



Assessing land suitability for leguminous crops in the okavango river basin: A multicriteria and machine learning approach

Kaleb Gizaw Negussie^{a,b,d,*}, Bisrat Haile Gebrekidan^c, Daniel Wyss^b, Martin Kappas^b

^a Department of Land and Spatial Sciences, Namibia University of Science and Technology, Windhoek 13388, Namibia

^b Department of Cartography, GIS and Remote Sensing, Georg-August University Goettingen, Goettingen 37077, Germany

^c International Maize and Wheat Improvement Center (CIMMYT), Addis Ababa, Ethiopia

^d Earth Observation and Satellite Application Research and Training Centre (EOSA-RTC), Namibia University of Science and Technology, Windhoek 13388, Namibia

ARTICLE INFO

Keywords:

Multi-criteria decision making (MCDM)
Ensemble Learning Models
Leguminous Crops (sunn hemp pigeon pea)
Land suitability
Okavango river basin
Namibia

ABSTRACT

This study aimed to create a model to identify land suitable for growing sunn hemp (*Crotalaria juncea*) and pigeon pea (*Cajanus cajan*) in the Okavango River basin of the Kavango East region of Namibia. Advanced tree-based ensemble learning models, including Random Forest, Extra Trees, Gradient Boosting, XGBoost and multivariate regression analysis, were employed to enhance analytical accuracy. The Random Forest and XGBoost models exhibited outstanding performance, as evidenced by their respective accuracy values of 0.97 and 0.96. In addition, this study proposed an innovative approach through the integration of subjective and objective analytical methods, which are independent of one another. The subjective component of the analysis employed a Multi-Criteria Decision Making-Analytic Hierarchy Process (MCDM-AHP). On the other hand, the objective component used a data-driven multivariate approach supported by tree-based learning algorithms. Twenty-two variables were considered, encompassing climatic conditions, hydro-geomorphologic features, soil characteristics, vegetation patterns, and socio-economic factors. These variables played a crucial role to identify the most suitable areas for growing the selected leguminous crops. The MCDM-AHP method utilised expert evaluations to rank the importance of variables, identifying water sources, slope, and soil properties as key factors. A suitability mapping analysis revealed that 17.63% of the area was highly suitable and 62.77% moderately suitable, while 10% was less suitable and 9.59% unsuitable for growing these two legumes. According to the data driven methodology, soil fertility and nitrogen content emerged as key determinants for land suitability. This is particularly relevant for nitrogen-fixing leguminous crops such as sunn hemp and pigeon pea, which play a central role in improving soil quality and ensuring food security.

1. Introduction

Food security poses a significant challenge in Sub-Saharan Africa. As the population rapidly expands, numerous households are unable to meet their essential nutritional requirements (FAO, 2023). Salgado Baptista et al. (2022) estimate that approximately 12 % of the population in this region is currently experiencing food insecurity. This is attributed to a combination of natural and human-induced factors, such as floods, droughts, rapid population growth, conflict, and the overall impact of climate change. The report by the FAO and UN Water (2021) underscores the growing risk of water scarcity, particularly in areas where agriculture relies on rainfall. This situation has already led to a 50 % reduction in crop production, significantly impacting food

availability (IPCC, 2014). In pursuit of their developmental objectives, nations have incorporated the sustainable management of land and water resources into their strategic frameworks, recognising these as pivotal elements in achieving food security and environmental stability.

Inter-annual climate variations, soil erosion and loss of nutrients have been identified as significant factors related to yield decline in this region (IFAD, 2021; Sharma et al., 2011). One of the challenges faced in many parts of southern Africa is the decrease in soil fertility in semi-arid regions (Ncube, 2007). Proper soil and water management could mitigate the effects of climate change on crop productivity. Predictions indicate that by mid-century, several countries including Angola, Malawi, Mozambique, Lesotho, and Zimbabwe will witness a substantial decrease in arable land, rendering it unsuitable for growing staple crops

* Corresponding author at: Department of Land and Spatial Sciences, Namibia University of Science and Technology, Windhoek 13388, Namibia.

E-mail address: kaleb.negussie@stud.uni-goettingen.de (K.G. Negussie).

<https://doi.org/10.1016/j.jag.2024.104284>

Received 30 April 2024; Received in revised form 1 November 2024; Accepted 20 November 2024

Available online 28 November 2024

1569-8432/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

such as beans, cassava, maize, millet, pigeon pea, sorghum, and wheat, as well as cash crops (IFAD, 2021).

The state of food insecurity in Namibia mirrors the conditions found across sub-Saharan Africa. According to the latest [Integrated Food Security Phase Classification \(2023\)](#) twelve out of the fourteen regions are experiencing severe food insecurity, with 22 % of the total population in dire need of food assistance. The main cause of the crisis is the occurrence of unpredictable climate shocks, including prolonged periods of drought, fluctuations in annual rainfall, and economic hardship. Kavango East and Kavango West, both located in the Okavango River Basin, were expected to face severe food insecurity from 2023 to 2024 ([Republic of Namibia Office of the Prime Minister, 2023](#)).

The transboundary freshwater ecosystem of the Okavango River basin located in the Southern Angola highlands receive more rain than its downstream counterparts in the semi-arid regions of northeastern Namibia and Northwest Botswana ([Mendelsohn and El Obeid, 2004](#)). The basin has experienced land degradation due to inappropriate land use, and inadequate water conservation strategies, from the prevalence of ad hoc farming practices along the basin, highlighting the need for efficient planning, development, and management of land and water resources in areas of high rainfall variability ([NEPAD Water Centres of Excellence, 2013](#)). Irrigation and flood recession agriculture in the Okavango River basin have enhanced crop yields for small-scale farmers, thereby increasing household food security, especially in the semi-arid regions of Namibia and Botswana ([Mosepele and Hambira, 2019](#)).

For better understanding of how to manage the interplay between supply and demand for water and land resources, it is helpful to conduct research that examines the behaviour of watersheds and land use practices in the region. One approach is the assessment of land suitability. This evaluates the potential of the land to support various uses by considering factors such as soil quality, topography, climate, and water availability ([Sys et al., 1991](#)). By determining the optimal land use for agriculture, agronomists, agricultural extension officers, land-use planners and policymakers can make informed decisions regarding resource allocation, promoting a balance between economic development and environmental sustainability.

Land suitability assessment is a crucial component of land use planning. It provides valuable information to support decision-making to optimise the utilisation of land resources in a specific area. This assessment considers a broad spectrum of characteristics and requirements ([FAO, 1993, 1976](#)). Furthermore, it can assist in identifying the locations of greater importance and quantifying the extent to which a particular piece of land is suitable for specific purposes such as agriculture, urban development or conservation ([Yalew et al., 2016](#)).

In most comparable studies of land suitability, commonly used parameters fall under categories of climate, hydro-geomorphology and soil features ([Bhullar et al., 2023; Sarkar et al., 2023](#)). Depending on the investigation, different combinations of parameters may be employed to determine the suitability of a piece of land for crop farming. Parameters receive a weighted score through a multi-criteria evaluation approach. A multi-criteria decision making (MCDM) approach such as the Analytic Hierarchy Process (AHP) from [Saaty \(1987\)](#) is commonly used to determine land suitability for agriculture. However, several studies have emphasised the lack of objectivity in the MCDM-AHP method, which relies on expert opinions to assess the agricultural land suitability ([Arab and Ahamed, 2022; Pramanik, 2016; Sarkar et al., 2023](#)). A subsequent multivariate model enables researchers to analyse the individual contributions of various factors and their interactions to the observed suitability class, uninfluenced by subjective views. The most frequently used methods for evaluating the suitability of crop land are the Analytic Hierarchy Process (AHP) and fuzzy sets. These are subjective in nature and define members of pixels within suitability classes and categorise pixels that belong to unsuitable classes ([Zadeh, 1965](#)). Using the AHP approach ([Moisa et al., 2022; Tashayo et al., 2020](#)) conducted crop suitability assessments in semi-arid environments, classifying the

suitability for maize production into three categories. However, they identified crop-specific suitability analysis as a limitation. Similarly, studies by [Adem et al. \(2023\)](#) and [Waleed et al. \(2020\)](#), successfully used AHP to identify irrigation potential for crop growth, but noted the subjective nature of relying on expert opinions. Additionally, a research by [Arab and Ahamed \(2022\)](#) and [Özkan et al. \(2020\)](#), used fuzzy expert system to identify suitable areas for agricultural production. While they achieved to establish a representative suitability ranking, uncertainties related to expert opinions were also highlighted.

Machine learning, combined with remote sensing and GIS, is an emerging technique to address the shortcomings in evaluation of MCDM models traditionally used in land suitability assessment. This avoids the subjective nature of commonly used approaches, such as AHP and fuzzy sets. Recent research has used ensemble learning to maintain objectivity in the evaluation of agricultural land suitability ([Bhullar et al., 2023; Ismaili et al., 2023; Mohammed et al., 2023](#)). Various ensemble learning models have been used to assess suitability of agricultural land based on random forest (RF) ([Dolsirittigul et al., 2022](#)), extreme gradient boosting (XGBoost) ([Hao et al., 2024](#)), support vector machine (SVM) ([Sarmadian et al., 2014](#)), and artificial neural networks (ANN) ([Ahmadi and Layegh, 2015](#)). Comparisons of machine learning models for land suitability assessment found that RF and XGBoost predicted better than SVM, ANN and k-Nearest Neighbor algorithms ([Ganesan et al., 2021; Ismaili et al., 2023](#)). Most importantly, tree-based ensemble models can handle categorical and continuous variables for classification and regression. Data-driven suitability models are becoming increasingly prevalent, due to their capacity to handle large amounts of data and their ability to capture complex non-linear relations. Consequently, machine learning algorithms enable researchers to model each variable's influence and their interactions towards an observed suitability classification based on data-driven attributes.

Namibia lacks crop suitability studies, especially in communal small-scale farming areas. Thus far, only a limited number of studies have focused on general agro-ecological zoning and suitability of livestock farming ([De Pauw and Coetzee, 1998; Strohbach, 2018](#)). One study on land evaluation highlighted the potential for growing multiple crop varieties through intercropping, using a year-round irrigated water supply ([FAO, 1984](#)). A separate study examined land evaluation of an individual commercial resettlement farm in the north central Oshikoto region of Namibia ([Mwazi, 2006](#)). To our knowledge, no previous studies on the Okavango River basin in Namibia have been conducted to assess leguminous crop suitability, despite a significant increase in crop growing in the area ([Fortunato and Enciso, 2023](#)). This indicates a knowledge gap of land uses and their feasibility to support crop farming activities in Namibia's north and northeastern communal areas. To determine suitable land for agriculture in the Rundu area of Namibia, located along the Okavango River basin, it is necessary to identify and utilise key variables representing biophysical and socio-economic elements in a multi-criteria assessment. This approach employs parameters that have been extracted from previous investigations in other parts of the world with comparable biophysical characteristics ([Roy and Saha, 2018; Tashayo et al., 2020](#)).

The objective of this study was to investigate the degree of land suitability for leguminous crops, namely sunn hemp (*Crotalaria juncea*) and pigeon pea (*Cajanus cajan*) along the Okavango River basin. The focus was on the sub-basins surrounding Rundu in Kavango East region of Namibia. The viability of land use for agriculture in the basin is crucial due to rapid population growth and high demand for food. This is particularly important because the basin already faces extreme climate conditions caused by climate change in an already water-scarce semi-arid environment. Additionally, there are more informal farming activities along the basin. Sunn hemp and pigeon pea are legume crops mainly cultivated in tropical and sub-tropical regions for human consumption and/or as livestock feed. These leguminous cover crops may help to suppress weeds by the allelochemicals found in their biomass ([Javaid et al., 2015](#)). Many smallholder farmers in semi-arid regions

view pigeon peas as a green manure alternative, as they are drought-resistant, high-protein cover crops that can improve soil nutrients (e. g., nitrogen fixing) (Cook et al., 2020; Sharma et al., 2011). In a similar vein, sunn hemp is renowned for its exceptional resilience to environmental stress and is employed as a green manure to enhance soil nutrients and increase livestock fodder supply. This legume crop can increase soil organic matter, raise nitrogen levels, and sequester carbon (Chaudhary et al., 2015; FAO, 2022). Since both these legumes improve soil moisture, enhance organic matter, and decrease soil erosion, they are typically interplanted with cash crops such as maize, sorghum, and pearl millet to increase yields. Smallholder farmers intercrop legumes as green manure cover crops with cereals in drought-prone semi-arid regions with low soil fertility. This improves soil nutrient contents and boosts output (Mashingaidze et al., 2012; Simon and Itanna, 2019). Following an in-depth literature review on suitable approaches in semi-arid regions of sub-Saharan Africa and discussions with agronomists, the two leguminous crops were selected to identify areas suitable to grow them in. The study was motivated by the numerous benefits that can be obtained from intercropping legumes, as demonstrated by several researchers (Odunze et al., 2004; Rani et al., 2019; Sharma et al., 2011; Tsubo et al., 2005). These benefits include soil rehabilitation, drought tolerance, as well as nutritional and economical values. The crop characteristics and information on environmental requirements for sunn hemp and pigeon pea were collected to determine optimal land for enhanced yields. Both legumes mainly grow in similar climatic regimes, demand similar soil properties and require matching topographical features (Cook et al., 2020; FAO, 2022; Orwa et al., 2009; Zimmermann, 2023).

In this study, our objectives are: i.) to subjectively assess and estimate land suitability for sunn hemp and pigeon pea using an MCDM-AHP approach; and ii.) to evaluate an objective data-driven non-linear multivariate approach, in conjunction with supervised learning models to understand the relationship amongst the selected variables and allocate crop suitability scores in the basin. In addition, this study seeks to evaluate the capability of the Okavango River basin to grow leguminous crops through investigation of the biophysical and socio-economic

characteristics of the area. This research contributes towards a better understanding of the interdependence of the Okavango River basin and its adjacent land use near Rundu.

2. Materials and methods

2.1. Study site

The study area, as shown in Fig. 1, is located in the semi-arid environments of seven sub-basins surrounding the Rundu municipality in the Kavango East region of northeastern Namibia. These sub-basins are part of the greater Okavango basin. The Okavango River flows from wetter highlands of Southern Angola, passing through Namibia, and dissipates downstream in northwestern Botswana. This transboundary river system is occupied by approximately 1 million inhabitants, predominantly engaged in small-scale farming (EPSMO-BIOKAVANGO Eflows Team, 2009).

The Rundu municipality and its surrounding sub-basins were chosen as a study site, as the area mainly subsists through crop production and shows the highest population density within the basin. According to the 2011 census, Rundu had a population of 63,431. With an annual growth rate of 5.4 %, the current estimated population is close to 120,000. This makes Rundu the second most populated town in Namibia (Namibia Statistics Agency, 2011). Rundu is a potential market for agricultural products, and is located along the Trans-Caprivi Corridor, a trade route linking land locked neighboring countries. On the Namibian side, the population along the basin mainly relies on farming pearl millet (known as mahangu in the local language), and to a lesser extent on sorghum, maize and vegetables such as cabbages and tomatoes. The major constraints for farming inland away from the river are the poorly suited soil types, which are mostly Arenosols characterised by low water and nutrient storing capacities. Along the river, Fluvisols and Gleysols are the prevailing soil types. These soils consist mainly of silty clay loam and sandy clay loam, which contribute to their high nutrient content and water storage capacity (Atlas of Namibia Team, 2022).

Approximately 70 % of the population in the Kavango East region,

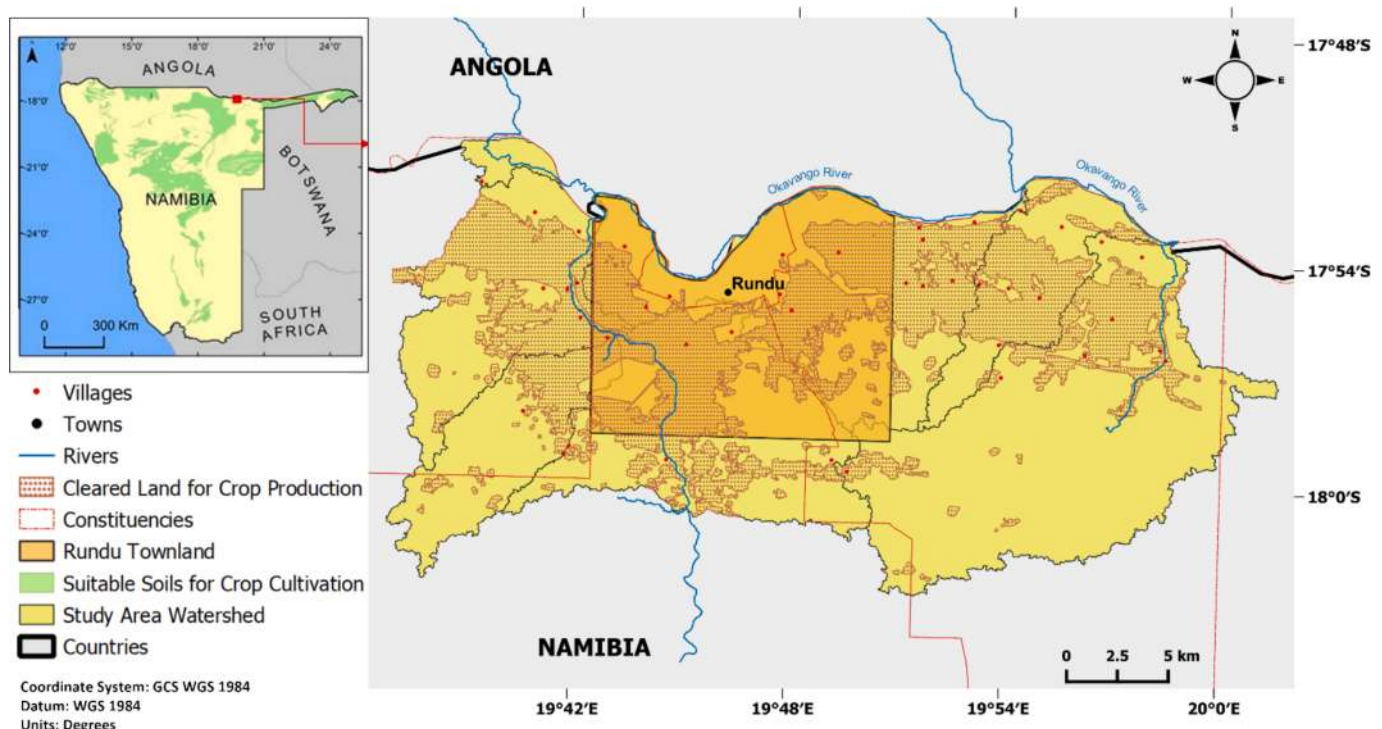


Fig. 1. Location of the study area: Okavango River sub-basins surrounding Rundu, Namibia.

where Rundu is located, are settled along the Okavango River basin. This is due to the availability of water, arable soil, and pastures suitable for agricultural activities (Brown et al., 2010). 53 % of households are engaged in crop farming which is higher than the national average of 32.8 % (Mwinga et al., 2018). The rainy season in northeast Namibia typically occurs between October and April, coinciding with cultivation activities. However, increased rainfall variability and intermittent drought occurrences can be observed in this semi-arid environment, causing reduced crop yields (Awala et al., 2019; Lu et al., 2016; Mendelsohn et al., 2002). As a result, small-scale farmers found in Rundu, the rapidly growing capital of Kavango East region are highly dependent on rainfed agriculture.

The study site has a mean annual rainfall ranging between 450 to 550 mm during the rainy season. Light rainfall occurs at the beginning of the rainy season in October and November and towards the end of April (The World Bank Group, 2021). Furthermore, an increase in rainfall is typically recorded between December and March. Most farming activities take place during this time to maximise yield. The typical size of a crop farm along the study area varies between 2 and 5 ha, and usually belongs to an individual household. Similar to other valleys found along rivers, the Okavango valley is dominated by floodplain grasslands and woodland vegetation (Mendelsohn and El Obeid, 2004). However, much of this land has been cleared for accelerated crop production resulting in soil degradation caused by erosion, which adversely affects the levels of organic matter and nutrients.

2.2. Data and methods

Variables used in this study were derived from previous literature, agricultural land evaluation guidelines and expert recommendations. Similar to typical land suitability assessments, a collective twenty-two climatic, hydro-geomorphologic, soil, vegetation and socio-economic variables (see Table 1) were identified to determine the suitability of leguminous crops in the study area. These variables were all resampled to 100 m spatial resolution considering the processing capacity requirements for the collective variables. The variables were grouped into five different categories, covering 22 significant biophysical and socio-economic characteristics for crop cultivation. These characteristics

were further classified into the following five subclasses: Highly suitable, suitable, moderately suitable, less suitable, and unsuitable. Primary and secondary data for the study site (see Table 1) were obtained from various sources and processed to identify suitable areas for intercropping sunn hemp and pigeon pea through multi criteria decision making approach and multivariate regression.

2.3. Climatic variables

The climate variables considered in this study were the daily amounts of precipitation and evaporation recorded over a ten-year period from 2013 to 2023. The months between October and April are usually considered as rainy season in Namibia, with the highest precipitation recorded in December, January and February (The World Bank Group, 2021). Crop growing season is also typically regarded to occur during these months with a relative abundance of rain. During this period, December, January, and February have the highest temperatures and evaporation rates. CHIRPS Daily precipitation dataset with 0.05° resolution from the Climate Hazards group Infrared Precipitation with Stations (Funk et al., 2015) is used to extract the daily mean, minimum and maximum of the study area. Daily aggregated ERA-5 Land ECMWF (European Centre for Medium-Range Weather Forecasts) climate reanalysis (Copernicus Climate Change Service, 2019), is used to extract mean, minimum and maximum evaporation at 0.1° resolution for the duration of the growing season.

2.4. Hydro-geomorphologic variables

The hydro-geomorphological variables considered in this study were elevation, slope, drainage density and ground water depth. The first three variables were generated using a TanDEM-X Digital Elevation Model with 12 m spatial resolution and relative vertical accuracy of 2 m, obtained from German Aerospace Center (DLR) (Wessel et al., 2018). Ground water depth was generated through inverse Distance Weighting (IDW) interpolation using observed data from Geohydrology Division at Ministry of Agriculture, Water and Land Reform in Namibia. Elevation for the study site ranges between 1063 and 1152 m above sea level, whereas a dominant gentle slope is observed in most of the study region,

Table 1

Summary of variables for land suitability of leguminous crops and their description, sources, spatial and temporal resolution.

Parameters	Description	Data Source	Res.	Time
Elevation	Elevation	TanDEM-X (HRTI-3) – TerraSAR-X DLR mission	12 m	Fixed
Slope	Slope	TanDEM-X (HRTI-3) – TerraSAR-X DLR mission	12 m	Fixed
LULC	Land use land cover	ESA WorldCover 2021	10 m	Fixed
Distance from river	Distance from river	TanDEM-X (HRTI-3) – TerraSAR-X DLR mission	12 m	Fixed
Drainage density	Drainage density	TanDEM-X (HRTI-3) – TerraSAR-X DLR mission	12 m	Fixed
Ground water depth	Ground water depth	Hydrological Services of Namibia		Fixed
Proximity to market	World Settlement Footprint	Deutsches Zentrum für Luft- und Raumfahrt (DLR)	10 m	Fixed
Proximity to road	Distance from road	Directorate of Survey and Mapping (DSM), Namibia		Fixed
Soil bulk density	Soil bulk density	Innovative Solutions for Decision Agriculture Ltd. (ISDA)	30 m	Fixed
Soil cation exchange capacity	Soil cation exchange capacity	Innovative Solutions for Decision Agriculture Ltd. (ISDA)	30 m	Fixed
Soil organic carbon	Soil organic carbon	Innovative Solutions for Decision Agriculture Ltd. (ISDA)	30 m	Fixed
Soil fertility capability classification	Soil fertility capability classification	Innovative Solutions for Decision Agriculture Ltd. (ISDA)	30 m	Fixed
Soil depth	Soil depth to bed rock	Innovative Solutions for Decision Agriculture Ltd. (ISDA)	30 m	Fixed
Soil nitrogen	Soil nitrogen	Innovative Solutions for Decision Agriculture Ltd. (ISDA)	30 m	Fixed
Soil phosphorus	Soil phosphorus	Innovative Solutions for Decision Agriculture Ltd. (ISDA)	30 m	Fixed
Soil pH	Soil pH	Innovative Solutions for Decision Agriculture Ltd. (ISDA)	30 m	Fixed
Soil texture	Soil texture	Innovative Solutions for Decision Agriculture Ltd. (ISDA)	30 m	Fixed
Soil type	Soil type	Updated Soil Map and Soil Profile Database of Namibia (Coetzee 2020)	250 m	Fixed
Soil water content	Soil water content	EnvirometriX Ltd	30 m	Fixed
NDVI	Mean Normalized Difference Vegetation Index (October – April)	NASA's Land Processes Distributed Active Archive Center (LP DAAC)	250 m	2013—2023
Precipitation	Mean of daily mean precipitation rate (October – April)	Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)	5566 m	2013—2023
Evaporation	Mean of daily mean evaporation rate (October – April)	ERA5-Land – Copernicus Climate Data Store	11132 m	2013—2023

between 2° to 5°. The drainage density varied between 0.4 and 4.5, with most of the area concentrated between 2 and 3.5 km/km².

2.5. Soil variables

This crop land suitability assessment took into account soil physical and chemical properties to determine the potential for increased crop yield, with a focus on soil nutrients such as nitrogen, organic carbon and phosphorus; soil pH measuring the level of alkalinity or acidity; cation exchange capacity measuring soil ability to store nutrients; soil bulk density, texture, and type determining its porosity and structure; water content defining the amount of soil moisture and soil fertility capability classifying its capacity to grow crops. All soil data were retrieved from Innovative Solutions for Decision Agriculture Ltd. (iSDA), except for soil water content from EnvirometriX Ltd (Hengl and Gupta, 2019) and soil type from updated soil map and soil profile database of Namibia (Coetzee, 2020). Their characteristics were divided into five different classes representing different categories of suitability as per reviewed literature and expert's opinion. Soil properties suffer negative consequences due to high intensity farming causing soil degradation and eventually reduction of crop productivity. Sunn hemp and pigeon pea, also referred to as leguminous green manure cover crops (GMCC), have a positive impact on soil chemical properties improving soil nutrients, holding capacity, and cation exchange capacity (Simon and Itanna, 2019). They also play an important role in soil nitrogen fixing, maintaining pH level, enhancing phosphorus uptake, and weed management through allelopathic potential. Furthermore, improved soil physical properties can be observed resulting in improved water holding capacity, decreased soil erosion, increased porosity and allowing aeration and percolation (Rani et al., 2019). According to Fan et al. (2016) and Hag Husein et al. (2021), a majority of crop's root biomass is located in the upper 30 cm of soil profile. However, in their study on data-driven multilayer crop suitability prediction, Bhullar et al. (2023) found that the median root depth ranged from 8.4 to 19.8 cm. They also emphasised that the model's accuracy was improved when considering a soil depth of 10 cm. In light of these observations, the extracted soil dataset for our study area were at the depth ranging between 0 – 20 cm mapped at 30 m spatial resolution, using the iSDA data assimilation of remotely sensed products (i.e., Sentinel-2, Landsat, DEM products, etc.), and over 100 thousand soil samples (Hengl et al., 2021).

2.6. Socioeconomic variables

The socioeconomic criteria considered in this study were the proximity to the nearest population settlements and the Euclidean distance to roads. These criteria determine potential marketplaces to connect buyers and sellers, as well as facilitate transport infrastructure for farming inputs and market accessibility for smallholders. The proximity to population settlements was extracted from the World Settlement Footprint (WSF) dataset by Deutsches Zentrum für Luft- und Raumfahrt (DLR), derived from Landsat-8 and Sentinel-1 products, and validated by ground-truth samples (Marconcini et al., 2020). Population settlement data in the study area were used as indicator for potential marketplace for crop yields to satisfy the demand and supply. The major roads dataset sourced from Namibia's Directorate of Survey and Mapping (DSM) were used to compute accessibility to agricultural extension services, access to technology, and access to market for smallholder farmers. The dataset was generated by applying the Euclidean distance metric to assign a distance value to each pixel within the study area based on its proximity to the major roads.

2.7. Vegetation variables

This study used Land Use Land Cover (LULC) product from ESA WorldCover (Zanaga et al., 2022), which directly correlates with the expected outcome of crop suitability. LULC is used to filter out the

constraints such as built-up area from the final suitability map during AHP. A multivariate setting crop land from LULC is used to see the probability of suitability classification and rate the importance of a variable for this suitability classifier. The Normalized Difference Vegetation Index (NDVI) was extracted from MOD13Q1.061 Terra Vegetation Indices collection (Didan, 2021), for the duration of ten years (2013–2023), and used as a proxy for suitability in the multivariate regression.

2.8. Analytic hierarchy process

The AHP method, developed by Saaty (1987), was used to determine the weights of different variables influencing growth of the two legumes was conducted through pair-wise comparison. Expert opinions collected from 5 experts to evaluate the relative importance of each variable was carried out, resulting with the weights of each variable ranging from 0 to 1. This was done to identify appropriate land for growing leguminous crops in the sub-basins surrounding Rundu along the Okavango River. Several criteria categorised under climatic, soil, hydro-geomorphic, socioeconomic and vegetation factors, as shown in Fig. 2, were ranked to compare the relative importance of each individual criterion. Based on Saaty (1987), the relative scale ranges between 1 indicating “both criteria contribute equally to the objective” and 9 representing “evidence favors a criterion over another in the highest possible order”. Thereafter, the pair-wise comparison matrix is computed based on the outcome of the relative importance scale using the {ahp} package in R (Glur, 2024), as shown in Eq. (1).

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \quad (1)$$

Once the comparison matrix is computed, the weight of each criterion is determined using a priority vector (i.e., normalized eigen vector) method. The derived normalized eigen vector represents the average values of each row of variables, showing the relative weights of each individual criterion being compared. At this point, a consistency check is crucial, as inconsistencies may occur due to randomly generated matrixes. The consistency index which measures the expert judgements of the variable comparison is measured using the consistency ratio (CR), as seen in Eq. (2).

$$CR = \frac{CI}{RI} \quad (2)$$

CI refers to the consistency index and RI stands for the random consistency index. The acceptable value for the CR must be equal to or lower than 10 %. If the value is less than 10 %, the inconsistency will be tolerable, and the parameter will further be used in the suitability analysis. However, if CR is greater than 10 %, the variable will be omitted, as it cannot be trusted for the analysis process.

2.9. Weighted overlay process

Finally, a crop suitability map will be generated based on the acceptable criteria weights. This is achieved using a weighted overlay analysis in ArcGIS 10.5 through the application of normalised criteria and their associated weights (Drobné and Lisec, 2009). A weighted overlay application creates a common scale for varied and heterogenous input values. This allows an integrated criteria analysis to address complex spatial problems in the classification of suitability. The application of the weighted overlay method, combined with the AHP, enables the integration of different thematic layers to assess how various factors affect crop suitability. The relative importance of individual criteria towards land suitability is resolved through a Pairwise comparison matrix. The weight of each variable is multiplied by the suitability score

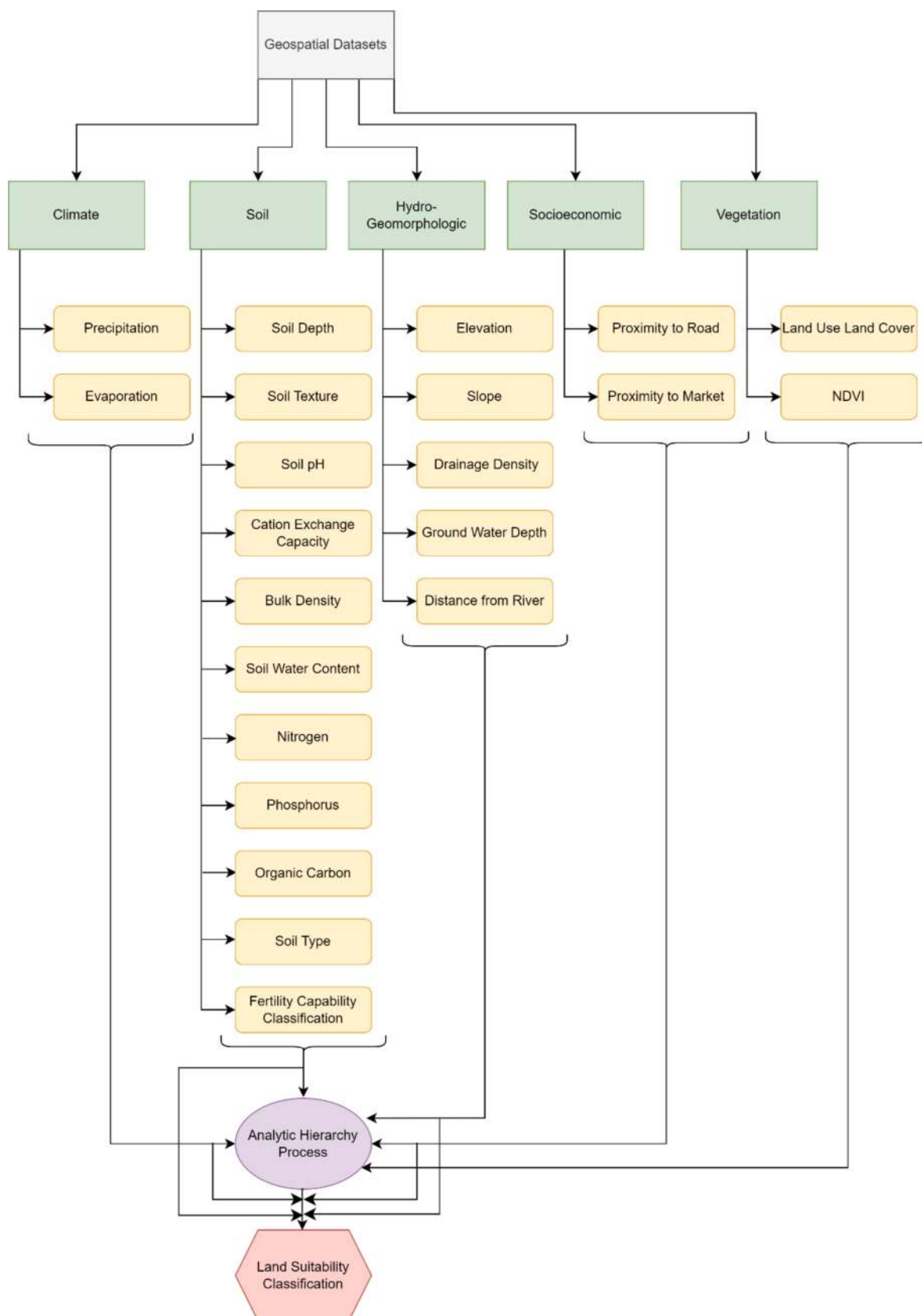


Fig. 2. Workflow diagram showing geospatial datasets used as input variables for AHP criteria weighting for land suitability classification.

of that variable, and the resulting values are summed to compute a crop suitability map, as illustrated by the equation (Cengiz and Akbulak, 2009; Owusu et al., 2017).

$$LS = \sum_{i=1}^n w_i x_i \quad (3)$$

where LS is suitability, w_i denotes weight of variable i , and x_i denotes score of variable i .

Given the subjective nature of the MCDM-AHP method, which employs expert opinion to assess the suitability of land for agricultural productivity, a subsequent multivariate statistical approach is proposed in order to assess the parameters objectively.

2.10. Data-driven crop-suitability approach

Our study utilised a dichotomous approach to assess crop suitability. While the AHP method was used to assess land suitability with expert-weighted variables, the machine learning models were trained exclusively on the original, unweighted environmental variables. No expert-derived weights or outputs from the AHP analysis were incorporated into the machine learning models. As such our machine learning approach is purely data-driven, without propagating any expert bias

(Figs. 2 and 3). The method allowed us to compare empirical data with subjective assessments, leading to a better understanding of the factors that affect crop suitability. We used the health and vigor of the vegetation using the NDVI as a proxy due to the lack of data from the two key crops. This index is closely related to crop suitability (Junges et al., 2017; Yan et al., 2023). Furthermore, we used land use classification data to analyse the various potential factors that influence crop cultivation.

The primary focus of the approach is to achieve two main objectives. First, we aimed to examine the complex relationships between different environmental factors and their effect on crop suitability. We focused on identifying and understanding how different elements of the natural and socio-economic environment interacted to influence crop suitability. Second, the goal was to build a robust set of predictive models capable of accurately identifying the most influential factors for crop suitability. The models aimed to provide precise predictions and offer insights into the relative weight of various environmental and socio-economic factors in determining crop cultivation suitability.

The machine learning model training and analysis were conducted using the Python programming language, utilising scikit-learn for machine learning models and XGBoost for gradient boosting algorithms.

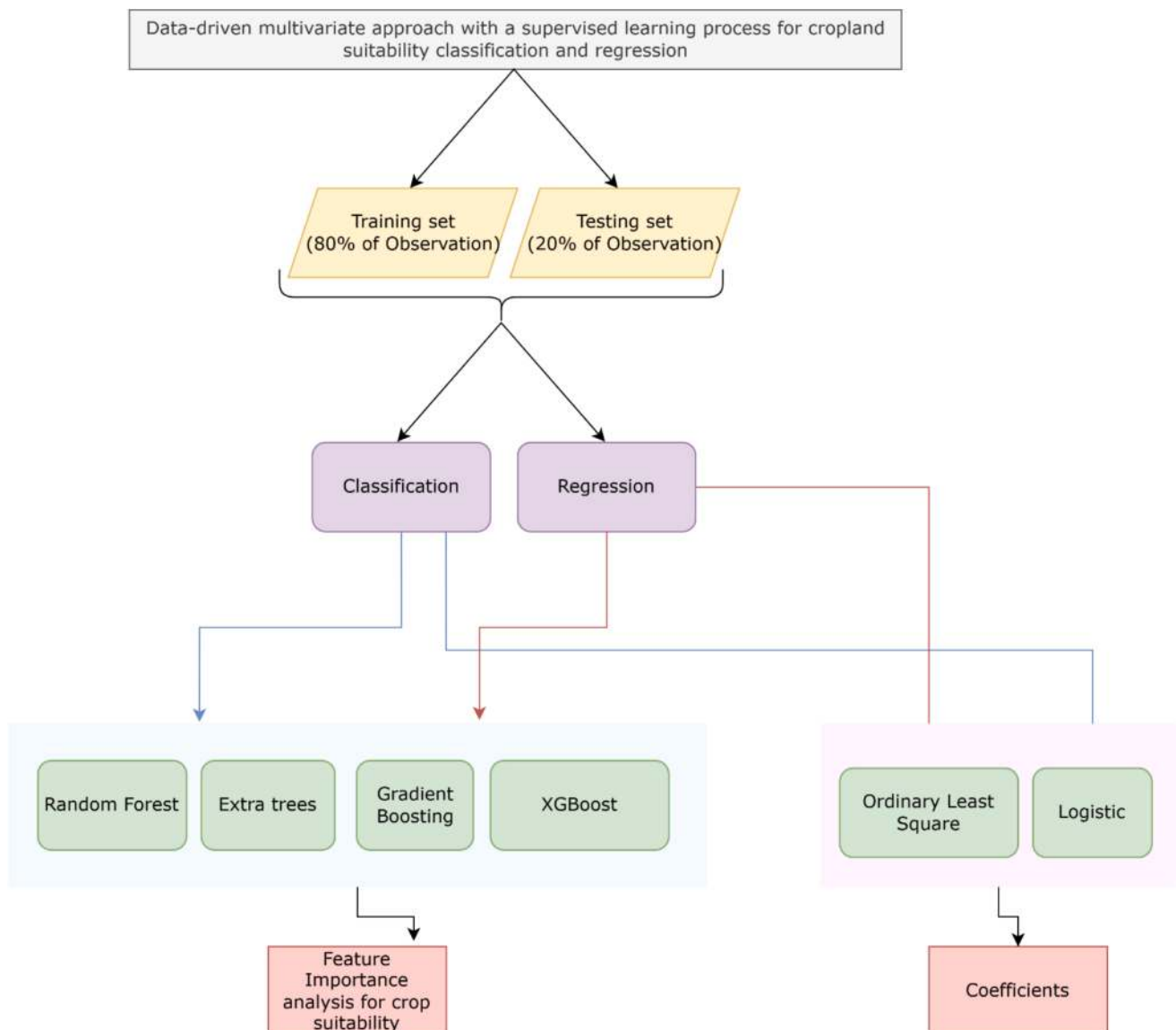


Fig. 3. Ensemble learning models for crop cultivation suitability.

The computations were executed on a High-Performance Computing Platform (HPCP) equipped with 128 GB of RAM and 64 CPU cores.

2.10.1. Ensemble machine learning models

Using decision trees, we created and evaluated ensemble learning models. These models incorporate numerous decision trees to improve prediction accuracy and flexibility, making them suitable for dealing with complex datasets. We employed tree-based ensemble learning models, including Random Forest, Extra Trees, Gradient Boosting, and XGBoost (see Fig. 3). During training, these models build several decision trees and return the mode of the classes for classification or the mean prediction for regression of the individual trees.

The Random Forest is an ensemble learning method that builds multiple decision trees on bootstrapped training samples and a random sample of predictors (James et al., 2013). It provides outputs based on the average or majority voting of the individual trees (Breiman, 2001). Extra Trees are another ensemble learning approach that, similar to Random Forest, generates numerous decision trees and outputs the average prediction of the individual trees. However, Extra Trees uses the full dataset and random thresholds for splitting (Geurts et al., 2006). On the other hand, Gradient Boosting combines weak predictive models to generate a powerful predictor by sequentially building an ensemble of trees, with each tree correcting the mistakes of the previous ones (Chen and Guestrin, 2016; Natekin and Knoll, 2013). XGBoost is an optimised distributed gradient boosting library that provides a scalable and accurate implementation of gradient boosting.

2.10.2. Multivariate regression analysis

In addition to using tree-based models, we conducted multivariate regression analyses to determine evidence of association between various variables and suitability. This approach yielded coefficients that indicate the strength of the relationship between each predictor and the outcome variable (Fig. 3). The basic goal of multivariate regression is to estimate the coefficients associated with the independent variables in order to minimise the difference between the observed and predicted values of the dependent variable, as indicated by the regression equations (James et al., 2013). Linear regression methods are used to express multivariate regression when dealing with continuous variables such as NDVI. On the other hand, when dealing with dummy variables (crop land), a logistic model is used to represent the regression.

$$y_i = \beta_0 + \sum_{i=1}^n \beta_i X_i + \varepsilon \quad (4)$$

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \sum_{i=1}^n X_i + \varepsilon \quad (5)$$

In Eq. (4), y_i represents the dependent variable, which refers to the suitability level of a particular pixel (NDVI). β_0 represents the intercept term, β_i are the coefficients associated with the independent variables $X_1, X_2, \dots, X_i$ respectively, and ε denotes the error term capturing unexplained variability (James et al., 2013). Similar, $\log(p(X)/(1-p(X)))$ in Eq. (5) represent the log odds or logit.

2.10.3. Data preprocessing and estimation

To ensure numerical stability during model estimation, we standardised continuous variables using mean normalisation and variance scaling. We also transformed categorical variables into a binary matrix using one-hot encoding, in order to make them compatible with the different learning algorithms.

We partitioned the dataset into five sets and used a 5-fold cross-validation method. Each set was used once for validation, while the remaining four datasets were used as training sets. This would mitigate overfitting risks and improve the generalisability of the results (James et al., 2013). We also used performance metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2). We

analysed the models that achieved the highest R2 values to evaluate their ability to capture variance within the dataset. Additionally, we optimised the ensemble models through hyperparameter tuning (see Table 2). We fine-tuned parameters such as subsample, learning rate, and tree characteristics to improve model performance.

The optimised model was assessed using training and testing datasets. Primary evaluation metrics included accuracy, precision, recall, and F1-score, providing a comprehensive outlook on the model's ability to reliably classify cropland. To identify potential overfitting or underfitting, the bias-variance trade-off analysis was crucial in evaluating this model. Overfitting occurs when a model performs better on the training set than the validation set, indicating that the model is too complex for the data. On the other hand, underfitting is suggested by equally high error rates on both sets, indicating that the model is too simple. A Feature Importance analysis was conducted for both regression and classification models to compute and visualise the relative importance of each characteristic. This analysis identified the most important factors which can be used to inform environmental policy decisions.

2.10.4. Model evaluation framework

To evaluate model performance, a variety of metrics were considered. Classification models were evaluated for accuracy, precision, recall, and F1-score. While regression models were assessed using R-squared and Root Mean Squared Error (RMSE). Tree-based models identified crucial variables affecting crop suitability. Furthermore, the multivariate regression technique provided an insight into the direction and magnitude of the effect of each factor.

Model performance in the Ordinary Least Squares (OLS) regression was evaluated with R2 to indicate the proportion of dependent variable variance explained by predictors, and RMSE to reflect the average magnitude of prediction errors. The logistic regression results display changes in the log odds of an area's suitability for cropland in response to predictor variable variations.

3. Results and discussion

3.1. Cropland suitability evaluation

A comprehensive literature review was conducted, and the input of five experts was sought based on their informed judgement. This was done to assign weights and rank selected variables according to their relative importance through AHP to determine suitable areas for sunn hemp and pigeon pea cultivation at pixel level. Based on the framework of land evaluation from FAO (1976), five suitability classes reflecting the degree of suitability for agricultural land use were defined (i.e., Highly Suitable, Suitable, Moderately Suitable, Less Suitable and Unsuitable). Land that is highly suitable for crop growth has no significant limitations. If minor limitations do occur, they will not increase inputs beyond an acceptable level to achieve expected productivity. Suitable class is associated with land that has few or no limitations that would negatively affect its productivity, and it requires minimal inputs to achieve optimal crop yield. Moderately suitable agricultural land is considered to have substantial input requirements to realise balanced productivity. Despite the fact that it may be attractive to farmers, it is still considered inferior to the Suitable class. The class of land that is less suitable represents land used for agricultural purposes. It has significant constraints that pose

Table 2
Hyperparameters of ML models.

Parameter	Values
<i>model_learning_rate</i>	0.01, 0.1, 0.2, 0.3, 0.4
<i>model_subsample</i>	0.5, 0.6, 0.7, 0.8, 0.9
<i>model_colsample_bytree</i>	0.5, 0.6, 0.7, 0.8, 0.9
<i>model_reg_alpha</i>	0.1, 0.001, 0.00001
<i>model_reg_lambda</i>	0.1, 0.001, 0.00001

challenges for sustainable use and requires substantial inputs to acquire benefits. Consequently, it is regarded as marginally viable. The unsuitable class represents a piece of agricultural land that is incapable of being used sustainably for its intended purpose. No amount of input

could justify its sustainable use. In the decision-making process, pairwise comparison was performed using a relative scale between 1 and 9, as stated in [section 2.8](#). This comparison aimed to determine the preference of one criterion over another and compute its comparison

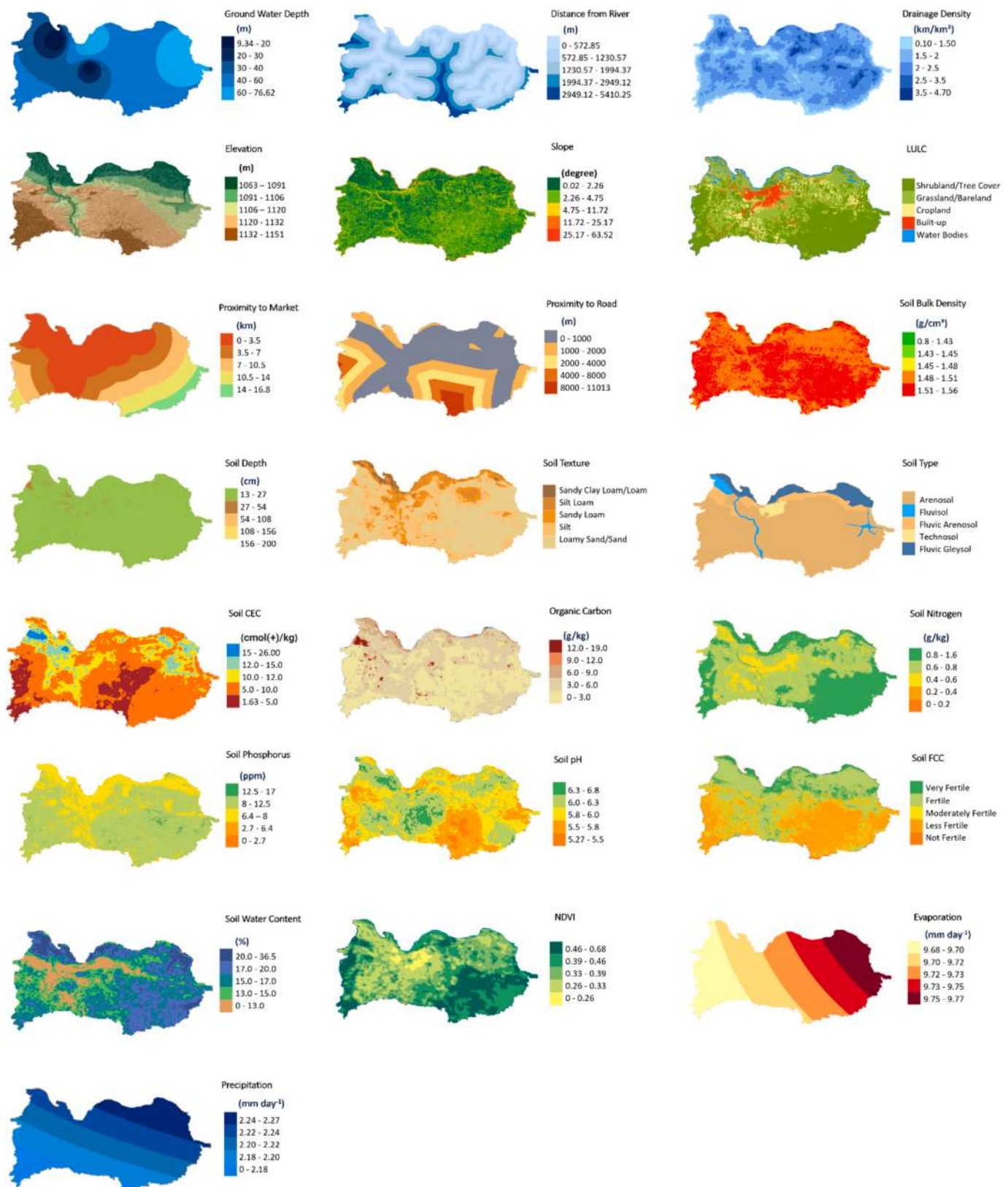


Fig. 4. Classification data layers of the study area with variables from climatic, hydro-geomorphologic, soil, socioeconomic and vegetation categories.

matrix. The weights of each criterion were then derived using a priority vector with an acceptable consistency ratio of 8.85 %.

3.2. Climatic factors

In search of appropriate location for leguminous crop production, two variables were specifically considered: precipitation and evaporation. These factors were used to determine the suitability of the climate conditions in the study region. The average daily rainfall over crop growing period ranged between 0 and 2.27 mm, as shown in Fig. 4. Local experts in conservation agriculture advised an average daily rainfall of at least 1.5 mm for intercropping legumes with pearl millet (Zimmermann, 2023). Moreover, the ECOCROP database indicates a minimum average daily precipitation requirement for sunn hemp and pigeon pea at 1 mm and 1.8 mm, respectively (FAO, 2022). Accordingly, the optimum precipitation rate was set to be between 2.27–2.24 (i.e., 15.6 % of the total study area size), while the least suitable class of 2.18–0 was assigned as unsuitable, representing 9.39 % of the whole study area. The potential mean evaporation rate in the study region is approximately four times higher than the mean rainfall. This is consistent with a high evaporation rate observed by The Permanent Okavango River Basin Water Commission (2011) and The World Bank Group (2021), and a mean daily maximum temperature recorded between 30–35 °C in the Namibian basin. However, both leguminous crops possess a high tolerance to withstand drought, which enables them to remain resilient to temperatures as high as 40–45 °C (FAO, 2022; Zimmermann, 2023).

3.3. Hydro-geomorphologic factors

In order to assess the hydrological and geomorphological conditions of the study site, five variables were analysed: slope, elevation, ground water depth, drainage density and distance from the river. The objective was to establish the suitable hydro-geomorphological characteristics for the crops under consideration. Areas with lower elevations were categorised as highly suitable, while their suitability decreases as the altitude increases. This was confirmed through compiled datasets by various sources from experts and availed for farmers, agricultural extension personnel and researchers (Cook et al., 2020; FAO, 2022; Orwa et al., 2009). *Crotalaria juncea* (aka sunn hemp) optimally grows between 0–300 m above sea level (Orwa et al., 2009; USDA, 1999). However, there have been recorded cases of sunn hemp and pigeon pea crops between 0–1500 m and 0–3000 m, respectively (Cook et al., 2020; FAO, 2022). Nevertheless, sluggish growth is observed due to lower temperatures at high altitude. Rundu and its surrounding basin are located between 1063–1152 m above sea level. This semi-arid sub-tropical climatic regime experiences warmer climate even if it has a high elevation. This makes most of the study area suitable for growing leguminous crops. The topography of the area is characterised as mainly gentle slope ranging between 2–12 degrees, representing over 98 % of the study area. Harmony (1999) demonstrated over three growing seasons that legume crops performed best on slopes between 2 and 10 degree. Guretzky et al. (2004) highlighted that the most successful slopes for legumes were observed between 8 and 12 degrees. Drainage density in the study area ranges from 0.1–4.6 km/km² and strikes a good balance between very high and low densities. Low to moderate drainage densities (85 % of the study area) were classified as suitable for optimal sunn hemp and pigeon pea cultivation. As noted by Dragičević et al. (2019), a higher drainage density is usually linked to low suitability due to erosive tendencies of the land. On the other hand, a low drainage could result in waterlogging adversely affecting the root zone of the leguminous crop (Kaur et al., 2020). Ground water depth is a predominant influencing factor for crop cultivation. The water table exhibits a range of 9.34 to 76.62 m, with more than 70 % of the study area experiencing groundwater levels between 30 and 60 m below the surface, which aligns with Worqlul et al.'s (2017) finding that there is a

moderate potential for irrigation using groundwater. One of the primary determinant factors for successful cultivation of the leguminous crops under investigation is their proximity to the river. Almost 60 % of the study area is within 1.5 km of the river, thus creating an optimum situation for extracting water for crop fields. In addition, a study by Zhang et al. (2018) conducted in an arid environment demonstrated consistent higher soil water content within 2 km distance from a river, followed by a significant decline beyond the 2 km limit. Considering this scenario, more than 80 % of the study region is located within the 2 km range of higher soil water content, thereby establishing conducive conditions for arable land.

3.4. Soil factors

This study examined soil physical and chemical properties to assess suitability for leguminous crop cultivation. The soil properties utilised for identifying regions with potential for enhanced crop yields include nitrogen content, soil organic carbon, phosphorus, pH, cation exchange capacity, bulk density, soil depth, soil water content, fertility capability, soil texture and soil type. Soil pH recorded in the study area is between 5.2 and 6.8, which is moderate to slightly acidic and neutral. The optimal pH range recommended by Cook et al. (2020) and Orwa et al. (2009) for cover crops sunn hemp and pigeon pea varies from 4.5 to 8.4. Nevertheless, these crops thrive in pH 6–7 (FAO (2022)), considering this 43.35 % of the entire area well suited for enhanced crop yields. In the sub-basins surrounding Rundu, the total nitrogen content ranges between 0 – 1.6 g/kg, with 99 % of the study area comprising nitrogen levels from poor to fairly moderate. This occurs due to lack of nutrients received from external sources, combined with extreme climatic conditions such as high temperature and low precipitation. Therefore, to sustain and improve soil fertility and nutrient levels such as nitrogen, smallholders must consider practicing cover crop leguminous crops such as sunn hemp and pigeon pea, as recommended by Awala et al. (2019), and Hag Husein et al. (2021). Phosphorus levels in the study site are significantly low, ranging from 0 to 17 ppm, with the majority of areas showing poor phosphorus levels ranging from 0 to 8 ppm. Soil Organic Carbon (SOC) holds an immense potential in determining the suitability of agricultural land, due to its ability to improve soil biological, chemical, and physical properties, water-holding capacity, and structural stability. It also plays an integral part in the formation of soil's organic acids that influence soil minerals dissolutions and availability to plants and nutrient leaching. The SOC in the research area varies in concentration between 0 and 19 g/kg, with dominant values ranging between 0 to 6 g/kg. Areas with moderate to rich carbon contents only amount to 8.5 % of the entire study area. Intensification of sunn hemp and pigeon pea has shown potential to increase the level of soil phosphorus and carbon (Singh et al., 2023). Moreover, leguminous crops have shown to demand 20 % less phosphorus than non-leguminous crops such as pearl millet, which is the dominant crop in the study area (Li et al., 2011). Cation exchange capacity in the study area falls in the range of 1.63 – 26 cmol(+)/kg. This variable indicates the overall potential of soil fertility. Only 8.85 % of the entire study area demonstrated moderate to high fertility, whereas the remaining 91.15 % exhibited low fertility. This aligns with the findings by Githae et al. (2011) that soil CEC < 16 cmol (+)/kg is low in fertility for leguminous crops. The spatial variation of bulk density in the study region ranged from 1.41 – 1.56 g/cm³ with 49.4 % displaying a moderate bulk density falling between 1.41 – 1.51 g/cm³, and the remaining exhibited soil with compacted texture spanning from moderate to high. A higher bulk density restricts root growth, which in turn reduces crop yield. Nevertheless, leguminous crops have shown enhanced resistance to higher bulk density improving compacted soil textures. This was confirmed by Odunze et al. (2004) in their study on grain legumes and their potential to improve soil bulk density. Sandy loam and silt with coarse and medium texture are the dominant classes with 82 % coverage. The primary soil types were Arenosols (78.73 %), followed by Fluvisol (10 %) and Fluvic gleysol (9 %). Legume crops such

as sunn hemp and pigeon pea perform best in sandy soil, with 5 – 10 % clay and sandy loam in semi-arid environments (Jacobi, 2008). Soil depth to bedrock varied from 13 – 200 cm, with over 97 % of the area observed to be 2 m or below the top surface. Classes indicating soil fertility capability reflect varying levels of constraints for potential farming area within the study region. Out of the entire area, only 9.3 % exhibited no constraints. Predominantly, areas with only one constraint constituted 43.3 %, followed by 36.4 % representing regions with two constraints.

3.5. Socioeconomic factors

In socioeconomic analyses for suitable agricultural land, key features such as population, infrastructure, and markets are the primary factors to determine optimal locations for legume crop cultivation. Two variables, proximity to market and accessibility to roads, were central to understand the social and economic influences to classify the suitability of croplands. Potential crop lands were classified to determine their proximity to major roads. The categories ranged from < 1000 m, which were the most accessible, to distances > 8000 m, which were the least accessible. Roads distances spanning from < 1000 to 4000 m were classified to be highly to moderately suitable, which represented approximately 77 % of the study area. This demarcates regions where economically feasible production of leguminous crops is viable, as well as facilitates easy access to transport raw agricultural inputs. Similar studies by Pramanik (2016), and Sarkar et al. (2023), also reported comparable findings. The distance to the market was measured using population settlement data signifying spatial distribution over the study site. This is significant to determine the proximity of markets to sell potential sunn hemp and pigeon pea yields within the study region. For this purpose, the spatial distribution of settlements was classified into five classes according to their distance from clustered settlements. Areas within 3.5 km from potential market had the highest coverage with 40.7 %. This is attributed to the urban settlement of Rundu found within our study site.

3.6. Vegetation factors

The vegetation parameters considered in this study were NDVI (used to measure vegetation density and health) and LULC (used to monitor the landscape and its various uses). The spatial distribution of vegetation health, as measured by the NDVI, was categorised into five classes with values ranging from 0 to 0.68. The robust with NDVI values between 0.39 and 0.68, accounted for 51.6 % of the entire study region. Vegetation with moderate health was observed in 25.9 % of the study area. The Land Use and Land Cover (LULC) data was primarily utilised as a criterion to identify potential limitations in the study area, such as built-up areas and water bodies, and exclude these classes from the final suitability map.

3.7. Land suitability evaluation using analytical hierarchical process

Through AHP approach, weights were assigned (Fig. 6) for twenty-two variables in five distinct categories (see Fig. 4). Distance to River (8.48 %), Precipitation (8.14 %) and Soil water content (8.02 %) scored the highest weights signifying that water is an important element for suitability ranking during the cultivation of leguminous crops, even if considered to be drought resistant. For crops to achieve maximum yield in the semi-arid study region, they need to be closer to the river for ease of irrigation. Moreover, as discussed by Zhang et al. (2018), soil water content is observed to be highest in soils usually found in the riverbanks. In areas further away from the river, smallholders are highly dependent on rainfall availability. This aligns with similar research findings (Adem et al., 2023; Ammar et al., 2016; Zhang et al., 2018), that showed crops in arid and semi-arid environments are highly dependent on rainfall, irrigation and availability of water in soil to maintain or increase yields.

Furthermore, Evaporation (1.89 %), Proximity to Roads (2.08 %) and LULC (2.70 %) had the lowest weights. Evaporation is reduced by shades created from legume cover crops reducing direct sun light reaching the soil. This further reduces soil temperature minimising soil moisture loss and consequently, making sunn hemp and pigeon pea drought resistant cover crops that could potentially be intercropped with maize and millet in semi-arid environments. Tsubo et al. (2005) developed a cereal-legume model to demonstrate the significance of intercropping for sustainable crop production in developing countries. During the expert's discussion, the importance of proximity to roads was not considered to be as significant as other variables such as soil physical and chemical properties. These properties were deemed to be more crucial for crop health and higher yields. For example, soil type and texture both recorded a weight of 5.1 % and were considered to be more important to grow legume cover crops in the study area. LULC also had one of the lowest weights, as this class is mainly used to deal with constraints such as built-up area and waterbodies. Built-up area was considered as a constraint, as intensive agricultural productivity cannot be implemented in this land-use type apart from small activities such as backyard gardens. Moreover, recession farming, which is a form of agricultural cultivation on a floodplain where farmers practice agriculture by successively planting in the flooded areas after the waters recede, is also not acceptable, as the two crop types considered do not cope well in a waterlogged environment. Soil nutrients such as nitrogen (4.57 %), organic carbon (4.39 %) and phosphorus (4.1 %) scored weights ranging between 4 % and 4.6 %. This indicates that the expert team considered leguminous crops as specialised crops that require fewer nutrients from the soil. These crops are commonly used to improve the nutrient content of the soil in the study area. This has been discussed in the specialised literature (Cook et al., 2020; Sharma et al., 2011; Simon and Itanna, 2019; USDA, 1999), emphasising the ability of leguminous crops such as sunn hemp and pigeon pea to rehabilitate soil physical and chemical properties.

Sunn hemp and pigeon pea are known for their soil rehabilitation traits, lowering soil temperature, preventing erosion and water retention capacity. This is highly influential as they can be treated as an intercropped legume with the widely cultivated mahangu (Pearl millet) and maize in the study region. This has been discussed by various studies (Rani et al., 2019; Sharma et al., 2011; Tsubo et al., 2005), where leguminous crops exhibited regenerative potential in poor soil quality regions, mostly in semi-arid environments. According to the AHP ranking, the study area suitability map resembled the characteristics of only four of the five classes mentioned above. The overall suitability map in Fig. 5 exhibited 17.63 % of the total study area to be suitable for leguminous crops, signifying the land to have minor constraints and requiring less inputs for enhanced productivity. A visual analysis of this outcome reveals that most of the suitable land is in close proximity to the river and its tributaries. This is attributed to the requirement of water for crop growth in this semi-arid environment and the fluvic-gleysol soil types with silty and sandy loam texture, which have an enhanced capacity to hold nutrients and water. Even though gleysols in Namibia are usually associated with waterlogged places such as rivers and dams (Atlas of Namibia Team, 2022), which is a detrimental factor for the cultivation of sunn hemp and pigeon pea, the river basin usually recedes by the end of the dry season. As a result, most of the floodplain becomes devoid of water but retains high soil moisture. This creates an opportunity for cultivating these legumes during the light rainy season in October and November, before waterlogging becomes a problem.

On the suitability map, most of the coverage is classified as Moderately Suitable, with 62.77 % of the entire study area falling within this class. This means that the constraints identified in this class limit its performance. To achieve optimal growth of legumes and maximise the expected yield from this land use, a higher level of input will be necessary compared to areas with a suitable classification. However, this remains appealing to farmers. The suitable class was the second largest with 17.63 % of the entire study area showing the highest potential for

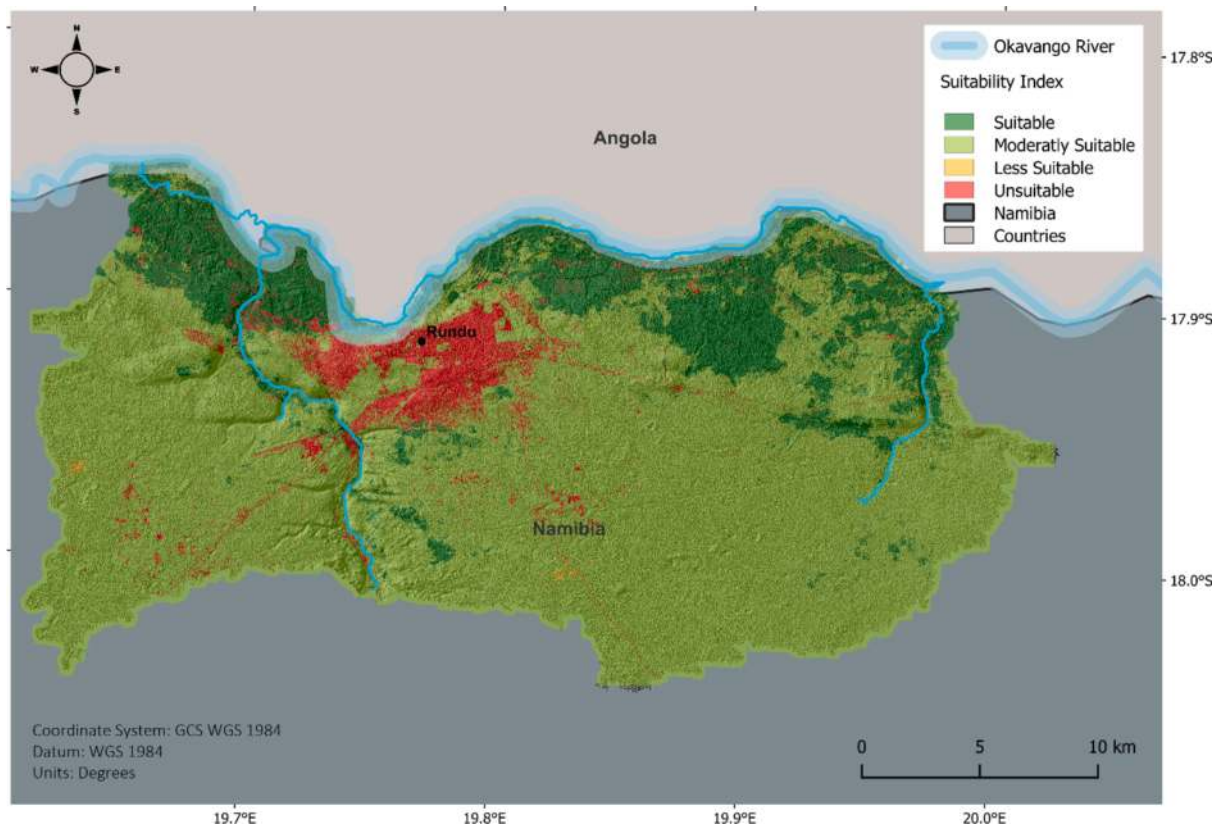


Fig. 5. Leguminous Crop suitability map using AHP method.

crop yield. Most of this class's coverage is within 2 km distance from the river and its tributaries, influencing soil moisture contents to be high and anticipating potentially higher nutrient contents. This becomes clear when examining Soil Nitrogen and Phosphorus classes, as defined in Fig. 4. Soil CEC and soil organic carbon seem to be moderately low in the moderately suitable class. However, two of the three highest weighted variables (i.e., distance to river and precipitation) seem to agree with the outcome of this class. Classes with less suitability are mainly located in the southern parts of the study area representing 10 % of the entire region. This class is characterised as having severe constraints for agricultural land use and reduced crop yield is expected. Land within this class is likely to demand a higher level of input for marginal gain. Several of the variables used to generate the suitability map agree with this classification: i) the least suitable areas have the lowest rainfall; and ii) areas located in higher elevation were the furthest away from waterbody representing the least fertile land and with the least capacity to hold soil nutrients. Unsuitable classifications were observed in 9.59 % of the study site and these areas have the highest limitations to practice leguminous crop cultivation with unsustainable use and require infeasible inputs to increase productivity. As seen in Fig. 5, these areas are mainly concentrated in the built-up area and within the water body. This is due to the unavailability of land use for farming. In addition, it is interesting to note that soil fertility classification is similar to the final suitability map.

3.8. Models' performance

In evaluating crop suitability classification models, comprehensive metrics were used to assess performance (Fig. 7), including Accuracy, Precision, Recall, F1 Score, and ROC AUC. The Random Forest Classifier and XGB Classifier demonstrated exceptional performance in all metrics. The Random Forest Classifier exhibited exceptional accuracy (0.97), precision (0.96), and recall (0.97), as well as an F1 Score and ROC AUC

of 0.97 and 0.99, respectively. These results indicate its remarkable capacity to accurately categorise both positive and negative cases with minimal error. The XGB Classifier exhibited a superior degree of precision (0.98) and accuracy (0.96) alongside a moderately lower recall (0.93), indicating a heightened emphasis on curbing false positives, while retaining elevated true positive count. The high value of ROC AUC (0.99) for both models provides evidence of their discriminatory power.

In contrast, the logistic regression had a moderate precision (0.76), a good balance between precision (0.73) and recall (0.83), which can be very important when trying to identify the most positive cases. Both classifiers had a balanced performance with high precision and recall, indicating their effectiveness in general classification tasks. The Gradient Boosting Classifier has a precision score of 0.97, and a recall score of 0.89. The Decision Tree Classifier has an accuracy score, a precision score, and a recall score all around 0.93. The Gradient Boosting Classifier and the Decision Tree Classifier both have strong classification abilities. The Gradient Boosting Classifier has a precision score of 0.97 and a recall score of 0.89, while the Decision Tree Classifier has accuracy, precision, and recall scores all around 0.93. The models also have different ROC AUC values, ranging from 0.83 to 0.99. This confirms their strong ability to distinguish between different aspects of the classification task.

Analysing the various crop suitability models yields relevant results that may be differentiated by performance indicators such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R²). These metrics indicate the accuracy of each model and its ability to highlight variations in the dataset. The ensemble models perform satisfactorily, especially the Random Forest Regressor and Extra Trees Regressor. The findings demonstrate their ability to capture complex, perhaps non-linear relationships in the data, as indicated by the lowest RMSEs (Fig. 9) and high R² values (Fig. 8). Ridge regression and linear regression have only moderate results, possibly due to their limitations in the representation of complex and nonlinear interactions. The slight

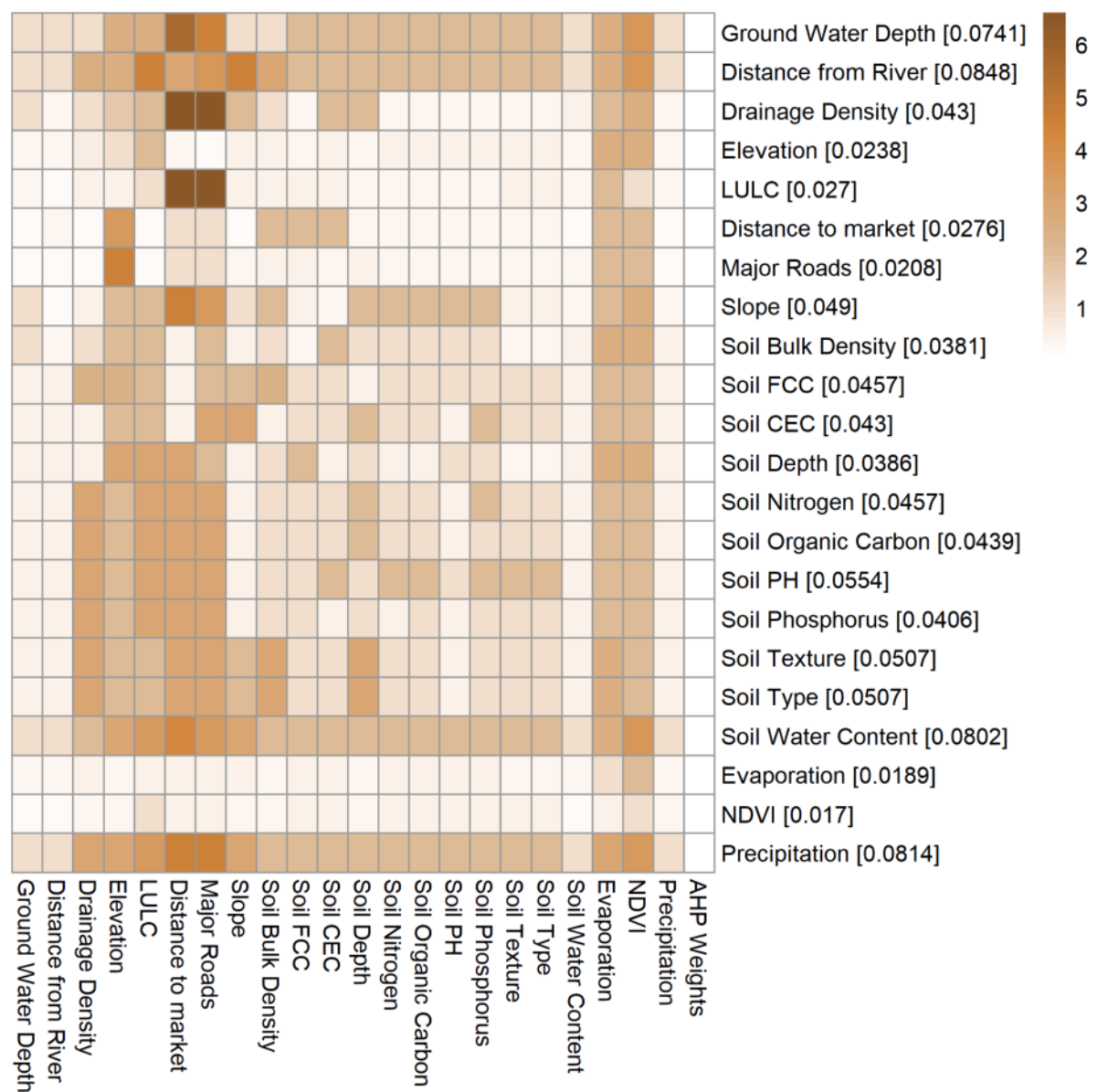


Fig. 6. AHP criteria ranking and weights.

variations in performance among the ensemble models can be attributed to differences in their methods of model construction, error minimisation, and sparse data management.

The tree ensemble models are clearly the better choice, which is an indication of their robustness and effectiveness in environmental modeling. This comparison shows the importance of selecting appropriate modeling methods depending on the characteristics of the dataset and the complexity of the relationships within it.

3.9. Covariate importance

The feature importance analysis of the crop suitability model (Fig. 10) provides significant insight into the predictors that are critical for determining crop suitability using NDVI to assess vegetation health. Feature importance scores in XGBoost reflect how each variable contributes to the model’s ability to distinguish between suitability classes. These scores are indicators of the influence each variable has on predicting suitability classes, providing insights into the variable’s relative effect on land suitability as determined by the model. The analysis shows a clear hierarchy of feature importance, with “Fertility Capacity Class 2”

as the top factor, with an importance score of 0.7884. According to the model, the fertility of the soil is the most important factor in determining crop suitability.

Other features also contribute to the assessment of crop suitability, but to a lesser extent. The second most influential feature is the level of soil nitrogen in Band 1, which has an importance score of 0.0764. The significance of soil physics in agricultural output is emphasised by presenting ‘Soil Texture’, and its corresponding classification of ‘Soil Texture class 9’ (soil_texture_9.0). The examination also highlights the importance of diverse soil types in the study, yielding importance scores of 0.0203 and 0.0177, respectively.

The suitability of crops is significantly affected by the specific soil classification of Cambisol (importance score: 0.0156), Fluvic Arenosol (0.0114), and Technosol (0.0087). Additionally, accessibility and logistics may impact crop cultivation viability, as Distance to Market is identified as a significant factor (importance score: 0.0124).

Soil fertility, especially nitrogen content, is crucial for crop suitability. This is emphasised by the presence of Soil Nitrogen Band 2, which has an importance score of 0.0076. Additionally, the soil’s cation exchange capacity, with a score of 0.0055, supports crop growth,

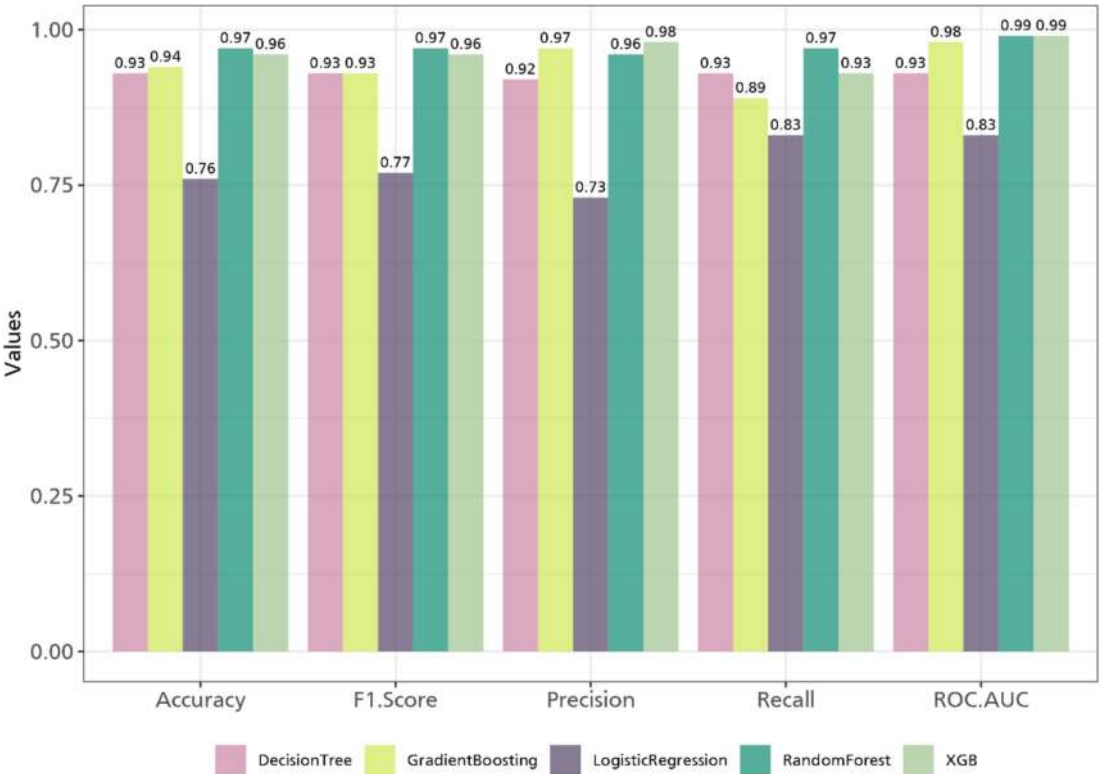


Fig. 7. Classification Metrics.

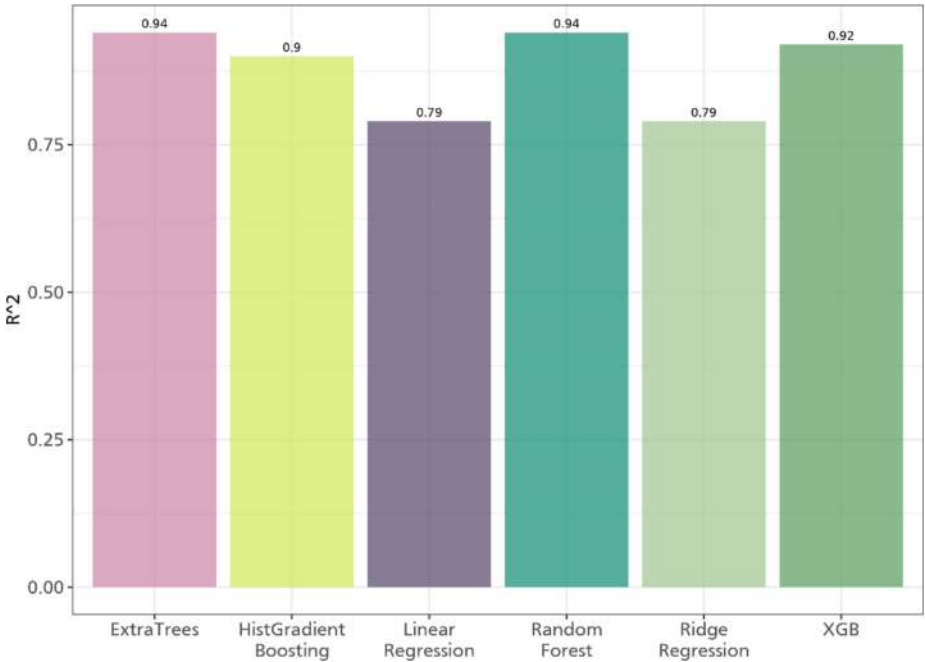


Fig. 8. Regression Metrics (R^2).

although it is less influential than other features. This analysis highlights that soil characteristics, such as fertility and texture, play a crucial role in determining crop suitability. Factors such as proximity to markets and specific soil types also contribute to this determination. The prevalence of specific aspects, particularly soil fertility classifications, highlights the complexity of evaluating land suitability for agriculture.

Fig. 11 shows the various soil and environmental factors with a significant impact on the crop production potential according to the

cropland classification. The highest-ranked feature is 'Fertility Capacity Class 1', with an importance score of 0.257. Soil fertility plays a crucial role in agricultural decision-making and the nutrient profile and soil health are critical factors in the feasibility of crop cultivation.

Specific soil classifications, such as Technosol and Cambisol, are more suitable for agriculture than others. Technosol has an importance score of 0.128, while Cambisol scores 0.033. These soil types have different characteristics that are important for crop growth, such as

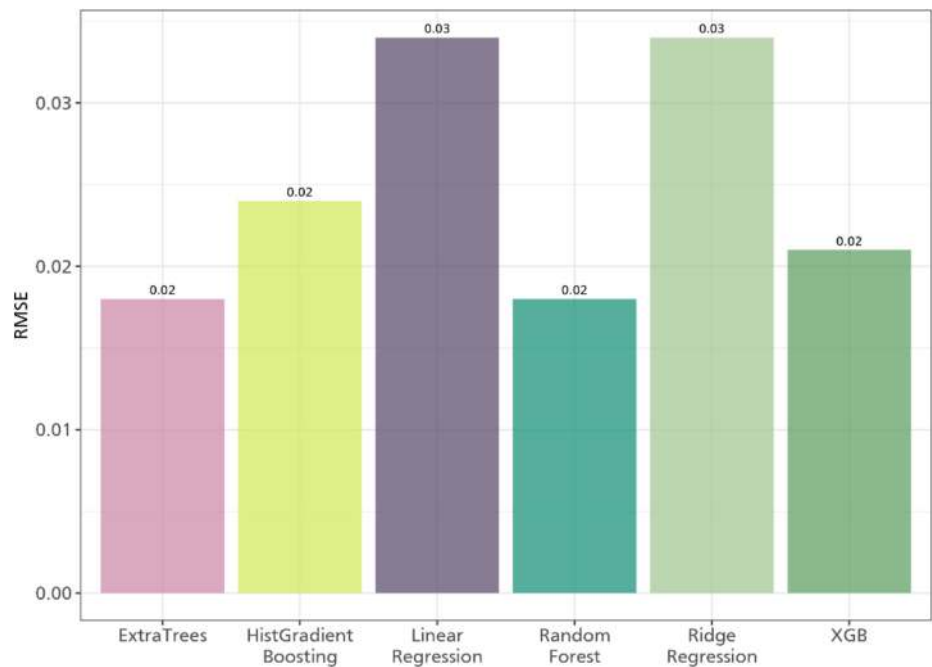


Fig. 9. Regression Metrics (RMSE).

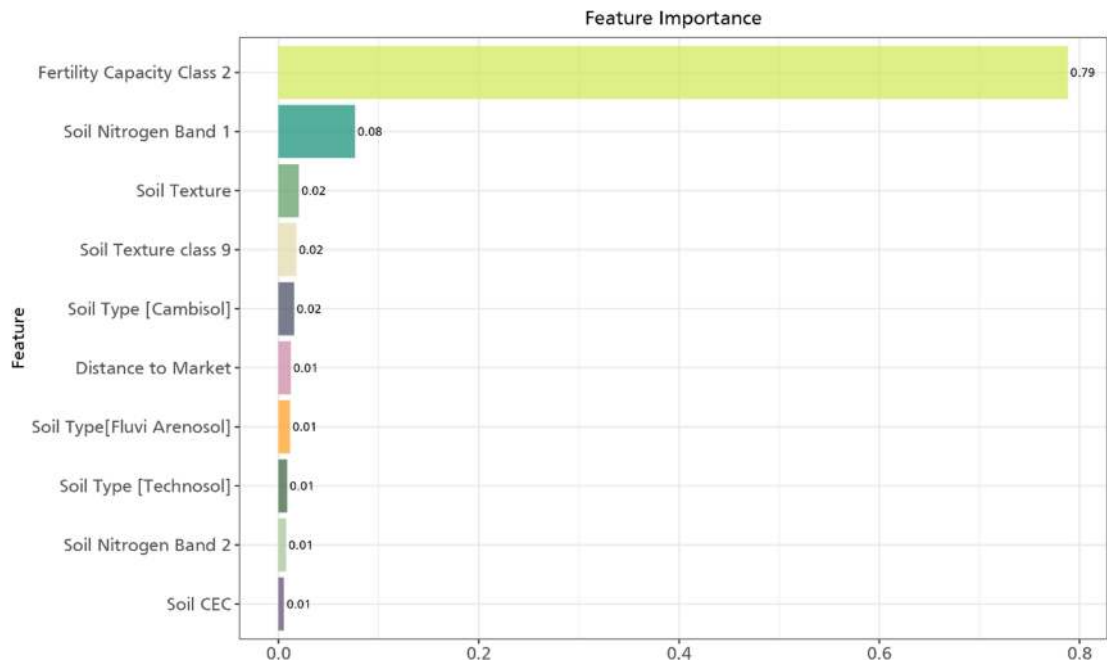


Fig. 10. Feature Importance (XGB Regressor).

water retention, nutrient availability, and structural stability.

Soil physical properties have a significant influence on plant growth. Bulk density and texture are particularly important, with importance scores of 0.085 and 0.075, respectively. These factors affect root penetration, water and air movement, and overall soil health. Soil chemistry is crucial in determining land suitability for agriculture due to the availability and presence of essential nutrients, such as phosphorus and nitrogen. These nutrients play a critical role in plant growth and crop yields, as indicated by their respective importance scores ranging from 0.063 to 0.066.

The model shows that elevation affects agricultural suitability (importance score of 0.035). This means that altitude-influenced

geographic and climatic conditions impact the ability to carry out agricultural activities. Considering elevation points to a wider environmental context, where temperature, precipitation patterns, and potential frost risks related to altitude can impact crop viability.

The regression and classification models show how various soil properties and environmental factors are interdependent in determining the suitability of land for agriculture. The research provides insight into the main factors that determine agricultural land suitability. This underscores the complexity of agricultural ecosystems and emphasises the need for a multifaceted approach to managing and assessing land for crop production. These findings can support effective and sustainable use of land resources, aiding in agricultural planning and policymaking.

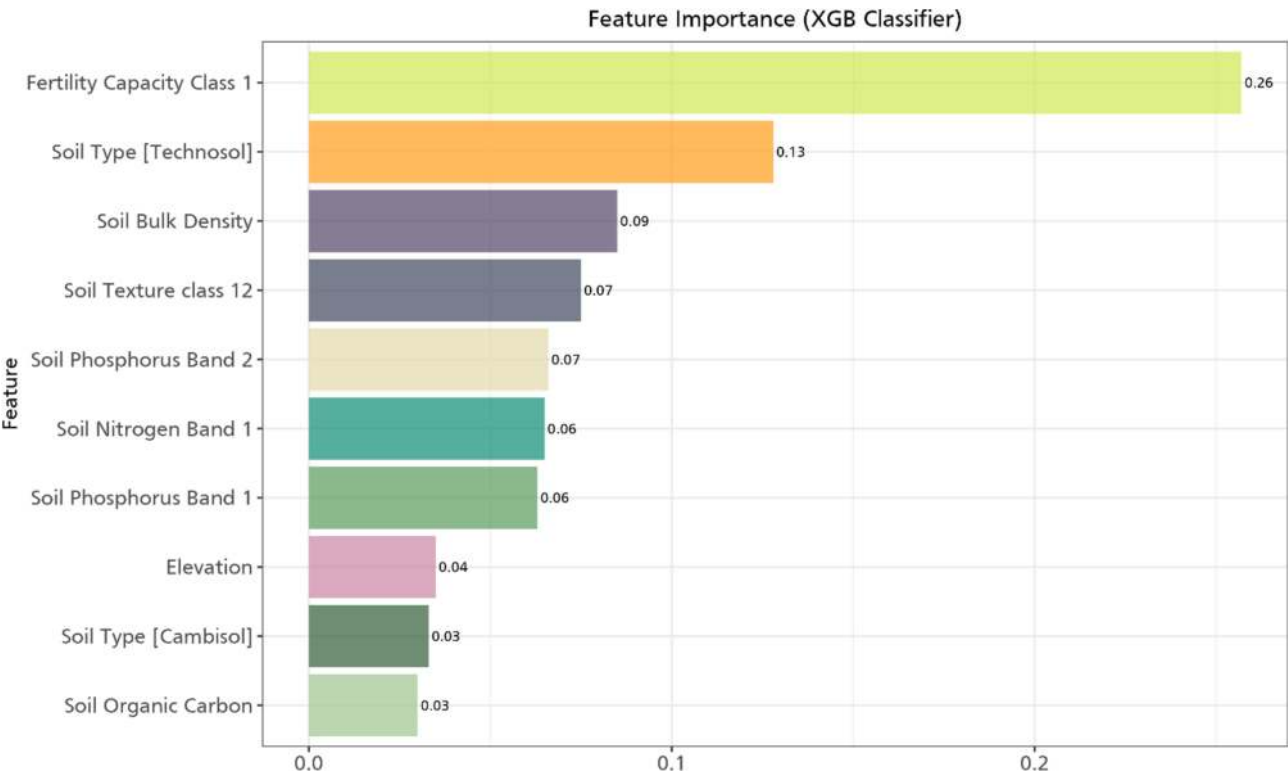


Fig. 11. Feature Importance (XGB Classifier).

3.10. Insights from logistic regression and OLS analysis

The logistic regression revealed correlations between environmental, geographical, and soil parameters and their impact on agricultural land

use (categorizing land as cropland).

Increased levels of precipitation (−0.5597) and evaporation (−0.4989) have a negative impact on the suitability of land for agriculture, indicating the influence of climatic conditions. The negative

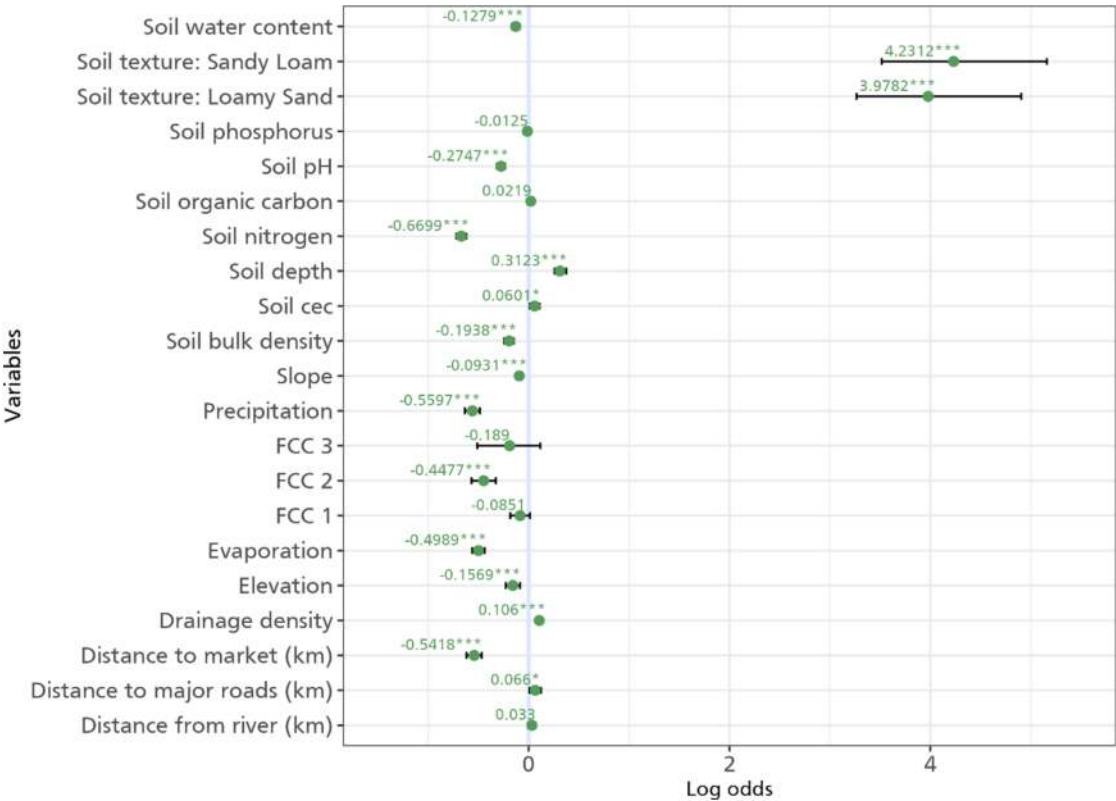


Fig. 12. Cropland Regression Analysis.

coefficient for precipitation indicates that crop production is currently practiced in areas with limited rainfall. Similarly, geographical aspects like Distance to Market (-0.5418) and Distance to Major Roads (-0.066) show negative relationships, indicating that increased remoteness reduces the likelihood of a parcel being cropland. The findings for Elevation (-0.1569) and Slope (-0.0931) underscore the preference for flatter and lower lands in agriculture, as steep and elevated terrains are less likely to be cultivated.

Soil Characteristics and Fertility Capacity: The regression analysis in Fig. 12 displays significant variations in how soil properties affect cropland probability. Soil Bulk Density and Soil pH, with their negative impacts, suggest that denser and more acidic soils are less favorable for agriculture. Conversely, positive coefficients for Soil CEC and Soil Depth indicate their beneficial roles. The notable negative coefficient for Soil Nitrogen (-0.6699) points to a complex relationship with land use, potentially influenced by factors such as nutrient management practices. Interestingly, specific soil textures, particularly Sandy Loam and Loamy Sand, with high positive coefficients (4.2312 and 3.9782), emerge as highly conducive for agricultural use. The Fertility Capacity Classes (FCC) further illuminate the nuances of soil fertility on land use, with FCC 2 showing a distinct negative effect. This highlights the variability in agricultural suitability across different fertility levels.

The OLS regression analysis (Fig. 13), with NDVI as the outcome variable, explores how different environmental, geographical, and soil factors impact vegetation health according to NDVI values.

In terms of environmental and geographical factors, evaporation and precipitation exhibit negative correlations with NDVI, with coefficients of -0.0021 and -0.0045 , correspondingly. These findings suggest that increased levels of evaporation and precipitation are connected to decreased NDVI values, which may signal reduced vegetative growth in regions with more severe hydrological conditions. The positive coefficient for Distance to Market (0.0096) is particularly interesting from a

geographical perspective, implying that areas situated farther away from markets may exhibit a higher NDVI, possibly due to less disturbance arising from anthropogenic activities. The slight positive effect of Distance from River (0.0005) and the significant positive impact of Drainage Density (0.0059) emphasizes the significance of water availability and drainage in vegetation support.

Furthermore, Topographical and Soil Factors show negative associations with NDVI. Elevation and Slope coefficients are -0.007 and -0.0013 , respectively, indicating that steeper slopes and higher elevations may hinder vegetative growth. Soil characteristics play a crucial role in vegetation health. Soil Bulk Density and Soil pH have negative effects, implying that denser soils and more acidic conditions could hinder vegetation growth. In contrast, Soil CEC displays a significant negative relationship (-0.0128), reflecting complex interactions between soil chemistry and plant growth. On the other hand, Soil Nitrogen and Soil Organic Carbon, exhibit positive coefficients (0.0201 and 0.0033), highlighting their role in promoting vegetation health.

Fertility Capacity Classes (FCC) and Soil Texture: The beneficial coefficients for FCC 1 and FCC 2 (0.0021 and 0.0075) suggest that these fertility levels promote higher NDVI values. Nevertheless, FCC 3 exhibits a negative impact (-0.0031), indicating irregularity in the influence of fertility on vegetation. The substantial negative coefficients for soil textures, such as Sandy Loam and Loamy Sand (-0.0307 and -0.015), indicate that these textures may be less conducive to vegetative growth compared to other soil types.

3.11. Perspectives of subjective Analytic Hierarchy process and objective, data-driven statistical approaches

The primary goal of the study was to present a detailed comparison between the data-driven machine learning models and the expert-based AHP approach. The aim was to highlight the strengths and differences of

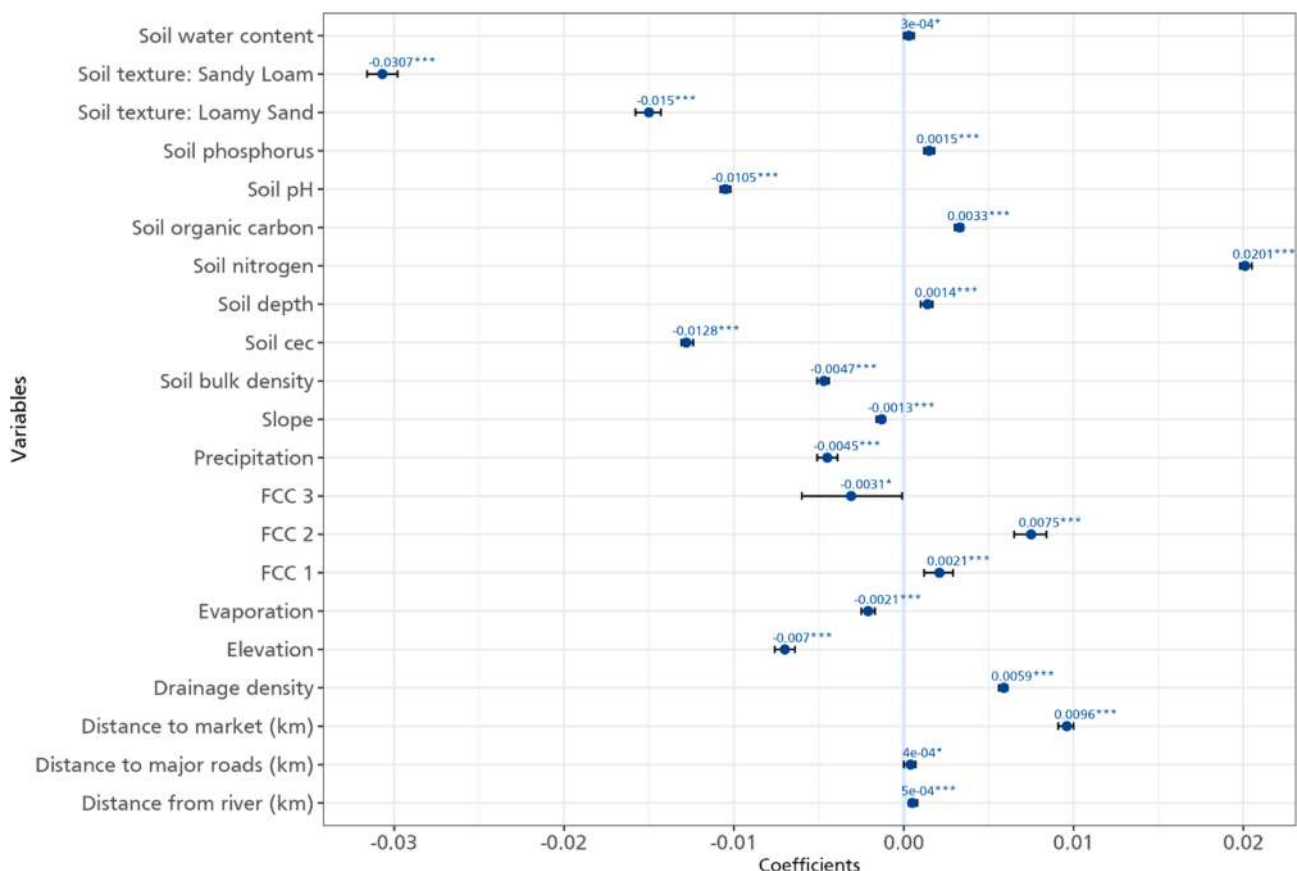


Fig. 13. NDVI Regression Analysis.

each method rather than using machine learning to add objectivity to the subjective nature of the AHP. By utilising an (AHP) approach and objective data-driven statistical methods, we successfully evaluated various variables and their level of importance in determining the agricultural potential in the study area. AHP criteria weights were influenced by characteristics of the semi-arid environment and biophysical requirements of legumes proposed to be grown in the study area. Source of water recorded the highest weights, ranging between 8–8.5 %, followed by slope, soil physical and chemical properties with weights varying from 3.8–5.5 %. Lowest weights were observed in socio-economic and vegetation criteria categories with varying weights less than 2.7 % determined by the process of expert valuation. The data-driven statistical approach exhibited unconventional results, with soil fertility capacity classification being the paramount variable with an importance score of 0.7884. The subsequent significant factor (soil nitrogen band 1) recorded much lower importance, with a score of 0.0764. Additional soil physical and chemical properties, such as soil texture and types, as well as cation exchange capacity, were also discovered to have a lesser impact on predicting suitable agricultural land.

A noteworthy observation of the data-driven approach is that the model has the capability to rank categorical variables and discern the importance of individual categories while the subjective based method solely assigns weights to the variables as a whole. Nevertheless, the subjective approach has the advantage to carry out a weighted ranking based on specific biophysical and socio-economic requirements of the crop types. In view of this, the data-driven method showed that soil physical and chemical factors were dominant, ranking at the top of the importance hierarchy. In contrast, the subjective assessment ranked water and soil factors as the highest. Section 3.9 describes the statistical method used to predict land for crop cultivation. The top ten features with the highest importance were soil physical and chemical factors, along with distance to market and elevation. These features were determined to have the greatest influence on the prediction process, while others made meaningful contributions, albeit to a lesser extent. Crop cultivation areas determined by AHP approach (section 3.7), considered source of water, as well as the physical and chemical properties of the soil, as the most influential factors. Even though subjective land suitability assessment using MCDM is most common, data-driven statistical models are gaining recognition in the discipline. This is due to the emergence of algorithms and processing capacity to handle complex data and their relations. Thus, both approaches could offer a better insight into land suitability assessment and enhance crop yields for smallholders.

4. Conclusion

Crop farming is the primary activity extracting water in the Rundu area of Namibia within the Okavango River basin. Therefore, the responsible use of this scarce resource promotes the practice of conservation agriculture to ensure the sustainable continuation of agricultural activities in the basin. The purpose of this research was to develop reference materials for agricultural extension officers, agronomists, land-use planners, and decision makers to facilitate the adoption of conservation agriculture strategies, ultimately leading to food security for smallholders. This study utilised climate, soil, hydrogeomorphology, socioeconomic and vegetation indicators to assess the suitability of potential agricultural land in the Rundu area, encompassing 722 Sq. km of rural and urban land. Two approaches of multi-criteria decision-making and data-driven multivariate models were employed to investigate suitability of agricultural land to grow legume crops, primarily relying on remotely sensed data. The study's findings are presented as follows:

- Machine learning data-driven approach avoids subjectivity and reliance on expert knowledge in the design of suitability analysis. However, it should be noted that machine learning models require

well-balanced datasets and may exhibit a higher degree of uncertainty in predicting suitability in areas where crops are not typically grown or underrepresented in the data. Furthermore, while machine learning methods provide a more robust and objective means of variable importance classification, AHP approach also demonstrated its ability to classify land suitability for leguminous crops through comprehensive and systemic process of expert consultation. Although, the AHP approach remained sensitive to expert knowledge, it was more susceptible to subjective biases.

- AHP assessment indicates that the moderately suitable class is geographically predominant, accounting for 62.77 % of the area, followed by the suitable class (17.63 %) and the less suitable and unsuitable classes (10 % and 9.59 %) of the area, respectively. Over 70 % of the suitable class coverage is observed within a 2 km radius from the river or its tributaries due to higher soil moisture and potentially higher nutrient levels found in soil types near rivers.
- Machine learning approach showed soil physical and chemical properties as the most important features to predict suitability. Among these properties, soil fertility capability class had the highest importance score as predictor for suitability of leguminous crops. The favourable condition for crop land use is established by the positive relationship between sandy loam and loamy sand textures. On the other hand, environmental and geographical factors, including evaporation, precipitation, elevation, slope, distance to market and proximity to major roads, are negatively associated with the probability of land being classified as suitable.

Future studies should focus on improvements of suitability analyses through remote sensing by incorporating extensive in-situ datasets such as soil moisture content, soil nutrients, as well as socio-economic variables such as labour income, financial capacity of smallholders, and accessibility to farming technology, among others. In developing sub-Saharan countries such as Namibia, Botswana and Angola, facing data scarcity, adopting an alternative solution such as reanalysis data will improve accuracy in crop suitability assessments. Additionally, machine learning results could further be used as inputs in AHP to enhance the analysis and remove subjective biases. The feature importance findings based on data-driven method could be used to determine the weights of criteria for the subjective approach and evaluate changes in the crop suitability map. These methods offer improved understanding into land suitability modelling through substitution of the subjective criteria weights in the AHP by a more objective, data-driven approach. Such an approach will enhance the framework for sustainable land and water resource management which could provide reliable tool for agricultural planning and environmental conservation.

This study examined the challenges and factors fundamental to the process of identifying suitable land for agricultural purposes in a comprehensive manner, primarily relying on reanalysis data as the most viable option in regions with limited data availability. The application of our proposed methodology to identify agricultural land can serve as a valuable decision-making tool for planning and assessment of land suitability. This approach could assist in optimal resource allocation, the development of irrigation strategies, the implementation of conservation agricultural practices and guiding future agricultural expansions while promoting sustainable land management.

CRedit authorship contribution statement

Kaleb Gizaw Negussie: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Bisrat Haile Gebrekidan:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Daniel Wyss:** Writing – review & editing, Supervision, Conceptualization. **Martin Kappas:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors acknowledge the Deutsche Zentrum für Luft- und Raumfahrt (DLR) for supplying the high-resolution TanDEM-X Digital Elevation Model datasets. Special thanks to Ms. Marina Coetzee for contributing soil profile data used in the research. Our gratitude goes to Dr. Ibo Zimmermann and Dr. Miguel Vallejo Orti for their feedback on the study. Gebrekidan acknowledge funding from the OneCGIAR Initiative on Excellence in Agronomy (INV-005431), as well as from the Guiding Acid Soil Management Investments in Africa (GAIA) project (Grant no: INV-029117), supported by the Bill & Melinda Gates Foundation (BMGF).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jag.2024.104284>.

Data availability

Data will be made available on request.

References

- Adem, A.A., Yibeltal, M., Mhiret, D.A., Tilahun, S.A., Zimale, F.A., Worqlul, A.W., Enku, T., Teferi, E.T., 2023. Identification of suitable land for supplemental surface irrigation in semi-arid areas of North-western Ethiopia. *Remote Sens. Appl.: Soc. Environ.* 29, 100899. <https://doi.org/10.1016/j.rsase.2022.100899>.
- Ahmadi, F.F., Layegh, N.F., 2015. Integration of artificial neural network and geographical information system for intelligent assessment of land suitability for the cultivation of a selected crop. *Neural Comput. Appl.* 26, 1311–1320. <https://doi.org/10.1007/s00521-014-1801-z>.
- Ammar, A., Riksen, M., Ouassar, M., Ritsema, C., 2016. Identification of suitable sites for rainwater harvesting structures in arid and semi-arid regions: a review. *Soil Water Conserv. Res.* 4, 108–120. <https://doi.org/10.1016/j.iswcr.2016.03.001>.
- Arab, S.T., Ahamed, T., 2022. Land suitability analysis for potential vineyards extension in Afghanistan at regional scale using remote sensing datasets. *Remote Sens. (basel)* 14, 4450. <https://doi.org/10.3390/rs14184450>.
- Atlas of Namibia Team, 2022. *Atlas of Namibia: its land, water and life*. Namibia Nature Foundation, Windhoek, Namibia.
- Awala, S.K., Hove, K., Wanga, M.A., Valombola, J.S., Mwandemele, O.D., 2019. Rainfall trend and variability in semi-arid northern Namibia: Implications for smallholder agricultural production. *Welwitschia Int. J. Agric. Sci.* 1. <https://doi.org/10.32642/wijas.v1i0.1364>.
- Bhullar, A., Nadeem, K., Ali, R.A., 2023. Simultaneous multi-crop land suitability prediction from remote sensing data using semi-supervised learning. *Sci. Rep.* 13, 6823. <https://doi.org/10.1038/s41598-023-33840-6>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Brown, C., Wamunyima, D., Kapere, A., Mwazi, M., Baker, A., Mujetenga, C., Mukuahima, E., Haub, O., Groeneveld, C., Tarr, P., Mendelsohn, J., Jones, B., Barnes, J., Tagg, J., Stuart-Hill, G., 2010. Land use planning framework for the Kavango region of Namibia within the Okavango River basin. *Namibia Nature Foundation and Ministry of Lands & Resettlement, Windhoek*.
- Cengiz, T., Akbulak, C., 2009. Application of analytical hierarchy process and geographic information systems in land-use suitability evaluation: a case study of Dümrek village (Çanakkale, Turkey). *Int. J. Sustain. Dev. World Ecol.* 16, 286–294. <https://doi.org/10.1080/13504500903106634>.
- Chaudhary, B., Tripathi, M.K., Bhandari, H.R., Pandey, S.K., Meena, D.R., Prajapati, S.P., 2015. Evaluation of sunnhemp (*Crotalaria juncea*) genotypes for high fibre yield. *Indian J. Agri Sci* 85, 850–853. <https://doi.org/10.56093/ijas.v85i6.49266>.
- Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Presented at the KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Francisco California USA, pp. 785–794. doi: 10.1145/2939672.2939785.
- Coetzee, M.E., 2020. Updated soil map and soil profile database of Namibia. *Commission, T.P.O.R.B.W. (Ed.), 2011. Cubango-Okavango River Basin Transboundary Diagnostic Analysis. OKACOM, Maun*.
- Cook, B., Pengelly, B., Schultze-Kraft, R., Taylor, M., Burkart, S., Cardoso Arango, J., González Guzmán, J., Cox, K., Jones, C., Peters, M., 2020. Tropical Forages: An interactive selection tool.
- Copernicus Climate Change Service, 2019. ERA5-Land monthly averaged data from 2001 to present. doi: 10.24381/CDS.68D2BB30.
- De Pauw, E., Coetzee, M.E., 1998. *Production of an Agro-Ecological Zones map of Namibia*. AGRICOLA, Windhoek, Namibia.
- Didan, K., 2021. MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V061. doi: 10.5067/MODIS/MOD13Q1.061.
- Dolsirittigul, P., Pichitlamken, J., Jaijit, S., 2022. Determination of land suitability for rice cultivation using random forest analysis in central region Thailand, in: *2022 3rd International Conference on Big Data Analytics and Practices (IBDAP)*. Presented at the 2022 3rd International Conference on Big Data Analytics and Practices (IBDAP), IEEE, Bangkok, Thailand, pp. 47–51. doi: 10.1109/IBDAP55587.2022.9907329.
- Dragičević, N., Karleuša, B., Ožanić, N., 2019. Different approaches to estimation of drainage density and their effect on the erosion potential method. *Water* 11, 593. <https://doi.org/10.3390/w11030593>.
- Drobne, S., Liseč, A., 2009. Multi-attribute decision analysis in GIS: weighted linear combination and ordered weighted averaging. *Informatica* 33.
- EPSMO-BIOKAVANGO Eflows Team, 2009. Hydrology report: Data and models (EPSMO/BIOKAVANGO Okavango basin environmental flows assessment project No. 05/2009). OKACOM, Maun, Botswana.
- Fan, J., McConkey, B., Wang, H., Janzen, H., 2016. Root distribution by depth for temperate agricultural crops. *Field Crop Res* 189, 68–74. <https://doi.org/10.1016/j.fcr.2016.02.013>.
- FAO, 1976. *A framework for land evaluation*. Soils Bulletin, Food and Agriculture Organization of the United Nations, Rome.
- FAO, 1993. *Guidelines for land-use planning*. FAO development, FAO, Rome.
- FAO and UN Water, 2021. Progress on level of water stress: Global status and acceleration needs for SDG indicator. FAO and UN Water. doi: 10.4060/cb6241en.
- FAO, 1984. Assessment of potential land suitability, Namibia. Land regions and land-use potential (AG: DP/NAM/78/004 No. Technical Report 2). Food and Agricultural Organization of the United Nations, Rome.
- FAO, 2022. *Crop Ecological Requirements Database (ECOCROP)*.
- FAO (Ed.), 2023. *Urbanization, agrifood systems transformation and healthy diets across the rural-urban continuum, The state of food security and nutrition in the world*. FAO, Rome. doi: 10.4060/cc3017en.
- Fortunato, A., Enciso, S., 2023. Food for growth: A diagnostics of Namibia's agriculture sector. Center for International Development at Harvard University, CID Research Fellow and Graduate Student: Working Paper.
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci. Data* 2, 150066. <https://doi.org/10.1038/sdata.2015.66>.
- Ganesan, M., Andavar, S., Raj, R.S.P., 2021. Prediction of land suitability for crop cultivation using classification techniques. e21200483 Braz. Arch. Biol. Technol. 64. <https://doi.org/10.1590/1678-4324-2021200483>.
- Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. *Mach. Learn.* 63, 3–42. <https://doi.org/10.1007/s10994-006-6226-1>.
- Githae, E.W., Gachene, C.K.K., Njoka, J.T., 2011. Soil physicochemical properties under *Acacia senegal* varieties in the dryland areas of Kenya. *Afr. J. Plant Sci.*
- Glur, C., 2024. {ahp}: A package for Analytical Hierarchy Process [R package].
- Guretzky, J.A., Moore, K.J., Burras, C.L., Brummer, E.C., 2004. Distribution of legumes along gradients of slope and soil electrical conductivity in pastures. *Agron. J.* 96, 547–555. <https://doi.org/10.2134/agronj2004.5470>.
- Hag Husein, H., Lucke, B., Bäuml, R., Sahwan, W., 2021. A contribution to soil fertility assessment for arid and semi-arid lands. *Soil Syst.* 5, 42. <https://doi.org/10.3390/soilsystems5030042>.
- Hao, Q., Zhang, T., Cheng, X., He, P., Zhu, X., Chen, Y., 2024. GIS-based non-grain cultivated land susceptibility prediction using data mining methods. *Sci. Rep.* 14, 4433. <https://doi.org/10.1038/s41598-024-55002-y>.
- Harmoney, K.R., 1999. Legume establishment and persistence at sites varying in landscape position, grazing method, and soil characteristics (Doctor of Philosophy). Iowa State University, Digital Repository, Ames. doi: 10.31274/rtd-180813-13411.
- Hengl, T., Gupta, S., 2019. Soil water content (volumetric %) for 33kPa and 1500kPa suction predicted at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution. doi: 10.5281/ZENODO.2629589.
- Hengl, T., Miller, M.A.E., Krizan, J., Shepherd, K.D., Sila, A., Kilibarda, M., Antonijević, O., Glušica, L., Dobermann, A., Haefele, S.M., McGrath, S.P., Acquah, G. E., Collinson, J., Parente, L., Sheykhoum, M., Saito, K., Johnson, J.-M., Chamberlin, J., Silats, F.B.T., Yemefack, M., Wendt, J., MacMillan, R.A., Wheeler, I., Crouch, J., 2021. African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning. *Sci. Rep.* 11, 6130. <https://doi.org/10.1038/s41598-021-85639-y>.
- IFAD, 2021. *What can smallholder farmers grow in a warmer world? Climate change and future crop suitability in East and Southern Africa*, Rome.
- Integrated Food Security Phase Classification, 2023. *IPC acute food insecurity analysis*. IPC, Namibia.
- IPCC (Ed.), 2014. *Climate change 2014: Impacts, adaptation, and vulnerability*. Part A: Global and sectoral aspects. Contribution of working group II to the fifth assessment report of the Intergovernmental Panel on Climate Change, Field, C.B., V.R. Barros, D. J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.). ed. Cambridge University Press, New York, NY.
- Ismail, M., Krimissa, S., Namous, M., Hitiou, A., Abdelrahman, K., Fnais, M.S., Lhissou, R., Eloudi, H., Faouzi, E., Benabdellouahab, T., 2023. Assessment of soil

- suitability using machine learning in arid and semi-arid regions. *Agronomy* 13, 165. <https://doi.org/10.3390/agronomy13010165>.
- Jacobi, I., 2008. Crop production. Joint Presidency Committee, Windhoek.
- James, G., Witten, D., Hastie, T., Tibshirani, R. (Eds.), 2013. *An Introduction to Statistical Learning: with Applications in R*, Springer Texts in Statistics. Springer, New York.
- Javaid, M.M., Bhan, M., Johnson, J.V., Rathinasapathi, B., Chase, C.A., 2015. Biological and chemical characterizations of allelopathic potential of diverse accessions of the cover crop Sunn Hemp. *J. Amer. Soc. Hort. Sci.* 140, 532–541. <https://doi.org/10.21273/JASHS.140.6.532>.
- Junges, A.H., Fontana, D.C., Anzanello, R., Bremm, C., 2017. Normalized Difference Vegetation Index obtained by ground-based remote sensing to characterize vine cycle in Rio Grande do Sul, Brazil. *Ciênc. Agrotec.* 41, 543–553. <https://doi.org/10.1590/1413-70542017415049016>.
- Kaur, G., Singh, G., Motavalli, P.P., Nelson, K.A., Orlowski, J.M., Golden, B.R., 2020. Impacts and management strategies for crop production in waterlogged or flooded soils: a review. *Agron. J.* 112, 1475–1501. <https://doi.org/10.1002/agj2.20093>.
- Li, S.-X., Wang, Z.-H., Stewart, B.A., 2011. Differences of some leguminous and nonleguminous crops in utilization of soil Phosphorus and responses to Phosphate fertilizers, in: *Advances in Agronomy*. Elsevier, pp. 125–249. doi: 10.1016/B978-0-12-385531-2.00003-7.
- Lu, X., Wang, L., Pan, M., Kaseke, K.F., Li, B., 2016. A multi-scale analysis of Namibian rainfall over the recent decade – comparing TMPA satellite estimates and ground observations. *J. Hydrol. Region. Stud.* 8, 59–68. <https://doi.org/10.1016/j.ejrh.2016.07.003>.
- Marconini, M., Metz-Marconini, A., Üreyen, S., Palacios-Lopez, D., Hanke, W., Bachofer, F., Zeidler, J., Esch, T., Gorelick, N., Kakarla, A., Paganini, M., Strano, E., 2020. Outlining where humans live, the World Settlement Footprint 2015. *Sci. Data* 7, 242. <https://doi.org/10.1038/s41597-020-00580-5>.
- Mashingaidze, N., Madakadze, C., Twomlow, S., Nyamangara, J., Hove, L., 2012. Crop yield and weed growth under conservation agriculture in semi-arid Zimbabwe. *Soil Tillage Res.* 124, 102–110. <https://doi.org/10.1016/j.still.2012.05.008>.
- Mendelsohn, J., El Obeid, S., 2004. *Okavango River: the flow of a lifeline*, 1st ed. Struik, Cape Town.
- Mendelsohn, J., Jarvis, A., Roberts, C., Robertson, T., 2002. *Atlas of Namibia: A Portrait of the Land and its People*, 1st. Editio. ed. David Philip, Cape Town, South Africa.
- Mohammed, M.A.A., Khleel, N.A.A., Szabó, N.P., Szűcs, P., 2023. Modeling of groundwater quality index by using artificial intelligence algorithms in northern Khartoum State. *Sudan. Model. Earth Syst. Environ.* 9, 2501–2516. <https://doi.org/10.1007/s40808-022-01638-6>.
- Moisa, M.B., Tiye, F.S., Dejene, I.N., Gemed, D.O., 2022. Land suitability analysis for maize production using geospatial technologies in the Didessa watershed, Ethiopia. *Artif. Intell. Agric.* 6, 34–46. <https://doi.org/10.1016/j.iaia.2022.02.001>.
- Mosepele, K., L. Hambira, W., E.J. Mogomotsi, G., K. Mogomotsi, P., Moses, O., Dhlwayo, M., Makati, A., Setomba, B., 2019. Water, ecosystem dynamics and human livelihoods in the Okavango River basin (ORB): competing needs or balanced use? A review, in: *Thanjavur Chandrasekaran, P. (Ed.), Water and Sustainability*. IntechOpen. doi: 10.5772/intechopen.80554.
- Mwazi, F.N., 2006. Spatial analysis of land suitability to support alternative land uses at Excelsior resettlement project, Oshikoto region, Namibia. International Institute for Geo-information Science and Earth Observation, Enschede, The Netherlands.
- Mwinga, M., Siboleka, M., Kavezuva, C., Amadhila, F., 2018. Economy profile of Rundu, Ossue 3.
- Namibia Statistics Agency, 2011. Namibia 2011 population and housing census.
- Natekin, A., Knoll, A., 2013. Gradient boosting machines, a tutorial. *Front. Neurobot.* 7. <https://doi.org/10.3389/fnbot.2013.00021>.
- Ncube, B., 2007. *Understanding cropping systems in the semi-arid environments of Zimbabwe: options for soil fertility management*. Wageningen University, Wageningen, The Netherlands.
- NEPAD Water Centres of Excellence, 2013. Country water resource profile. Botswana.
- Oduze, A.C., Tarawali, S.A., de Haan, N.C., Iwuafor, E.N.O., Katung, P.D., Akoueguo, G.E., Amadi, A.F., Schultze-Kraft, R., Atala, T.K., Ahmed, B., Adamu, A., Babalola, A.O., Ogunwale, J.O., Alimi, A., Ewansiha, S.U., Adediran, S.A., 2004. Grain legumes for soil productivity improvement in the Northern Guinea savanna of Nigeria. *J. Food Agric. Environ.* 2, 218–226.
- Orwa, C., Mutua, A., Kindt, R., Jamnadass, R., Simons, A., 2009. *Agroforestry Database: a tree reference and selection guide*.
- Owusu, S., Mul, M.L., Ghansah, B., Osei-Owusu, P.K., Awotwe-Pratt, V., Kadyampakeni, D., 2017. Assessing land suitability for aquifer storage and recharge in northern Ghana using remote sensing and GIS multi-criteria decision analysis technique. *Model. Earth Syst. Environ.* 3, 1383–1393. <https://doi.org/10.1007/s40808-017-0360-6>.
- Özkan, B., Dengiz, O., Turan, İ.D., 2020. Site suitability analysis for potential agricultural land with spatial fuzzy multi-criteria decision analysis in regional scale under semi-arid terrestrial ecosystem. *Sci. Rep.* 10, 22074. <https://doi.org/10.1038/s41598-020-79105-4>.
- Pramanik, M.K., 2016. Site suitability analysis for agricultural land use of Darjeeling district using AHP and GIS techniques. *Model. Earth Syst. Environ.* 2, 56. <https://doi.org/10.1007/s40808-016-0116-8>.
- Rani, K., Sharma, P., Kumar, S., Wati, L., Kumar, R., Gurjar, D.S., Kumar, D., Kumar, R., 2019. Legumes for sustainable soil and crop management, in: Meena, R.S. (Ed.), *Sustainable Management of Soil and Environment*. Springer Singapore, Singapore, pp. 193–215. doi: 10.1007/978-981-13-8832-3_6.
- Republic of Namibia Office of the Prime Minister, 2023. *Namibia livelihood vulnerability assessment and analysis (VAA)*.
- Roy, J., Saha, S., 2018. Assessment of land suitability for the paddy cultivation using analytical hierarchical process (AHP): a study on Hinglo river basin, Eastern India. *Model. Earth Syst. Environ.* 4, 601–618. <https://doi.org/10.1007/s40808-018-0467-4>.
- Saaty, R.W., 1987. The analytic hierarchy process—what it is and how it is used. *Mathemat. Modell.* 9, 161–176. [https://doi.org/10.1016/0270-0255\(87\)90473-8](https://doi.org/10.1016/0270-0255(87)90473-8).
- Salgado Baptista, D.M., Farid, M., Fayad, D., Kemoe, L., Lanci, L., Mitra, P., Muehlschlegel, T., Okou, C., Spray, J., Tuitoeck, K., Unsal, F., 2022. Climate change and chronic food insecurity in sub-Saharan Africa. *Departmental Papers* 2022, 1. <https://doi.org/10.5089/9798400218507.087>.
- Sarkar, D., Saha, S., Mondal, P., 2023. Modelling agricultural land suitability for vegetable crops farming using RS and GIS in conjunction with bivariate techniques in the Uttar Dinajpur district of Eastern India. *Green Technol. Sustainability* 1, 100022. <https://doi.org/10.1016/j.grets.2023.100022>.
- Sarmadian, F., Keshavarzi, A., Rooien, A., Zahedi, G., Javadikia, H., Iqbal, M., 2014. Support Vector Machines based-modeling of land suitability analysis for rainfed agriculture. *J. Geosci. Geom.* 2, 165–171. <https://doi.org/10.12691/jgg-2-4-4>.
- Sharma, S., Agarwal, N., Verma, P., 2011. Pigeon pea (*Cajanus cajan* L.): A hidden treasure of regime nutrition. *J. Funct. Environ. Bot.* 1, 91. <https://doi.org/10.5958/j.2231-1742.1.2.010>.
- Simon, S.A., Itanna, F., 2019. *Effects of green manure cover crops on weed population, pearl millet and maize productivity under conservation agriculture in Liselo and Mashare Namibia*. University of Namibia, Namibia.
- Singh, A.K., Singh, J.B., Singh, R., Kantwa, S.R., Jha, P.K., Ahamad, S., Singh, A., Ghosh, A., Prasad, M., Singh, S., Singh, S., Prasad, P.V.V., 2023. Understanding soil Carbon and Phosphorus dynamics under grass-legume intercropping in a semi-arid region. *Agronomy* 13, 1692. <https://doi.org/10.3390/agronomy13071692>.
- Strohbach, B.J., 2018. Making more of vegetation classification results: A livestock farming suitability index as tool for land-use planning in Namibia. *phyto* 48, 7–22. doi: 10.1127/phyto/2017/0182.
- Sys, C., van Ranst, E., Debaveye, J., 1991. Land evaluation. Part 2: Methods in land evaluation, Agricultural Publications no. 7. ed. General Administration for Development Cooperation, Brussels.
- Tashayo, B., Honarbakhsh, A., Akbari, M., Eftekhari, M., 2020. Land suitability assessment for maize farming using a GIS-AHP method for a semi- arid region, Iran. *J. Saud. Soc. Agric. Sci.* 19, 332–338. <https://doi.org/10.1016/j.jssas.2020.03.003>.
- The World Bank Group, 2021. *Climate risk profile: Namibia*.
- Tsubo, M., Walker, S., Ogindo, H.O., 2005. A simulation model of cereal-legume intercropping systems for semi-arid regions. *Field Crop Res* 93, 10–22. <https://doi.org/10.1016/j.fcr.2004.09.002>.
- USDA, 1999. *Sunn Hemp: A cover crop for Southern and tropical farming systems*.
- Waleed, M., Ahmad, S.R., Javed, M.A., Samiullah, S., 2020. Identification of irrigation potential areas, using multi-criteria analysis in Khyber District, Pakistan. *Environ. Sci. Pollut. Res.* 27, 39832–39840. <https://doi.org/10.1007/s11356-020-08967-y>.
- Wessel, B., Huber, M., Wohlfart, C., Marschall, U., Kosmann, D., Roth, A., 2018. Accuracy assessment of the global TanDEM-X Digital Elevation Model with GPS data. *ISPRS J. Photogramm. Remote Sens.* 139, 171–182. <https://doi.org/10.1016/j.isprsjprs.2018.02.017>.
- Worqlul, A.W., Jeong, J., Dile, Y.T., Osorio, J., Schmitter, P., Gerik, T., Srinivasan, R., Clark, N., 2017. Assessing potential land suitable for surface irrigation using groundwater in Ethiopia. *Appl. Geogr.* 85, 1–13. <https://doi.org/10.1016/j.apgeog.2017.05.010>.
- Yalew, S.G., Van Griensven, A., Mul, M.L., Van Der Zaag, P., 2016. Land suitability analysis for agriculture in the Abay basin using remote sensing, GIS and AHP techniques. *Model. Earth Syst. Environ.* 2, 101. <https://doi.org/10.1007/s40808-016-0167-x>.
- Yan, Z., Li, Z., Li, P., Zhao, C., Xu, Y., Cui, Z., Sun, H., 2023. Spatial and temporal variation of NDVI and its driving factors based on geographical detector: A case study of Guanzhong plain urban agglomeration. *Remote Sens. Appl.: Soc. Environ.* 32, 101030. <https://doi.org/10.1016/j.rsase.2023.101030>.
- Zadeh, L.A., 1965. Fuzzy sets. *Inf. Control* 8, 338–353. [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X).
- Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.-E., Xu, P., Ramoino, F., Arino, O., 2022. *ESA WorldCover 10 m 2021 v200*. doi: 10.5281/ZENODO.7254221.
- Zhang, X.-N., Yang, X.-D., Li, Y., He, X.-M., Lv, G.-H., Yang, J.-J., 2018. Influence of edaphic factors on plant distribution and diversity in the arid area of Xinjiang, Northwest China. *Arid Land Res. Manage.* 32, 38–56. <https://doi.org/10.1080/15324982.2017.1376004>.
- Zimmermann, I., 2023. Leguminous crop suitability requirement - Personal interview.