The R Project for Statistical Computing’, [Http://Www.R-Project.Org/](http://www.R-project.org/), pp. 1–12.

R Core Team (2021) ‘R: A language and environment for statistical computing’, *R Foundation for statistical computing, Vienna*.

Rahmati, O. *et al.* (2017) ‘Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion’, *Geomorphology*, 298, pp. 118–137.

Rahmati, O. *et al.* (2022) ‘Contribution of physical and anthropogenic factors to gully erosion initiation’, *Catena*, 210, p. 105925.

Rahmati, O., Pourghasemi, H. R. and Zeinivand, H. (2016) 'Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran', *Geocarto International*, 31(1), pp. 42–70.

Rensburg, L. D. van (2010) 'Advances in soil physics: Application in irrigation and dryland crop production', *South African Journal of Plant and Soil*, 27(1), pp. 9–18.

Reynolds, J. F. *et al.* (2007) 'Global desertification: building a science for dryland development', *science*, 316(5826), pp. 847–851.

Richards, John A and Richards, J A (1999) *Remote sensing digital image analysis*. Springer.

Ridgeway, G. (2007) 'Generalized Boosted Models: A guide to the gbm package', *Update*, p. 2007.

Roberts, M. E. *et al.* (2022) 'Modelling classical gullies—A review', *Geomorphology*, p. 108216.

Rodriguez-Galiano, V. F. *et al.* (2012) 'An assessment of the effectiveness of a random forest classifier for land-cover classification', *ISPRS Journal of Photogrammetry and Remote Sensing*, 67(1), pp. 93–104. doi: 10.1016/j.isprsjprs.2011.11.002.

Le Roux, J. J. *et al.* (2008) 'Water erosion prediction at a national scale for South Africa', *Water Sa*, 34(3), pp. 305–314.

Le Roux, J. J. and Sumner, P. D. (2012) 'Factors controlling gully development: Comparing continuous and discontinuous gullies', *Land Degradation and Development*, 23(5), pp. 440–449. doi: 10.1002/ldr.1083.

Roy, J. and Saha, S. (2019) 'GIS-based gully erosion susceptibility evaluation using frequency ratio, cosine amplitude and logistic regression ensembled with fuzzy logic in Hinglo River Basin, India', *Remote Sensing Applications: Society and Environment*, 15, p. 100247.

Roy, J. and Saha, S. (2021) 'Integration of artificial intelligence with meta classifiers for the gully erosion susceptibility assessment in Hinglo river basin, Eastern India', *Advances in Space Research*, 67(1), pp. 316–333.

Roy, J. and Saha, S. (2022) 'Ensemble hybrid machine learning methods for gully erosion susceptibility mapping: K-fold cross validation approach', *Artificial Intelligence in Geosciences*, 3, pp. 28–45.

Safriel, U. *et al.* (2006) 'Ecosystems and human well-being: current state and trends, vol 1, p 623–662'. Island Press, Washington, DC.

Saha, A. *et al.* (2022) 'Land degradation risk dynamics assessment in red and lateritic zones of eastern plateau, India: a combine approach of K-fold CV, data mining and field validation', *Ecological Informatics*, 69, p. 101653.

Sahour, H. *et al.* (2021) 'Machine learning applications for water-induced soil erosion modeling and mapping', *Soil and Tillage Research*, 211, p. 105032.

Sepuru, T. K. and Dube, T. (2018) 'Understanding the spatial distribution of eroded areas in the former rural homelands of South Africa: Comparative evidence from two new non-commercial multispectral sensors', *International Journal of Applied Earth Observation and Geoinformation*, 69(February), pp. 119–132. doi: 10.1016/j.jag.2018.02.020.

Shafapour Tehrany, M. *et al.* (2019) 'Evaluating the application of the statistical index method in flood susceptibility mapping and its comparison with frequency ratio and logistic regression methods', *Geomatics, Natural Hazards and Risk*, 10(1), pp. 79–101.

Shrestha, A. and Mahmood, A. (2019) 'Review of deep learning algorithms and architectures', *IEEE access*, 7, pp. 53040–53065.

Shruthi, R. B. V *et al.* (2015) 'Quantifying temporal changes in gully erosion areas with object oriented analysis', *Catena*, 128, pp. 262–277.

Sörensen, R., Zinko, U. and Seibert, J. (2006) 'On the calculation of the topographic wetness index: evaluation of different methods based on field observations', *Hydrology and Earth System Sciences*, 10(1), pp. 101–112.

Svoray, T. *et al.* (2012) 'Predicting gully initiation: comparing data mining techniques, analytical hierarchy processes and the topographic threshold', *Earth Surface Processes and Landforms*, 37(6), pp. 607–619.

Szabó, S. *et al.* (2016) 'Possibilities of land use change analysis in a mountainous rural area: a methodological approach', *International Journal of Geographical Information Science*, 30(4), pp. 708–726. doi: 10.1080/13658816.2015.1092546.

Tallarida, R. J. and Murray, R. B. (1987) 'Dunnett's Test (Comparison with a Control)', in *Manual of Pharmacologic Calculations*. Springer, pp. 145–148.

Tang, L. *et al.* (2014) 'A new method combining LDA and PLS for dimension reduction', *PloS one*, 9(5), p. e96944.

- Thanh Noi, P. and Kappas, M. (2017) ‘Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery’, *Sensors (Basel, Switzerland)*, 18(1). doi: 10.3390/s18010018.
- Unay, D. and Gosselin, B. (2007) ‘Stem and calyx recognition on ‘Jonagold’ apples by pattern recognition’, *Journal of food Engineering*, 78(2), pp. 597–605.
- Usama, M. *et al.* (2019) ‘Unsupervised machine learning for networking: Techniques, applications and research challenges’, *IEEE access*, 7, pp. 65579–65615.
- Utsumi, A. G. *et al.* (2020) ‘Gully mapping using geographic object-based image analysis: A case study at catchment scale in the Brazilian Cerrado’, *Remote Sensing Applications: Society and Environment*, 20, p. 100399.
- Vaidyanathan, N. S. *et al.* (2002) ‘Mapping of erosion intensity in the Garhwal Himalaya’, *International Journal of Remote Sensing*, 23(20), pp. 4125–4129.
- Valentin, C., d’Herbès, J.-M. and Poesen, J. (1999) ‘Soil and water components of banded vegetation patterns’, *Catena*, 37(1–2), pp. 1–24.
- Valentin, C., Poesen, J. and Li, Y. (2005) ‘Gully erosion: Impacts, factors and control’, *Catena*, 63(2–3), pp. 132–153. doi: 10.1016/j.catena.2005.06.001.
- Valiant, L. G. (1984) ‘A theory of the learnable’, *Communications of the ACM*, 27(11), pp. 1134–1142.
- Vanmaercke, M. *et al.* (2021) ‘Measuring, modelling and managing gully erosion at large scales: A state of the art’, *Earth-Science Reviews*, 218, p. 103637.
- Vapnik, V. (1999) *The nature of statistical learning theory*. Springer science & business media.
- Varga, O. G. *et al.* (2021) ‘Validation of visually interpreted corine land cover classes with spectral values of satellite images and machine learning’, *Remote Sensing*, 13(5), p. 857.
- Vatcheva, K. P. *et al.* (2016) ‘Multicollinearity in regression analyses conducted in epidemiologic studies’, *Epidemiology (Sunnyvale, Calif.)*, 6(2).
- Venables, W. N. and Ripley, B. D. (2013) *Modern applied statistics with S-PLUS*. Springer Science & Business Media.

- Vetrivel, A. *et al.* (2018) ‘Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning’, *ISPRS journal of photogrammetry and remote sensing*, 140, pp. 45–59.
- Vrieling, A. *et al.* (2007) ‘Automatic identification of erosion gullies with ASTER imagery in the Brazilian Cerrados’, *International Journal of Remote Sensing*, 28(12), pp. 2723–2738. doi: 10.1080/01431160600857469.
- Vrieling, A., Sterk, G. and de Jong, S. M. (2010) ‘Satellite-based estimation of rainfall erosivity for Africa’, *Journal of hydrology*, 395(3–4), pp. 235–241.
- Wang, R. *et al.* (2016) ‘Gully erosion mapping and monitoring at multiple scales based on multi-source remote sensing data of the Sancha River Catchment, Northeast China’, *ISPRS International Journal of Geo-Information*, 5(11), p. 200.
- Wang, S.-C. (2003) ‘Artificial neural network’, in *Interdisciplinary computing in java programming*. Springer, pp. 81–100.
- Wehrens, R. and Mevik, B.-H. (2007) ‘The pls package: principal component and partial least squares regression in R’, *Journal of statistical software*, 18(2), pp. 1–24.
- Welch, B. L. (1939) ‘Note on discriminant functions’, *Biometrika*, 31(1/2), pp. 218–220.
- Williams, J. R. (1995) ‘The EPIC model.’, *Computer models of watershed hydrology.*, pp. 909–1000.
- Wilson, G. V (2009) ‘Mechanisms of ephemeral gully erosion caused by constant flow through a continuous soil-pipe’, *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group*, 34(14), pp. 1858–1866.
- Wischmeier, W. H. and Smith, D. D. (1978) *Predicting rainfall erosion losses: a guide to conservation planning*. Department of Agriculture, Science and Education Administration.
- Wold, H. (1966) ‘Estimation of principal components and related models by iterative least squares’, *Multivariate analysis*, pp. 391–420.
- Wu, W. *et al.* (1996) ‘Comparison of regularized discriminant analysis linear discriminant analysis and quadratic discriminant analysis applied to NIR data’, *Analytica Chimica Acta*, 329(3), pp. 257–265.

- Xiong, K. *et al.* (2020) 'Comparison of Different Machine Learning Methods for Debris Flow Susceptibility Mapping: A Case Study in the Sichuan Province, China', *Remote Sensing*, 12(2), p. 295. doi: 10.3390/rs12020295.
- Zhang, L. *et al.* (2016) 'Deep Learning for Remote Sensing Image Understanding', *Journal of Sensors*, 2016(june). doi: 10.1155/2016/7954154.
- Zhang, T. *et al.* (2016) 'Spatial distribution and morphologic characteristics of gullies in the Black Soil Region of Northeast China: Hebei watershed', *Physical Geography*, 37(3–4), pp. 228–250. doi: 10.1080/02723646.2016.1184079.
- Zhang, Y. *et al.* (2007) 'Characteristics and factors controlling the development of ephemeral gullies in cultivated catchments of black soil region, Northeast China', *Soil and Tillage Research*, 96(1–2), pp. 28–41.
- Van Zijl, G. M., Le Roux, P. A. L. and Turner, D. P. (2013) 'Disaggregation of land types using terrain analysis, expert knowledge and GIS methods', *South African Journal of Plant and Soil*, 30(3), pp. 123–129.
- Žížala, D. *et al.* (2019) 'Mapping soil degradation using remote sensing data and ancillary data: South-East Moravia, Czech Republic', *European Journal of Remote Sensing*, 52(sup1), pp. 108–122. doi: 10.1080/22797254.2018.1482524.