JARS Journal of Applied Remote Sensing

RESEARCH PAPER

Machine learning-based maize (*Zea mays L.*) extraction at parcel level using Sentinel 2A-derived spectral indices

Bere Benjamin Bantchina[a,*] **and Kemal Sulhi Gündoğdu**[b]

[a]Bursa Uludağ University, Institute of Natural and Applied Sciences, Department of Biosystems Engineering, Bursa, Turkey

[b]Bursa Uludağ University, Department of Biosystems Engineering, Faculty of Agriculture, Bursa, Turkey

ABSTRACT. Crop type classification is crucial for policymaking and precision agriculture applications. This study aimed to develop a parcel-based maize (*Zea mays L.*) extraction approach using Sentinel 2A-derived spectral indices and machine learning (ML) in two distinct study sites: Yeşilova and Ormankadı villages in Bursa Province, Turkey. Employing 13 widely recognized spectral indices, the investigation implemented 4 ML classifiers: support vector machines, random forest, K-nearest neighbors, and bootstrap aggregating. The training-test methodology was explored using two scenarios: Yeşilova as the training set and Ormankadı as the test set, and vice versa. The models calibrated on Yeşilova and validated on Ormankadı maintained the accuracy of the model, with an overall accuracy (OA) ranging from 79.3% to 89.9%, precision between 72.8% and 80.1%, recall between 82.1% and 84.9%, F1-score between 77.4% and 82.2%, and a Matthews correlation coefficient (MCC) ranging from 58.9% to 68.3%. Furthermore, the models consistently demonstrated good performance when Ormankadı served as the training set and Yeşilova as the test set, with commendable OA (78.7% to 84.8%), precision (85.5% to 88.0%), recall (88.0% to 91.1%), F1-score (86.2% to 89.5%), and MCC (68.2% to 76.0%). This study demonstrated the potential of using high-resolution remote sensing and ML for effective maize crop extraction using diverse datasets.

© 2024 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JRS.18.034515]

Keywords: crop type classification; machine learning; maize; hyperspectral remote sensing; Sentinel 2A; spectral indices

Paper 230664G received Dec. 4, 2023; revised Jun. 25, 2024; accepted Jul. 31, 2024; published Aug. 23, 2024.

1 Introduction

The emergence of remote sensing, specifically Sentinel and Landsat satellites, has provided opportunities for detailed research using high-resolution data. Land use/land cover and crop type classifications using satellite remote sensing data and machine learning (ML) approaches have recently gained significant attention. Remote sensing technology is a valuable data source for crop classification and has the ability to detect wide areas.[1] The availability of satellite imagery, such as Sentinel 2 and Landsat 8, with high spatial and spectral resolution, has further enhanced the potential for accurate crop type and land cover classification.[2]

ML algorithms have been widely used for crop type and land cover classification. Support vector machines (SVMs), random forest (RF), and K-nearest neighbors (KNNs) are popular supervised ML algorithms used in crop type classification. These algorithms leverage the spectral

*Address all correspondence to Bere Benjamin Bantchina, ben.bantchina@gmail.com

1931-3195/2024/$28.00 © 2024 SPIE

information extracted from remote sensing data to classify different land cover classes and identify specific crop types.[3–5] For instance, the RF algorithm has been particularly successful in classification tasks, achieving high overall classification accuracy, as demonstrated by Refs. 6–8. The combination of hyperspectral remote sensing data and ML has enabled accurate crop type and land cover classification. Using time-series imagery from multiple Earth observation satellites has improved classification accuracy, especially for monitoring crop growth dynamics. Integrating polarimetric synthetic-aperture radar data with optical remote sensing data has shown promising crop monitoring and classification results.[2] Increasing the computing power and decreasing the cost of consumer computing have contributed to the popularity of ML methods in remote sensing.[9]

Despite advancements in remote sensing technology and ML techniques, crop type and land cover classification still need to be improved. The presence of persistent cloud cover and small agricultural field sizes can limit the improvement in classification accuracy.[10] The mixed-pixel problem, in which a single pixel contains multiple land cover or crop types, adds complexity to the classification process. In addition, seasonal variations in crop growth patterns and land-cover dynamics require robust algorithms that can adapt to temporal changes.

Numerous studies have focused on the classification of crop type and land cover using datasets obtained from the same locations.[11–15] Therefore, the examined literature demonstrates the significance of employing ML and classification algorithms and integrating remote sensing data with crop models to classify crop types. Hyperspectral imaging and deep learning algorithms have been identified as viable methods for crop classification. Various technologies and methodologies exist for accurately mapping crops using remote sensing data. These studies collectively emphasize the significance of utilizing diverse remote sensing datasets, including satellite imagery and aerial photography, to accurately classify crop types and land cover.

However, only a few studies, like the one referenced in Ref. 16, have explored the classification approach using datasets from different geographical locations. To overcome this research gap, further research is required to develop effective crop type classification approaches using high-resolution remote sensing data and advanced ML algorithms.

Different remote sensing approaches can be used to monitor vegetation and primary production dynamics. Land cover and crop type classifications using remote sensing data and ML techniques have gained considerable and continuous attention in the scientific community. This study aims to explore a new approach to enhance remote sensing research using high-resolution satellite-derived vegetation indices (VIs). The study evaluates the capability of well VIs, including the normalized difference vegetation index (NDVI), normalized difference water index (NDWI), green NDVI (GNDVI), atmospherically resistant vegetation index (ARVI), soil adjusted vegetation index (SAVI), optimized SAVI (OSAVI), modified SAVI (MSAVI), visible atmospherically resistant index (VARI), structure insensitive pigment index (SIPI), simple ratio (SR), red edge chlorophyll index (ReCI), green chlorophyll index (GCI), and green ratio vegetation index (GRVI), to extract maize crop fields in two distinct agricultural land. The classification is performed using RF (RF), SVM, KNNs, and bootstrap aggregating (Bagging) ML algorithms. The models are trained at a specific location and validated at another location. This crop type classification approach is adopted in the current study because it has yet to be widely used in the literature.

2 Materials and Methods

2.1 Study Sites

This study was conducted in two different sites in Bursa province in northwest Turkey, which lies approximately between latitudes 40°5'26.39''N and 40°11'19.03''N and longitudes 28°21'35.43''E and 28°29'25.84''E (see Fig. 1). The study site included two villages, Yeşilova and Ormankadı, located in the Mustafakemalpaşa district, with an area of 13.64 and 17.60 km^2 respectively.

The climate in the study area is characteristic of a Marmara transition Mediterranean climate. Precipitation occurs mainly during spring and winter. The precipitation during the winter typically occurs in the form of snow. The coldest month of the year is generally February, and the hottest month is July. Although the vegetation cover and the soils of the study area provide general characteristics of the Marmara Region, the zones close to the sea and rural areas show

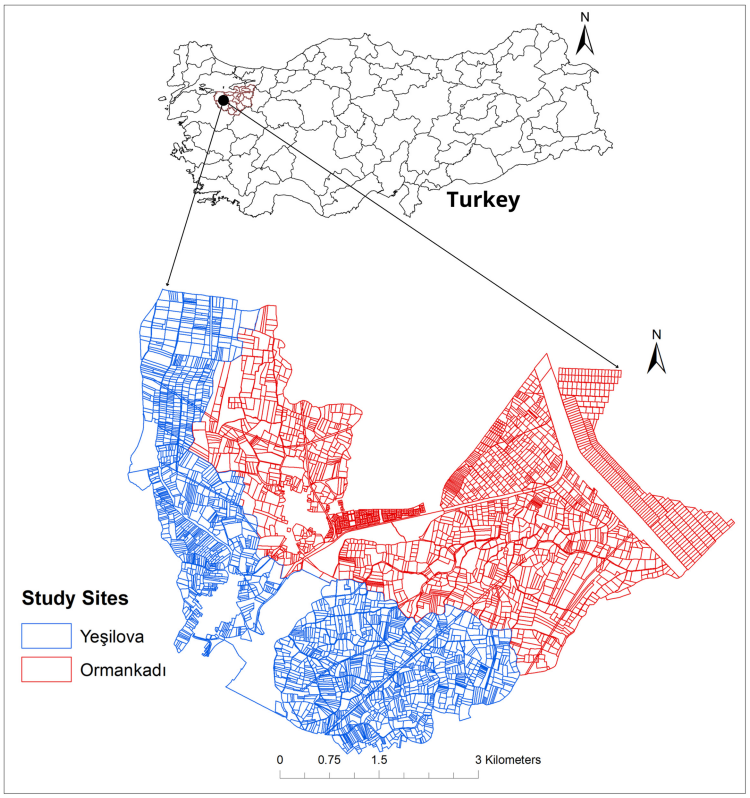


Fig. 1 Investigated fields in the study area.

differences. The soils in the study area are alluvial, calcareous brown, red-brown Mediterranean, and large Rendzina soil groups formed *in situ*.[17] Maize, pulses, tomatoes, forage plants, orchards, sugar beets, meadows, and pastures are generally cultivated in the study area, with minor changes from year to year. Maize (*Zea mays L.*) is the major cereal crop cultivated in the study area, which motivated this study's focus on maize crop extraction.

2.2 Methodology

In this study, Sentinel 2A-derived spectral indices (NDVI, NDWI, GNDVI, ARVI, SAVI, OSAVI, MSAVI, VARI, SIPI, SR, ReCI, GCI, and GRVI) and ML algorithms (RF, KNN, SVM, and Bagging) were selected to assess the feasibility of accurate maize crop extraction. The overall methodology and steps followed in the study are illustrated in Fig. 2.

2.3 Ground Truth

This study used data obtained from the Mustafakemalpaşa and Karacabey Irrigation Unions, where crop types declared by farmers on a parcel basis during the cultivation period were collected. These declarations were cross-referenced with onsite observations to validate their accuracy. Field investigations were conducted during the 2022 crop season. Plots with maize crops were investigated in Yeşilova and Ormankadı villages in the Mustafakemalpaşa district. A total of 2000 parcels were investigated. The plot distribution of the ground truth in the study area is shown in Fig. 3. The field investigations demonstrated that data obtained from the Mustafakemalpaşa and Karacabey Irrigation Unions were accurate.

The established crop patterns from field observations and parcel maps were processed using ArcGIS ArcMap 10.8 (ESRI, Redlands, California, United States) software. The vector-based data were further transformed into raster pixels for the analysis. After the preprocessing steps, 1747 parcels were considered in each village (Ormankadı and Yeşilova). The data were exported in an Excel worksheet format, with each row containing information on a distinct pixel and a column encompassing the parcel number. The dataset was labeled into two classes: maize and other (for the parcels with crops other than maize). The index values were averaged per parcel, and a dataset with 1747 parcels/rows was curated for both training and testing purposes.

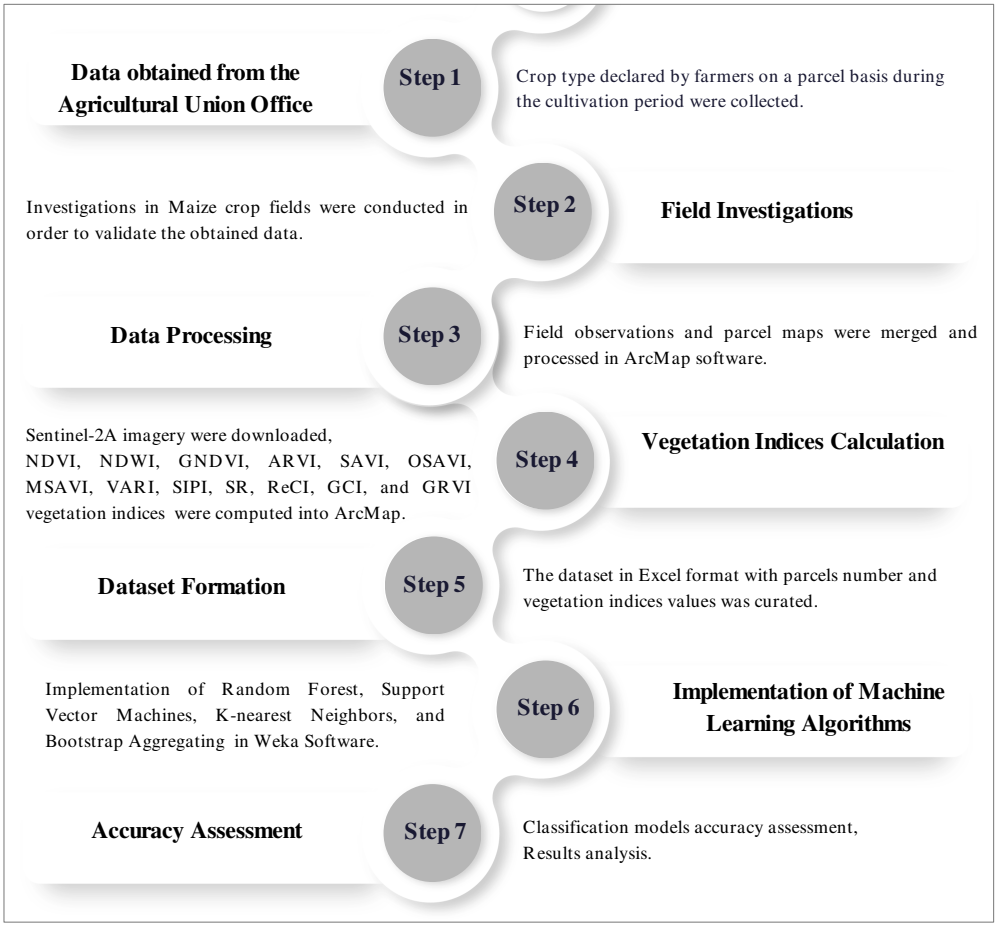


Fig. 2 Study methodology flowchart.

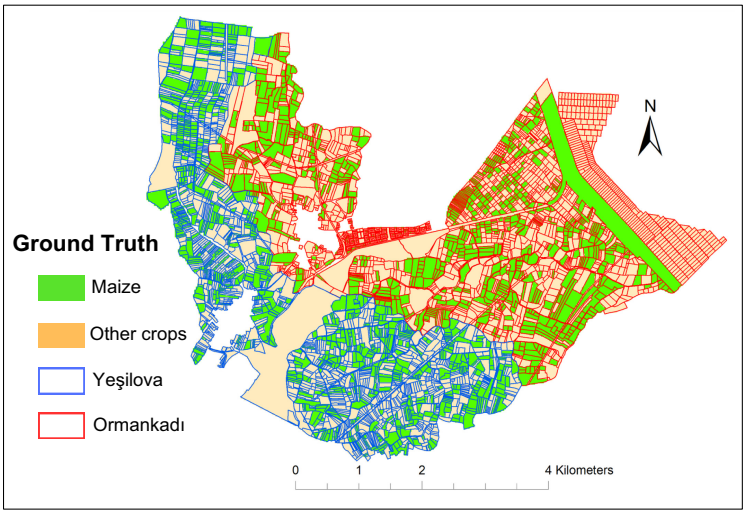


Fig. 3 Plot distribution of the ground truth.

2.4 Sentinel 2A-Derived Spectral Indices

Sentinel 2A data were used in this study. The Sentinel fleet of satellites was designed to deliver remote sensing data central to the European Commission's Copernicus program. The Sentinel 2A satellite was launched by the European Space Agency on June 23, 2015 and operates in a sun-synchronous orbit with a 10-day repeat cycle.

Table 1 Sentinel 2A-derived VIs used in this study.

VIs	Description	Formula	Reference
NDVI	Normalized difference vegetation index	(Band8 – Band4)/(Band8 + Band4)	25
NDWI	Normalized difference water index	(Band3 – Band8)/(Band3 + Band8)	26
GNDVI	Green normalized difference vegetation index	(Band8 – Band3)/(Band8 + Band3)	27
ARVI	Atmospherically resistant vegetation index	[Band8 – (2 × Band4) + Band2]/[Band8 + (2 × Band4) + Band2]	28
SAVI	Soil adjusted vegetation index	[(Band8 – Band4)/(Band8 + Band4 + 0.5)] × (1 + 0.5)	29
OSAVI	Optimized soil adjusted vegetation index	(Band8 – Band4)/(Band8 + Band4 + 0.16)	30
MSAVI	Modified soil-adjusted vegetation index	$[2 \times \text{Band4} + 1 - \text{sqrt}((2 \times \text{Band4} + 1)^2 - 8 \times (\text{Band4} - \text{Band3}))]/2$	31
VARI	Visible atmospherically resistant index	(Band3 – Band4)/(Band3 + Band4 – Band2)	32
SIPI	Structure insensitive pigment index	(Band8 – Band2)/(Band8 – Band4)	33
SR	Simple ratio	Band8/Band4	34
ReCI	Red edge chlorophyll index	(Band8/Band4) – 1	35
GCI	Green chlorophyll index	(Band8/Band3) – 1	36
GRVI	Green ratio vegetation index	Band8/Band3	37

Sentinel 2A-derived VIs have a significant utility in agricultural and environmental monitoring. These indices are valuable tools for evaluating the vegetation's health, growth, and overall productivity. In this study, Sentinel 2A imagery was used to derive VIs. The images were downloaded from Ref. 18. Six images from May 12, June 18, July 18, August 10, September 16, and October 09, 2022, covering the study area, all atmospherically corrected and cloudless, were selected. These six images were selected considering one image per month to cover the phenological stages of the maize crop. This number of images was chosen to evaluate whether it was possible to achieve a high accuracy with few images. In the study area, maize crops are generally seeded in May and harvested in October or November.

In this study, 13 well-known and commonly used VIs were selected. These indices have been used in many studies and have been recognized to improve the accuracy of crop type classification.[19–24] The VIs using bands with a 10 m resolution were chosen to obtain high-resolution data. For each scene, 10 m spatial resolution bands (2, 3, 4, and 8, respectively, and blue, green, red, and near-infrared spectral bands) were selected. The VIs and formulae used in this study are listed in Table 1.

2.5 Machine Learning Classifier Algorithms

ML classifier algorithms are used in supervised learning tasks to classify data points into predefined categories or classes. These algorithms learn patterns from labeled training data and use this knowledge to predict new, unseen data. The four classifier algorithms of RF, SVMs, KNNs, and Bagging are used in this study.

(a) Support vector machines: SVM are powerful and widely used supervised ML algorithms for classification tasks. SVMs have been particularly effective in handling high-dimensional data with a limited number of training samples, making them suitable for applications such as hyperspectral image classification.[38]

(b) Random forest: RF is an ML technique that utilizes an ensemble of decision trees to generate predictions.[39] Every decision tree in the RF is trained on a random subset of the training data and a random subset of the input features.[40]

(c) K-nearest neighbor**:** KNN is a simple yet effective method that determines the class of a query example by identifying its nearest neighbors in the training dataset.[41] The KNN algorithm assigns the class label of most of its KNNs to the query example, where k is a parameter defined by the user.[42]

(d) Bootstrap aggregating: Bagging produces numerous iterations of a predictor by generating bootstrap replicas of training data. This can provide substantial gains in accuracy, particularly when the prediction method is unstable. [43]

2.6 Classification Models Implementation

Waikato Environment for Knowledge Analysis (WEKA), a popular open-source ML software toolkit that provides a wide range of tools and algorithms for data mining and ML tasks, was used to perform the classification task in this study.[44] The datasets were split into training and test sets using a 70:30 partition. The selection of the best hyperparameters of ML models is very important in remote sensing image classification.

In this study, the grid search method was used for the hyperparameters tuning. For each ML model, different hyperparameters were evaluated. Using the grid search method for hyperparameter tuning in WEKA involves systematically exploring a manually specified subset of the hyperparameter space of an ML algorithm to identify the optimal set of parameters. In this study, the meta/multisearch/gridsearch option was selected after loading the dataset to search through various combinations of hyperparameters for a classifier and identify the best combination for training the model. Before starting the grid search, a range of values for each parameter is defined. Once the grid search is completed, the results are examined to determine the best combination of parameters, which is then used to train the final model. The hyperparameters of each algorithm used in this study are listed in Table 2.

2.7 Machine Learning Models Accuracy Assessment

The evaluation metrics used to assess the performance of the classification models include true positives (TPs), overall accuracy (OA), precision (P), recall (R), F1-score, Matthews correlation coefficient (MCC), receiver operating characteristic area (ROC area), and precision-recall curve area (PRC area).

TPs are instances correctly predicted as positive by the model. In other words, the model correctly identifies instances from the positive class. True negatives (TNs) are instances correctly predicted as negative by the model: the model correctly identified instances from the negative class. False positives (FPs) are instances incorrectly predicted as positive by the model but belong to the negative class. False negatives (FNs) are instances incorrectly predicted as negative by the model but belong to the positive class. Precision (P) measures the number of predicted positive instances that are actually positive. Recall (R) indicates how well the model captures all positive instances. The F1-score is the harmonic mean of precision and recall, providing a balanced

Table 2 ML hyperparameters used for modeling in this study.

ML models	Best hyperparameters used
RF	{'bagSizePercent': 100, 'numTrees': 100, 'batchSize': 100, 'max_depth': 0, 'numDecimalPlaces': 2, 'numExecutionSlots': 1, 'numFeatures': 0, 'numIterations': 100, 'seed': 1}
SVMs	{'batchSize': 100, 'c': 1.0, 'calibrator': 'Logistic regressor', 'kernel': 'PolyKernel', 'numDecimalPlaces': 2, 'numFolds': -1, 'randomSeed': 1, 'tolranceParameter': 0.001}
KNNs	{'KNN': 5, 'c': 1.0, 'batchSize': 100, 'classifier': 'SVM-M 2-V 0.001-N 3', 'numDecimalPlaces': 2}
Bagging	{'bagSizePercent': 100, 'batchSize': 100, 'max_depth': 0, 'numDecimalPlaces': 2, 'numExecutionSlots': 1, 'numIterations': 10, 'seed': 1}

Table 3 Model performance evaluation metrics.

Performance evaluation metrics	Formula
OA	$\frac{TP+TN}{(TP+TN+FP+FN)}$
Precision (P)	$\frac{TP}{(TP+FP)}$
Recall (R)	$\frac{TP}{(TP+FN)}$
F1-score	$2*\frac{P*R}{(P+R)}$
MCC	$MCC=\frac{TP\times TN-FP\times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$
ROC area	$\text{ROC area}=\int_0^1 \text{True positive rate}\, d\,(\text{false positive rate})$
PROC area	$\text{PROC area}=\int_0^1 \text{Precision}\, d(\text{recall})$

TP, true positives; TN, true negatives; FP, false positives; FN, false negatives; MCC, Matthews correlation coefficient; ROC area, receiver operating characteristic area; and PRC area, precision-recall curve area.

measure between the two. The MCC measures the quality of binary (two-class) classification. It considers TPs, TNs, FPs, and FNs. The ROC represents the trade-off between the TP and FP rates at various thresholds. The ROC area, also known as the area under the curve (AUC), is a numerical measure of the performance of a classification model at various classification thresholds. The PRC represents the trade-off between precision and recall at various thresholds. The PRC area, also known as the area under the precision-recall curve (AUC-PR), is a numerical measure of the model's ability to balance precision and recall.

Collectively, these metrics help to understand the strengths and weaknesses of the classification model used in this research and make informed decisions about its performance and potential improvements.[44] The formulas for these performance metrics are presented in Table 3.

3 Results

3.1 Analysis of the Studied Vegetation Indices

Temporal analysis of Sentinel 2-derived VIs provides valuable insights into crop types and their temporal dynamics. The temporal analysis of Sentinel 2A-derived VIs used in this study is shown in Fig. 4. The VIs had higher values during the growing stage and lower values during the early and late stages of the maize crop.

3.2 Quantitative Classification Results

The quantitative results of the four ML classification algorithms used to classify maize crop in Yeşilova and Ormankadı are presented in this section. The results are presented regarding high values of the OA, TP rate, precision, recall, F1-score, MCC, ROC area, and PRC area and low values of the FP rate. Furthermore, the analysis is explicitly focused on the maize crop, which is the aim of this study.

3.2.1 *Random forest model accuracy*

The RF model exhibited good accuracy in maize classification when considering Yeşilova to be the training set and Ormankadı to be the test set (Table 4). Notably, the model achieved a TP rate of 89.1%, indicating a high proportion of accurately identified maize crop parcels. The low FP rate of 9.3% underscored the model's proficiency in minimizing misclassifications. Precision, at 91.4%, reflected the model's precision in labeling instances as maize, and recall (89.1%) signified its ability to capture a substantial portion of actual maize occurrences. The F1-score of 90.3% indicated a harmonious balance between precision and recall. An MCC of 79.8% highlighted the model's performance, considering both FPs and FNs. The ROC and PRC area values

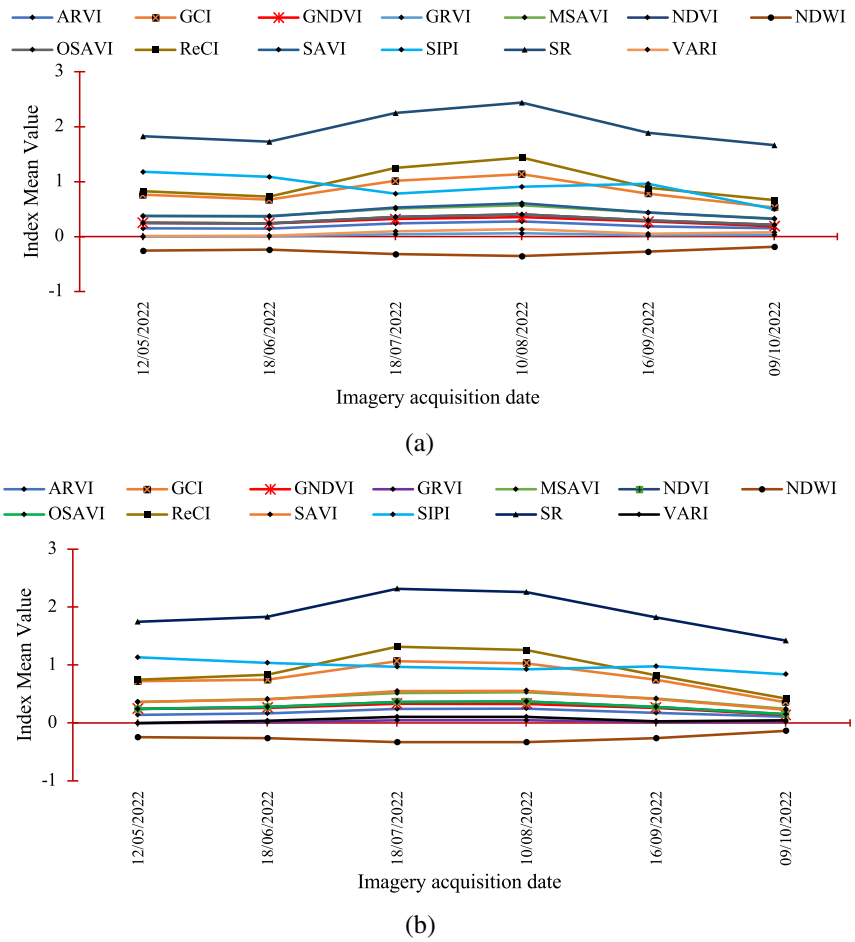


Fig. 4 Temporal distribution of Sentinel 2A-derived VIs of maize crop in (a) Ormankadı and (b) Yeşilova.

Table 4 Performance metrics of the RF model.

RF models	TP rate (%)	FP rate (%)	Precision (%)	Recall (%)	F1-score (%)	MCC (%)	ROC area (%)	PRC area (%)	CLASS
Yeşilova training	89.1	9.3	91.4	89.1	90.3	79.8	95.1	95.6	Maize
	90.7	10.9	88.2	90.7	89.5	79.8	95.1	92.7	Other
	89.9	10.0	89.9	89.9	89.9	79.8	95.1	94.2	OA
Testing on Ormankadı	84.9	17.6	78.3	84.9	81.5	66.7	91.3	88.8	Maize
	82.4	15.1	87.9	82.4	85.1	66.7	91.3	92.3	Other
	83.5	16.2	83.8	83.5	83.5	66.7	91.3	90.8	OA
Ormankadı training	91.1	15.5	88.0	91.1	89.5	76.0	94.1	94.3	Maize
	84.5	8.9	88.3	84.5	86.4	76.0	94.1	93.4	Other
	88.2	12.6	88.2	88.2	88.1	76.0	94.1	93.9	OA
Testing on Yeşilova	88.7	20.2	81.6	88.7	85.0	68.8	91.1	89.6	Maize
	79.8	11.3	87.5	79.8	83.5	68.8	91.1	91.6	Other
	84.3	15.8	84.5	84.3	84.2	68.8	91.1	90.6	OA

of 95.1% and 95.6%, respectively, affirmed the model's discriminative capacity and precision-recall trade-off. The OA for maize and other crops was 89.9%. The RF model validation on Ormankadı maintained the model's performance by achieving an OA of 83.5%, with high TP rate (84.9%), precision (78.3%), recall (84.9%), F1-score (81.5%), and MCC (66.7%) values, indicating its resilience across diverse datasets. The ROC and PRC area achieved accuracies of 91.3% and 88.8%, respectively.

Furthermore, the RF model consistently demonstrated robust performance in maize classification when Ormankadı served as the training set and Yeşilova as the test set. The model calibration on the Ormankadı dataset resulted in a high overall accuracy (OA) of 88.2%, along with a true positive (TP) rate of 91.1%, precision of 88.0%, recall of 91.1%, F1-score of 89.5%, and Matthews correlation coefficient (MCC) of 76.0%. These results confirm the model's effectiveness in accurately capturing maize patterns at this study site. Subsequently, testing on Yeşilova maintained the model's accuracy, with an 84.3% OA and TP rate (88.7%), precision (81.6%), recall (88.7%), F1-score (85.0%), and MCC (68.8%) values, indicating its resilience across diverse datasets.

3.2.2 *Support vector machines model accuracy*

The results of the SVM model are presented in Table 5. The SVM model exhibited good accuracy in maize classification when considering Yeşilova to be the training set and Ormankadı to be the test set. The training model achieved an OA of 90.1% and a TP rate of 88.8%, indicating a high proportion of accurately identified maize crop parcels. The low false positive rate of 8.5% underscored the model's proficiency in minimizing misclassifications. Precision, at 92.1%, reflected the model's precision in labeling instances as maize, and recall (88.8%) signified its ability to capture a substantial portion of actual maize occurrences. The F1-score of 90.4% indicated a harmonious balance between precision and recall. An MCC of 80.2% highlighted the model's overall performance, considering both FPs and FNs. The ROC and PRC area values of 90.2% and 87.7%, respectively, affirmed the model's discriminative capacity and precision-recall trade-off. The SVM model validation on Ormankadı maintained the accuracy of the model with an OA of 84.3%. The TP rate (84.4%), precision (80.1%), recall (84.4%), F1-score (82.2%), and MCC (68.3%) values indicated model resilience across diverse datasets. The ROC and PRC area values achieved accuracies of 84.3% and 74.2%, respectively.

Table 5 Performance metrics of the SVM model.

SVM models	TP rate (%)	FP rate (%)	Precision (%)	Recall (%)	F1-score (%)	MCC (%)	ROC area (%)	PRC area (%)	CLASS
Yeşilova training	88.8	8.5	92.1	88.8	90.4	80.2	90.2	87.7	Maize
	91.5	11.2	88.0	91.5	89.7	80.2	90.2	84.5	Other
	90.1	9.8	90.2	90.1	90.1	80.2	90.2	86.2	OA
Testing on Ormankadı	84.4	15.7	80.1	84.4	82.2	68.3	84.3	74.2	Maize
	84.3	15.6	87.8	84.3	86.0	68.3	84.3	83.0	Other
	84.3	15.7	84.5	84.3	84.4	68.3	84.3	79.2	OA
Ormankadı training	90.4	17.2	86.8	90.4	88.6	73.7	86.6	83.8	Maize
	82.8	9.6	87.3	82.8	85.0	73.7	86.6	80.0	Other
	87.0	13.8	87.0	87.0	87.0	73.7	86.6	82.1	OA
Testing on Yeşilova	90.4	20.9	81.3	90.4	85.6	70.0	84.8	78.4	Maize
	79.1	9.6	89.1	79.1	83.8	70.0	84.8	80.9	Other
	84.8	15.3	85.2	84.8	84.7	70.0	84.8	79.6	OA

The SVM model consistently performed well in the maize classification when Ormankadı served as the training set and Yeşilova as the test set. Training on Ormankadı showed an OA of 87.0%, TP rate of 90.4%, precision of 86.8%, recall of 90.4%, F1-score of 88.6%, and MCC of 73.7%, confirming the model's ability to classify maize crop in Ormankadı effectively. Subsequently, testing on Yeşilova maintained the accuracy of the model with an OA of 84.8%. The high values of TP rate (90.4%), precision (81.3%), recall (90.4%), F1-score (85.6%), and MCC (70.0%) indicated the model performance.

3.2.3 *K-nearest neighbors model accuracy*

Table 6 summarizes the results of the KNN model. This table found that the KNN model exhibited good accuracy in maize classification when considering Yeşilova to be the training set and Ormankadı to be the test set. The training model achieved an OA of 88.0% and a TP rate of 86.2%, indicating a high proportion of accurately identified maize crop parcels. The low FP rate of 10.1% underscored the model's proficiency in minimizing misclassifications. The precision, at 90.5%, reflected the model's precision in labeling instances as maize, and recall, at 86.2%, signified its ability to capture a substantial portion of actual maize occurrences. The F1-score of 88.3% indicated a harmonious balance between precision and recall. An MCC of 76.0% highlighted the model's overall performance, considering both FPs and FNs. The ROC and PRC area values of 88.1% and 85.3%, respectively, affirmed the model's discriminative capacity and precision-recall trade-off. The KNN testing model on Ormankadı maintained the accuracy of the model with an OA of 79.3%. The TP rate (82.6%), precision (72.8%), recall (82.6%), F1-score (77.4%), and MCC (58.9%) values demonstrated the resilience of the model across diverse datasets. ROC and PRC area values were 79.7% and 67.6%, respectively.

When Ormankadı served as the training set and Yeşilova as the test set, the KNN model consistently performed well for maize crop classification. For instance, training on Ormankadı showed a high OA of 84.4%, and the TP rate (88.0%), precision (84.5%), recall (88.0%), F1-score (86.2%), and MCC (68.2%) values confirmed the model's ability to classify maize crop in Ormankadı effectively. The ROC and PRC area values were 83.9% and 81.0%, respectively. Testing on Yeşilova maintained the accuracy of the model, with an OA of 78.7% and a TP rate of 81.1%. The precision (77.5%), recall (81.1%), F1-score (79.3%), and MCC (57.5%) values indicated good performance of the model. The ROC and PRC area values were 78.7% and 72.4%, respectively.

Table 6 Performance metrics of the KNN model.

KNN models	TP rate (%)	FP rate (%)	Precision (%)	Recall (%)	F1-score (%)	MCC (%)	ROC area (%)	PRC area (%)	CLASS
Yeşilova training	86.2	10.1	90.5	86.2	88.3	76.0	88.1	85.3	Maize
	89.9	13.8	85.4	89.9	87.6	76.0	88.1	81.6	Other
	88.0	11.8	88.1	88.0	88.0	76.0	88.1	83.5	OA
Testing on Ormankadı	82.6	23.1	72.8	82.6	77.4	58.9	79.7	67.6	Maize
	76.9	17.4	85.5	76.9	81.0	58.9	79.7	79.0	Other
	79.3	19.8	80.1	79.3	79.4	58.9	79.7	74.1	OA
Ormankadı training	88.0	20.2	84.5	88.0	86.2	68.2	83.9	81.0	Maize
	79.8	12.0	84.2	79.8	81.9	68.2	83.9	76.2	Other
	84.4	16.6	84.3	84.4	84.3	68.2	83.9	78.8	OA
Testing on Yeşilova	81.1	23.7	77.5	81.1	79.3	57.5	78.7	72.4	Maize
	76.3	18.9	80.0	76.3	78.1	57.5	78.7	72.8	Other
	78.7	21.3	78.8	78.7	78.7	57.5	78.7	72.6	OA

Table 7 Performance metrics of the Bagging model.

Bagging models	TP rate (%)	FP rate (%)	Precision (%)	Recall (%)	F1-score (%)	MCC (%)	ROC area (%)	PRC area (%)	CLASS
Yeşilova training	88.0	10.9	90.0	88.0	89.0	77.1	93.5	94.6	Maize
	89.1	12.0	87.0	89.1	88.0	77.1	93.5	90.4	Other
	88.5	11.4	88.6	88.5	88.6	77.1	93.5	92.6	OA
Testing on Ormankadı	82.1	19.0	76.4	82.1	79.1	62.6	89.1	86.0	Maize
	81.0	17.9	85.8	81.0	83.3	62.6	89.1	90.1	Other
	81.5	18.4	81.8	81.5	81.5	62.6	89.1	88.3	OA
Ormankadı training	90.7	19.7	85.2	90.7	87.9	71.8	92.2	92.7	Maize
	80.3	9.3	87.4	80.3	83.7	71.8	92.2	91.3	Other
	86.1	15.1	86.1	86.1	86.0	71.8	92.2	92.1	OA
Testing on Yeşilova	83.5	20.3	80.5	83.5	82.0	63.2	88.3	87.0	Maize
	79.7	16.5	82.7	79.7	81.1	63.2	88.3	87.0	Other
	81.6	18.4	81.6	81.6	81.6	63.2	88.3	87.0	OA

3.2.4 *Bagging model accuracy*

When considering Yeşilova to be the training set and Ormankadı to be the test set, the Bagging model exhibited good accuracy in maize classification (Table 7). The training model on Yeşilova achieved an OA of 88.5% and a TP rate of 88.0%, indicating a high proportion of accurately identified maize crop parcels. The low FP rate of 10.9% underscored the model's proficiency in minimizing misclassifications. Precision, at 90.0%, reflected the model's precision in labeling instances as maize, and recall, at 88.0%, signified its ability to capture a substantial portion of actual maize occurrences. The F1-score of 89.0% indicated a harmonious balance between precision and recall. The MCC of 77.1% highlighted the model's overall performance, considering both FPs and FNs. The ROC and PRC area values of 93.5% and 94.6%, respectively, affirmed the model's discriminative capacity and precision-recall trade-off. The Bagging testing model on Ormankadı maintained the accuracy of the model, with an OA of 81.5% and a TP rate of 82.1%. The values of precision (76.4%), recall (82.1%), F1-score (79.1%), and MCC (62.6%) indicated the good performance of the model. The ROC and PRC area values achieved 89.1% and 86% accuracy, respectively.

Furthermore, the model consistently demonstrated robust performance in maize classification when Ormankadı served as the training set and Yeşilova as the test set. Training in Ormankadı achieved a high OA of 86.1% with a TP rate of 90.7%. The precision (85.2%), recall (90.7%), F1-score (87.9%), MCC (71.8%), ROC area (92.2%), and PROC area (92.7%) values demonstrated the model's ability to capture the intricacies of maize patterns effectively. Testing the model on Yeşilova maintained high accuracy, with an OA of 81.6% and a TP rate of 83.5%. The ROC and PRC area values achieved accuracies of 88.3% and 87.0%, respectively, and the precision (80.5%), recall (83.5%), F1-score (82.0%), and MCC (63.2%) values indicated the resilience of the model across diverse datasets.

3.3 Qualitative Classification Results and Crop Type Mapping

Figure 5 shows the distribution of maize extracted by different algorithms in the study area. The presented maps in the figure visually represent the classification outcomes of best models, offering insights into the spatial distribution of the crop types. These maps serve as valuable tools for comprehending the effectiveness of our classification approach and its practical implications for agricultural monitoring and decision making. Upon analyzing the maps, it becomes evident that the ML models have succeeded in delineating distinct regions corresponding to maize and other crops.

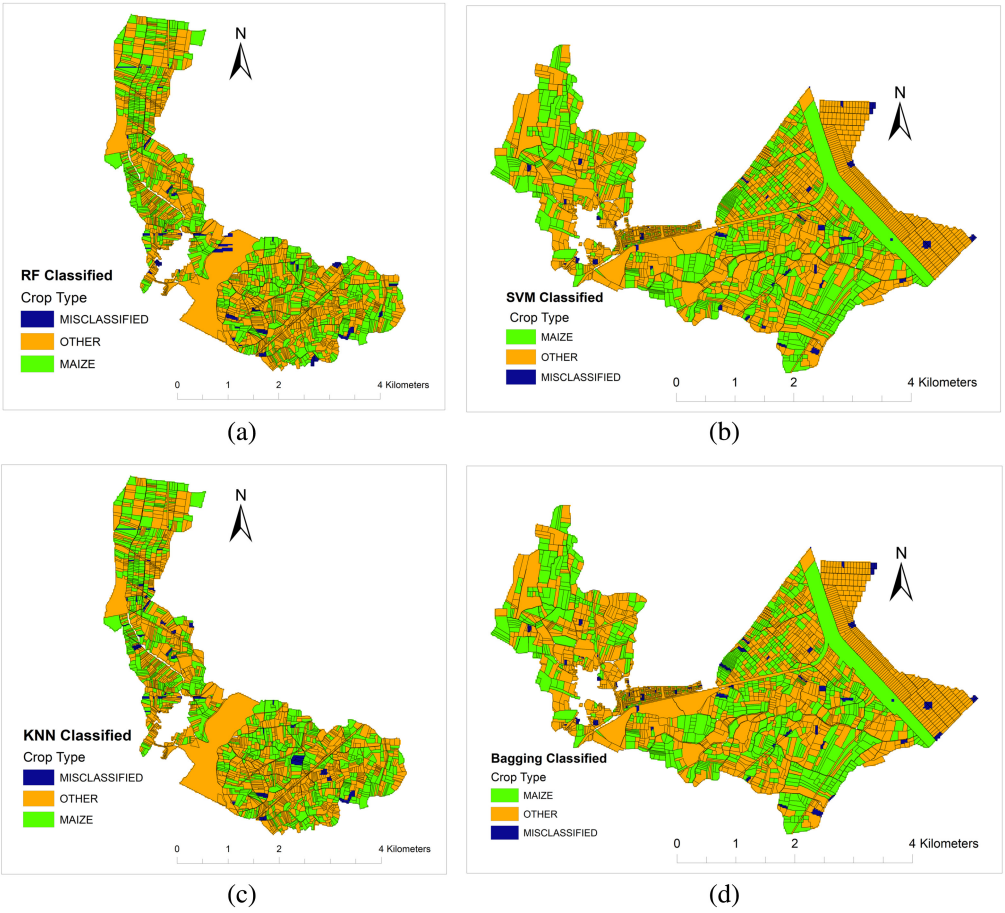


Fig. 5 Distribution of the maize crop extracted by different ML models. (a) RF model validation in Yeşilova, (b) SVMs model validation in Ormankadı, (c) KNNs validation in Yeşilova, and (d) Bagging model validation in Ormankadı.

The regions where maize is identified exhibit consistent spectral patterns that align with the expected characteristics of this crop. Although the maps reflect a high level of accuracy, some localized variations might be observed due to factors such as varying soil conditions, land management practices, and natural variations in vegetation health. This underscores the need to refine and adapt the classification models to account for these nuanced differences.

In summary, when the models were calibrated on Yeşilova and validated on Ormankadı, a ranking of the models based on their performance for maize and other crops classification showed that the SVM model outperformed, with an OA rate of 84.3% and an MCC of 68.3%. It was followed by the RF model with an OA of 83.5% and 66.7% for MCC, the Bagging model with an OA of 81.5% and 62.6% for MCC, and the KNN model with an OA of 79.3% and 58.9% for MCC.

The models maintained their leading position order when calibrated on Ormankadı and validated on Yeşilova. The SVM consistently led with an OA rate of 84.8% and an MCC of 70.0%, followed by the RF model with an OA of 84.3% and 68.8% for MCC, the Bagging model with an OA of 81.6% and 63.2% for MCC, and the KNN model with an OA of 78.7% and 57.5% for MCC.

4 Discussion

This study underscores the commendable performance consistently exhibited by ML models in maize crop extraction across the two distinct investigated villages. Using the widely used spectral indices of NDVI, NDWI, GNDVI, ARVI, SAVI, OSAVI, MSAVI, VARI, SIPI, SR, ReCI, GCI, and GRVI, four ML classifiers, including SVM, RF, KNN, and Bagging, were implemented. The

training-test methodology was explored using two scenarios: Yeşilova as the calibration set and Ormankadı as the validation set, and vice versa. The models calibrated on Yeşilova and validated on Ormankadı maintained the accuracy of the model, with an OA ranging from 79.3% to 89.9%. Furthermore, the models consistently demonstrated robust performance when Ormankadı served as the calibration set and Yeşilova as the validation set, with high OA (78.7% to 84.8%).

These models, characterized by good accuracy, are viable tools for crop science research within diverse regional contexts. In the two scenarios in this study, the OA of the testing models ranged between 78.7% and 89.8%, and the MCC ranged between 58.9% and 76.0%. The ML algorithms used in the two scenarios demonstrated similar results in terms of the model's performance. The SVM model performed over RF, Bagging, and KNN. The findings are acceptable as they align comparably with the results documented in the relevant literature. For instance, in one investigation of deep learning for crop type classification,[16] researchers reported a 70.0% overall classification accuracy achieved by models trained on one geographical area and subsequently validated it in other areas within Nebraska (United States).

Furthermore, hyperspectral data, when appropriately pre-processed and analyzed, can effectively be valuable in maize crop management. For instance, estimating chlorophyll content in maize is crucial for monitoring plant health and photosynthetic activity. Various methods, including spectral indices and wavelet transform, have been employed to achieve accurate estimations using hyperspectral reflectance data. A study conducted by Ref. 45 demonstrated that wavelet features in the red-edge region were more effective in estimating chlorophyll content than spectral indices, with a determination coefficient (R^2) of 0.905 compared with 0.819 for the best spectral index. In addition, discriminating maize varieties is essential for crop monitoring, high-throughput phenotyping, and yield forecasting. In one study,[46] researchers explored the effectiveness of using multitemporal hyperspectral data and multivariate techniques for *in situ* varietal discrimination. They demonstrated that the identified spectral bands and phenological stages are critical for accurate and high-throughput varietal discrimination.

To the best of our knowledge, few studies in the literature have utilized our approach. Furthermore, integrating ML algorithms with Sentinel 2-derived spectral indices has emerged as a focal point of interest in remote sensing and agricultural research. Numerous studies, including those conducted by Refs. 47–51, have contributed valuable insights into utilizing ML for crop classification and mapping. Other researchers have achieved higher accuracy using a common approach (calibration and validation in the same study area). For instance,[8] researchers highlighted the effectiveness of RF and SVM as the most effective classification approaches for identifying vegetation types using remote sensing data, with an OA of 92.1%. High overall accuracies of 96.93% and 97.44% were achieved by Ref. 52 for maize classification using RF, SVMs, simple Bayes, ML algorithms, and Sentinel 2-derived VIs. In one study on crop type classification,[53] researchers achieved the highest OA of 96% to 98% using Sentinel 2 and the RF algorithm.

Our proposed approach is unique because it suggests exploring alternative methods other than common techniques of crop type classification. The vegetation indices used in this study and the number of satellites images selected (six images covering all phenological stages) have shown that it was possible to reach acceptable accuracy in maize crop classification.

5 Conclusion

This study successfully addressed the objective of developing an approach for the classification of maize (*Zea mays L.*) crops across heterogeneous agricultural areas. The utilization of Sentinel 2A-derived VIs in conjunction with ML classifier algorithms yielded commendable results. The results validated that the hyperparameters selection, the studied VIs, and the quality of the dataset greatly influenced the performance of ML models. The SVM, RF, and Bagging models especially exhibited encouraging potential for maize crop extraction from the investigated fields. The integration of high-resolution satellite imagery and advanced ML algorithms has underscored the success of accurate parcel-based maize crop classification.

The results indicated that the SVM, RF, and Bagging ML models accurately extracted maize across different agricultural lands. The combination of the NDVI, NDWI, GNDVI, ARVI, SAVI, OSAVI, MSAVI, VARI, SIPI, SR, ReCI, GCI, and GRVI spectral indices was found to be highly

effective in accurately extracting maize crops at all phenological stages. Our approach and findings provide a comprehensive overview of alternative methodologies used in such classifications, contributing to the existing body of knowledge and offering insights into sustainable land management practices. The conceptual framework for crop type classification developed here can help in developing suitable land cover and land use monitoring systems to adapt to climate change, assisting policymakers in safeguarding crop management.

Based on the results of this study on maize crop extraction, future work is expected to be performed considering other crops, more remote data, and deep learning techniques to improve classification accuracy. In future research, it may be necessary to conduct studies using synthetic-aperture radar data and images taken only in some specific phenological stages. Thus, crop type classification could be performed without following all phenological stages to investigate whether the classification accuracy improved or not. These additional research efforts should help understand the factors contributing to suboptimal classification performance within specific test areas for other crop types within the same geographical region.

Code and Data Availability

The data used in this study will be made available upon request from the corresponding author.

Acknowledgments

The authors express their gratitude to Mr. Hakan ABALIOĞLU from the General Directorate of State Hydraulic Works (DSİ 12. Şübe Müdürlüğü) in Karacabey for his assistance with fieldwork and data collection. This study did not receive any funding, and the authors declare that they have no conflicts of interest.

References

1. T. Zhou et al., "Mapping winter wheat with multi-temporal SAR and optical images in an urban agricultural region," *Sensors* **17**(6), 1210 (2017).
2. Q. Xie et al., "Crop monitoring and classification using polarimetric radarSAT-2 time-series data across growing season: a case study in Southwestern Ontario, Canada," *Remote Sens.* **13**(7), 1394 (2021).
3. A. Htitiou et al., "The performance of random forest classification based on phenological metrics derived from Sentinel-2 and Landsat 8 to map crop cover in an irrigated semi-arid region," *Remote Sens. Earth Syst. Sci.* **2**(4), 208–224 (2019).
4. S. Wang et al., "Mapping crop types in Southeast India with smartphone crowdsourcing and deep learning," *Remote Sens.* **12**(18), 2957 (2020).
5. A. Jarayee et al., "Oil palm plantation land cover and age mapping using Sentinel-2 satellite imagery and machine learning algorithms," in *IOP Conf. Ser.: Earth Environ. Sci.*, Vol. **1**, p. 012024 (2022).
6. M. Hassan et al., "Rohingya refugee crisis and forest cover change in Teknaf, Bangladesh," *Remote Sens.* **10**(5), 689 (2018).
7. R. Sonobe et al., "Crop classification from Sentinel-2-derived vegetation indices using ensemble learning," *J. Appl. Remote Sens.* **12**(2), 026019 (2018).
8. R. Sonobe, "Parcel-based crop classification using multi-temporal Terrasar-x dual polarimetric data," *Remote Sens.* **11**(10), 1148 (2019).
9. A. Abdi, "Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data," *GISci. Remote Sens.* **57**(1), 1–20 (2019).
10. D. Vijayasekaran, "Sen2-agri – crop type mapping pilot study using Sentinel-2 satellite imagery in India," *Int. Arch. Photogramm. Remote Sens. Sp. Inf. Sci.* **XLII-3/W6**, 175–180 (2019).
11. K. Jia et al., "Land cover classification of Landsat data with phenological features extracted from time series Modis NDVI data," *Remote Sens.* **6**(11), 11518–11532 (2014).
12. M. Wu et al., "Evaluation of orthomosics and digital surface models derived from aerial imagery for crop type mapping," *Remote Sens.* **9**(3), 239 (2017).
13. R. Jagadeeswaran, A. Poornima, and R. Kumaraperumal, "Mapping and classification of crops using high resolution satellite image," *J. Appl. Nat. Sci.* **10**(3), 818–825 (2018).
14. R. Valcarce-Diñeiro et al., "Multi-temporal dual- and quad-polarimetric synthetic aperture radar data for crop-type mapping," *Remote Sens.* **11**(13), 1518 (2019).
15. I. Fernández and N. Morales, "One-class land-cover classification using maxent: the effect of modelling parameterization on classification accuracy," *PeerJ* **7**, e7016 (2019).

16. A. H. Khan, M. M. Fraz, and M. Shahzad, "Deep learning-based land cover and crop type classification: a comparative study," in *Int. Conf. Digit. Futures and Transform. Technol. (ICoDT2)*, IEEE, pp. 1–6 (2021).
17. B. B. Bantchina, U. Mucan, and K. S. Gündoğdu, "Land availability analysis in bursa using geographic information systems," in *Proc. 5th Int. Participation Soil and Water Resour. Congr.*, 12–15 September 2017, Kırklareli, Turkey; Atatürk Soil Water and Agricultural Meteorology Research Institute Kırklareli, Merkez, Turkey, Vol. **1**, pp. 65–74 (2017).
18. Copernicus Data Space Ecosystem, http://dataspace.copernicus.eu/ (accessed 30 October 2022).
19. P. S. Thenkabail, R. B. Smith, and E. De Pauw, "Hyperspectral vegetation indices and their relationships with agricultural crop characteristics," *Remote Sens. Environ.* **71**, 158–182 (2000).
20. Z. Li et al., "Evaluation of spectral and texture features for object-based vegetation species classification using support vector machines," in *ISPRSTCVII Symp.–100 Years SPRS*, Vienna, Austria, 5–7 July 2010, W. Wagner and B. Székely, Eds., IAPRS, Vol. **XXXVIII000000**, Part 7A (2010).
21. M. Ustuner et al., "Crop type classification using vegetation indices of RapidEye imagery," *Int. Arch. Photogramm. Remote Sens. Sp. Inf. Sci.* **XL-7**, 195–198 (2014).
22. S. S. Sawant and M. Prabukumar, "Band fusion based hyper spectral image classification," *Int. J. Pure Appl. Math.* **117**, 71–76 (2017).
23. S. Felegari et al., "Integration of Sentinel 1 and Sentinel 2 satellite images for crop mapping," *Appl. Sci.* **11**, 10104 (2021).
24. L. Agilandeeswari et al., "Crop classification for agricultural applications in hyperspectral remote sensing images," *Appl. Sci.* **12**, 1670 (2022).
25. C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sens. Environ.* **8**, 127–150 (1979).
26. B.-C. Gao, "NDWI - a normalized difference water index for remote sensing of vegetation liquid water from space," *Remote Sens. Environ.* **58**, 257–266 (1996).
27. A. A. Gitelson, Y. J. Kaufman, and M. N. Merzlyak, "Use of a green channel in remote sensing of global vegetation from EOS-MODIS," *Remote Sens. Environ.* **58**, 289–298 (1996).
28. Y. J. Kaufman and D. Tanre, "Atmospherically resistant vegetation index (ARVI) for EOS-MODIS," *IEEE Trans. Geosci. Remote Sens.* **30**, 261–270 (1992).
29. A. R. Huete, "A soil-adjusted vegetation index (SAVI)," *Remote Sens. Environ.* **25**, 295–309 (1988).
30. G. Rondeaux, M. Steven, and F. Baret, "Optimization of soil-adjusted vegetation indices," *Remote Sens. Environ.* **55**, 95–107 (1996).
31. J. Qi, "A modified soil adjusted vegetation index," *Remote Sens. Environ.* **48**, 119–126 (1994).
32. A. A. Gitelson et al., "Vegetation and soil lines in visible spectral space: a concept and technique for remote estimation of vegetation fraction," *Int. J. Remote Sens.* **23**(13), 2537–2562 (2002).
33. J. Penuelas, F. Baret, and I. Filella, "Semi-empirical indices to assess carotenoids/chlorophyll-a ratio from leaf spectral reflectance," *Photosynthetica* **31**, 221–230 (1995).
34. G. Birth and G. McVey, "Measuring the color of growing turf with a reflectance spectrophotometer," *Agron. J.* **60**, 640–643 (1968).
35. A. A. Gitelson and M. N. Merzlyak, "Quantitative estimation of chlorophyll using reflectance spectra," *J. Photochem. Photobiol.* **22**, 247–252 (1994).
36. A. A Gitelson, Y. Gritz, and M. Merzlyak, "Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves," *J. Plant Physiol.* **160**, 271–282 (2003).
37. R. P. Sripada et al., "Aerial color infrared photography for determining early in-season nitrogen requirements in corn," *Agron. J.* **98**(4), 968–977 (2006).
38. P. Ghamisi et al., "Advanced spectral classifiers for hyperspectral images: a review," *IEEE Geosci. Remote Sens. Mag.* **5**(1), 8–32 (2017).
39. A. K. Arora et al. "Random forest analysis of impact of abiotic factors on Culex pipiens and Culex quinquefasciatus occurrence," *Front. Ecol. Evol.* **9**, 773360 (2022).
40. C. Strobl et al., "Bias in random forest variable importance measures: illustrations, sources and a solution," *BMC Bioinf.* **8**(1), 25 (2007).
41. P. Cunningham and S. Delany, "K-nearest neighbour classifiers - a tutorial," *ACM Comput. Surv.* **54**(6), 1–25 (2021).
42. R. Jensen and C. Cornelis, "A new approach to fuzzy-rough nearest neighbour classification," *Lect. Notes Comput. Sci.* **5306**, 310–319 (2008).
43. L. Breiman, "Bagging predictors," *Mach. Learn.* **24**(2), 123–140 (1996).
44. M. Hall et al., "The WEKA data mining software: an update," *SIGKDD Explor.* **11**, 1–18 (2009).
45. Q. Liao et al., "Comparison of spectral indices and wavelet transform for estimating chlorophyll content of maize from hyperspectral reflectance," *J. Appl. Remote Sens.* **7**(1), 073575 (2013).
46. W. Chivasa, O. Mutanga, and C. Biradar, "Phenology-based discrimination of maize (Zea mays L.) varieties using multitemporal hyperspectral data," *J. Appl. Remote Sens.* **13**(1), 017504 (2019).

47. M. Muntean and F. D. Militaru, "Metrics for evaluating classification algorithms," in *Education, Research and Business Technologies, Smart Innovation, Systems and Technologies*, C. Ciurea, P. Pocatilu, and F. G. Filip, Eds., Vol. **321**. Springer, Singapore (2023).
48. M. Cuenca, M. Campo-Bescós, and J. Álvarez-Mozos, "Crop classification based on temporal signatures of Sentinel-1 observations over Navarre province, Spain," *Remote Sens.* **12**(2), 278 (2020).
49. E. Ibrahim et al., "Mapping crop types and cropping systems in Nigeria with Sentinel-2 imagery," *Remote Sens.* **13**(17), 3523 (2021).
50. A. Moumni and A. Lahrouni, "Machine learning-based classification for crop-type mapping using the fusion of high-resolution satellite imagery in a semiarid area," *Scientifica* **2021**(1), 8810279 (2021).
51. M. Hudait and P. Patel, "Crop-type mapping and acreage estimation in smallholding plots using Sentinel-2 images and machine learning algorithms: some comparisons," *Egypt. J. Remote Sens. Space Sci.* **25**(1), 147–156 (2022).
52. G. Abubakar et al., "Mapping maize fields by using multi-temporal Sentinel-1A and Sentinel-2A images in Makarfi, Northern Nigeria, Africa," *Sustainability* **12**(6), 2539 (2020).
53. J. Fan et al., "Evaluation of crop type classification with different high resolution satellite data sources," *Remote Sens.* **13**(5), 911 (2021).

Bere Benjamin Bantchina received his BSc degree in agricultural engineering from Parakou University, Benin, in 2014 and his MSc degree and PhD in biosystems engineering from Bursa Uludağ University, Turkey, in 2018 and 2024, respectively. His research interests include precision agriculture, proximal soil sensing, site-specific fertigation, variable rate technology, hyperspectral remote sensing, GIS applications, crop modeling, and machine learning.

Kemal Sulhi Gündoğdu received his BSc and MSc degrees in agricultural structures and irrigation from Çukurova University, Turkey, in 1985 and 1988, respectively, and his PhD in agricultural structures and irrigation from Bursa Uludağ University, Turkey, in 1993. He is currently a professor in the Department of Biosystems Engineering at this same university. His research interests include groundwater in agricultural areas, hydrology, meteorology, planning and design of land-water resources, and remote and proximal sensing technologies.