ISPRS Journal of Photogrammetry and Remote Sensing 204 (2023) 117–130

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Fine-scale characterization of irrigated and rainfed croplands at national scale using multi-source data, random forest, and deep learning algorithms

Kudzai S. Mpakairi [a], Timothy Dube [a], Mbulisi Sibanda [b], Onisimo Mutanga [c]

[a] Institute of Water Studies, Department of Earth Sciences, University of the Western Cape, Bellville, Cape Town, South Africa
[b] Department of Geography, Environmental Studies and Tourism, Faculty of Arts and Humanities, University of the Western Cape, Bellville, Cape Town, South Africa
[c] Discipline of Geography and Environmental Science, School of Agricultural Earth and Environmental Sciences, University of KwaZulu-Natal, Scottsville, Pietermaritzburg, South Africa

ARTICLE INFO

Keywords:
Cropland areas
Earth observation
Large scale
South Africa
National scale mapping
Wall-to-wall mapping

ABSTRACT

Knowledge of the extent and distribution of irrigated and rainfed croplands is critical in providing the necessary baseline data for enhancing agricultural efficiency and making informed policy decisions. Accurately identifying and mapping irrigated and rainfed croplands can hasten the attainment of Sustainable Development Goals (SDGs) 1 and 2, aimed at reducing poverty and hunger, respectively. However, traditional methods employed to identify and map cropland areas are expensive and require substantial labour, particularly in extensive environments. As a result, this study presents a comprehensive and spatially explicit methodological framework for identifying and mapping national-scale irrigated and rainfed croplands in South Africa. This framework leverages low-cost earth observation technologies (Sentinel-2 MSI) and employs highly accurate classification algorithms, namely Deep Learning Neural Network (DNN) and Random Forest (RF). The proposed methodology strategically integrates data from multiple sources, including public repositories (e.g., cropland data, evapotranspiration), ongoing research (e.g., land cover maps), and field data, to enhance the accuracy and reliability of the results. The methodology begins by employing a robust random forest model to classify the study area into distinct land cover types. Leveraging the power of a deep learning neural network (DNN), the method accurately distinguishes between irrigated and rainfed croplands in South Africa. The random forest model achieved a notable classification accuracy of 0.77 when identifying the main land-use and land cover types. Meanwhile, the deep learning neural network (DNN) model achieved an accuracy of 0.71 in differentiating rainfed and irrigated croplands at a national scale. These results highlight the effectiveness of the proposed methodology in providing baseline information relevant to crop monitoring, yield forecasting, and understanding agricultural food supply systems. Furthermore, the proposed methodology has the potential to offer timely and accurate information on cropland areas and their extent which could assist in implementing targeted interventions for optimising agricultural productivity. With its potential to be upscaled to other sub-Saharan countries, this methodology enriches agricultural decision-making and plays a vital role in bolstering food security and advancing the attainment of SDGs.

1. Introduction

Agriculture plays a vital role in addressing food insecurity in most developing countries, considering how food systems are currently under pressure from population growth, climate change, pandemics, and land degradation (Baiphethi & Jacobs, 2009, Hendriks, 2014, Waldner et al., 2017). In addition to tackling food insecurity, agriculture also supports livelihoods by contributing towards socio-economic activities, as observed in 2019 when 53 % of the working population in sub-Saharan Africa was employed in the agricultural sector (Bank, 2021). Despite these contributions, 220 million people in sub-Saharan Africa are still undernourished, and less than 20 % of countries have yet accelerated agricultural productivity as committed under the Malabo Declaration (Shimeles et al., 2018, Ayanlade & Radeny, 2020). Subsequently, the first step towards improving agricultural productivity in sub-Saharan Africa involves an inventory of all cropland types (e.g., irrigated or rainfed). By conducting an inventory of cropland types, including irrigated and rainfed lands, governments in sub-Saharan Africa can understand their food supply potential and accelerate progress towards achieving SDG 1 (no poverty) and SDG 2 (zero hunger).

Most agricultural landscapes in sub-Saharan Africa are dominated by rainfed croplands while irrigated croplands constitute a small portion of the available arable land (Landmann et al., 2019, Ayanlade & Radeny, 2020). The scales at which irrigated and rainfed croplands address food insecurity are different but aligned. Irrigated croplands are usually associated with commercial farms and can potentially improve national

https://doi.org/10.1016/j.isprsjprs.2023.09.006
Received 5 May 2023; Received in revised form 18 August 2023; Accepted 8 September 2023
Available online 13 September 2023

food security, whilst rainfed croplands are typically linked with smallholder farms and can improve household food security (Baiphethi & Jacobs, 2009, Akbari et al., 2020). Consequently, understanding the magnitude of these croplands contribution towards redressing food insecurity requires a robust inventory of their productivity, size, and distribution. An inventory of the extent of irrigated and rainfed croplands will likely assist with crop monitoring and yield forecasting (de Graaff et al., 2011, Blair et al., 2018). This will also help with monitoring the expansion of agricultural landscapes into protected and sensitive ecosystems (Potapov et al., 2022).

Several attempts have been made by different institutions to derive global cropland data, using freely accessible remote sensing technologies (Pittman et al., 2010, Teluguntla et al., 2018). For example, the global cropland data are provided by the Global Food Security-support Analysis Data (GFSAD) project under NASA (Congalton et al., 2017). The GFSAD provides global cropland data at a spatial resolution of 30 m, which is adequate to detect most croplands at a national scale. Although the dataset captures most croplands, the dataset does not provide information on whether croplands are irrigated or rainfed (Congalton et al., 2017, Oliphant et al., 2022). On the other hand, existing global datasets on irrigated croplands are provided at a low spatial resolution and might not be suitable for local and national scale planning purposes (Suyker & Verma, 2009, Portmann et al., 2010). For instance, the Global Map of Irrigation Areas (GMIA) published by FAO is provided with a spatial resolution of ~10 km, which might be coarse for yield forecasting and crop monitoring in small-holder farming areas (Siebert et al., 2005, Portmann et al., 2010). Accurately assessing the extent of irrigated and rainfed croplands is crucial particularly for sub-Saharan African countries, yet data availability presents several constraints to this task at both a finer and national scale (Waldner et al., 2016, Waldner et al., 2017). The absence of reliable datasets on irrigated and rainfed croplands could result in drawing up unreliable decisions and policy initiatives that could retard agricultural productivity and potentially increasing food insecurity.

Efforts to map national-scale irrigated and rainfed croplands at higher resolution have been limited by the trade-off between spatial resolution and the extent of coverage in remotely sensed data (Duveiller & Defourny, 2010, Hu et al., 2021). Previous attempts to spatially quantify the extent of irrigated and rainfed croplands at a higher spatial resolution have mainly focused on smaller geographic extents (Suyker & Verma, 2009, Van Koppen et al., 2017). This has led to the lack of high-resolution national scale maps on the extent of irrigated and rainfed croplands which are necessary for accurately mapping and forecasting yield and crop productivity in most countries (e.g., in South Africa) (Li et al., 2022). Understanding the distribution and extent of national scale irrigated and rainfed croplands is essential for identifying areas of potential expansion or intensification of agricultural production (Hu et al., 2021, Xia et al., 2022). As a result, the lack of national-scale maps of irrigated and rainfed croplands makes it challenging to address the looming food insecurity and improve agricultural productivity in most sub-Saharan African countries (Waldner et al., 2016, Waldner et al., 2017, Useya & Chen, 2019, Useya et al., 2019).

In 2008, 14 million people in South Africa were estimated to be susceptible to food insecurity (Labadarios et al., 2009). However, the necessary information on the extent and distribution of irrigated and rainfed croplands, required in addressing the issue of food insecurity in South Africa, was scanty and outdated (SA, 2017, SA, 2021). The latest information available showing the extent and distribution of irrigated and rainfed croplands in South Africa was published in 2019 (DFFE, 2020), and it is outdated since the COVID-19 pandemic negatively impacted all sectors including agriculture and resulted in reduced food production and increased farm abandonment (Paganini et al., 2020). Between 2019 and 2021, the combined impact of COVID-19 and crop abandonment resulted in 17.3 % of South Africans experiencing moderate to severe food insecurity, with 7.0 % facing severe food insecurity (SA, 2021). Timely information on the cropland extent, crop yield and the available agricultural food supply could have served as a support system, guiding decision-making processes aimed at implementing effective measures to redress food insecurities. Therefore, there is an urgent need to come up with robust time efficient methods for generating accurate cropland extent information at national scales.

Producing national-scale maps requires rigorous sampling strategies that are costly and take time to complete (Schoeman et al., 2013, Waldner et al., 2017, Landmann et al., 2019, Useya et al., 2019). Owing to these constraints, data collection can be restricted to accessible areas and optimistically biased samples can be used. The trade-off between data sampling costs or using optimistically biased samples can be avoided through the use of multisource data (Millard & Richardson, 2015). Multisource data from several open-source repositories can be useful in accurately mapping extensive or complex environments. For example, Funk and Budde (2009) used phenologically tuned NDVI-based production anomaly estimates to characterise agricultural land and crop production in Zimbabwe. Schoeman et al. (2013) used multisource data to map land cover changes in South Africa at a national scale. In another study, Waldner et al. (2017) used multisource data to accurately estimate the cropland extent in South Africa. These studies demonstrate how multisource datasets can be equally reliable and relevant whenmapping croplands at a national scale.

National scale rainfed and irrigated cropland maps for countries that are expansive in areal extent require high computational power (Waldner et al., 2017, Landmann et al., 2019, Useya et al., 2019). In addition, discriminating irrigated and rainfed croplands in extensive landscapes with commonly used classifiers such as maximum likelihood classifier (MLC) or Naïve Bayes might return maps with low accuracies because of the high spectral mixing existent in heterogeneous land uses and land cover types (Zhu et al., 2017, Van Niekerk et al., 2018, Thompson et al., 2020). Recent advancements leveraging clouding computing technologies such as Google Earth Engine (GEE) and deep learning algorithms present a novel approach for accurately mapping and monitoring irrigated and rainfed croplands at higher spatial resolutions (Sun et al., 2019, Magidi et al., 2021, Saleem et al., 2021, Sun et al., 2023). Deep learning algorithms, such as deep neural networks (DNN), convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown promising results in remote sensing applications, including land cover classification and crop type mapping (Sidike et al., 2019, Sun et al., 2019, Saleem et al., 2021, Sun et al., 2023). These approaches leverage the ability of deep neural networks to automatically learn features from large amounts of remote sensing data, allowing for more accurate and detailed mapping of croplands, including their types and irrigation status (Zhu et al., 2017, Ma et al., 2019). The ability of DNN to learn from data also allows the use of the algorithm with multisource data (e.g., photo-interpreted points and secondary data) and this has increased its popularity (Liu et al., 2018). When combined with highly accurate algorithms such as random forest, the DNN has been proven to be robust and has achieved relatively higher prediction accuracies, particularly with complex and high-dimensional datasets, due to the complementary strengths of both approaches (Dong et al., 2019, Zamani Joharestani et al., 2019, Kavhu et al., 2022). Random forests can handle categorical variables and noisy data, while deep neural networks can capture intricate patterns and relationships in the data (Dong et al., 2019, Dang et al., 2021). Together, they can provide a powerful ensemble that combines the benefits of both techniques and can lead to better overall performance in various tasks, such as image recognition, natural language processing, and predictive modelling (Liu et al., 2018, Dang et al., 2021).

The main objective of this study is to create an accurate and comprehensive national scale inventory of irrigated and rainfed croplands, using multisource data. The goal is to provide reliable and up-to-date information to support decision-making in addressing food insecurity and improving agricultural productivity, particularly in sub-Saharan African countries such as South Africa. The integrated approach includes a two-step approach that involved identifying all

cultivated cropland areas with the random forest algorithm and reclassifying the cultivated areas into rainfed and irrigated croplands with a DNN. The use of deep neural networks presents a novel approach for accurately mapping and monitoring irrigated and rainfed croplands at a national scale in South Africa. The approach has the potential to address the data gap and provide timely and up-to-date information for addressing food insecurity and improving agricultural productivity. By incorporating deep learning techniques into the mapping process, this study aims to overcome the limitations of traditional remote sensing methods and provide a more accurate and comprehensive understanding of the extent and distribution of irrigated and rainfed croplands in South Africa. The methodological framework presented in this study provides baseline data and information necessary for timely decision making as a pathway towards addressing food security. This framework can also be upscaled to other sub-Saharan countries for improved decision making in the agriculture sector.

2. Methods and materials

2.1. Study area

Crop and livestock production are the main agricultural activities in South Africa and crop production is mainly dominated by maize and wheat, whilst livestock production is dominated by cattle ranching and a developed poultry sector (Netshipale et al., 2020). The distribution of these agricultural activities in the nine administrative provinces: Eastern Cape, Free State, Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, Northern Cape, Northwest, and Western Cape is shown in Fig. 1. These activities exist in subsistence or commercial farms, using rainfall or irrigation water as the primary water source (Giannecchini et al., 2007, Schoeman et al., 2013, Waldner et al., 2017).

The climate in South Africa ranges from subtropical to Mediterranean, allowing different farming systems across the landscape (Nyam et al., 2020, Meza et al., 2021). Although the climate and landscape support agricultural activity, the contribution of agriculture to gross domestic product (GDP) has declined over the years from a high of 10 % in 1960 (Division, 2010) to 2.3 % in 2020 (Department of Agriculture, 2020).

2.2. Collection of training data

Training data (n = 18,520) were collected from multiple sources (fieldwork, photo-interpreted data, and secondary data) for image classification. The training points were distributed across the study area and represented a range of land-cover classes (irrigated, rainfed, cultivated, shrubland, barren land, built-up, forested land, grassland, and water), as defined in Table 1.

The fieldwork for this study was carried out during the wet and dry seasons between June 2021 and December 2022. To collect the field-based training points (n = 520), 400 m^2 plots were randomly placed in 500 km^2 grids that subdivided the study area following the methodology outlined in DFFE (2020). The plot size was based on the pixel size of the Sentinel-2 MSI satellite image used in the study. The main land-cover type in the 400 m^2 plots was observed and recorded for image classification. Due to the extensive geographic scope of the study area, it was not possible to collect field data simultaneously with the acquisition of all images covering the area.

In addition to the field-collected training points, data from photo-

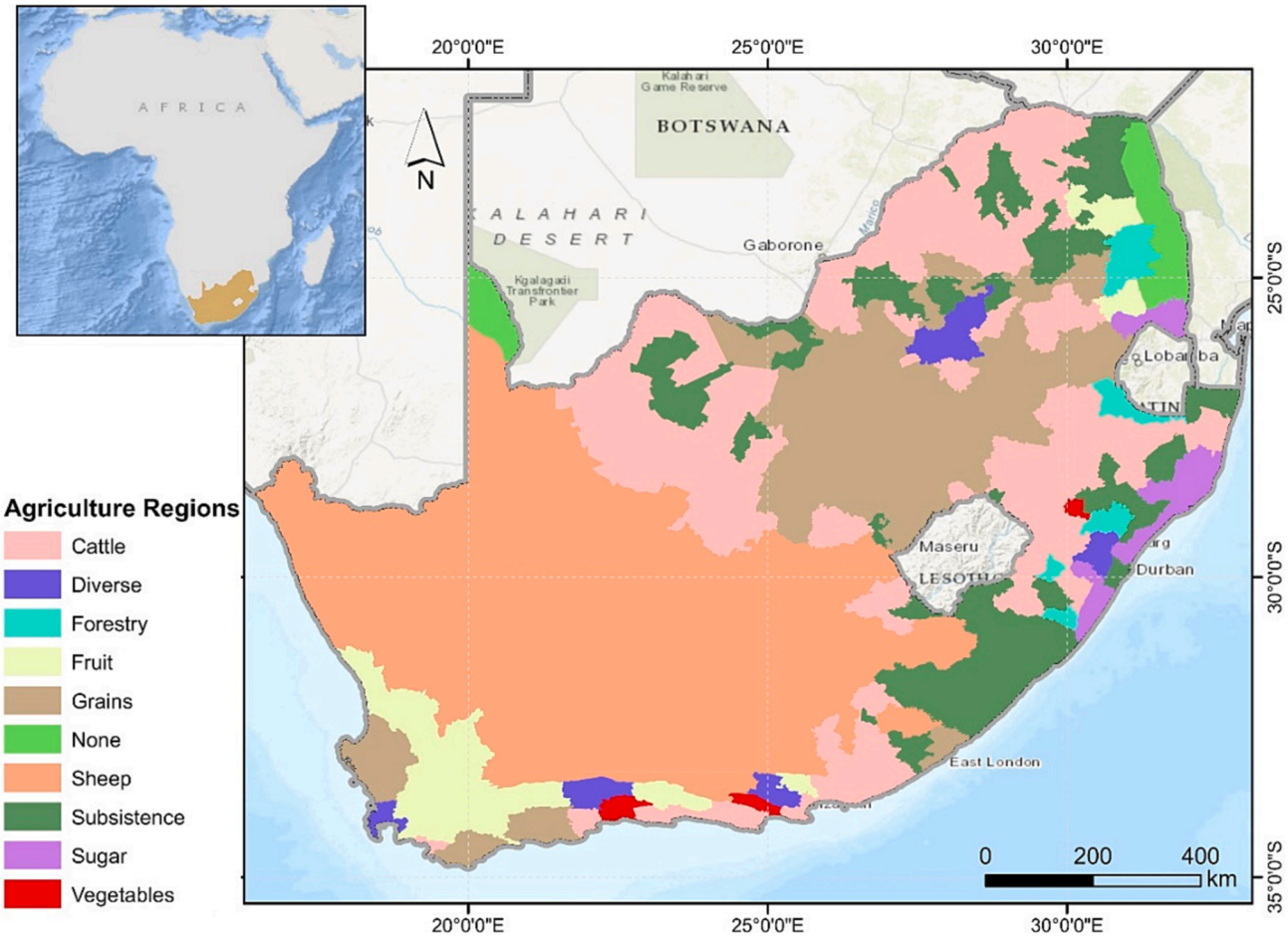

Fig. 1. Map of South Africa and the main agricultural regions across the study area (Waldner et al., 2017).

Table 1
The description used to define the landcover types used in this study.

Landcover	Description	Source
Barren/Bare (n = 2500)	Area with minimal to zero vegetation cover	(Sharma et al., 2016)
Forested (n = 2890)	Land predominantly covered by trees	(Zhu & Waller, 2003)
Grassland (n = 2024)	Area with minimal tree cover but extensive grass	(DFFE, 2020)
Water (n = 1023)	Areas containing open surface water e.g., rivers, lakes, dams	(Ritchie & Roser, 2013)
Cultivated (n = 5620)	An area with agricultural activity	(DFFE, 2020)
Built-up (n = 2400)	An urban area or an area characterised by buildings	(Mpakairi & Muvengwi, 2019)
Shrubland (n = 2063)	Areas with extensive shrubs intermixed with bare soil and scarcely distributed trees	(DFFE, 2020)
Irrigated (n = 3120)	Agricultural areas with evidence of winter farming and irrigation water use (e.g., pivot agriculture)	(Ritchie & Roser, 2013, DFFE, 2020)
Rainfed (n = 2500)	Agricultural areas depended on rainfall with minimal winter cropping	(Ritchie & Roser, 2013, DFFE, 2020)

interpreted sources (n = 8,000) and secondary data (n = 10,000) were also used to train the image classifiers. For photo-interpreted points, the high-resolution (less than0.5 m) imagery in Google Earth (https://www.googleearth.com) was systematically scanned, using the 500 km^2 grids as a guide, with the land-cover for each training point defined based on the observed land-cover class the image. To verify the land-cover of training points derived from secondary data, each point was overlaid on high-resolution imagery in Google Earth, and points with matching land-cover were retained while those with mismatching land-cover were excluded. This process enabled the visual confirmation of the landcover classification of the training points, which was essential to ensure the accuracy and quality of the data used for the analysis.

2.3. Satellite data acquisition and processing

To detect and distinguish different land cover types as well as irrigated and rainfed croplands, this study used surface reflectance from Sentinel-2 Multispectral Instrument (MSI) accessed via the Google Earth Engine platform (GEE) (https://code.earthengine.google.com). The images (n = 34 373) were cloud-free and spanned three years (2019–2021). The cropping season in the study area varies with region owing to the climate that exists, therefore, when identifying the different types of land cover, this study used surface reflectance images that were acquired throughout the year to ensure that all land cover types were captured. For example, crops are grown during different months i.e., winter and summer months in different regions of the study area, therefore images for those regions were captured during the corresponding months to capture the variations in land cover types. This approach allows for a more accurate and comprehensive understanding of the land cover and land use in the study area. The surface reflectance images provided in the GEE are already processed for atmospheric and radiometric errors; however, to remove remnant clouds, this study used the quality assessment band (QA60) to mask out areas with smoke and cirrus clouds following Mpakairi et al. (2022a). The resulting data cube with several images was stacked. The median function in the GEE was then used as a reducer to composite the data cube into only ten spectral bands (i.e., blue, green, red, near infrared (NIR), four red-edge, and two short–wave infrared). This was done to address the high variability owing to landscape heterogeneity (Gxokwe et al., 2022, Mpakairi et al., 2022a). The resulting data cube was then resampled to 20 m resolution.

Three spectral indices (Normalised Difference Vegetation Index (NDVI), Modified Soil Adjusted Vegetation Index (MSAVI 2) and Modified Triangular Vegetation Index (MTVI2)) were calculated from the median composite Sentinel-2 MSI imagery and used in the classification process (Table 2). NDVI and MSAVI2 were chosen based on their sensitivity to vegetation greenness even in areas with interfering soil reflectance (Mudereri et al., 2021). Lastly, MTVI2 was selected based on how it is sensitive to leaf chlorophyll content (Xing et al., 2019). In addition to vegetation indices the Shuttle Radar Topographic Mission (SRTM) Digital Elevation Model was also used in the classification processes.

To identify all the cultivated cropland areas in the study area, the median composite Sentinel-2 MSI images were combined with the vegetation indices and the DEM. The data cube was then classified into distinct landcover classes (i.e., cultivated, shrubland, barren land, built-up, forested land, grassland, and water). Particularly, the data cube was comprised of 14 variables which were of the 10 spectral bands (i.e., blue, green, red, near infrared (NIR), four red-edge, and two short–wave infrared), three spectral indices (NDVI, MSAVI2, MTVI2) and elevation illustrated in Table 3. This final image composite was used for the land cover classification in GEE, using the random forest algorithm.

2.4. Image classification

To identify rainfed and irrigated cropland, this study used a two-step approach that included (1) classifying the entire study area with the random forest model, masking out the non-croplands and (2) reclassifying the random forest generated cultivated areas into rainfed, and irrigated croplands using the deep learning algorithm. These steps are summarised broadly in Fig. 2.

Before testing the proposed methodology at a national-scale, we compared the robustness of the random forest, gradient tree boosting, deep learning in classifying irrigated and rainfed cropland amongst other classes. In addition, we also compared the proposed two step approach of classifying the cultivated land and reclassifying it into irrigated and rainfed using both random forest and gradient tree boosting combined with the deep learning algorithm. The performance of these algorithms in classifying irrigated and rainfed cropland was evaluated at a smaller spatial extent. The method that exhibited the highest overall accuracy was then adopted to characterise irrigated and rainfed cropland at a national scale. Using insights gained from this localized testing, the two-step approach that used random forest and deep learning had the highest accuracy hence it was adopted and applied at a national scale, fulfilling our overarching goal of generating accurate and comprehensive cropland inventory data.

The random forest algorithm was used because it is a reliable and highly predictive classifier capable of dealing with non-linear data (Mpakairi & Muvengwi, 2019, Mpakairi et al., 2022b). It has also been observed to outperform other classifiers such as Support Vector Machines and Naïve Bayes (Gxokwe et al., 2022). The random forest algorithm was implemented in GEE. In addition, the gradient tree boosting algorithm was selected as a robust classification method due to its ability to handle complex relationships and capture intricate patterns within the data (Georganos et al., 2018, Jozdani et al., 2019, McCarty et al., 2020). This algorithm is well-suited for non-linear data and has exhibited superior performance compared to other classifiers, such as Support Vector Machines and Naïve Bayes, as demonstrated in previous studies (Jozdani et al., 2019). The gradient tree boosting algorithm was also implemented within the Google Earth Engine (GEE) platform, allowing for efficient and effective classification of land cover types in the study area. Lastly, the deep learning was executed in R (TEAM, 2014), using the h2o package (LeDell et al., 2018). Deep learning in h2o uses a multilayer feed-forward artificial neural network that relies on error back-propagation (Candel et al., 2016, Costache et al., 2022). The multilayer feed-forward artificial neural network structure comprises input layers connected to the output layer through hidden layers. The hidden layers transform the input data to output data through neurons with tanh, rectifier, and activation functions (Aiello et al., 2016). Deep learning relies on parameterisation to improve the prediction accuracy, and the key parameters that need to be parametrized include the

Table 2
The formulas used to calculate the ancillary spectral indices were used to enhance classification. Where NIR, Red and Green represent the near-infrared, red and green spectral bands respectively.

Spectral index	Formulae	Source
Normalised Difference Vegetation Index (NDVI),	$\frac{(Nir - Red)}{(Nir + Red}$	Mutanga et al. (2012)
Modified Soil Adjusted Vegetation Index (MSAVI 2)	$\frac{2NIR + 1 - \sqrt{(2NIR + 1)^2 - 8(NIR - Red)}}{2}$	Gholami Baghi & Oldeland (2019)
Modified Triangular Vegetation Index (MTVI2)	$1.5\frac{1.2(NIR - Green) - 2.5(Red - Green)}{\sqrt{(2NIR + 1)^2 - (6NIR - 5\sqrt{Red}) - 0.5}}$	Xing et al. (2019)

Table 3
The data sources used in this study.

Name	Resolution	Date range	Source
Digital Elevation Model (DEM)	30 m		Shuttle Radar Topographic Mission (SRTM)
Sentinel-2 MSI	20 m	January 2019–December 2021	
Modified Soil Adjusted Vegetation Index (MSAVI 2)	20 m	January 2019–December 2021	Derived from Sentinel-2 MSI
Modified Triangular Vegetation Index (MTVI2)	20 m	January 2019–December 2021	Derived from Sentinel-2 MSI
Normalised Difference Vegetation Index (NDVI)	20 m	January 2019–December 2021	Derived from Sentinel-2 MSI
Potential evapotranspiration (PET)	20 m	January 2019–December 2021	Derived from Sentinel-2 MSI using SEBAL model

activation function (activation), dropout, number of hidden layers, the number of times to iterate (epoch) and adaptive learning rate (Aiello et al., 2016, Costache et al., 2022). For example, increasing the hidden layers and the learning rate allows deep learning to solve complex situations (Aiello et al., 2016).

2.5. The application of the two-step approach at a national scale

2.5.1. Classification of cultivated land using random forest at a national scale

The random forest classification was conducted using eighty percent (80 %; n = 14,816) of the total sampled points representing cultivated, shrubland, barren land, built-up, forested land, and grassland as well as the water classes. Next, the remaining twenty percent (20 %; n = 3,704) of the total sampled points were used as the testing dataset to assess the accuracy of the classification model. The performance of the model was assessed using the Overall Accuracy (OA) whilst the accuracy of each land cover class was evaluated using the F-1 statistic, producer, and user accuracy. These accuracy measures are fully explained in Mudereri et al. (2021) and Ngadze et al. (2020).

After classifying the study area into different landcover types, all the

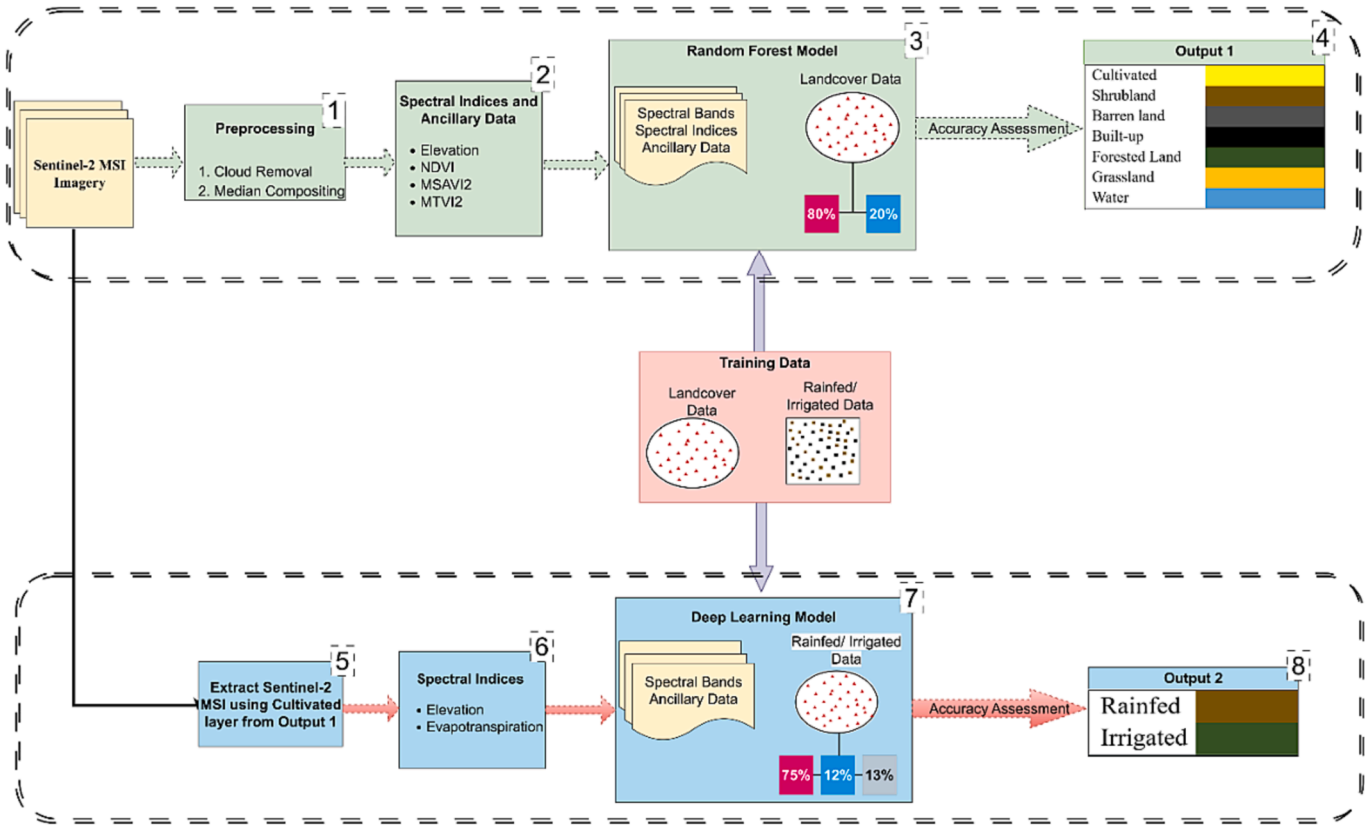

Fig. 2. Summarised flowchart showing the steps and processes followed in identifying irrigated and rainfed croplands in the study area.

cultivated areas were extracted and converted into a shapefile. The cultivated area shapefile was used to extract pixels from the median composite Sentinel-2 MSI image, and these pixels and their spectral information were deemed to represent all the cultivated areas in the study area.

2.5.2. Identifying irrigated and rainfed croplands at a national scale

To identify irrigated and rainfed crops, all areas identified as cultivated land by the random forest model were used. Other landcover classes were masked out retaining areas covered by the cultivated areas. Potential evapotranspiration (ET) was then added into the data cube generated in section, 2.3. Potential evapotranspiration (ET) was included to enhance the characterisation of irrigated croplands since they use more water when compared to rainfed croplands (Attia et al., 2015). Potential evapotranspiration (ET) was derived from the GEE using the Surface Energy Balance Algorithm for Land (SEBAL) following (Mhawej & Faour, 2020). The Surface Energy Balance Algorithm for Land (SEBAL) estimates potential evapotranspiration (ET) by computing the difference between the available energy at the land surface and the energy used for sensible heat flux, and then scaling this difference using a surface resistance parameter (Jassas et al., 2015, Mhawej & Faour, 2020). The algorithm employs remotely sensed data, such as surface temperature and vegetation indices, to derive the necessary inputs for the energy balance equation (Jassas et al., 2015). The ET was calculated for each satellite image acquired over the study period (2019–2021) and averaged to obtain the annual ET values for each cropland. The data cube was then clipped using the RF generated cultivated land extent. The resultant median composite image representing all the cultivated areas was then classified again into rainfed and irrigated areas. Then classification was done using training data representing irrigated and rainfed areas. All the datasets used in this study are listed in Table 3.

2.5.3. Deep learning modelling framework

To discriminate between irrigated and rainfed croplands using the deep learning algorithm, parameterisation was done using grid searching on a set of hyperparameter options shown in Supplementary 1. The optimal hyperparameters were identified through a cross-validation (n = 10 times) process with 80 % (n = 4496) of the training data. The remaining dataset was used for model validation (10 %; n = 562) and model testing (10 %; n = 562). During model calibration, the best-performing model had the lowest log loss even after increasing the number of epochs. The results from the DNN model were evaluated using the Overall Accuracy (OA) and area under the receiver operating curve (ROC) and area under the curve (AUC). Class accuracy was evaluated using the F-1 statistic, producer, and user accuracy.

3. Results

3.1. Exploratory comparison of different classification algorithms

The exploratory comparison of different classification algorithms on a small area showed that the model using the random forest combined with deep learning had a higher accuracy compared to the other algorithms tested (Table 4). This was followed by the gradient tree boosting combined with deep learning (OA = 0.82) and the deep learning model (OA = 80). The classification output from these algorithms is shown in Supplementary Fig. 1.

Table 4
The model performance for different models and model combinations.

Model	Model performance (Overall Accuracy)
Random forest	0.78
Gradient tree boosting	0.67
Deep learning	0.80
Random forest combined with deep learning	0.89
Gradient tree boosting combined with deep learning	0.82

3.2. National scale classification using random forest

The random forest was able to classify different landcover types at a national scale and had an overall accuracy of 0.77. The proximal classification output demonstrating the performance of the random forest model for different land cover types is shown in Fig. 3. The performance of the random forest model was influenced mainly by the shortwave infrared (B11), and blue (B2) spectral bands and one ancillary band (elevation). On the other hand, the least contributing bands were NDVI, MSVAI2, and the red (B4) spectral band (Fig. 4).

The per-class accuracy for all the land cover types was relatively high based on the F-1 score, user and producer accuracy (Fig. 5). Specifically, water (F-1 score = 0.97 and UA = 0.96), shrublands (F-1 score = 0.87 and UA = 0.92) and vegetation/forested (F-1 score = 0.73 and UA = 0.67) had the highest per-class accuracy whilst barren land (F-1 score = 0.71 and UA = 0.68), cultivated (F-1 score = 0.69 and UA = 0.63), and grassland (F-1 score = 0.71 and UA = 0.79) had the lowest per-class accuracy. Although cultivated areas had a lower per-class accuracy, the output had high spatial fidelity to the underlying landcover type (Fig. 3). In addition, shrublands (37 %) and grasslands (30 %) covered most of the study area whilst the combined area covered by water and built-up areas was less than 12,000 ha (Fig. 6).

3.3. Discriminating rainfed and irrigated areas using the DNN algorithm

The deep neural network model was able to classify irrigated and rainfed farms with high accuracy. The area under the receiver operating curve (AUC) for the DNN was high during model training (AUC = 0.80), validation (AUC = 0.71), and testing (AUC = 0.71) (Fig. 7(a)). The red-edge-3 (B7), red (B4), red-edge-4 (B8A) and one ancillary variable (elevation) were the most influential discrimination variables, and these contributed ~40 % to the overall performance of the model (Fig. 7(b)). (The second SWIR (B12), blue (B2) and red-edge 1 (B5) spectral bands contributed the least to the overall model performance with a combined contribution of ~21 %. The classified output of rainfed and irrigated croplands had high spatial fidelity and corresponded with the underlying farms when overlaid on high-resolution imagery (Fig. 9).

The overall per-class accuracy of the classified rainfed and irrigated farms was high (UA greater than 0.6). Specifically, irrigated farms had a higher accuracy (F-1 score = 0.81 and UA = 0.71) when compared to rainfed farms (F-1 score = 0.36 and UA = 0.68) (Fig. 8).

The distribution of the classified croplands varied across the study area (Fig., 10). The Free State, Northwest and Western Cape provinces had the highest number of croplands and accounted for ~60 % of all the croplands in South Africa (Fig. 11). In addition, these provinces also accounted for the highest number of irrigated croplands whilst the Eastern Cape, KwaZulu Natal and Western Cape provinces had the highest number of rainfed croplands. In all the nine provinces, irrigated croplands were higher than rainfed croplands except for KwaZulu Natal and Eastern Cape provinces where rainfed croplands were more than irrigated croplands.

4. Discussion

The extent and distribution of irrigated and rainfed croplands provides baseline data and information necessary for tackling food insecurity in South Africa. The current methods of identifying these cropland areas are costly for extensive environments. Alternatively, multisource data compiled from public repositories, ongoing research, and agricultural databases can be used. The study aimed to present an integrated approach towards mapping irrigated and rainfed croplands with the use of multisource data and machine learning algorithms. Results from this study demonstrate how this could be achieved a national scale in the

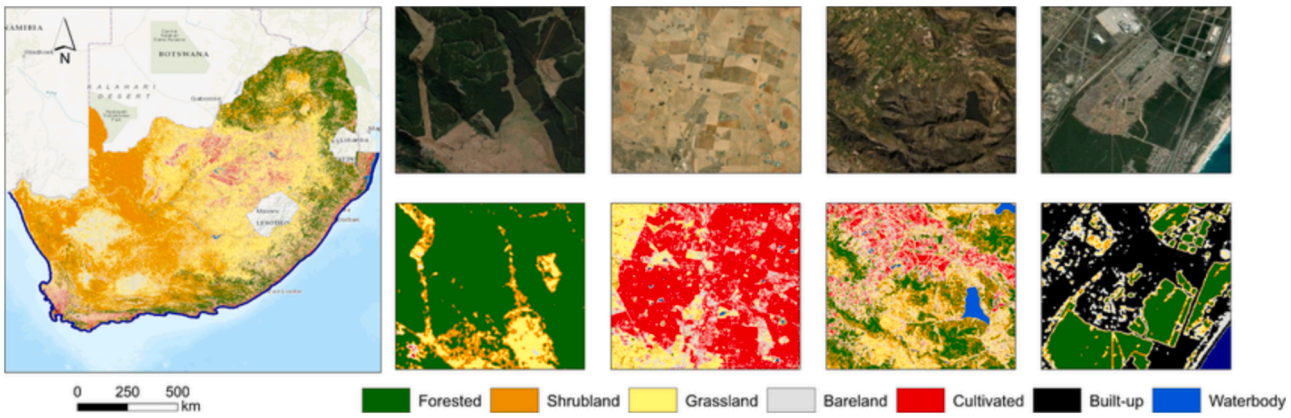

Fig. 3. Selected areas demonstrating the proximal classification output of different land cover classes produced by the random forest model for South Africa.

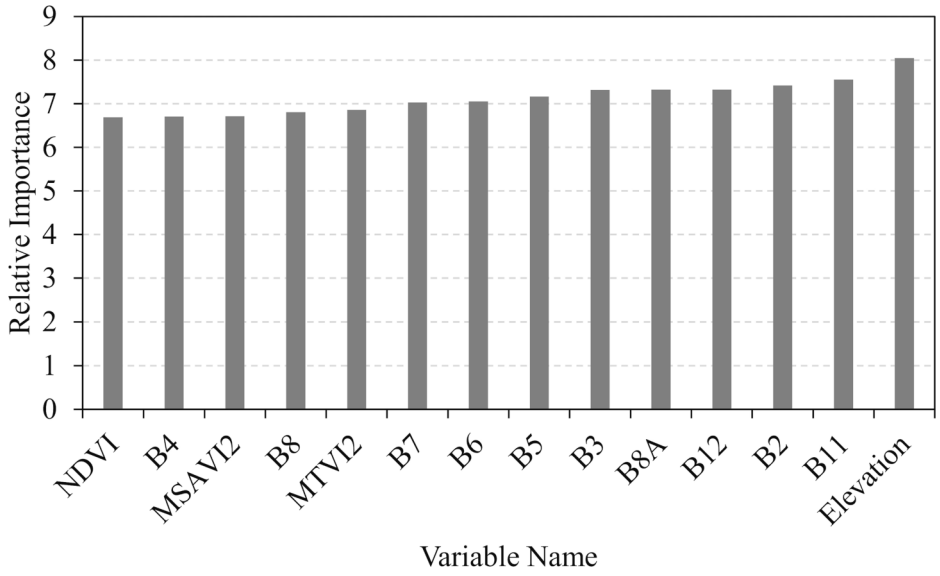

Fig. 4. Variable importance of the spectral bands and indices used to classify the landcover using the random forest model.

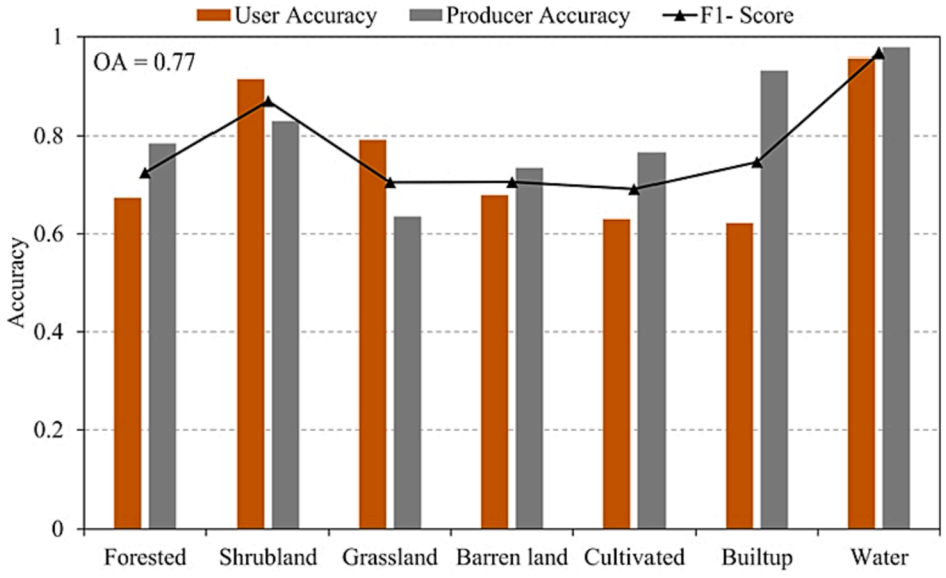

Fig. 5. The overall and per-class accuracy of all the landcover types classified using the random forest model.

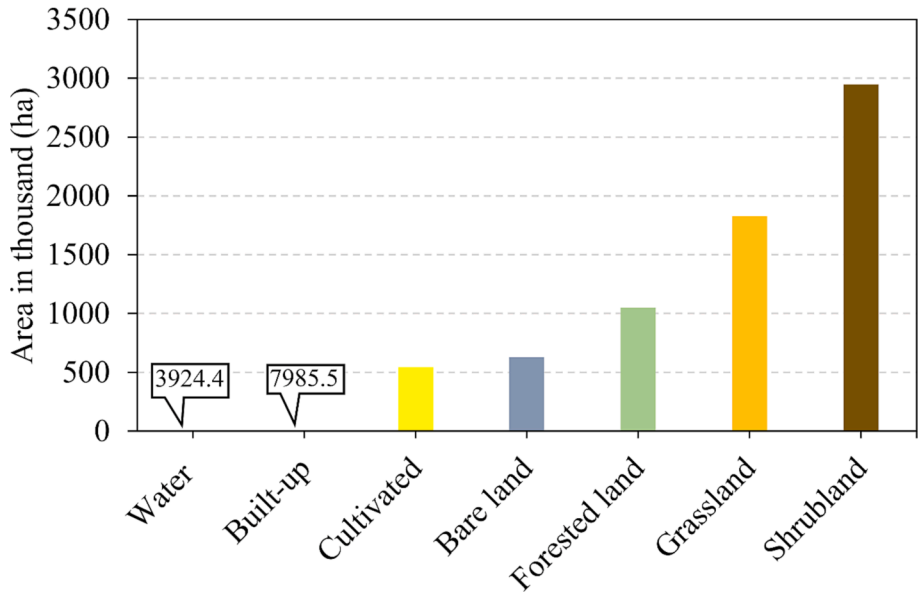

Fig. 6. The extent of different landcover classes derived from the random forest model for the study area.

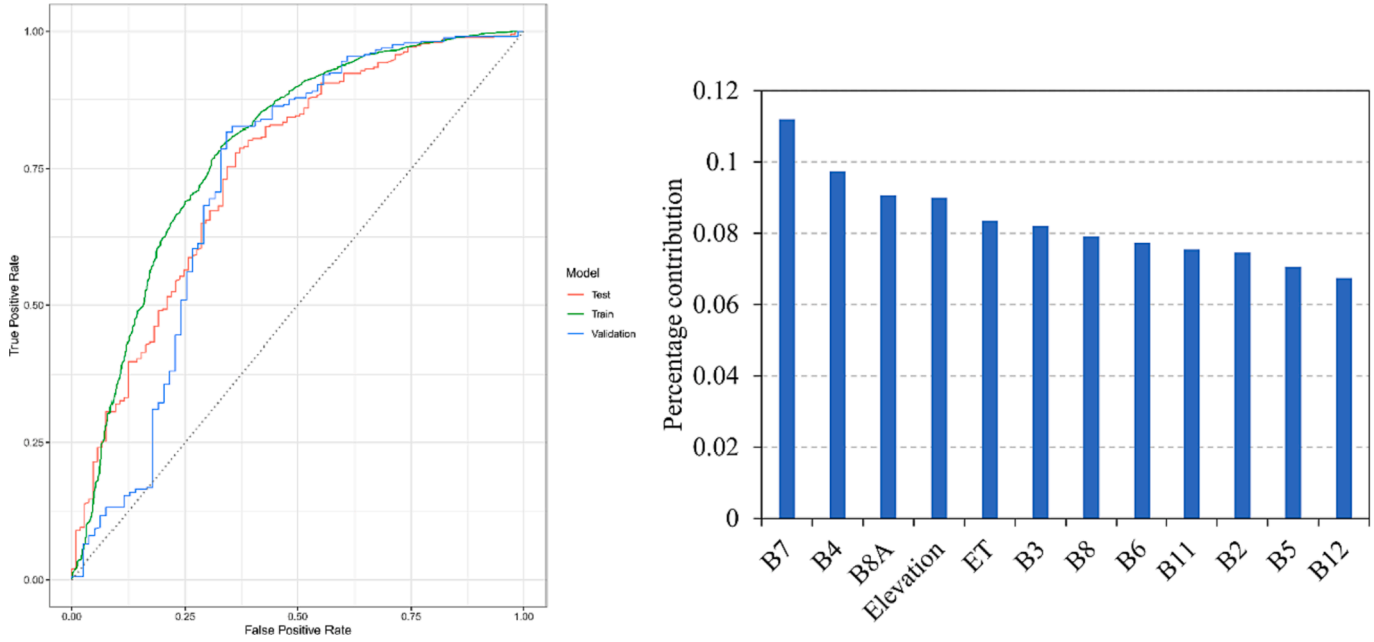

Fig. 7. (a)The area under the receiver operating curve (AUC) for the training, validation and test datasets used during model training, validation, and testing with the DNN model and (b) Variable importance scores of the spectral bands and indices used the DNN classification.

context of South Africa.

4.1. Algorithms performance in identifying irrigated and rainfed croplands

The sequential approach of using of random forest to map cultivated areas and further reclassifying the cultivated areas into rainfed and irrigated croplands using deep learning had a relatively higher accuracy (OA = 0.89) compared to all the other models. This justified its adoption for the identification of rainfed and irrigated croplands on a national level.

The high performance of the combined random forest and deep learning model is not surprising since the random forest model has exhibited higher accuracies in different studies that the algorithm has been used (e.g., in Rodriguez-Galiano et al. (2012), Zeng et al. (2020) and Chen et al. (2019)). The robustness of the random forest algorithm, characterized by its capacity to counter model overfitting, its adeptness in managing unbalanced datasets – including nationwide landcover data – has established its prominence in classification studies (Rodriguez-Galiano et al., 2012, Akbari et al., 2020, Phalke et al., 2020, Rahman et al., 2020). When combined with other algorithms, the random forest algorithm, can not only enhance classification accuracy, but also significantly enhances the spatial fidelity of the generated landcover maps. This was evident from the improved spatial fidelity after the random forest algorithm was combined with the deep learning model while exploring diverse algorithmic approaches. The optimal performance of the two-step approach could also be plausibly attributed to the DNN's capacity to effectively capture intricate patterns within complex

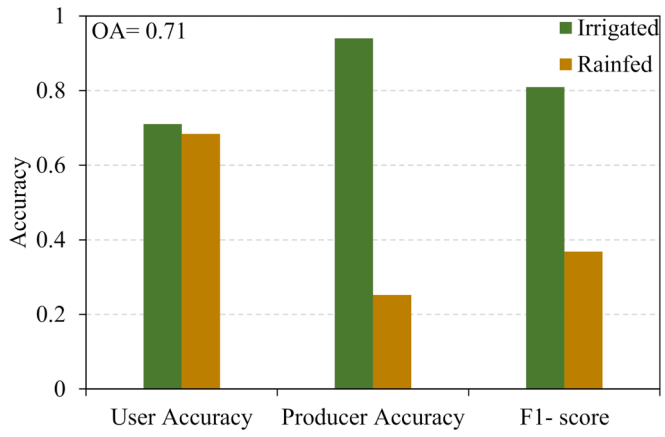

Fig. 8. The overall and per-class accuracy of irrigated and rainfed croplands classified using the deep DNN model.

and diverse multisource data thereby enabling accurate discrimination of irrigated and rainfed croplands (Landmann et al., 2019, Useya et al., 2019, Kavhu et al., 2022).

The results showed that the two-step approach exhibited an OA of 0.89 at a smaller extent whereas at a nation scale it exhibited an OA of 0.71. This minor difference in the accuracies could be attributed to the fact that at a national-scale mapping an added layer of complexity is introduced due to extensive landscape and crop heterogeneity, which could be impacting the discrimination of irrigated and rainfed crop spectral signatures (Smith et al., 2003, Waldner et al., 2017, Landmann et al., 2019, Useya et al., 2019). These findings are similar to those of (Bey et al., 2020), who noted that increased landscape extent affects the overall classification of croplands. In addition, the different agroclimatic characteristics at a national scale affect the crops phenological stages thus influencing the spectral separability of rainfed and irrigated croplands (Waldner et al., 2015, Yang et al., 2019, Bey et al., 2020). Lastly, the performance of the two-step approach could be attributed to the input variables that were included in model calibration.

The red-edge and red spectral bands were identified as the most influential variables in the discrimination of rainfed and irrigated croplands. These spectral bands are suitable for identifying agricultural landscapes because the biophysical and biochemical parameters of crops are easily detectable in these spectral regions (Mariotto et al., 2013, Waldner et al., 2016, Csillik & Belgiu, 2017). For example, crops and vegetated environments are sensitive to the red spectral region due to light absorption by chlorophyll, a pigment that is essential for photosynthetic activities, and whose spectral signature is often associated with the red edge spectral region (Clevers & Gitelson, 2013, Ding et al., 2014, Aneece & Thenkabail, 2021). The concentration of chlorophyll in crops can also influence the variations in the absorption levels of irrigated and rainfed crops (Morley et al., 2020). The chlorophyll content of irrigated crops is usually high since these crops are not water-stressed and have nutrients available, whilst rainfed crops might be water-stressed and nutrient deficient from subsistence cultivation with minimum agricultural inputs available (Arunyanark et al., 2008, Chen et al., 2016, Hailemichael et al., 2016). This then facilitates the discrimination of the rainfed crops, using the red and red-edge spectral regions owing to the lower levels of chlorophyll concentrations (Peng & Gitelson, 2012, Clevers & Gitelson, 2013, Ding et al., 2014). In addition, the inclusion of potential evapotranspiration allowed enhanced discrimination of rainfed and irrigated croplands. Areas with irrigation farming have a low water deficit because of the high plant and soil moisture content, which allows a high evaporative demand (Yue et al., 2019, Liu et al., 2021). However, for rainfed croplands, the plant and soil moisture content are low and a high evaporative demand results in a water deficit leading to low evapotranspiration and stunted crops (Cammalleri et al., 2014, Yue et al., 2019).

These dynamics related to plant and soil moisture content are important in discriminating between irrigated and rainfed farming systems. Albeit the spectral signature of irrigated and rainfed croplands can be similar, especially during the wet season since the soil moisture in these different farming systems can be similar. In this case, using time-series analysis with multiple dates of acquisition and incorporating ancillary information such as land use and crop management practices can improve the accuracy of discriminating between rainfed and irrigated croplands with similar spectral signatures (Peng & Gitelson, 2012).

Also, the inclusion of elevation could have provided *apriori* information for reducing spectral mixing caused by topographic shadows thus enhancing the discrimination of irrigated and rainfed croplands (Wulder et al., 2004, Wulder et al., 2018, Akbari et al., 2020, Phalke et al., 2020). Other ancillary variables that have also been observed to improve classification accuracy include the topography derivatives (e.g., slope) (Hermosilla et al., 2022). These variables are often used because they improve the spectral separability of features consequently improving the classification accuracy (Rodriguez-Galiano et al., 2012, Chen et al., 2019, Zeng et al., 2020).

4.2. *Implications of mapping irrigated and rainfed croplands in South Africa*

Most studies using remote sensing for cropland mapping focused mainly on identifying cultivated areas, not the water source behind the type of agriculture (Rainfed or irrigation agricultural areas) (Clevers & Gitelson, 2013, Csillik & Belgiu, 2017, Xu et al., 2018). The two-step approach using random forest and DNN addressed this challenge. The methodology utilised ccurate classification algorithms to present a national-scale map identifying irrigated and rainfed farming systems in South Africa. Future studies could improve this approach by including OBIA variables since complementary data from object-based image analysis (OBIA) can be helpful and improve the classification of these farming systems. This could improve the classification accuracies considering that irrigated areas are sometimes circular or rectangular and more extensive than rainfed farms (Li et al., 2015, De Castro et al., 2018). Although using OBIA might be possible, the main setback with regional scale mapping is the limited computational capacity (Thompson et al., 2020). Analysing remotely sensed data at a national scale requires high computational power, especially when coupled with algorithms such as DNN which are computationally intensive (Christophe et al., 2011). Regardless of the availability of GEE, some datasets, and algorithms (e.g., DNN) may not be available and might require integration with other advanced softwares (e.g., R).

The extent and distribution of rainfed and irrigated croplands varies across South Africa owing to several factors (e.g., soil, climate, and historical inequality). The 2017 report published on the census of commercial agriculture in South Africa (SA, 2017, Nortjé, 2020) showed that the highest commercial farms were in Free State, Western Cape, Northwest and Northern Cape provinces. The findings from the methodological framework used in this study are coherent to the report since irrigated croplands, which are mainly commercial farms, were observed to be highest in these provinces (SA, 2017, SA, 2021). The results also show that the KwaZulu Natal and Eastern Cape provinces had the highest number of rainfed croplands. These provinces have the highest number of smallholder farmers owing to the soil fertility and rainfall regime which supports subsistence farming requiring minimum agricultural inputs (Ntshangase et al., 2018, Du Preez et al., 2019). The agricultural produce from these smallholder farms is usually meant to support their livelihoods and is rarely available for export (Yobe et al., 2019).

There exists limited information on the extent and distribution of irrigated and rainfed croplands in South Africa. Conventional methods of mapping irrigated and rainfed croplands are costly, laborious and

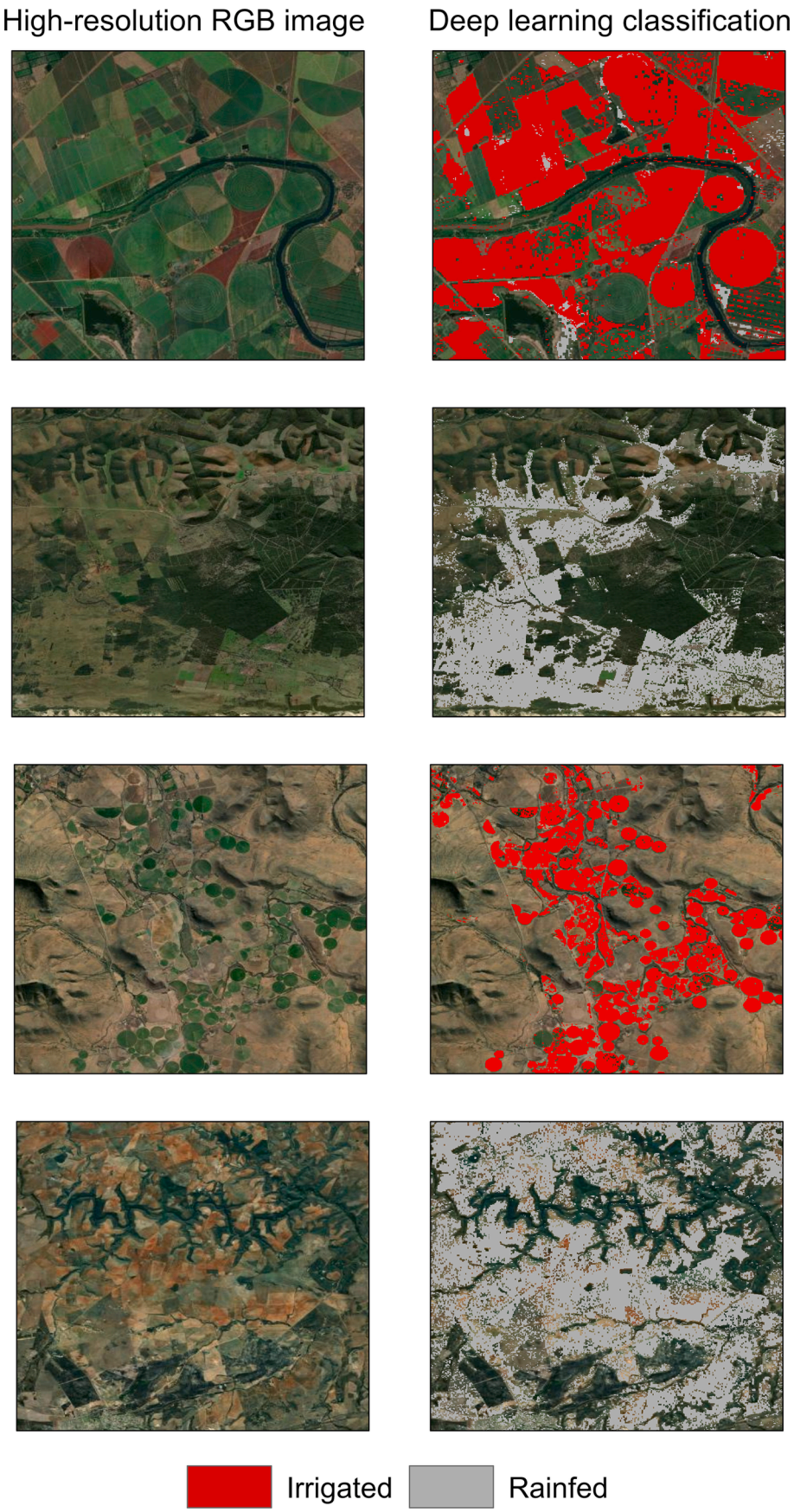

Fig. 9. Selected areas demonstrating the classification output of the DNN model for irrigated and rainfed croplands in South Africa.

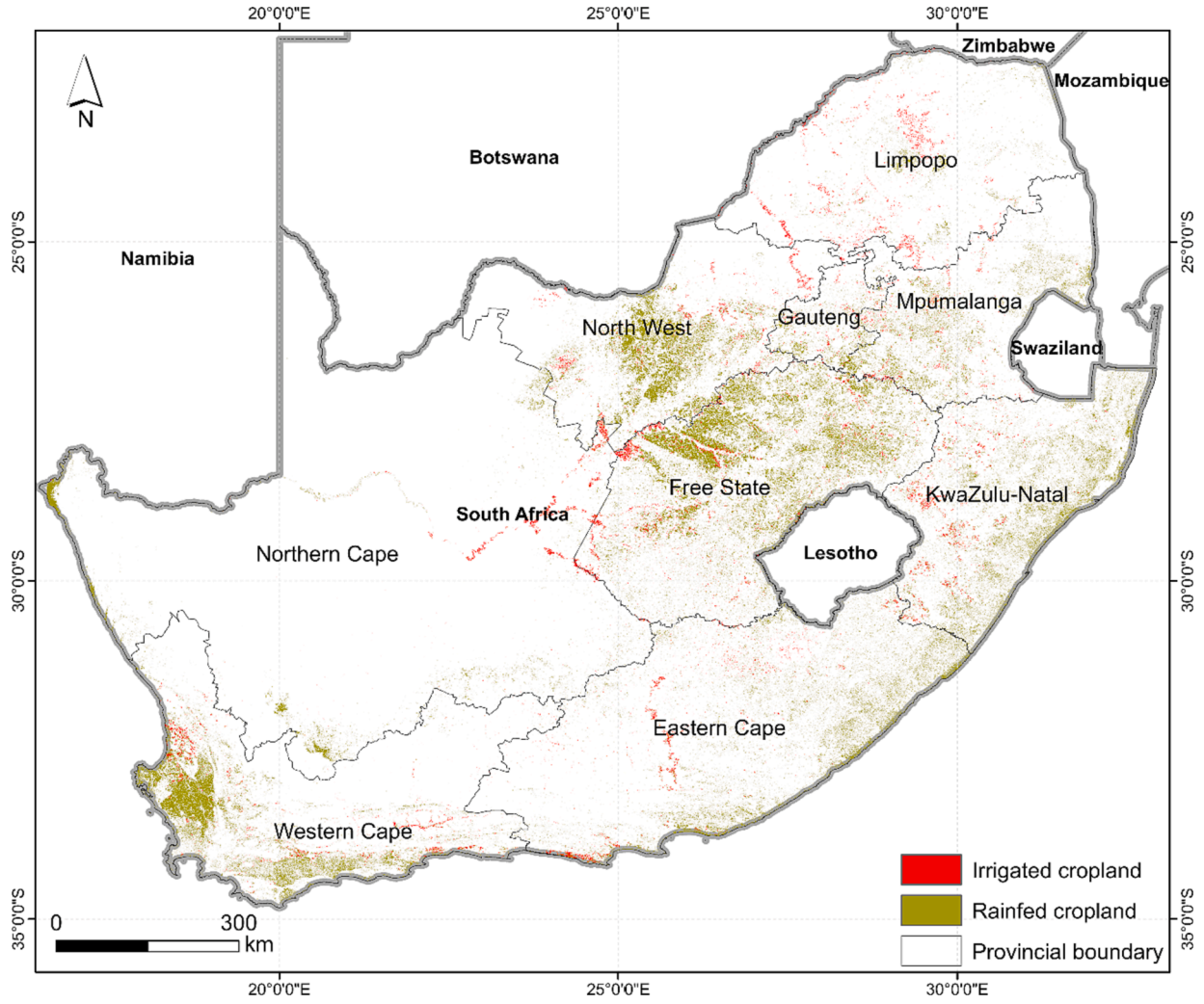

Fig. 10. National scale map of irrigated and rainfed croplands classified using deep neural network (DNN) model in South.

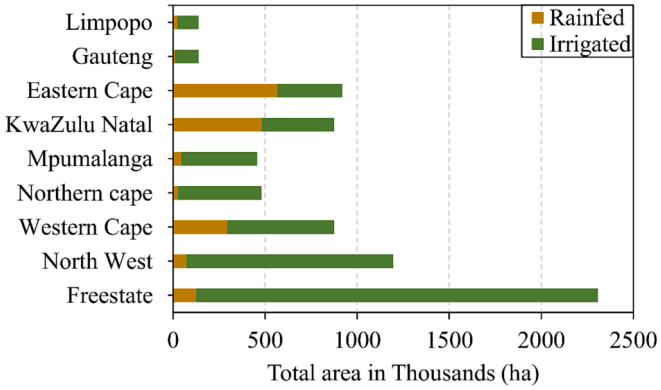

Fig. 11. Provincial statistics derived from the deep neural network (DNN) model showing the extent of irrigated and rainfed croplands in South Africa.

take time, however, leveraging on earth observation technologies allows timely food supply decisions to be made and reduce instances of food insecurity in South Africa (Labadarios et al., 2009). The study provides baseline data and information necessary for understanding agriculture food supply, yield forecasting, crop monitoring and trade policies.

The methodology presented in this study can be scaled to other sub-Saharan countries since baseline data and information on croplands is either outdated or non-existent. The methodology will likely improve agriculture productivity and support more than 330 million people affected by food insecurity especially in sub-Saharan rural households (Fao, 2017, Bjornlund et al., 2022).

4.3. Limitations of the proposed methodological framework

While the proposed methodologies offer valuable insights into mapping irrigated and rainfed croplands using multisource data and machine learning algorithms, it also faces some potential setbacks and limitations. One major limitation is the extensive scale of the study area, which can introduce challenges in accurately classifying croplands due to increased landscape heterogeneity (Landmann et al., 2019, Useya et al., 2019). The reliance on multisource data from various repositories could lead to inconsistencies or data gaps, affecting the overall accuracy of the classification (Waldner et al., 2017). Moreover, the methodology's applicability may vary across different sub-Saharan countries with varying environmental and agricultural characteristics, potentially limiting its generalizability. Despite these limitations, the study's integrated approach and detailed analysis hold promise for addressing food insecurity and improving agricultural productivity in the region.

5. Conclusions

The study aimed to present an integrated approach towards mapping irrigated and rainfed croplands in South Africa, using multisource data

and machine learning algorithms. The random forest model used for landcover classification showed high accuracy, comparable to similar studies, and was optimised by the inclusion of ancillary data such as elevation. The utilization of the deep learning neural network model on the random forest output facilitated the mapping and discrimination between rainfed and irrigated croplands, showcasing the potential of this approach for national scale national scale mapping activities. The inclusion of red-edge and red spectral bands, as well as potential evapotranspiration, were found to be important in improving the performance of the DNN model. The findings from this study are significant for addressing food insecurity in South Africa by providing baseline data and information on the extent and distribution of irrigated and rainfed croplands, which can inform agricultural planning and resource allocation. Furthermore, results from this study can be replicated in other extensive environments where current mapping methods are costly.

Future research can expand on these findings by employing time-series analysis and incorporating more ancillary data to improve cropland mapping accuracy, particularly in areas with similar spectral signatures. Overall, this study highlights the potential of machine learning algorithms and multisource data for cropland mapping and has important implications for food security and sustainable development.

Funding

The authors received support from the Water Research Commission (C2022/2023-00902).

CRediT authorship contribution statement

Kudzai S. Mpakairi: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Timothy Dube:** Conceptualization, Supervision, Writing – review & editing. **Mbulisi Sibanda:** Supervision, Writing – review & editing, Software, Validation. **Onisimo Mutanga:** Conceptualization, Supervision, Methodology, Writing – review & editing.

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.

Data availability

Data used in this research is freely available online and upon request.

Acknowledgments

The authors are grateful for the support offered by Water Research Commission. The authors would also like to extend their gratitude to the anonymous reviewers who improved the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.isprsjprs.2023.09.006.

References

Aiello, S., Eckstrand, E., Fu, A., Landry, M., Aboyoun, P., 2016. Machine Learning with R and H2O. H2O booklet 550.
Akbari, E., Darvishi Boloorani, A., Neysani Samany, N., Hamzeh, S., Soufizadeh, S., Pignatti, S., 2020. Crop mapping using random forest and particle swarm optimization based on multi-temporal Sentinel-2. Remote Sens. (Basel) 12, 1449.
Aneece, I., Thenkabail, P.S., 2021. Classifying crop types using two generations of hyperspectral sensors (Hyperion and DESIS) with machine learning on the cloud. Remote Sens. (Basel) 13, 4704.
Arunyanark, A., Jogloy, S., Akkasaeng, C., Vorasoot, N., Kesmala, T., Nageswara Rao, R. C., Wright, G.C., Patanothai, A., 2008. Chlorophyll stability is an indicator of drought tolerance in peanut. J. Agron. Crop Sci. 194, 113–125.
Attia, A., Rajan, N., Ritchie, G., Cui, S., Ibrahim, A., Hays, D., Xue, Q., Wilborn, J., 2015. Yield, quality, and spectral reflectance responses of cotton under subsurface drip irrigation. Agron. J. 107, 1355–1364.
Ayanlade, A., Radeny, M., 2020. COVID-19 and food security in Sub-Saharan Africa: implications of lockdown during agricultural planting seasons. NPJ Sci. Food 4, 13.
Baiphethi, M.N., Jacobs, P.T., 2009. The contribution of subsistence farming to food security in South Africa. Agrekon 48, 459–482.
Bank, W. 2021. Employment in agriculture (% of total employment) (modeled ILO estimate) - Sub-Saharan Africa.
Bey, A., Jetimane, J., Lisboa, S.N., Ribeiro, N., Sitoe, A., Meyfroidt, P., 2020. Mapping smallholder and large-scale cropland dynamics with a flexible classification system and pixel-based composites in an emerging frontier of Mozambique. Remote Sens. Environ. 239, 111611.
Bjornlund, V., Bjornlund, H., van Rooyen, A., 2022. Why food insecurity persists in sub-Saharan Africa: A review of existing evidence. Food security 14, 845–864.
Blair, D., Shackleton, C.M., Mograbi, P.J., 2018. Cropland abandonment in South African smallholder communal lands: Land cover change (1950–2010) and farmer perceptions of contributing factors. Land 7, 121.
Cammalleri, C., Anderson, M.C., Gao, F., Hain, C.R., Kustas, W.P., 2014. Mapping daily evapotranspiration at field scales over rainfed and irrigated agricultural areas using remote sensing data fusion. Agric. For. Meteorol. 186, 1–11.
Candel, A., Parmar, V., LeDell, E., Arora, A., 2016. Deep learning with H2O. H2O ai Inc, 1–21.
Chen, D., Wang, S., Cao, B., Cao, D., Leng, G., Li, H., Yin, L., Shan, L., Deng, X., 2016. Genotypic variation in growth and physiological response to drought stress and re-watering reveals the critical role of recovery in drought adaptation in maize seedlings. Front. Plant Sci. 6.
Chen, B., Xu, B., Zhu, Z., Yuan, C., Suen, H.P., Guo, J., Xu, N., Li, W., Zhao, Y., Yang, J., 2019. Stable classification with limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. Sci. Bull. 64, 370–373.
Christophe, E., Michel, J., Inglada, J., 2011. Remote sensing processing: From multicore to GPU. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 4, 643–652.
Clevers, J.G., Gitelson, A.A., 2013. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and-3. Int. J. Appl. Earth Obs. Geoinf. 23, 344–351.
Congalton, R., K. Yadav, K. McDonnell, J. Poehnelt, B. Stevens, M. Gumma, P. Teluguntla, and P. Thenkabail. 2017. Global Food Security-support Analysis Data (GFSAD) Cropland Extent 2015 Validation 30 m V001.
Costache, R., Tin, T.T., Arabameri, A., Crăciun, A., Ajin, R., Costache, I., Islam, A.R.M.T., Abba, S., Sahana, M., Avand, M., 2022. Flash-flood hazard using deep learning based on H2O R package and fuzzy-multicriteria decision-making analysis. J. Hydrol. 609, 127747.
Csillik, O., and Belgiu, M. (2017). Cropland mapping from Sentinel-2 time series data using object-based image analysis. In: Proceedings of the 20th AGILE International Conference on Geographic Information Science Societal Geo-Innovation Celebrating, Wageningen, The Netherlands.
Dang, C., Liu, Y., Yue, H., Qian, J., Zhu, R., 2021. Autumn crop yield prediction using data-driven approaches:-support vector machines, random forest, and deep neural network methods. Can. J. Remote. Sens. 47, 162–181.
De Castro, A.I., Six, J., Plant, R.E., Peña, J.M., 2018. Mapping crop calendar events and phenology-related metrics at the parcel level by object-based image analysis (OBIA) of MODIS-NDVI time-series: A case study in central California. Remote Sens. (Basel) 10, 1745.
de Graaff, J., Kessler, A., Nibbering, J.W., 2011. Agriculture and food security in selected countries in Sub-Saharan Africa: diversity in trends and opportunities. Food Secur. 3, 195–213.
Department of Agriculture L. R. a. R. D, 2020. Economic review of the South African agriculture. Department of Agriculture, Land reform and Rural Development Pretoria.
DFFE, 2020. South African national land-cover (SANLC) 2020. Department of Forest, Fisheries and the Environment.
Ding, Y., Zhao, K., Zheng, X., Jiang, T., 2014. Temporal dynamics of spatial heterogeneity over cropland quantified by time-series NDVI, near infrared and red reflectance of Landsat 8 OLI imagery. Int. J. Appl. Earth Obs. Geoinf. 30, 139–145.
Division, E.R. 2010. The declining contribution of agriculture to GDP: is the role of agriculture becoming less important? DIRECTORATE: ECONOMIC SERVICES.
Dong, L., Du, H., Mao, F., Han, N., Li, X., Zhou, G., Zheng, J., Zhang, M., Xing, L., Liu, T., 2019. Very high resolution remote sensing imagery classification using a fusion of random forest and deep learning technique—Subtropical area for example. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 113–128.
Du Preez, C.C., van Huyssteen, C.W., Amelung, W., 2019. Changes in soil organic matter content and quality in South African arable land. In: Soil degradation and restoration in Africa. CRC Press, pp. 110–143.
Duveiller, G., Defourny, P., 2010. A conceptual framework to define the spatial resolution requirements for agricultural monitoring using remote sensing. Remote Sens. Environ. 114, 2637–2650.
Fao, F., 2017. The future of food and agriculture–Trends and challenges. Annu. Rep. 296, 1–180.
Funk, C., Budde, M.E., 2009. Phenologically-tuned MODIS NDVI-based production anomaly estimates for Zimbabwe. Remote Sens. Environ. 113, 115–125.

Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., Wolff, E., 2018. Very high resolution object-based land use–land cover urban classification using extreme gradient boosting. IEEE Geosci. Remote Sens. Lett. 15, 607–611.
Gholami Baghi, N., Oldeland, J., 2019. Do soil-adjusted or standard vegetation indices better predict above ground biomass of semi-arid, saline rangelands in North-East Iran? Int. J. Remote Sens. 40, 8223–8235.
Giannecchini, M., Twine, W., Vogel, C., 2007. Land-cover change and human–environment interactions in a rural cultural landscape in South Africa. Geogr. J. 173, 26–42.
Gxokwe, S., Dube, T., Mazvimavi, D., 2022. Leveraging Google Earth Engine platform to characterize and map small seasonal wetlands in the semi-arid environments of South Africa. Sci. Total Environ. 803, 150139.
Hailemichael, G., Catalina, A., González, M., Martin, P., 2016. Relationships between water status, leaf chlorophyll content and photosynthetic performance in Tempranillo vineyards. S. Afr. J. Enol. Vitic. 37, 149–156.
Hendriks, S., 2014. Food security in South Africa: Status quo and policy imperatives. Agrekon 53, 1–24.
Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., 2022. Land cover classification in an era of big and open data: Optimizing localized implementation and training data selection to improve mapping outcomes. Remote Sens. Environ. 268, 112780.
Hu, Q., Yin, H., Friedl, M.A., You, L., Li, Z., Tang, H., Wu, W., 2021. Integrating coarse-resolution images and agricultural statistics to generate sub-pixel crop type maps and reconciled area estimates. Remote Sens. Environ. 258, 112365.
Jassas, H., Kanoua, W., Merkel, B., 2015. Actual evapotranspiration in the Al-Khazir Gomal Basin (Northern Iraq) using the surface energy balance algorithm for land (SEBAL) and water balance. Geosciences 5, 141–159.
Jozdani, S.E., Johnson, B.A., Chen, D., 2019. Comparing deep neural networks, ensemble classifiers, and support vector machine algorithms for object-based urban land use/land cover classification. Remote Sens. (Basel) 11, 1713.
Kavhu, B., Mashimbye, Z.E., Luvuno, L., 2022. Characterising social-ecological drivers of landuse/cover change in a complex transboundary basin using singular or ensemble machine learning. Remote Sens. Appl.: Soc. Environ. 27, 100773.
Labadarios, D., Davids, Y.D., Mchiza, Z., Weir-Smith, G., 2009. The Assessment of Food Insecurity in South Africa. Centre for Poverty, Employment and Growth Human Sciences Research Council, Pretoria, South Africa.
Landmann, T., Eidmann, D., Cornish, N., Franke, J., Siebert, S., 2019. Optimizing harmonics from Landsat time series data: The case of mapping rainfed and irrigated agriculture in Zimbabwe. Remote Sens. Lett. 10, 1038–1046.
LeDell, E., N. Gill, S. Aiello, A. Fu, A. Candel, C. Click, T. Kraljevic, T. Nykodym, P. Aboyoun, and M. Kurka. 2018. Package 'h2o'. dim 2:17.
Li, Z., Ding, L., Xu, D., 2022. Exploring the potential role of environmental and multi-source satellite data in crop yield prediction across Northeast China. Sci. Total Environ. 815, 152880.
Li, Q., Wang, C., Zhang, B., Lu, L., 2015. Object-based crop classification with Landsat-MODIS enhanced time-series data. Remote Sens. (Basel) 7, 16091–16107.
Liu, T., Abd-Elrahman, A., Morton, J., Wilhelm, V.L., 2018. Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. GIScience Remote Sens. 55, 243–264.
Liu, Y., Qian, J., Yue, H., 2021. Comprehensive evaluation of Sentinel-2 red edge and shortwave-infrared bands to estimate soil moisture. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 7448–7465.
Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS J. Photogramm. Remote Sens. 152, 166–177.
Magidi, J., Nhamo, L., Mpandeli, S., Mabhaudhi, T., 2021. Application of the random forest classifier to map irrigated areas using google earth engine. Remote Sens. (Basel) 13, 876.
Mariotto, I., Thenkabail, P.S., Huete, A., Slonecker, E.T., Platonov, A., 2013. Hyperspectral versus multispectral crop-productivity modeling and type discrimination for the HyspIRI mission. Remote Sens. Environ. 139, 291–305.
McCarty, D.A., Kim, H.W., Lee, H.K., 2020. Evaluation of light gradient boosted machine learning technique in large scale land use and land cover classification. Environments 7, 84.
Meza, I., Eyshi Rezaei, E., Siebert, S., Ghazaryan, G., Nouri, H., Dubovyk, O., Gerdener, H., Herbert, C., Kusche, J., Popat, E., Rhyner, J., Jordaan, A., Walz, Y., Hagenlocher, M., 2021. Drought risk for agricultural systems in South Africa: Drivers, spatial patterns, and implications for drought risk management. Sci. Total Environ. 799, 149505.
Mhawej, M., Faour, G., 2020. Open-source Google Earth Engine 30-m evapotranspiration rates retrieval: The SEBALIGEE system. Environ. Model. Softw. 133, 104845.
Millard, K., Richardson, M., 2015. On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping. Remote Sens. (Basel) 7, 8489–8515.
Morley, P.J., Jump, A.S., West, M.D., Donoghue, D.N.M., 2020. Spectral response of chlorophyll content during leaf senescence in European beech trees. Environmental Research Communications 2, 071002.
Mpakairi, K.S., Dube, T., Dondofema, F., Dalu, T., 2022a. Spatial characterisation of vegetation diversity in groundwater-dependent ecosystems using in-situ and Sentinel-2 MSI satellite data. Remote Sens. (Basel) 14, 2995.
Mpakairi, K.S., Dube, T., Dondofema, F., Dalu, T., 2022b. Spatio–temporal variation of vegetation heterogeneity in groundwater dependent ecosystems within arid environments. Eco. Inform. 69, 101667.
Mpakairi, K.S., Muvengwi, J., 2019. Night-time lights and their influence on summer night land surface temperature in two urban cities of Zimbabwe: A geospatial perspective. Urban Clim. 29, 100468.
Mudereri, B.T., Abdel-Rahman, E.M., Dube, T., Niassy, S., Khan, Z., Tonnang, H.E., Landmann, T., 2021. A two-step approach for detecting Striga in a complex agroecological system using Sentinel-2 data. Sci. Total Environ. 762, 143151.
Mutanga, O., Adam, E., Cho, M.A., 2012. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. Int. J. Appl. Earth Obs. Geoinf. 18, 399–406.
Netshipale, A.J., Oosting, S.J., Mashiloane, M.L., Van Reenen, C., De Boer, I.J., Raidimi, E.N., 2020. Agriculture in land reform farms: Impact on livelihoods of beneficiaries in the Waterberg district, South Africa. Land Use Policy 97, 104710.
Ngadze, F., Mpakairi, K.S., Kavhu, B., Ndaimani, H., Maremba, M.S., 2020. Exploring the utility of Sentinel-2 MSI and Landsat 8 OLI in burned area mapping for a heterogenous savannah landscape. PLoS One 15, e0232962.
Nortjé, C., 2020. Census 2017: an overview of commercial agriculture. Oilseeds Focus 6, 46–47.
Ntshangase, N.L., Muroyiwa, B., Sibanda, M., 2018. Farmers' perceptions and factors influencing the adoption of no-till conservation agriculture by small-scale farmers in Zashuke, KwaZulu-Natal Province. Sustainability 10, 555.
Nyam, Y., Kotir, J., Jordaan, A., Ogundeji, A., Turton, A., 2020. Drivers of change in sustainable water management and agricultural development in South Africa: a participatory approach. Sustain. Water Resources Manage. 6, 1–20.
Oliphant, A., P. Thenkabail, and P. Teluguntla. 2022. Global food-security-support-analysis data at 30-m resolution (GFSAD30) cropland-extent products—Download Analysis. US Geological Survey, pp. 2331–1258.
Paganini, N., Adinata, K., Buthelezi, N., Harris, D., Lemke, S., Luis, A., Koppelin, J., Karriem, A., Ncube, F., Nervi Aguirre, E., 2020. Growing and eating food during the COVID-19 pandemic: Farmers' perspectives on local food system resilience to shocks in Southern Africa and Indonesia. Sustainability 12, 8556.
Peng, Y., Gitelson, A.A., 2012. Remote estimation of gross primary productivity in soybean and maize based on total crop chlorophyll content. Remote Sens. Environ. 117, 440–448.
Phalke, A.R., Özdoğan, M., Thenkabail, P.S., Erickson, T., Gorelick, N., Yadav, K., Congalton, R.G., 2020. Mapping croplands of Europe, Middle East, Russia, and Central Asia using Landsat, Random Forest, and Google Earth Engine. ISPRS J. Photogramm. Remote Sens. 167, 104–122.
Pittman, K., Hansen, M.C., Becker-Reshef, I., Potapov, P.V., Justice, C.O., 2010. Estimating global cropland extent with multi-year MODIS data. Remote Sens. (Basel) 2, 1844–1863.
Portmann, F.T., Siebert, S., Döll, P., 2010. MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. Global Biogeochem. Cycles 24.
Potapov, P., Turubanova, S., Hansen, M.C., Tyukavina, A., Zalles, V., Khan, A., Song, X.-P., Pickens, A., Shen, Q., Cortez, J., 2022. Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century. Nature Food 3, 19–28.
Rahman, A., Abdullah, H.M., Tanzir, M.T., Hossain, M.J., Khan, B.M., Miah, M.G., Islam, I., 2020. Performance of different machine learning algorithms on satellite image classification in rural and urban setup. Remote Sens. Appl.: Soc. Environ. 20, 100410.
Ritchie, H., and Roser, M. (2013). Land use. Our World in Data.
Rodriguez-Galiano, V.F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P.M., Jeganathan, C., 2012. Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. Remote Sens. Environ. 121, 93–107.
SA, S. 2017. Census of Commercial Agriculture, 2017. Statistics South Africa.
Sa, S., 2021. Towards measuring the extent of food security in South Africa: An examination of hunger and food adequacy. Statistics South Africa, Pretoria, South Africa.
Saleem, M.H., Potgieter, J., Arif, K.M., 2021. Automation in agriculture by machine and deep learning techniques: A review of recent developments. Precis. Agric. 22, 2053–2091.
Schoeman, F., Newby, T., Thompson, M., Van den Berg, E.C., 2013. South African national land-cover change map. South African J. Geomat. 2, 94–105.
Sharma, R.C., Tateishi, R., Hara, K., 2016. A biophysical image compositing technique for the global-scale extraction and mapping of barren lands. ISPRS Int. J. Geo Inf. 5, 225.
Shimeles, A., Verdier-Chouchane, A., Boly, A., 2018. Introduction: Understanding the Challenges of the Agricultural Sector in Sub-Saharan Africa. In: Shimeles, A., Verdier-Chouchane, A., Boly, A. (Eds.), Building a Resilient and Sustainable Agriculture in Sub-Saharan Africa. Springer International Publishing, Cham, pp. 1–12.
Sidike, P., Sagan, V., Maimaitijiang, M., Maimaitiyiming, M., Shakoor, N., Burken, J., Mockler, T., Fritschi, F.B., 2019. dPEN: Deep Progressively Expanded Network for mapping heterogeneous agricultural landscape using WorldView-3 satellite imagery. Remote Sens. Environ. 221, 756–772.
Siebert, S., Döll, P., Hoogeveen, J., Faures, J.-M., Frenken, K., Feick, S., 2005. Development and validation of the global map of irrigation areas. Hydrol. Earth Syst. Sci. 9, 535–547.
Smith, J.H., Stehman, S.V., Wickham, J.D., Yang, L., 2003. Effects of landscape characteristics on land-cover class accuracy. Remote Sens. Environ. 84, 342–349.
Sun, Z., Di, L., Fang, H., 2019. Using long short-term memory recurrent neural network in land cover classification on Landsat and Cropland data layer time series. Int. J. Remote Sens. 40, 593–614.
Sun, H., Wang, B., Wu, Y., Yang, H., 2023. Deep learning method based on spectral characteristic rein-forcement for the extraction of winter wheat planting area in complex agricultural landscapes. Remote Sens. (Basel) 15, 1301.

Suyker, A.E., Verma, S.B., 2009. Evapotranspiration of irrigated and rainfed maize–soybean cropping systems. Agric. For. Meteorol. 149, 443–452.
Team, R.C. 2014. R: A Language and Environment for Statistical Computing. http://www.R-project.org.
Teluguntla, P., Thenkabail, P.S., Oliphant, A., Xiong, J., Gumma, M.K., Congalton, R.G., Yadav, K., Huete, A., 2018. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. ISPRS J. Photogramm. Remote Sens. 144, 325–340.
Thompson, N. C., K. Greenewald, K. Lee, and G. F. Manso. 2020. The computational limits of deep learning. arXiv preprint arXiv:2007.05558.
Useya, J., Chen, S., 2019. Exploring the potential of mapping cropping patterns on smallholder scale croplands using Sentinel-1 SAR Data. Chin. Geogr. Sci. 29, 626–639.
Useya, J., Chen, S., Murefu, M., 2019. Cropland mapping and change detection: toward Zimbabwean cropland inventory. IEEE Access 7, 53603–53620.
Van Koppen, B., Nhamo, L., Cai, X., Gabriel, M.J., Sekgala, M., Shikwambana, S., Tshikolomo, K., Nevhutanda, S., Matlala, B., Manyama, D., 2017. Smallholder irrigation schemes in the Limpopo Province. International Water Management Institute (IWMI), South Africa.
Van Niekerk, A., Jarmain, C., Goudriaan, R., 2018. An earth observation approach towards mapping irrigated area and quantifying water use by irrigated crops in South Africa. Water Research Commission.
Waldner, F., Canto, G.S., Defourny, P., 2015. Automated annual cropland mapping using knowledge-based temporal features. ISPRS J. Photogramm. Remote Sens. 110, 1–13.
Waldner, F., De Abelleyra, D., Verón, S.R., Zhang, M., Wu, B., Plotnikov, D., Bartalev, S., Lavreniuk, M., Skakun, S., Kussul, N., 2016. Towards a set of agrosystem-specific cropland mapping methods to address the global cropland diversity. Int. J. Remote Sens. 37, 3196–3231.
Waldner, F., Hansen, M.C., Potapov, P.V., Löw, F., Newby, T., Ferreira, S., Defourny, P., 2017. National-scale cropland mapping based on spectral-temporal features and outdated land cover information. PLoS One 12, e0181911.
Wulder, M.A., Franklin, S.E., White, J.C., Cranny, M.M., Dechka, J.A., 2004. Inclusion of topographic variables in an unsupervised classification of satellite imagery. Can. J. Remote. Sens. 30, 137–149.
Wulder, M.A., Coops, N.C., Roy, D.P., White, J.C., Hermosilla, T., 2018. Land cover 2.0. Int. J. Remote Sens. 39, 4254–4284.
Xia, T., He, Z., Cai, Z., Wang, C., Wang, W., Wang, J., Hu, Q., Song, Q., 2022. Exploring the potential of Chinese GF-6 images for crop mapping in regions with complex agricultural landscapes. Int. J. Appl. Earth Obs. Geoinf. 107, 102702.
Xing, N., Huang, W., Xie, Q., Shi, Y., Ye, H., Dong, Y., Wu, M., Sun, G., Jiao, Q., 2019. A transformed triangular vegetation index for estimating winter wheat leaf area index. Remote Sens. (Basel) 12, 16.
Xu, Y., Yu, L., Peng, D., Cai, X., Cheng, Y., Zhao, J., Zhao, Y., Feng, D., Hackman, K., Huang, X., Lu, H., Yu, C., Gong, P., 2018. Exploring the temporal density of Landsat observations for cropland mapping: experiments from Egypt, Ethiopia, and South Africa. Int. J. Remote Sens. 39, 7328–7349.
Yang, N., Liu, D., Feng, Q., Xiong, Q., Zhang, L., Ren, T., Zhao, Y., Zhu, D., Huang, J., 2019. Large-scale crop mapping based on machine learning and parallel computation with grids. Remote Sens. (Basel) 11, 1500.
Yobe, C.L., Mudhara, M., Mafongoya, P., 2019. Livelihood strategies and their determinants among smallholder farming households in KwaZulu-Natal province, South Africa. Agrekon 58, 340–353.
Yue, J., Tian, J., Tian, Q., Xu, K., Xu, N., 2019. Development of soil moisture indices from differences in water absorption between shortwave-infrared bands. ISPRS J. Photogramm. Remote Sens. 154, 216–230.
Zamani Joharestani, M., Cao, C., Ni, X., Bashir, B., Talebiesfandarani, S., 2019. PM2. 5 prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data. Atmos. 10, 373.
Zeng, H., Wu, B., Wang, S., Musakwa, W., Tian, F., Mashimbye, Z.E., Poona, N., Syndey, M., 2020. A synthesizing land-cover classification method based on Google Earth Engine: A case study in Nzhelele and Levhuvu catchments, South Africa. Chin. Geogr. Sci. 30, 397–409.
Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geosci. Remote Sens. Mag. 5, 8–36.
Zhu, Z., Waller, E., 2003. Global forest cover mapping for the United Nations Food and Agriculture Organization forest resources assessment 2000 program. For. Sci. 49, 369–380.