Article

A Comparison of Machine Learning Models for Mapping Tree Species Using WorldView-2 Imagery in the Agroforestry Landscape of West Africa

Muhammad Usman [1], Mahnoor Ejaz [2], Janet E. Nichol [3], Muhammad Shahid Farid [2,*], Sawaid Abbas [1,4] and Muhammad Hassan Khan [2]

1   Centre for Geographical Information System, University of the Punjab, Lahore 54590, Pakistan; sawaid.gis@pu.edu.pk or sawaid.abbbas@connect.polyu.hk (S.A.)
2   Department of Computer Science, Faculty Computing & Information Technology, University of the Punjab, Lahore 54590, Pakistan
3   Department of Geography, School of Global Studies, University of Sussex, Brighton BN1 9RH, UK
4   Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong
*   Correspondence: shahid@pucit.edu.pk

Abstract: Farmland trees are a vital part of the local economy as trees are used by farmers for fuelwood as well as food, fodder, medicines, fibre, and building materials. As a result, mapping tree species is important for ecological, socio-economic, and natural resource management. The study evaluates very high-resolution remotely sensed WorldView-2 (WV-2) imagery for tree species classification in the agroforestry landscape of the Kano Close-Settled Zone (KCSZ), Northern Nigeria. Individual tree crowns extracted by geographic object-based image analysis (GEOBIA) were used to remotely identify nine dominant tree species (*Faidherbia albida, Anogeissus leiocarpus, Azadirachta indica, Diospyros mespiliformis, Mangifera indica, Parkia biglobosa, Piliostigma reticulatum, Tamarindus indica, and Vitellaria paradoxa*) at the object level. For every tree object in the reference datasets, eight original spectral bands of the WV-2 image, their spectral statistics (minimum, maximum, mean, standard deviation, etc.), spatial, textural, and color-space (hue, saturation), and different spectral vegetation indices (VI) were used as predictor variables for the classification of tree species. Nine different machine learning methods were used for object-level tree species classification. These were Extra Gradient Boost (XGB), Gaussian Naïve Bayes (GNB), Gradient Boosting (GB), K-nearest neighbours (KNN), Light Gradient Boosting Machine (LGBM), Logistic Regression (LR), Multi-layered Perceptron (MLP), Random Forest (RF), and Support Vector Machines (SVM). The two top-performing models in terms of highest accuracies for individual tree species classification were found to be SVM (overall accuracy = 82.1% and Cohen's kappa = 0.79) and MLP (overall accuracy = 81.7% and Cohen's kappa = 0.79) with the lowest numbers of misclassified trees compared to other machine learning methods.

Keywords: agroforestry; machine learning; Sudano-Sahelian; tree species mapping; WorldView-2

1. Introduction

In the semi-arid Sudano-Sahelian ecological zone of West Africa, trees maintained by farmers on their farmed plots are an important element of the local livelihood [1]. Farmers use trees for fuelwood for their own use as well as for sale to supplement farm incomes. Wood fuel in Kano has traditionally been derived from trees grown and maintained by farmers in the farmed parklands surrounding the city. The 'parkland' landscape is defined by the large variety of trees grown and maintained on farmland, which are used for a very wide variety of purposes, including fuel wood, timber for building materials, food, fodder, fibre, and medicines [1,2]. Additionally, the large areal extent of farmed parkland landscapes in the Sudano-Sahelian ecological zone makes them an important component of the global climate system, as they store and sequester large amounts of carbon in the woody

biomass and soils [3,4]. Higher demand for fuelwood due to high population growth combined with predictions of higher temperatures and decreased rainfall pose a serious challenge to tree stocks. Therefore, spatial, and quantitative assessments of tree species are especially urgent since climate change and intensified land use in recent decades have put increasing pressure on tree cover.

In addition to pixel-based approaches to classifying tree species, object-based image analysis is effective for classifying objects at multiple scales. This means that tree crowns of different sizes can be delineated separately, from individual tree crowns to large clusters of tree crowns. Numerous studies have found high accuracy and low error for the classification of tree species using GEOBIA compared to pixel-based approaches [5,6]. Over the last few years, there has been enormous development in remote sensing with the launch of high-spatial-resolution commercial satellites. The use of traditional statistical analysis of single pixels is not appropriate for high resolution satellite images, as the pixel under consideration and its neighbouring pixels may differ spectrally but belong to the same land cover class [7]. This high spectral variability within the same land cover class creates a "salt-and-pepper" effect during classification. As human beings normally recognise patterns in a landscape by their spatial relationship to neighbourhood objects, it is useful to use spatial and contextual information for the characterisation of land use classes, along with spectral information [8]. Spatial relationships between adjacent pixels in the form of texture provide important information for identification of individual objects, which are building blocks of the original features of interest [9]. In this way, homogeneous objects based on spatially connected groups of pixels with similar spectral characteristics can be identified. Image segmentation is the process by which homogeneous image objects are created by aggregating groups of pixels with regard to spectral and spatial characteristics. The term 'homogeneous' implies that within-object variance is low compared to that between objects, and those identified objects also contain additional information about geometry (size and shape), contextual, and textural aspects besides spectral information [10]. These homogeneous objects reflect real-world objects of interest [7].

Many studies have used high-spatial-resolution satellite images for tree crown delineation [5,11,12]. Bunting and Lucas [5] extracted and classified different tree crown species in Australian mixed forests using the Compact Airborne Spectrographic Imager (CASI) hyperspectral data through GEOBIA. Rasmussen et al. [12] used QuickBird imagery for extracting tree crowns in Northern Senegal, and Karlson et al. [11] used WorldView-2 data for tree cover extraction in Burkina Faso using GEOBIA. In an agroforestry landscape, there is a variety of deciduous trees with varying crown sizes and ages; therefore, GEOBIA is well suited for such tree crown cover mapping. Remote sensing has been successfully used for tree species mapping using airborne hyperspectral systems [13,14], but the high cost and small footprint of these airborne systems restrict their usage for large areas. Therefore, there has been a growing interest in the use of very high resolution space-based satellite remote sensing images for the identification of tree species [6,13,15–17] because they provide timely, repetitive, and large area coverage from local to global scales. Karlson et al. [15] investigated the capability of multi-seasonal WorldView-2 imagery to map five dominant tree species at the object level in central Burkina Faso using the Random Forest (RF) classifier.

There have been many studies using satellite data with different machine learning methods for tree species classification from satellite data, including Support Vector Machine (SVM) [6,18], K-nearest neighbours (KNN) [19–21], Random Forest (RF) [6,22,23], Logistic Regression (LR) [20,24], Extra Gradient Boosting (XGBoost) [25–27], Multi-Layered Perceptron (MLP) [28–30], Light Gradient Boosting (LightGBM) [25], Gaussian Naïve Bayes (GNB) [31], Gradient Boosting (GB) [32]. However, most studies are based on very few species, and the accuracy levels achieved are generally not above 80%. For example, Karlson et al. [15] tested five (only four native) tree species with one machine-learning classifier, Random Forest. Producer accuracy in the dry season was below 80%, except for the distinctive species *M. indica*, which has dark green shiny leaves, and a non-native

Eucalyptus species, which has a very distinctive compact crown and blue-green leaves. Lelong et al. [16] examined two machine learning algorithms, SVM and RF, but achieved a relatively low kappa index of 0.71 for identifying four tree species in Senegal. Most previous studies have also combined different sensing systems, such as optical satellite images combined with Lidar or airborne images, or combined multiple dates. To the best of our knowledge, no study has compared different machine-learning methods for tree species classification. Moreover, our study uses only a single sensor.

The objective of the study is to test and evaluate a cost-effective method for detailed tree species classification in the agroforestry landscape of West Africa, using Kano, Nigeria, as a case study. WorldView-2 imagery is used, as a single image covers a large area at a high level of detail, and airborne or UAV imagery is not generally available in a developing country environment. An evaluation of nine different machine learning methods is performed to suggest the best-performing and most cost-effective method for detailed classification of tree species over large areas. This will enable effective rural afforestation programmes, which need an accurate inventory of existing tree stocks, and contribute to a sustainable rural economy where farm trees have multiple and diverse uses.

2. Materials and Methods

2.1. Study Area

The Sudan zone of West Africa is densely populated, with rural population densities of 300–500 persons per km^2 surrounding Kano, Nigeria's second city and the largest city in savanna Africa. Kano has some of the highest rural population densities in the world. The number of persons per km^2 almost doubled from 169 in 1991 [33] to 308 in 2006 [34]. The 'Kano Close-Settled Zone' (KCSZ) of Northern Nigeria [35] surrounding the city describes the densely populated agricultural region influenced by the proximity of Kano and serving as its hinterland in terms of interdependency of products and trade, goods, and services.

The study was conducted over an area of 100 km^2 in the intensively farmed parklands surrounding Kano, extending westwards from Kano city at 11.97° N, 8.39° E (Figure 1). The study site is typical of the Northern Sudan zone of West Africa, where over 80% of the land is cultivated in the April–September rainy season. The main crops are the cereals: maize (Guinea corn), millet, and sorghum, which are grown for subsistence, along with a few field crops of root vegetables, beans, and a few vegetables. The mean annual rainfall of 750 mm at Kano supports a natural vegetation of tree savanna, with flat-topped trees browsed by savanna fauna and livestock when the grassy ground cover dries during the winter dry season [36,37]. A wide variety of tree species on farmland is necessary because each species provides a particular product, ranging from food, fodder, medicine, building materials, and fencing to wood fuel. Although most species can be burned for cooking and heating, the different burning properties of species also have different purposes. The large tree numbers, with up to 25 trees per hectare in the study area [36], make field surveys of large areas impossible. However, due to the importance of trees in the local economy and culture, species inventories are highly desirable in order to understand trends in the face of external climatic and economic threats.

2.2. WorldView-2 Satellite Data

This study used a cloud-free WorldView-2 image from February 2, 2014. WV-2 has eight multispectral bands at 2 m and a panchromatic band at 0.5 m resolution (Table 1). The image covers an area of 100 km^2. For this study, pan-sharpened images at 0.5 m resolution were produced using the Hyperspherical Colour Space (HCS) method [38] by fusing multispectral bands at 2 m with the panchromatic band at 0.5 m.
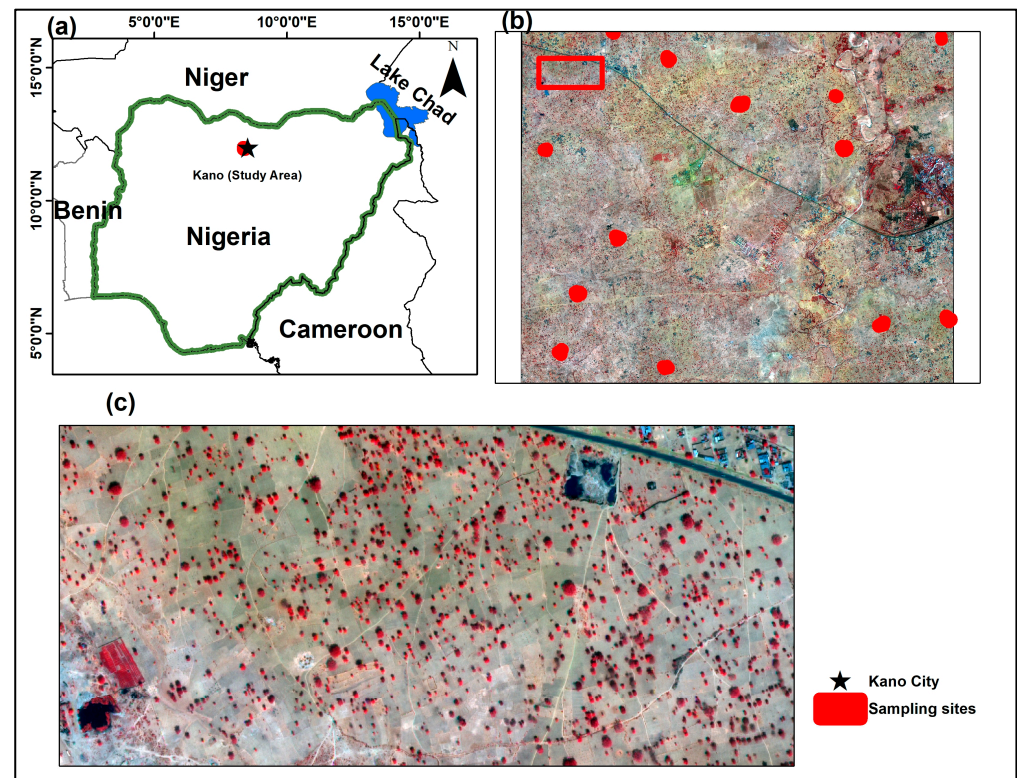
Figure 1. (**a**) Location of study area; (**b**) WV-2 image of study area with location of the sampled sites in the field; (**c**) zoomed-in view of a pansharpened WV-2 image.

Table 1. Specifications of the WorldView-2 satellite image.

Image Parameters		Bands (μm)
Acquisition date	2 February 2014	Coastal Blue (0.40–0.45)
Acquisition time	10:24:00	Blue (0.45–0.51)
Off-nadir angle	26.06	Green (0.51–0.58)
Mean sun azimuth	139.50	Yellow (0.58–0.62)
Mean sun elevation	60.40	Red (0.63–0.69)
Cloud cover (%)	0	Red Edge (0.705–0.745)
Map projection	UTM WGS 84	NIR 1 (0.77–0.89)
Location:		
NW (Lat, Long)	(12.01, 8.34)	NIR 2 (0.86–1.04) Pan (0.45–0.80)
SE (Lat, Long)	(11.92, 8.43)	

2.3. Reference Field Inventory Data

Fieldwork conducted in the study area during the 2015–16 dry season provides the basis for the identification of individual tree species. A fallow agricultural field with different tree species in the study area is shown in Figure 2. Field data collected included the enumeration of 210 trees (for trees >5 cm in diameter), the height of trees, the measurement of Diameter at Breast Height (DBH), as well as species identification. Nine of the most common tree species sampled in the west area were *Faidherbia albida*, *Anogeissus leiocarpus*, *Azadirachta indica*, *Diospyros mespiliformis*, *Mangifera indica*, *Parkia biglobosa*, *Piliostigma reticulatum*, *Tamarindus indica, and Vitellaria paradoxa* (Table 2). Every individual tree was manually located on a colour print of a Worldview-2 (WV-2) pansharpened image of 0.5 m resolution. Clusters of tree crowns were not included. From this, a GIS-based point shapefile of field tree locations was generated for further analysis.

Figure 2. A fallow agricultural field with different tree species in the Kano Close Settled Zone, Northern Nigeria, during the dry season (January 2016).

Table 2. Dominant agroforestry tree species, number of tree species in the sample, and their crown dimensions.

Scientific Name	Common Name	Number	Min (m^2)	Max (m^2)	Mean (m^2)	Stdev (m^2)
Faidherbia albida	Gawo	15	17.75	172.8	73.5	51.5
Anogeissus leiocarpus	Marke	24	2.625	267.2	33.3	51.5
Mangifera indica	Mango	11	30.6	230.4	92.8	50.6
Azadirachta indica	Neem	70	6.4	131.3	54.7	30.5
Parkia biglobosa	African locust bean	19	13.125	464.6	99.3	105.2
Tamarindus indica	Tsamiya	16	11.5	158	66.9	43.7
Vitellaria paradoxa	Kadanya	13	8.75	85.4	43.9	21.7
Piliostigma reticulatum	Kalgo	25	4.25	116.1	34.8	26.9
Diospyros mespiliformis	African ebony	17	8.37	107.7	45.1	30.5

2.4. Tree Crown Delineation Using GEOBIA

In this study, tree crown areas were extracted from WV-2 data using Geographic Object-Based Image Analysis (GEOBIA) (Figure 3) by modifying the method of Bunting and Lucas [5], which they proposed for tree crown delineation in Australian mixed-species forests. For accuracy assessment of tree crown area delineated through GEOBIA using WV-2 data, an independent reference tree crown area measured during a field survey was compared with satellite image-based tree crown area by a linear regression line. A significant value of $R^2 = 0.88$ was found [37].

Individual tree crowns extracted by GEOBIA were overlaid onto the WV-2 image, and mean spectral values of the eight bands were extracted for a species-specific spectral library. Different vegetation indices based on different combinations of the novel bands of WV-2 were used as predictors in combination with the original spectral bands. Previous studies have shown improvements in the accuracy of tree species discrimination by using vegetation indices [15,39,40]. For every tree object in the reference datasets, the predictor variables for classification were the eight original spectral bands of the WorldView-2 image, their spectral statistics (minimum, maximum, mean, standard deviation, etc.), spatial, textural, color-space (hue, saturation), and different spectral vegetation indices (Table 3). The selection of appropriate feature variables plays a vital role in obtaining excellent

classification results. In order to screen the computed variables in our proposed algorithm, we used Principal Component Analysis (PCA) which reduces the dimension of the feature vector by mapping the original high-dimensional data onto a compact low-dimensional space. Similar to [41,42], we investigated the impact of PCA to reduce the dimensionality of our features by preserving the maximum variance in the data. However, we observed that reducing the dimensions of the proposed features would result in dropping some important information, hence a decline in the recognition accuracy. Therefore, we used all the computed features in the classification.
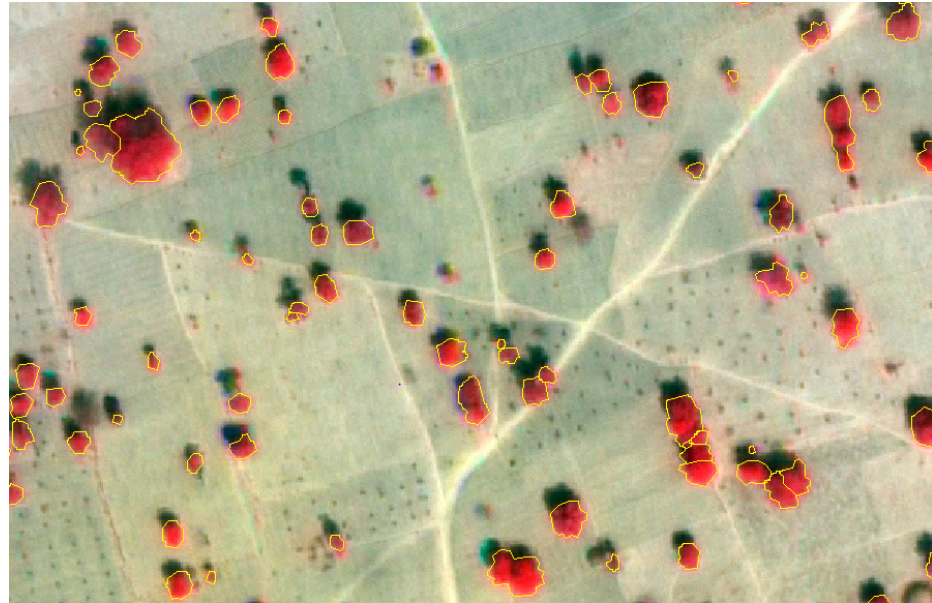

Figure 3. A subset of the WorldView-2 imagery (dry season) acquired over KCSZ is shown as a false-colour RGB composite consisting of NIR 1, red, and green bands, and tree crowns extracted using GEOBIA.

Table 3. Vegetation indices-based predictor used for tree species classification.

Vegetation Indices	Formula	Source
Normalised Difference Vegetation Index	$\dfrac{NIR - Red}{NIR + Red}$	[43]
Visible Atmospherically Resistant Index	$\dfrac{G - Red}{G + Red - B}$	[44]
Normalised Difference Vegetation Index–Red Edge	$\dfrac{RE - Red}{RE + Red}$	[45]
Anthocyanin Reflectance Index	$\dfrac{1}{G} - \dfrac{1}{Red}$	[46]
Modified Anthocyanin Reflectance Index	$\left(\dfrac{1}{G} - \dfrac{1}{Red}\right) * NIR$	[47]
Chlorophyll Index Green	$\dfrac{NIR2}{G} - 1$	[48]
Chlorophyll Index Red Edge	$\dfrac{NIR2}{RE} - 1$	[49]

Table 3. *Cont.*

Vegetation Indices	Formula	Source
Normalised Difference NIR/Red	$\dfrac{\text{NIR2} - \text{Red}}{\text{NIR2} + \text{Red}}$	[50]
Pigment Specific Simple Ratio	$\dfrac{\text{NIR1}}{\text{Red}}$	[51]
Chlorophyll Vegetation Index	$\left(\dfrac{\text{Red}}{\text{G}^2}\right) * \text{NIR2}$	[44]
Green Difference Vegetation Index	$\dfrac{\text{NIR2} - \text{G}}{\text{NIR2} + \text{G}}$	[52]
Enhanced Vegetation Index	$\dfrac{2.5 * (\text{NIR2} - \text{Red})}{(\text{NIR2} + 6 * \text{Red} - 7.5 * \text{Coastal Blue}) + 1}$	[53]
Enhanced Vegetation Index 2	$\dfrac{2.4 * (\text{NIR2} - \text{Red})}{\text{NIR2} + \text{Red} + 1}$	[54]
Enhanced Vegetation Index 2-2	$\dfrac{2.5 * (\text{NIR2} - \text{Red})}{\text{NIR2} + 2.4 * \text{Red} + 1}$	[53]
Green Atmospherically Resistant Vegetation Index	$\dfrac{\text{NIR2} - (\text{Green} - (\text{Coastal Blue} - \text{Red}))}{\text{NIR2} - (\text{Green} + (\text{Coastal Blue} - \text{Red})}$	[55]
Green Optimised Soil Adjusted Vegetation Index	$\dfrac{\text{NIR2} - \text{Green}}{\text{NIR2} + \text{Green} + 0.16}$	[56]
Infrared Percentage Vegetation Index	$\dfrac{2 * \text{NIR2}}{\text{NIR2} + \text{Red}} * \left(\left(\dfrac{\text{Red} - \text{Green}}{\text{Red} + \text{Green}}\right) + 1\right)$	[57]
Blue-Wide Dynamic Range Vegetation Index	$\dfrac{0.1 * \text{NIR2} - \text{Coastal Blue}}{0.1 * \text{NIR2} + \text{Coastal Blue}}$	[58]
Optimised Soil-Adjusted Vegetation Index	$\dfrac{(1 + 0.16)(\text{NIR1} - \text{Red})}{\text{NIR1} + \text{Red} + 0.16}$	[59]
Modified Soil-Adjusted Vegetation Index	$\dfrac{2 * \text{NIR2} + 1 - \sqrt{(2 * \text{NIR2} + 1)^2 - (\text{NIR2} - \text{Red})}}{2}$	[47]
Discriminant Normalised Vegetation Index	$\dfrac{\text{Coastal Blue} - \text{Blue}}{\sqrt{\text{Coastal Blue} + \text{Coastal Blue}}}$	[60]
Modified Normalised Difference Vegetation Index	$\dfrac{\text{NIR1} - \text{Red}}{\text{NIR1} + \text{Red} - 2 * \text{Coastal Blue}}$	[47]
Plant Pigment Ratio	$\dfrac{\text{G} - \text{Coastal Blue}}{\text{G} + \text{Coastal Blue}}$	[61]
Structure Intensive Pigment Index	$\dfrac{\text{NIR1} - \text{Coastal Blue}}{\text{NIR1} + \text{Red}}$	[62]
Modified Simple Ratio	$\dfrac{\text{NIR1}}{\text{Red}}$	[60]
Photosynthetic Vigour Ratio	$\dfrac{\text{G} - \text{Red}}{\text{G} + \text{Red}}$	[61]

2.5. Preprocessing

As all the predictor features were numeric and scaled differently, some features with higher values could have dominated the results of the machine learning algorithms. So, scaling was applied to bring all the values to a single scale. Standard scaling was applied,

which considers the data to be normally distributed for each feature and then scales it using the following formula:

$$\text{Xnew} = \frac{x - \mu}{\sigma}$$

such that the distribution of values becomes centered at 0 and its standard deviation becomes 1. Finally, each feature's scaling is performed independently by calculating the mean (μ) and standard deviation (σ). After standard scaling, the standard deviations of predictors become one, which allows the min-max scalar to perform better.

The label encoding technique is applied to convert reference data of tree species names to a numeric format, which is a more readable format for the machines. In this process, each tree species in the label data is assigned a value from zero onwards. As the reference tree species data were few in quantity and imbalanced in the frequency of each species, as well as the high occurrence of certain tree species, this could cause wrong predictions and overfitting of machine learning models. Therefore, Synthetic Minority Over-sampling Technique (SMOTE) was used to up-sample the data and add more data points for scarce classes [63,64]. In SMOTE, synthetic data for the minority class were generated from its nearest neighbours using Euclidean distance. Each label in the data was sampled according to its occurrence in the original data. On the basis of features, newly generated data is similar to the original data. As a result of resampling, the number of data points increased from 210 to 325 (Figure S1). Individual tree species with low frequencies increased after applying SMOTE sampling, as shown in Supplementary Table S1. Due to the difference in frequency of tree species, the stratified k-fold cross-validation technique was used to handle the imbalanced distribution of classes in training and testing. Because of the limited reference dataset (n = 325), we performed successive random stratified 4-fold cross-validation, which randomly divides the reference dataset into 4 folds or groups having equal proportions of different tree species. The first fold (80 trees, or 25% of the trees in each class) is treated as a validation set, and the model is trained on the remaining 75% of trees of different species. The process is repeated 4 times, considering a different fold each time for the validation dataset, and the final accuracies are derived using summed values. Validation of tree species classification employed the producer, user, and overall accuracies and Cohen's kappa coefficient K and F1 scores as measures for the different machine learning methods.

2.6. Machine Learning Methods

The study is designed to provide a comparison of machine learning classification algorithms to assess which performs best for classifying tree species in the agroforestry landscape of the Sudan zone of West Africa. Machine learning tasks were performed in Python using libraries such as scikit-learn, geopandas, rasterio, earthpy, etc., that are known for the processing of geographic data, and Jupyter notebooks.

2.6.1. Support Vector Machine (SVM)

Support Vector Machine (SVM) [65] is suitable for regression, classification, and outlier detection problems. The SVM takes an n-dimensional space, where n refers to the number of features, and a data point is plotted so that the combined features give out the coordinate. Further, classification is done by finding a hyperplane that segregates the classes. It is also known as a large margin classifier. For tree species classification, SVM is among the best-performing models [6,18]; therefore, it was chosen for this study. As parameters of the implementation are provided by sklearn, its default Radial Basis kernel function is used in this study, and the value of regularisation, C, is set to 100, as the best results were obtained from this combination through tuning.

2.6.2. K-Nearest Neighbours (KNN)

The K-Neighbour Classifier (KNN) [66] is a supervised machine learning technique for both regression and classification challenges. It is advantageous in terms of easy

implementation, but its performance is degraded when data size increases. It works on the simple rule of identifying the k-data points nearest to the test data and assigning this test point to the cluster to which most of its neighbours belong. KNN is one of the algorithms widely used for tree classification [19–21]. In our case, we have used grid search to select the optimal k-value, which is 3. Apart from that, the weights parameter was set to 'uniform', and the metric for distance computation was set to Minkowski.

2.6.3. Random Forest (RF)

The RF [67] is an ensemble machine learning technique that combines regression and classification models with bagging. It comprises several combined decision trees, giving each tree a class prediction. In the end, the results from all trees are combined, and the class with the greatest vote count is given as the final prediction. The RF classifier has been used with and compared with various machine learning algorithms for tree classification [6,22,23]. As for the parametric values, the number of trees was set to 100, the criterion to measure the split quality was 'gini', max_depth was set to none, and max_features to be considered was set to 'sqrt', because tuning showed these values worked best for our dataset.

2.6.4. Logistic Regression (LR)

LR is a predictive analysis algorithm [68] that uses a sigmoid as a cost function and is based on the simple concept of probabilities to classify data into various categories, either binary or multi-class, by using some extensions. Due to its simple implementation, this algorithm has been used for tree species classification [20,24]. We have used 'l2' as a penalty, tolerance for stopping criteria as, 1×10^{-4} the inverse of regularisation strength C as 1.0, and the solver as 'lbfgs' for optimisation. The selection of these values was based on results obtained through grid-search.

2.6.5. Extra Gradient Boosting (XGBoost)

A member of the gradient boosting library, XGBoost is an optimised, portable, and flexible machine learning algorithm built on the framework of Gradient Boosting. In addition, tree-boosting in parallel is provided by XGBoost, which gives out accurate and fast performance. The algorithm was introduced as a research project as a part of the Distributed (Deep) Machine Learning Community [69]. Tree-ensemble techniques are one of the best-performing models for tree classification [25–27]. In terms of the input parameters, through grid-search, max_depth was set to 3, learning_rate to 0.1, number of estimators to 100, and objective to 'binary:logistic', as they were shown to work the best.

2.6.6. Multi-Layer Perceptron (MLP)

MLP provides a way to map input and output in a non-linear way. It comprises one or more hidden layers, apart from the input and output layers. Moreover, the neurons must have an activation function to apply a threshold. In addition, these algorithms can be categorised as feed-forward algorithms. MLP has been used for tree classification [28–30] and has produced higher accuracy. For this particular dataset, the MLP classifier provided by the sklearn library has been used, whereas, through hyperparameter tuning, the hidden_layer_sizes were selected to be (200,100), the activation_function was set to 'relu', 'adam' as a solver for weight optimisation, the strength of L2 regularisation, i.e., alpha, to 0.0001, and learning_rate to 0.001 as a constant.

2.6.7. Light Gradient Boosting (LightGBM)

Another method from the gradient-boosting family, LightGBM, increases decision tree efficiency while minimising memory usage. Two characteristic techniques in this algorithm, namely, Exclusive Feature Bundling (EFB) and gradient-based one-side sampling, boost the training speed by 20 times [70]. As tree-based algorithms performed well for tree classification [25], Ge et al. [71] made use of a variant of LightGBM for a similar study in which they classified oolong tea using hyperspectral imaging data and were able to attain

97.33% prediction accuracy. For this study, the parameter values for light GBM were set to 0.1 for the learning rate, 150 for the number of estimators, and 20 for the subsample binning.

2.6.8. Gaussian Naïve Bayes (GNB)

Gaussian Naïve Bayes is a simple probabilistic algorithm based on the Bayes principle. It assumes that each class follows the normal or gaussian distribution and that there is no dependence between dimensions. Although simple, this algorithm is found to work effectively for sophisticated datasets and classification problems [31]. The parameters of this model were selected through tuning because there were no prior probabilities of classes available, such as 1e-9 for var_smoothing.

2.6.9. Gradient Boosting (GB)

The gradient boosting algorithm works on the tree-based boosting approach and builds up an additive model, moving forward iteratively and optimising the loss functions, which are log losses in the case of classification. Additionally, boosting techniques were found to work well on tree species classification datasets [32]. For the implementation provided by the sklearn library, the learning_rate was set to 0.1, the number of estimators to 100, the max_depth of the tree was set to 3, and the loss function was chosen as log_loss.

3. Results

Machine Learning Model Comparison

Tree species classification was performed by using different machine learning methods and incorporating different predictors, including spectral, spatial, colour-based (hue, saturation), and vegetation indices (Table 4). Initial classification using only spectral band information gave very low accuracy.

Table 4. Overall tree species classification accuracy, kappa coefficient, and total misclassified trees by using different machine learning methods.

Method	Overall Accuracy (OA)	Kappa	Number of Trees Misclassified (%)
XGB	76.2	0.72	23.4
GNB	52.4	0.46	46.5
RF	77.60	0.74	21.8
GB	75.1	0.71	24.3
LR	77.3	0.74	22.2
KNN	67.2	0.63	32
SVM	82.1	0.79	17.5
MLP	81.7	0.79	17.8
LGBM	79.8	0.76	19.7

Figure 4 compares the accuracy of different machine learning methods by using spatial, textural, and colour features only, spectral features only, and combining spectral, spatial, textural, and colour features together. Results show that using all predictor variables together gave the best overall result. Among the models, the highest accuracy values were achieved by MLP and SVM, and logistic regression was the third-best method. Gaussian Naive Bayes shows consistently low performance in all cases and appears to be the worst-performing method. Overall, it is observed that the tree-based models perform best for such datasets, along with the MLP.
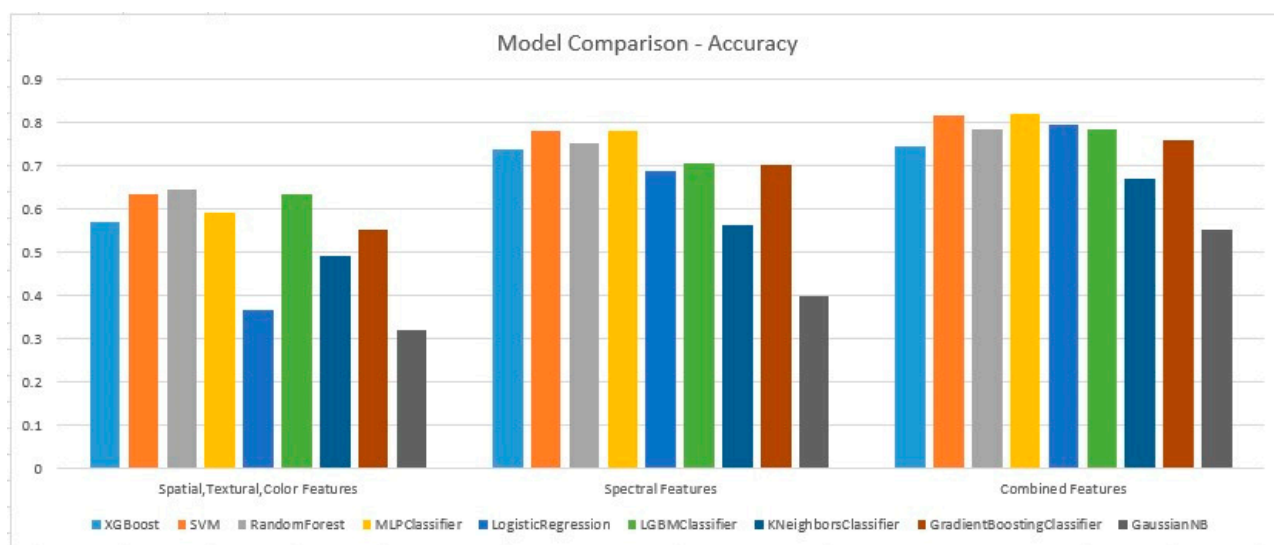
Figure 4. Comparison of the accuracy of different machine learning methods using different predictors. The *y*-axis represents the kappa coefficient.

The confusion matrices (Tables 5 and 6) were derived using the random stratified 4-fold resampling cross-validation technique, which randomly divides the reference dataset into 4 folds, where 25% of the reference data is treated as a validation set and the model is trained on the remaining 75% of reference trees. By using different validation datasets of about 80 trees each four times, the resulting confusion matrices are based on a total of 320 reference tree crowns. The results show that among the nine machine learning methods, the MLP and SVM methods were the most accurate (OA = 81.7% for MLP and OA = 82.1 for SVM), with the least number of misclassified trees. The producers' accuracy in Table 5 represents what percentage of each species were correctly classified, and the user's accuracy represents reliability, or the percentage of other species that were wrongly included in a species class. The MLP and SVM classifiers gave similar results, showing very good agreement between reference and classification for most of the tree species. The exceptions were *Parkia biglobosa* and *Piliostigma reticulatum*, with low producer's accuracy of 71.4%, and 72.5%, respectively, meaning that only 71.4% and 72.5% of known trees of these species were identified as such. This is in line with Karlson et al.'s findings in Burkina Faso [15], which show low producer accuracy of 73% for *P. biglobosa* in the dry season, whereas higher producer accuracy (81%) was observed for this species on wet season WV-2 imagery. The three species *Mangifera indica*, *Azadirachta indica*, and *Vitellaria paradoxa* were well identified by both classifiers, with producer's accuracies of over 85%, along with *Anogeissus leicarpus* for SVM. For the remaining 4/3 species for MLP/SVM, both classifiers had high producer accuracy of at least 80%, representing the percentage of trees correctly classified as that species. In terms of user accuracy, for both classifiers, *Mangifera indica* had the highest user accuracy of 88.4%/92% for MLP/SVM, meaning that only 11.6%/8% of this species were wrongly classified as other species, or 88.4%/92% of this species identified as *M. indica* do belong to that category on the ground. This tree is distinctive in its solid crown and dark green, shiny leaves (Figure S3). Significant confusion occurred between *P. biglobosa* and *A. indica*, which resulted in lower user accuracy for both classifiers. In fact, 20% of *P. biglobosa* trees were omitted from the class and wrongly classified mainly as *A. indica* and *P. reticulatum*. These three species have similar-shaped, rounded crowns that are flat at the base (Figure S3), and although *P. biglobosa* is generally a much larger tree, the younger of this species may appear similar to these two generally smaller trees. Overall results show that the species least identified by the two classifiers, MLP and SVM, were *P. biglobosa*, followed by *A. leiocarpus*. Those species most commonly identified as other species were *A. indica* and *P. reticulatum*.

Table 7 compares results for all classifiers for all tree species. It indicates that the poorest classifiers were GNB and KNN, in terms of both user and producer accuracy, which were only able to recognise approximately half and two thirds of species, respectively. GNB in particular wrongly allocated approximately half of all species to other classes. Another five classifiers, XGB, RF, GB, LR, and LGBM, achieved above 70% accuracy, and SVM and MLP achieved over 80% accuracy in tree species identification.

The species *Anogeissus leiocarpus* and *Mangifera indica* were mostly correctly classified by the algorithms. The pale green leaves and bark of the former and the dark, shiny leaves of the latter (Figure S3) may explain their distinctiveness on images. Those least correctly classified were *Faidherbia albida* and *Diospyros mespiliformis*. For all species except *F. albida*, *P. biglobosa*, and *P. reticulatum*, the highest F1 score is not less than 80%, indicating that six out of the nine species were highly classified by one model or the other.

The SVM classifier can be considered robust, and we applied it to a set of trees extracted from WV-2 imagery over the whole study area. Figure 5 shows a map of the spatial distribution of tree species in a small portion of the KCZS.
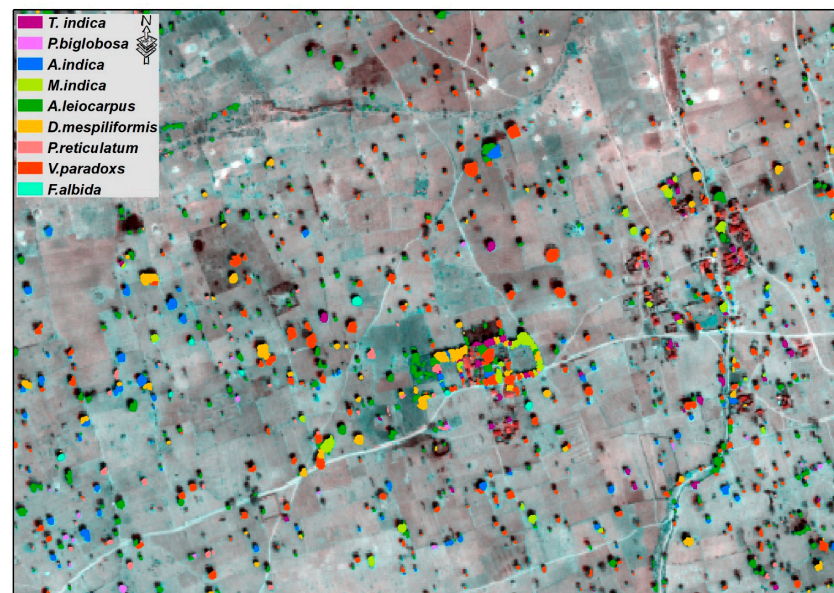

Figure 5. Map of the spatial distribution of tree species in a small portion of the study area.

Table 5. MLP-based confusion matrix showing producer and user accuracy for nine dominant tree species. Row sums are equal to the number of sampled tree species after applying SMOTE.

Reference Classified	Producer's Accuracy (%)	User's Accuracy (%)	*Faidherbia albida*	*Anogeissus leiocarpus*	*Mangifera indica*	*Azadirachta indica*	*Parkia biglobosa*	*Tamarindus indica*	*Vitellaria paradoxa*	*Piliostigma reticulatum*	*Diospyros mespiliformis*
Faidherbia albida	80	77.4	24	1	1	3	1	0	0	0	0
Anogeissus leiocarpus	80	77.7	2	28	0	1	0	0	1	2	1
Mangifera indica	92	88.4	0	0	23	1	1	0	0	0	0
Azadirachta indica	87.1	87.1	1	0	2	61	5	1	0	0	0
Parkia biglobosa	71.4	78.1	2	0	0	4	25	0	0	3	1
Tamarindus indica	83.3	80.6	2	0	0	0	0	25	0	0	3
Vitellaria paradoxa	90	85.7	0	0	0	0	0	1	18	0	1
Piliostigma reticulatum	72.5	80.5	0	6	0	0	0	1	2	29	2
Diospyros mespiliformis	81.2	76.4	0	1	0	0	0	3	0	2	26

Table 6. Support Vector Machine (SVM)-based confusion matrix showing producer and user accuracy for nine dominant tree species.

Reference Classified	Producer's Accuracy (%)	User's Accuracy (%)	*Faidherbia albida*	*Anogeissus leiocarpus*	*Mangifera indica*	*Azadirachta indica*	*Parkia biglobosa*	*Tamarindus indica*	*Vitellaria paradoxa*	*Piliostigma reticulatum*	*Diospyros mespiliformis*
Faidherbia albida	80	72.7	24	1	1	2	2	0	0	0	0
Anogeissus leiocarpus	85.7	83.3	2	30	0	1	0	0	1	1	0
Mangifera indica	92	92	0	0	23	2	0	0	0	0	0
Azadirachta indica	85.7	81	3	0	1	60	5	0	0	0	1
Parkia biglobosa	71.4	78.1	1	0	0	7	25	0	0	2	0
Tamarindus indica	83.3	89.2	2	0	0	0	0	25	0	0	3
Vitellaria paradoxa	90	85.7	0	0	0	0	0	1	18	1	0
Piliostigma reticulatum	72.5	82.8	1	4	0	2	0	0	1	29	3
Diospyros mespiliformis	81.2	78.7	0	1	0	0	0	2	1	2	26

Table 7. Species-wise user, producer accuracy, and F1-score by different learning methods.

Models	Accuracy	Faidherbia albida	Anogeissus leiocarpus	Mangifera indica	Azadirachta indica	Parkia biglobosa	Tamarindus indica	Vitellaria paradoxa	P. retculatum	Diospyros mespiliformis	Mean
XGB	Producer Accuracy	56.6	88.5	80	81.4	74.2	76.6	85	77.5	59.3	75.4
	User Accuracy	73.9	79.4	76.9	80.2	68.4	82.1	80.9	73.8	65.5	75.6
	F1-score	0.64	0.83	0.78	0.8	0.71	0.79	0.82	0.75	0.62	0.74
GNB	Producer Accuracy	10	68.5	88	62.8	37.1	73.3	65	45	21.8	52.3
	User Accuracy	10.7	53.3	88	78.5	36.1	52.3	86.6	40	28	52.6
	F1-score	0.1	0.6	0.88	0.69	0.36	0.61	0.74	0.42	0.24	0.51
RF	Producer Accuracy	60	88.5	80	87.1	71.4	83.3	75	75	65.6	76.2
	User Accuracy	69.2	86.1	83.3	81.3	78.1	80.6	71.4	68.1	75	77
	F1-score	0.64	0.87	0.81	0.84	0.74	0.81	0.73	0.71	0.7	0.76
GB	Producer Accuracy	63.3	80	72	88.5	71.4	63.3	70	80	65.6	72.6
	User Accuracy	63.3	87.5	90	75.6	67.5	90.4	82.3	69.5	65.6	76.8
	F1-score	0.63	0.83	0.8	0.81	0.69	0.74	0.75	0.74	0.65	0.73
LR	Producer Accuracy	66.6	74.2	92	77.1	77.1	86.6	85	75	68.7	78
	User Accuracy	64.5	72.2	88.4	88.5	69.2	74.2	85	75	75.8	76.9
	F1-score	0.65	0.73	0.9	0.82	0.72	0.8	0.85	0.75	0.72	0.77
KNN	Producer Accuracy	70	68.5	92	70	68.5	73.3	80	37.5	59.3	68.7
	User Accuracy	51.2	64.8	67.6	87.5	64.8	66.6	72.7	68.1	54.2	66.3
	F1-score	0.59	0.66	0.77	0.77	0.66	0.69	0.76	0.48	0.56	0.66

Table 7. *Cont.*

Models	Accuracy	*Faidherbia albida*	*Anogeissus leiocarpus*	*Mangifera indica*	*Azadirachta indica*	*Parkia biglobosa*	*Tamarindus indica*	*Vitellaria paradoxa*	*P. retculatum*	*Diospyros mespiliformis*	Mean
SVM	Producer Accuracy	80	85.7	92	85.7	71.4	83.3	90	72.5	81.2	82.4
	User Accuracy	72.7	83.3	92	81	78.1	89.2	85.7	82.8	78.7	82.6
	F1-score	0.76	0.84	0.92	0.83	0.74	0.86	0.87	0.77	0.8	0.82
MLP	Producer Accuracy	80	80	92	87.1	71.4	83.3	90	72.5	81.2	81.9
	User Accuracy	77.4	77.7	88.4	87.1	78.1	80.6	85.7	80.5	76.4	81.3
	F1-score	0.78	0.78	0.9	0.87	0.74	0.81	0.87	0.76	0.78	0.81
LGBM	Producer Accuracy	76.6	85.7	88	82.8	74.2	76.6	85	72.5	78.1	79.9
	User Accuracy	76.6	81	81.4	81.6	74.2	79.3	89.4	76.3	80.6	80
	F1-score	0.76	0.83	0.84	0.82	0.74	0.77	0.87	0.74	0.79	0.79

4. Discussion

Previous research has shown that remote sensing based tree species mapping in tropical dryland ecosystems is possible with the help of machine learning methods. The accuracy obtained by this study exceeds that of most other studies, uses only a single date remote sensing image (WorldView-2), and examines a larger number of different species than previous studies. The study identified SVM and MLP as the most accurate machine learning methods, with an OA of 82% and K = 0.79 that were substantially higher than the other methods tested. A study in Senegal by Lelong et al. [16] examined two machine learning methods, SVM and RF, and found SVM to have higher accuracy. Additionally, similar to our study, they found SVM to be superior to RF for a small number and unequal distribution of samples. The accuracies achieved by our study compare well with other similar studies, such as Lelong et al. [16], whose results were based on only four species and had a highest kappa index of 0.71, and Karlson et al. [11], whose results were limited to only four indigenous tree species and used a single machine learning method, Random Forest. They achieved accuracies of OA = 78% and K = 0.74 for dry season imagery. While our results are significantly better than these, Karlson et al. [11] did obtain accuracies comparable to ours (with an OA of 83% and K = 0.76) when multi-seasonal imagery was used, but for only four different tree species.

The current study demonstrates that accurate species identification can be achieved with machine learning methods for a range of species in agroforestry landscapes. The tree species studied here are among the most important species in the West African agroforestry landscape. *Parkia biglobosa*, which is used for soup stock as well as fibre, and *Faidherbia albida*, used for dry-season fodder, have been shown in a recent study by Usman et al. [72] to be fast declining.

The sample data for some species were limited (Table 2) in this study. However, the SMOTE sampling was used to increase the frequency of individual tree species with low frequencies (Supplementary Table S1), thereby avoiding overfitting of machine learning models. For example, four tree species—Mangifera indica, Vitellaria paradoxa, Faidherbia albida, and Tamarindus indica—have tree counts less than 30 trees. After applying SMOTE, the tree counts for those species substantially increased (Supplementary Table S1), which avoided wrong predictions and overfitting of machine learning models. In previous studies, e.g., Lelong et al. [16] had less than 30 field samples for three out of six species sampled, and Karlson et al. [11] had less than 10 field samples for three out of five native species sampled. In our study, out of nine species that we sampled, the average number of field samples was 24, although the method of compensation (SMOTE) we used increased the overall number from 210 to 325, as mentioned in Supplementary Figure S1. Nonetheless, that study was restricted in scope as it used field measurement to obtain data over limited areas, compared with the over 100 km^2 covered by the single WV2 image used in this study.

The other species studied here, including *Azadirachta indica*, *Piliostigma reticulatum*, *Anogessus leiocarpus*, and *Diospiros mespiliformis*, were shown by Usman et al.'s [36] study to be actively regenerating. These four important fuelwood species were identified by MPL with 87/72/80/81% accuracy by MLP, and 86/72/86/81% accuracy, respectively, by the SVM classifier. The huge dependence on wood as fuel in Nigeria, where the latest available figures (National Bureau of Statistics, 2011) suggest that 95% of the energy used for cooking is from wood, may explain the increased abundance of these species. The *Vitellaria paradoxum*, or shea butter, tree provides emollients and fats for a wide range of modern food, medicinal, and cosmetic products. This species was identified with 90% accuracy by the two best classifiers, MLP and SVM. Such trends need to be documented accurately over large areas in order to understand and manage possible threats to the local economy, as well as identify opportunities for growth.

WorldView-2 and WorldView-3, with their unique spectral band configurations including red edge, near infrared, and shortwave infrared bands at very high spatial resolution, have proved to be capable of mapping tree species in the West African agroforestry landscape [15,16]. As other recent very high resolution sensors such as WV3 and Pleiades have

similar spectral bands, spatial resolution, and swath width to the WV2 images used in this study, little advantage is expected from using them. This study has demonstrated that a single image, along with a robust machine-learning tool such as MLP or SVM, can provide highly accurate tree species inventories over large areas.

5. Conclusions

The comprehensive examination of methods for tree species classification presented here can assist state and rural authorities in undertaking rapid and cost-effective rural surveys of the agroforestry landscape in Nigeria. This will permit a better understanding of the pressures currently facing Nigeria's dryland ecosystems. Violent outbreaks in recent years among migrant pastoralists stemming from land shortages are related to trends in tree species, as declines in *Faidherbia albida*, traditionally used as dry-season fodder, are removed from farmland to counter predation by cattle. Rural households are susceptible to climatic fluctuations and trends as well as the current massive growth in the rural population. Land fragmentation due to traditional inheritance customs requires more farmland trees to supply additional households with wood fuel, as farmers indicate that they rarely buy wood. The disappearance of non-fuelwood species in recent decades is of concern due to their importance in the local household economy.

This study demonstrates two machine-based learning models that provide over 80% accuracy and can be applied to a single date of WorldView-2 imagery, to identify the most common farmland tree species in West African farmed parkland. Rural afforestation programmes would benefit from accurate inventories of current stocks of tree species and their regional variations.

Due to the longevity of trees, it is unlikely that such tree species inventories as described by this study would need to be repeated on a regular basis. More important would be to extend the survey to wider areas and repeat it perhaps once per decade. However, because a farmed parkland landscape with the same tree species exists throughout the semi-arid zone of West Africa and, to a lesser extent, southern Africa, the findings should have broader applicability.

References

1. Boffa, J.M. *Agroforestry Parkland in Sub-Saharan Africa: FAO Conservation Guide 34*; Food and Agriculture Organization (FAO): Rome, Italy, 1999.
2. Timberlake, J.; Chidumayo, E.; Sawadogo, L. Distribution and Characteristics of African Dry Forests and Woodlands. In *The Dry Forests and Woodlands of Africa: Managing for Products and Services*; Routledge: Oxford, UK, 2010; pp. 11–41.
3. Karlson, M.; Ostwald, M.; Reese, H.; Sanou, J.; Tankoano, B.; Mattsson, E. Mapping Tree Canopy Cover and Aboveground Biomass in Sudano-Sahelian Woodlands Using Landsat 8 and Random Forest. *Remote Sens.* **2015**, *7*, 10017–10041. [CrossRef]
4. Lal, R. Carbon Sequestration in Dryland Ecosystems. *Environ. Manag.* **2004**, *33*, 528–544. [CrossRef] [PubMed]
5. Bunting, P.; Lucas, R. The Delineation of Tree Crowns in Australian Mixed Species Forests Using Hyperspectral Compact Airborne Spectrographic Imager (CASI) Data. *Remote Sens. Environ.* **2006**, *101*, 230–248. [CrossRef]
6. Immitzer, M.; Atzberger, C.; Koukal, T. Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. *Remote Sens.* **2012**, *4*, 2661–2693. [CrossRef]
7. Blaschke, T.; Strobl, J. What's Wrong with Pixels? Some Recent Developments Interfacing Remote Sensing and GIS. *GIS—Zeitschrift Geoinformationssysteme* **2001**, *14*, 12–17.
8. Blaschke, T. Object Based Image Analysis for Remote Sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16. [CrossRef]
9. Thomas, N.; Hendrix, C.; Congalton, R.G. A Comparison of Urban Mapping Methods Using High-Resolution Digital Imagery. *Photogramm. Eng. Remote Sens.* **2003**, *69*, 963–972. [CrossRef]
10. Laliberte, A.S.; Rango, A.; Havstad, K.M.; Paris, J.F.; Beck, R.F.; McNeely, R.; Gonzalez, A.L. Object-Oriented Image Analysis for Mapping Shrub Encroachment from 1937 to 2003 in Southern New Mexico. *Remote Sens. Environ.* **2004**, *93*, 198–210. [CrossRef]
11. Karlson, M.; Reese, H.; Ostwald, M. Tree Crown Mapping in Managed Woodlands (Parklands) of Semi-Arid West Africa Using WorldView-2 Imagery and Geographic Object Based Image Analysis. *Sensors* **2014**, *14*, 22643–22669. [CrossRef]
12. Rasmussen, M.O.; Göttsche, F.M.; Diop, D.; Mbow, C.; Olesen, F.S.; Fensholt, R.; Sandholt, I. Tree Survey and Allometric Models for Tiger Bush in Northern Senegal and Comparison with Tree Parameters Derived from High Resolution Satellite Data. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 517–527. [CrossRef]
13. Cho, M.A.; Mathieu, R.; Asner, G.P.; Naidoo, L.; van Aardt, J.; Ramoelo, A.; Debba, P.; Wessels, K.; Main, R.; Smit, I.P.J.; et al. Mapping Tree Species Composition in South African Savannas Using an Integrated Airborne Spectral and LiDAR System. *Remote Sens. Environ.* **2012**, *125*, 214–226. [CrossRef]
14. Feret, J.B.; Asner, G.P. Tree Species Discrimination in Tropical Forests Using Airborne Imaging Spectroscopy. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 73–84. [CrossRef]
15. Karlson, M.; Ostwald, M.; Reese, H.; Bazié, H.R.; Tankoano, B. Assessing the Potential of Multi-Seasonal WorldView-2 Imagery for Mapping West African Agroforestry Tree Species. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *50*, 80–88. [CrossRef]
16. Lelong, C.C.D.; Tshingomba, U.K.; Soti, V. Assessing Worldview-3 Multispectral Imaging Abilities to Map the Tree Diversity in Semi-Arid Parklands. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *93*, 102211. [CrossRef]
17. Madonsela, S.; Cho, M.A.; Mathieu, R.; Mutanga, O.; Ramoelo, A.; Kaszta, Ż.; Van De Kerchove, R.V.; Wolff, E. Multi-Phenology WorldView-2 Imagery Improves Remote Sensing of Savannah Tree Species. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *58*, 65–73. [CrossRef]
18. Li, D.; Ke, Y.; Gong, H.; Li, X. Object-Based Urban Tree Species Classification Using Bi-Temporal Worldview-2 and Worldview-3 Images. *Remote Sens.* **2015**, *7*, 16917–16937. [CrossRef]
19. Wu, Y.; Zhang, X. Object-Based Tree Species Classification Using Airborne Hyperspectral Images and LiDAR Data. *Forests* **2020**, *11*, 32. [CrossRef]
20. Terryn, L.; Calders, K.; Disney, M.; Origo, N.; Malhi, Y.; Newnham, G.; Raumonen, P.; Åkerblom, M.; Verbeeck, H. Tree Species Classification Using Structural Features Derived from Terrestrial Laser Scanning. *ISPRS J. Photogramm. Remote Sens.* **2020**, *168*, 170–181. [CrossRef]
21. Zhang, C.; Xia, K.; Feng, H.; Yang, Y.; Du, X. Tree Species Classification Using Deep Learning and RGB Optical Images Obtained by an Unmanned Aerial Vehicle. *J. For. Res.* **2021**, *32*, 1879–1888. [CrossRef]
22. Raczko, E.; Zagajewski, B. Comparison of Support Vector Machine, Random Forest and Neural Network Classifiers for Tree Species Classification on Airborne Hyperspectral APEX Images. *Eur. J. Remote Sens.* **2017**, *50*, 144–154. [CrossRef]
23. Sabat-Tomala, A.; Raczko, E.; Zagajewski, B. Comparison of Support Vector Machine and Random Forest Algorithms for Invasive and Expansive Species Classification Using Airborne Hyperspectral Data. *Remote Sens.* **2020**, *12*, 516. [CrossRef]
24. Waser, L.T.; Küchler, M.; Jütte, K.; Stampfer, T. Evaluating the Potential of Worldview-2 Data to Classify Tree Species and Different Levels of Ash Mortality. *Remote Sens.* **2014**, *6*, 4515–4545. [CrossRef]
25. Łoś, H.; Mendes, G.S.; Cordeiro, D.; Grosso, N.; Costa, H.; Benevides, P.; Caetano, M. Evaluation of XGBoost and LGBM Performance in Tree Species Classification with Sentinel-2 Data. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 5803–5806.
26. You, H.; Huang, Y.; Qin, Z.; Chen, J.; Liu, Y. Forest Tree Species Classification Based on Sentinel-2 Images and Auxiliary Data. *Forests* **2022**, *13*, 1416. [CrossRef]
27. Wan, H.; Tang, Y.; Jing, L.; Li, H.; Qiu, F.; Wu, W. Tree Species Classification of Forest Stands Using Multisource Remote Sensing Data. *Remote Sens.* **2021**, *13*, 144. [CrossRef]

28. Nezami, S.; Khoramshahi, E.; Nevalainen, O.; Pölönen, I.; Honkavaara, E. Tree Species Classification of Drone Hyperspectral and RGB Imagery with Deep Learning Convolutional Neural Networks. *Remote Sens.* **2020**, *12*, 1070. [CrossRef]

29. Sumsion, G.R.; Bradshaw, M.S.; Hill, K.T.; Pinto, L.D.G.; Piccolo, S.R. Remote Sensing Tree Classification with a Multilayer Perceptron. *PeerJ* **2019**, *2019*, e6101. [CrossRef]

30. Cetin, Z.; Yastikli, N. The Use of Machine Learning Algorithms in Urban Tree Species Classification. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 226. [CrossRef]

31. Padao, F.R.F.; Maravillas, E.A. Using Naïve Bayesian Method for Plant Leaf Classification Based on Shape and Texture Features. In Proceedings of the 2015 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), Cebu, Philippines, 9–12 December 2015. [CrossRef]

32. Mäyrä, J.; Keski-Saari, S.; Kivinen, S.; Tanhuanpää, T.; Hurskainen, P.; Kullberg, P.; Poikolainen, L.; Viinikka, A.; Tuominen, S.; Kumpula, T.; et al. Tree Species Classification from Airborne Hyperspectral and LiDAR Data Using 3D Convolutional Neural Networks. *Remote Sens. Environ.* **2021**, *256*, 112322. [CrossRef]

33. Tiffen, M. *Profile of Demographic Change in the Kano-Maradi Region, 1960–2000*; Drylands Research: Riverside, CA, USA, 2001.

34. National Population Commission (NCP). *Nigerian Population Census Report*; National Population Commission (NCP): Calverton, MD, USA, 2006.

35. Wilson, M.M.; Wilson, J. *Land and People in the Kano Close-Settled Zone: A Survey of Some Aspects of Rural Economy in Ungogo District, Kano Province: A Report to the Greater Kano Planning Authority*; Paper No. 1; Ahmadu Bello University, Department of Geography: Zaria, Nigeria, 1965.

36. Usman, M.; Nichol, J.E. Remarkable Increase in Tree Density and Fuelwood Production in the Croplands of Northern Nigeria. *Land Use Policy* **2018**, *78*, 410–419. [CrossRef]

37. Usman, M. Modelling Woody Vegetation in Sudano-Sahelian Zone of Nigeria Using Remote Sensing. Ph.D. Thesis, The Hong Kong Polytechnic University, Hong Kong, China, 2018.

38. Padwick, C.; Scientist, P.; Deskevich, M.; Pacifici, F.; Smallwood, S. WorldView-2 Pan-Sharpening. In Proceedings of the Asprs 2010, San Diego, CA, USA, 26–30 April 2010; Volume 48, pp. 26–30.

39. Naidoo, L.; Cho, M.A.; Mathieu, R.; Asner, G. ISPRS Journal of Photogrammetry and Remote Sensing Classification of Savanna Tree Species, in the Greater Kruger National Park Region, by Integrating Hyperspectral and LiDAR Data in a Random Forest Data Mining Environment. *ISPRS J. Photogramm. Remote Sens.* **2012**, *69*, 167–179. [CrossRef]

40. Pu, R.; Landry, S. Remote Sensing of Environment A Comparative Analysis of High Spatial Resolution IKONOS and WorldView-2 Imagery for Mapping Urban Tree Species. *Remote Sens. Environ.* **2012**, *124*, 516–533. [CrossRef]

41. Peng, X.; Wang, L.; Wang, X.; Qiao, Y. Bag of Visual Words and Fusion Methods for Action Recognition: Comprehensive Study and Good Practice. *Comput. Vis. Image Underst.* **2016**, *150*, 109–125. [CrossRef]

42. Khan, M.H.; Farid, M.S.; Grzegorzek, M. A Comprehensive Study on Codebook-Based Feature Fusion for Gait Recognition. *Inf. Fusion* **2023**, *92*, 216–230. [CrossRef]

43. Gitelson, A.A.; Merzlyak, M.N. Remote Estimation of Chlorophyll Content in Higher Plant Leaves. *Int. J. Remote* **1997**, *18*, 2691–2697. [CrossRef]

44. Raymond Hunt, E.; Daughtry, C.S.T.; Eitel, J.U.H.; Long, D.S. Remote Sensing Leaf Chlorophyll Content Using a Visible Band Index. *Agron. J.* **2011**, *103*, 1090–1099. [CrossRef]

45. Ehammer, A.; Fritsch, S.; Conrad, C.; Lamers, J.; Dech, S. Statistical Derivation of FPAR and LAI for Irrigated Cotton and Rice in Arid Uzbekistan by Combining Multi-Temporal RapidEye Data and Ground Measurements. *Remote Sens. Agric. Ecosyst. Hydrol. XII* **2010**, *7824*, 782409. [CrossRef]

46. Gitelson, A.A. Non-Destructive and Remote Sensing Techniques for Estimation of Vegetation Status. In Proceedings of the 3rd European Conference on Precision Agriculture, Montpelier, France, 18–20 June 2001; pp. 205–210.

47. Main, R.; Cho, M.A.; Mathieu, R.; O'Kennedy, M.M.; Ramoelo, A.; Koch, S. An Investigation into Robust Spectral Indices for Leaf Chlorophyll Estimation. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 751–761. [CrossRef]

48. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a Green Channel in Remote Sensing of Global Vegetation from EOS- MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [CrossRef]

49. Gitelson, A.; Merzlyak, M.N. Quantitative Estimation of Chlorophyll-a Using Reflectance Spectra: Experiments with Autumn Chestnut and Maple Leaves. *J. Photochem. Photobiol. B Biol.* **1994**, *22*, 247–252. [CrossRef]

50. Underwood, E.; Ustin, S.; DiPietro, D. Mapping Nonnative Plants Using Hyperspectral Imagery. *Remote Sens. Environ.* **2003**, *86*, 150–161. [CrossRef]

51. Zarco-Tejada, P.J.; Miller, J.R.; Noland, T.L.; Mohammed, G.H.; Sampson, P.H. Scaling-up and Model Inversion Methods with Narrowband Optical Indices for Chlorophyll Content Estimation in Closed Forest Canopies with Hyperspectral Data. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 1491–1507. [CrossRef]

52. Buschmann, C.; Nagel, E. In Vivo Spectroscopy and Internal Optics of Leaves as Basis for Remote Sensing of Vegetation. *Int. J. Remote Sens.* **1993**, *14*, 711–722. [CrossRef]

53. Jiang, Z.; Huete, A.R.; Didan, K.; Miura, T. Development of a Two-Band Enhanced Vegetation Index without a Blue Band. *Remote Sens. Environ.* **2008**, *112*, 3833–3845. [CrossRef]

54. Miura, T.; Yoshioka, H.; Fujiwara, K.; Yamamoto, H. Inter-Comparison of ASTER and MODIS Surface Reflectance and Vegetation Index Products for Synergistic Applications to Natural Resource Monitoring. *Sensors* **2008**, *8*, 2480–2499. [CrossRef]

55. Gitelson, A.A.; Vina, A.; Arkebauer, T.J.; Rundquist, D.C.; Keydan, G.; Leavitt, B. Remote Estimation of Leaf Area Index and Green Leaf Biomass in Maize Canopies. *Geophys. Res. Lett.* **2003**, *30*, 4–7. [CrossRef]
56. Rondeaux, G.; Steven, M.; Baret, F. Optimization of Soil-Adjusted Vegetation Indices. *Remote Sens. Environ.* **1996**, *55*, 95–107. [CrossRef]
57. Kooistra, L.; Leuven, R.S.E.; Wehrens, R.; Nienhuis, P.H.; Buydens, L.M.C. A Comparison of Methods to Relate Grass Reflectance to Soil Metal Contamination. *Int. J. Remote Sens.* **2003**, *24*, 4995–5010. [CrossRef]
58. Hancock, D.W.; Dougherty, C.T. Relationships between Blue- and Red-Based Vegetation Indices and Leaf Area and Yield of Alfalfa. *Crop Sci.* **2007**, *47*, 2547–2556. [CrossRef]
59. Wu, C.; Niu, Z.; Tang, Q.; Huang, W. Estimating Chlorophyll Content from Hyperspectral Vegetation Indices: Modeling and Validation. *Agric. For. Meteorol.* **2008**, *148*, 1230–1241. [CrossRef]
60. Manna, S.; Raychaudhuri, B. Mapping Distribution of Sundarban Mangroves Using Sentinel-2 Data and New Spectral Metric for Detecting Their Health Condition. *Geocarto Int.* **2020**, *35*, 434–452. [CrossRef]
61. Metternicht, G. Vegetation Indices Derived from High-Resolution Airborne Videography for Precision Crop Management. *Int. J. Remote Sens.* **2003**, *24*, 2855–2877. [CrossRef]
62. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A Modified Soil Adjusted Vegetation Index. *Remote Sens. Environ.* **1994**, *48*, 119–126. [CrossRef]
63. Rustogi, R.; Prasad, A. Swift Imbalance Data Classification Using SMOTE and Extreme Learning Machine. In Proceedings of the 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 21–23 February 2019; pp. 1–6. [CrossRef]
64. Umer, M.; Sadiq, S.; Missen, M.M.S.; Hameed, Z.; Aslam, Z.; Siddique, M.A.; NAPPI, M. Scientific Papers Citation Analysis Using Textual Features and SMOTE Resampling Techniques. *Pattern Recognit. Lett.* **2021**, *150*, 250–257. [CrossRef]
65. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1999.
66. Peterson, L.E. K-Nearest Neighbor. *Scholarpedia* **2009**, *21*, 1883. [CrossRef]
67. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
68. Cramer, J.S. The Origins of Logistic Regression. *SSRN Electron. J.* **2002**. Tinbergen Institute Working Paper No. 2002-119/4. [CrossRef]
69. Chen, T.; Guestrin, C. Xgboost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
70. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T. Lightgbm: A Highly Efficient Gradient Boosting Decision Tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 52.
71. Ge, X.; Sun, J.; Lu, B.; Chen, Q.; Xun, W.; Jin, Y. Classification of Oolong Tea Varieties Based on Hyperspectral Imaging Technology and BOSS-LightGBM Model. *J. Food Process Eng.* **2019**, *42*, e13289. [CrossRef]
72. Usman, M.; Nichol, J. Trends in Farmland Tree Stocks in the Agroforestry Landscape of Northern Nigeria: Reconciling Scientific and Stakeholder Perceptions. *J. Rural Stud.* **2019**, *66*, 87–94. [CrossRef]