Potato Res. (2026) 69:14
https://doi.org/10.1007/s11540-025-09985-4

Detection of Potato Fields Using Sentinel-2 and Landsat 8 Data-Based Machine Learning Models in Semi-arid Region of Central Anatolia, Türkiye

Emre Tokluoğlu[1] · Bedri Kurtuluş[1,3] · Çağdaş Sağır[2,3,9] · Günseli Erdem Altın[4] · Elifnur Yurdakul[5] · Ersin Ateş[6] · Mustafa Can Canoğlu[7] · Sevim Seda Yamaç[8]

Received: 12 November 2024 / Accepted: 8 October 2025 / Published online: 19 January 2026

Abstract
Accurate mapping of crop types is essential for agricultural monitoring, resource management, and food security. This study evaluates the performance of two ensemble machine learning algorithms—random forest (RF) and gradient tree boosting (GTB)—for classifying potato fields using multispectral satellite imagery from Landsat 8 and Sentinel-2 in the Konya Plain, Türkiye. The methodology involved generating median composite images from the 2020 growing season (June–August), followed by feature extraction from training samples collected via ground truth, satellite, and synthetically generated data. Model performances were assessed using overall accuracy, kappa coefficient, *F*-score, and user's and producer's accuracy metrics. Five-fold cross-validation was employed to evaluate model generalizability. A sensitivity analysis was conducted on the number of trees, and 100 was selected for model training. Results indicate that the Sentinel-2 random forest model achieved the highest average overall accuracy (0.94) and kappa coefficient (0.86) across folds, demonstrating robust and consistent classification. Feature importance analysis showed that red-edge and SWIR bands were the most effective for model performance. Sentinel-2 imagery combined with random forest provided a reliable and efficient approach for potato classification in arid agricultural regions. These findings support the integration of remote sensing and machine learning for operational agricultural monitoring and suggest avenues for future research involving temporal analysis and multi-sensor data fusion. This approach can be extended to other crop types and regions, enhancing the use of satellite-based crop mapping in precision agriculture. It enables policymakers and agricultural agencies to implement timely, data-driven strategies for agricultural planning and food security.

Keywords Crop mapping · Gradient tree boost · Landsat 8 · Potato · Random forest · Sentinel-2

Extended author information available on the last page of the article

Introduction

Ranked as the world's fourth largest food crop following maize, wheat, and rice, the potato, often hailed as the "king of vegetables", serves as a fundamental dietary staple in numerous regions. Potato is cultivated primarily for its nutritious and edible tubers, which play a vital role in global food security and culinary traditions alike (Thapa and Thapa 2019; Zhao et al. 2022). Potato (*Solanum tuberosum* L.) holds a substantial position in global agriculture, boasting a vast production potential of approximately 470 million tons harvested from 23.5 million hectares worldwide in 2022 (FAOSTAT 2024). Potato is one of the cornerstones of agricultural production in Türkiye, as it is the 6th plant product grown in the country (FAOSTAT 2024). According to 2022 data, approximately 5.2 million tons of product were obtained from 0.14 million hectares of cultivated area (TÜİK 2024). With this amount, Türkiye ranks 15th in the world's potato production, and the country's potato production corresponds to 1.12% of the world's total production in tons (FAOSTAT 2024). While potato production in Türkiye was 5.1 million tons in 0.14 million hectares of cultivated area in 2021, in 2023, the cultivated area increased to 0.15 million hectares and total production increased to 5.7 million tons (TÜİK 2024). This increase highlights the critical role of potato cultivation in the Turkish agricultural environment and makes a significant contribution to both local food security and economic growth (Yavuz et al. 2012). Türkiye's top five potato producing cities are Konya, Niğde, Afyonkarahisar, Kayseri, and İzmir. Konya, which ranked first with 622,435 tons of potato production in 2021, fell to 3rd place with 518,677 tons of potato production in 2022 and to 5th place with 462,545 tons of potato production in 2023 (TÜİK 2024).

In recent years, numerous technological applications, including artificial intelligence, remote sensing, and precision agriculture tools, have been developed and employed to improve the cultivation, monitoring, and management of potato crops (Radwan et al. 2024; Eed et al. 2025). Accurate high-resolution mapping of crop types is essential for a range of agricultural and food security applications. This includes monitoring crop yields, assessing the impact of natural disasters on agricultural production, and analysing the effects of climate change on agriculture. However, effectively mapping crop types across large spatial scales remains a significant challenge (Wang et al. 2023). At present, classification and recognition of crops in agricultural lands is a necessity. The automatic process of determining the type of product grown in agricultural land, facilitating remote monitoring and efficient use of existing technology by utilizing the data provided by observation satellites from the soil, is an important input for calibrating models that will be a basis for subsequent analyses or a reference for decision-making systems (Bertellini et al. 2023). Traditional crop mapping methods are time-consuming, labour-intensive, and limited in scope (Gilbertson et al. 2017). In contrast, the integration of machine learning and satellite imagery offers fast, scalable, and highly accurate crop mapping over large areas (Bourriz et al. 2025). This approach significantly supports precision farming practices and data-driven decision-making processes by enabling more effective monitoring of agricultural areas.

Machine learning models have been increasingly applied across various fields in recent years (El-Kenawy et al. 2024a, b; Yassen et al. 2024). These advancements have also influenced the agricultural sector, where the integration of machine learning with remote sensing technologies presents significant potential for improving agricultural practices. The use of remote sensing methods, which are an output of developing technologies, together with machine learning, provides significant contributions to scientists in meeting this need. In recent times, the rapid advancement of sensor technology, computer science, and vision detection technology has significantly hastened the progress of agricultural automation and intelligence. These cutting-edge technologies are revolutionizing farming practices, enhancing efficiency, and leading to smarter, more sustainable agricultural systems (Zhao et al. 2022).

Many scientists have explored various techniques for detecting agricultural products and their derivatives. Among the studied methods are remote sensing analysis, spectral recognition, and machine vision techniques, all of which contribute to advancing the field of agricultural product detection and analysis. These methodologies play a crucial role in enhancing accuracy and efficiency in agricultural monitoring and quality control processes. These methodologies have paved the way for the development of a practical real-time weed identification system suitable for precise weeding in potato fields, even with constrained hardware resources (Zhao et al. 2022). Towfek and Khodadadi (2023) developed a deep convolutional neural network, enhanced through six distinct data augmentation methods, to detect diseases across 39 plant species from leaf images, achieving a classification accuracy of 83.12%. Abdelhamid et al. (2025) introduced an improved ResNet-59 deep learning model that accurately predicts potato harvests using global production data from 1961 to 2021, demonstrating superior performance over five other architectures with a mean squared error of 0.0083 and an R^2 value of 99.05%. Scientists used remote sensing and machine learning methods in mapping by evaluating the matching accuracy of three main classifier methods with four types of Landsat satellite-related data input (Yang et al. 2021). They used Landsat 8 and Landsat 9 satellite imagery to identify another significant agricultural product, specifically winter wheat. Artificial neural networks of varying capabilities were employed to analyse uncertainties associated with this identification process (Zhang et al. 2023). They investigated the efficacy of various machine learning classification techniques, such as random forest (RF), gradient tree boost (GTB), classification and regression trees (CART), and support vector machine (SVM), for crop classification purposes (Neetu2019, Savitha and Talari 2025). In another application, since some semantic segmentation models in product type classification have been found to have some deficiencies in processing the temporal dimension of time series images and in processing multiple spectral bands present in satellite images, methodologies have been proposed with appropriate band combinations for time series visual data (Khan et al. 2023). Another application is to map rice crops from Sentinel-1 time series (C band) using the long short-term memory (LSTM) model and its improvements such as Bi-directional LSTM (Bi-LSTM) of recurrent neural network (RNN). The results found in the study were compared with machine learning techniques such as SVM, RF, k-nearest neighbours (k-NN), and normal Bayes (NB) and found to have high overall accuracy (Crisóstomo de Castro Filho et al. 2020). In another application, Bertellini et al. (2023) have studied the task of product classification with indices derived from image data acquired by Sentinel 1 and 2 satellites. For this purpose, they focused on improving a binary classification for each of the products of interest, aiming to correctly distinguish the target product from other possible products, using traditional machine learning methods (Bertellini et al. 2023).

In the applications of remote sensing methods, Sentinel-1 (SAR) and Sentinel-2 (optical) satellite images were also used on the Google Earth Engine (GEE) platform at the regional scale, and a vegetation type mapping method was proposed by combining single-time feature images and time-series feature images (Wang et al. 2023). Some studies to improve the remote sensing method have been conducted by reviewing the use of RF classifier. For this, the variable importance measure provided by the RF classifier has been extensively observed in different scenarios, such as reducing the number of dimensions of hyperspectral data, determining the most relevant multi-source remote sensing and geospatial data, and selecting the most appropriate season to classify certain target classes (Belgiu and Drăguţ 2016).

Random forest (RF) has proven to be an effective tool for crop classification in various studies leveraging remote sensing data. Similarly, Monsalve-Tellez et al. (2022) explored the use of RF for classifying oil palm crop cover by fusing SAR and optical images. Their findings indicated that RF outperformed other classifiers in terms of accuracy and robustness. Another notable study by Iwańska et al. (2018) applied RF to examine the factors influencing winter wheat yield variation across fields in Poland. This research underscored RF's capability to manage complex interactions between environmental and management factors, offering valuable insights for agricultural management and planning.

Sharma and Ghosh (2023) evaluated the potential of the 8-band PlanetScope dataset for crop classification in the Haridwar district using RF and gradient tree boosting (GTB) algorithms within Google Earth Engine, finding that random forest achieves higher accuracy and that the dataset is effective for classifying crops, especially wheat. The highest accuracy was achieved for wheat, followed by sugarcane and mustard, highlighting the effectiveness of the PlanetScope 8-band dataset for crop classification in heterogeneous fields. Another study compared the performance of different machine learning algorithms for large area crop classification in ten Indian districts, using a range of satellite and climatic data, and finds that random forest is the best performer, followed by gradient boosting (Shukla et al. 2018). Husain et al. (2024) studied a hybrid approach combining the multi-module spatial semantic network (MSSNet) with the gradient boosting classifier (GBC) proposed to improve crop classification accuracy, leveraging strengths in feature extraction and pixel mapping to outperform traditional methods with an accuracy of 98.67% using Sentinel-2 data. Another study compared the classification performance of SAR, optical, and fused data for crop classification using different machine learning algorithms, finding that GBT yielded good results due to their ability to handle imbalanced data and the fused data yielded the highest accuracy and post-processing improved SAR classification accuracy (Dimov et al. 2017). RF and GTB have emerged as effective methods for crop classification around the world and in Türkiye. Studies have shown RF outperforms traditional methods like maximum likelihood classification, achieving overall accuracies of 85.89% for agricultural crops

(Ok et al. 2012) and 83.54% for tea and hazelnut plantations (Akar and Gungor 2015). The integration of texture information and vegetation indices with RF can significantly improve classification accuracy (Akar and Güngör 2015). Evrendilek and Gulbeyaz (2011) investigated the impact of different MODIS and ancillary data combinations on land cover classification accuracy in Turkey using boosted decision trees (BDT), finding that certain combinations improved accuracy, particularly for specific land cover types like grasslands and mixed forests, and concluding that BDTs could enhance national-scale land cover classification accuracy. Şimşek (2025) compared the performance of different machine learning algorithms for agricultural crop classification using corrected farmer-declared parcels as ground truth data, revealing XGBoost achieved the highest overall accuracy (92.14%) in crop classification, followed by RF, SVM, and ANN. Time-series satellite imagery, such as Landsat data, has proven valuable for crop classification using RF, with overall accuracies reaching 81% (Tatsumi et al. 2015). The RF algorithm has also been applied to predict forest fires in Türkiye, demonstrating superior performance compared to other models (Garg et al. 2023). These studies highlight the versatility and effectiveness of RF in various agricultural and environmental applications, making it a valuable tool for land cover mapping and monitoring in Türkiye and beyond.

Recent studies have explored the application of random forest (RF) algorithms in potato-related research in various contexts. RF has shown promise in potato disease detection, achieving 97% accuracy in classifying early and late blight in potato leaves using image segmentation (Iqbal and Talukder 2020). In Spain, RF was employed to classify potato yield from Sentinel-2 satellite data, achieving 65–77% accuracy across multiple cropping seasons (Uribeetxebarria et al. 2023). Additionally, RF was evaluated alongside other machine learning algorithms for detecting potato late blight using UAV-based multispectral imagery, demonstrating good performance in terms of accuracy and run time (Rodriguez et al. 2021). El-Kenawy et al. (2025) studied the use of machine learning and deep learning models to predict potato yield and found that GNNs and LSTMs showed higher accuracy and better captured complex spatial and temporal patterns compared to gradient boosting and XGBoost. Another study evaluated using an ensemble of gradient boosting machines to classify disease severity of potato Verticillium wilt from UAV-based multispectral images, showing potential and challenges in improving disease identification in commercial potato crops in Colombia (Lizarazo et al. 2023).

The decrease in water resources due to climate change significantly impacts crop mapping accuracy by altering phenological cycles, planting dates, and plant development on a seasonal scale. This variability complicates the differentiation of crop types in agricultural areas on satellite images and induces spectral confusion. However, crop mapping studies utilizing satellite imagery with temporal and spatial continuity offer enhanced monitoring of land dynamics, providing more detailed and accurate insights into the effects of climate change. This study introduces a multi-sensor approach for mapping potato crops in a semi-arid region. Potato fields in the Çumra district of Konya, Türkiye, were delineated using the random forest (RF) and gradient tree boost (GTB) algorithms implemented within the Google Earth Engine platform, based on Sentinel-2 and Landsat 8 imagery. This study tests whether an effective potato field classification can be achieved using a very limited amount of ground-truth data and only the raw bands of satellite imagery. The proposed methodology enables fast, robust, scalable, and accurate crop monitoring and is transferable to other geographic regions and crop types, thereby supporting data-driven decision-making in precision agriculture and food security strategies.

Materials and Methods

Study Area

The study area is in Çumra district of Konya city. The area is located in the Konya Plain, southeast of the Konya city centre, within the Central Anatolia region of Türkiye (Fig. 1). The exact study area is Çumra district border provided from the Directorate General for Mapping (Ministry of National Defence, Republic of Türkiye) (Fig. 1). This region is recognized as one of Türkiye's most vital agricultural zones and is characterized by a semi-arid climate. Monthly total precipitation and monthly average temperature data for the city of Konya between 2020 and 2023 are presented in Fig. 2 to provide a general understanding of the regional climate. Monthly temperature and precipitation data for the study area were obtained from the TerraClimate (Abatzoglou et al. 2018) and CHIRPS (Funk et al. 2015) datasets, respectively. Temperature and precipitation patterns were quite similar between 2020 and 2022. The only notable difference was observed in 2023, when spring precipitation occurred with a one-month delay. The average annual rainfall in the

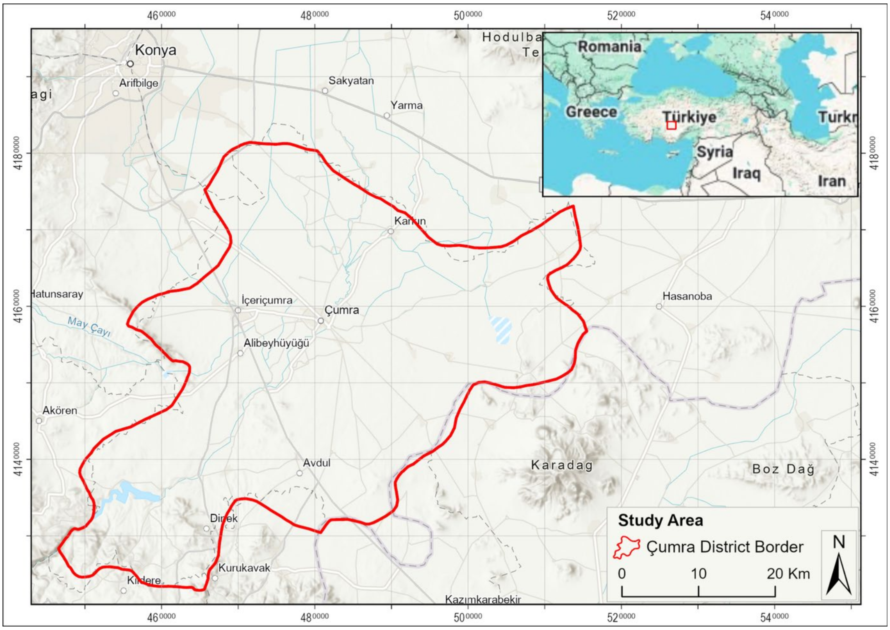

Fig. 1 Location map of the study area

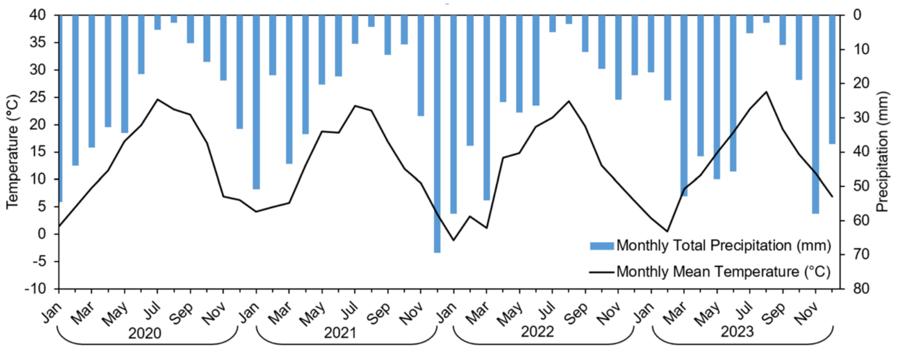

Fig. 2 Monthly precipitation and temperature values for Konya city for 2020, 2021, 2022, and 2023

area ranges from 280 to 500 mm, with significant variations across different parts of the plain (Fig. 2). Summers are typically hot and dry, with average temperatures often exceeding 30 °C, while winters can be cold, with temperatures dropping below freezing. The annual mean temperature ranges between 10 and 12 °C (Yamaç 2018). Çumra, a district located in the southeast of Konya province, is known for its significant agricultural production. Situated in the fertile lands of the Konya Plain, Çumra stands out with its irrigation facilities and extensive agricultural areas. Çumra has a continental climate characterized by cold, harsh winters and hot, dry summers. These climatic conditions provide a suitable environment for the cultivation of many crops. The soil in the district is generally clayey and calcareous, making it favourable for agriculture.

Data

Ground Data

The data on fields planted with potatoes were collected through field studies and are available in polygon shapefile format for the year 2020. In the study area, 44 fields were planted with potatoes. Within this dataset, the planting date information for each polygon is available. The planting dates for the potato fields generally range between April 14 and April 15. There are six fields where planting was not done within this date range. Among these, three were planted on March 26, 28, and 31, while the other three were planted on May 30. The potato planting and harvesting periods in the region do not exhibit significant year-to-year variation. Planting typically occurs between mid-April and early May, while harvesting generally takes place between late August and early September.

Data Generation

In machine learning applications, the performance and accuracy of prediction algorithms are generally enhanced with an increased amount of training data and features. However, in real-world field studies, it is often challenging to obtain an ideal quantity of ground truth data. In this study, real field data were used, and the number of potato-labelled samples was limited to 44 for the entire study area. The primary novelty of this work lies in investigating whether high-accuracy predictions can be achieved despite the limited number of ground truth samples. Recognizing that 44 labelled potato fields alone would be insufficient to effectively train a machine learning model, larger sample datasets were created. These larger sample datasets were deliberately incorporated to compensate for this limitation. It was expected that this would enhance the model's training performance by providing a broader context of land cover classes.

In addition to potato fields data, two other classes were created as "other fields" and "others". "Other fields" class represents the fields other than potato fields, while the "others" class represents all the other land cover types. Since the field data includes only locations where potatoes are present, it is necessary to obtain data from other land cover types to enable the machine learning (ML) algorithm to learn about the "other fields" and "others" classes. A total of 3649 sample points for "other fields" were created with the combination of "211: Non-irrigated arable land" and "212: Permanently Irrigated Land" classes of CORINE Land Cover 2018 data. To create "others" class data, approximately 800 points were manually created in ArcGIS Pro software through visual inspection of non-field areas using relevant satellite images. These points were carefully placed in locations outside agricultural fields, such as bare soil/rock, buildings, roads, surface water, and wooded areas. Additionally, around 489 extra points were randomly and automatically generated in ArcGIS Pro, covering an area slightly larger than the study area (Fig. 3). In total, 1289 sample points for "other" class were created. A total of 4982 data samples were used overall.

Some of the generated points for "other fields" and "other" classes may have fallen within potato fields identified by field data. To address this issue, these points located within known potato field polygons were identified and removed. However, it is important to note that some of these generated points might have fallen within actual potato fields that were not included in the field data. As a result, some points that are indeed potato fields may have been incorrectly labelled as "other fields" or "others".

Satellite Images

It is essential that the potato plant leaves were visible in the satellite imagery to perform accurate classification. Therefore, the classification process was conducted using images captured in June, July, and August, when the leaf development of the plants was sufficiently advanced. The June–August compositing window was selected based on the phenological development of the potato crop, which in the study area is typically planted in mid-April and harvested by the end of August. This period corresponds to the peak vegetative growth and early tuber bulking stages, generally occurring between 60 and 90 days after planting. During this phase, the potato canopy reaches its maximum leaf area and chlorophyll content, resulting in strong, stable spectral reflectance characteristics. These phenological traits

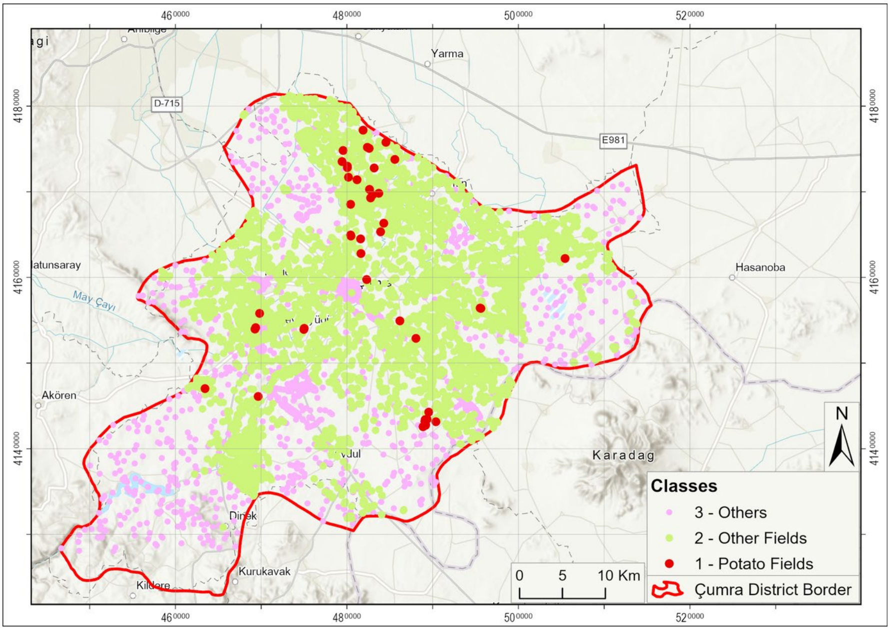

Fig. 3 Sample point classes and their spatial distribution

enhance the spectral separability of potato fields from other crop types, many of which are either in early vegetative stages or entering senescence. Consequently, the June–August window provides optimal conditions for image compositing and crop classification, improving the reliability and accuracy of potato mapping efforts. Images from September were not included because the crops were harvested by that time, rendering the plants invisible in the satellite images.

In this study, Landsat 8 and Sentinel-2 satellite images were employed to classify potato fields. Landsat 8 OLI/TIRS Collection 2 Tier 1 Surface Reflectance products and Sentinel-2 Multi Spectral Instrument Level-2A Surface Reflectance products were accessed and used in the Google Earth Engine environment. These products do not need any pre-processing procedure since they were already orthorectified and atmospherically corrected on the Google Earth Engine Catalog.

For June–August 2020, 36 and 12 images were found and used from Sentinel-2 and Landsat 8, respectively. The temporal distribution of the images is presented in Fig. 4. The used spectral bands for Sentinel-2 images included the following: blue, green, red, red edge 1, red edge 2, red edge 3, NIR, red edge 4, SWIR 1, and SWIR 2 (corresponding to B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12). The used spectral bands for Landsat 8 images included the following: blue, green, red, near-infrared, shortwave infrared 1, shortwave infrared 2 (corresponding to SR_B2, SR_B3, SR_B4, SR_B5, SR_B6, and SR_B7). The spatial resolution of the Sentinel-2 bands ranged between 10 and 20 m, while the Landsat 8 bands had a resolution of 30 m. Both satellite image band values were scaled according to the provider's declared specifications. A median composite image for each

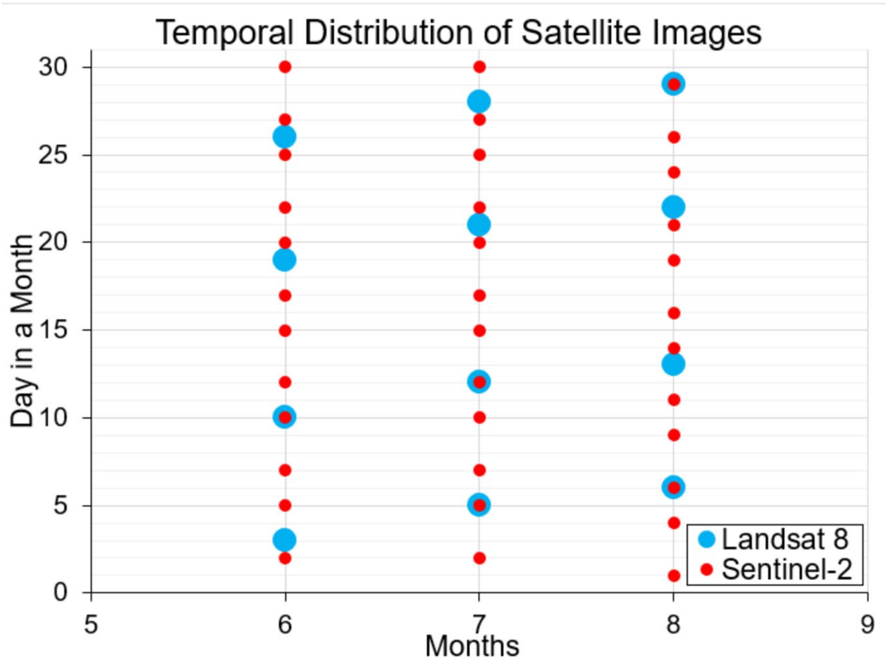

Fig. 4 Temporal distribution of the used satellite images

satellite was created using multiple images to be used in machine learning processes. Only the cloud-free pixels were used for the compositions. The Sentinel-2 composite image was rescaled to 20 m due to the spatial resolution of its bands varying between 10 and 20 m. However, since all bands in the Landsat 8 imagery have the same spatial resolution, they were used as-is without any rescaling.

This study aimed to explore the potential of achieving high prediction accuracy using a limited number of ground truth samples in training machine learning algorithms. In line with this objective, and to test the feasibility of generating reliable predictions under data-constrained conditions, we intentionally relied solely on readily available spectral band values from satellite imagery without incorporating additional derived indices. While it is well established in the literature that incorporating vegetation indices such as NDVI or EVI can enhance classification performance, the focus of this study was to demonstrate the potential accuracy that can be achieved using the most direct and computationally efficient input features, namely the raw spectral bands. Additionally, considering that vegetation indices are themselves derived from combinations of spectral bands, it was also of interest to observe whether specific combinations of these bands would emerge as more prominent contributors to the machine learning model's predictive performance. This approach was chosen to research the predictive power of spectral bands alone, without the need for further data processing or computational overhead, particularly in scenarios with limited resources or data availability.

Method

Random Forest (RF) Classification

Random forest (RF) is an ensemble learning algorithm extensively utilized for classification and regression tasks, particularly in remote sensing and agricultural studies (Watts and Lawrence 2008; Venkatanaresh and Kullayamma 2022; Millard and Richardson 2015). The algorithm operates by constructing numerous decision trees during the training phase and then making predictions based on the majority vote for classification tasks or the average prediction for regression tasks (Breiman 2001; Monsalve-Tellez et al. 2022). RF leverages the simplicity of decision trees while enhancing their performance through ensemble learning, which improves accuracy and mitigates the risk of overfitting (Liaw and Wiener 2002; Elghazel et al. 2011). Specifically, RF generates a "forest" of decision trees by employing bootstrap sampling of the training data and selecting a random subset of features at each node split. Each tree is constructed using a different subset of the data, with distinct features considered at each split, promoting diversity among the trees. This diversity strengthens the algorithm's robustness, leading to superior performance when aggregating the predictions of the individual trees. The final prediction in classification tasks is typically determined by majority voting (Breiman 2001; Sharma and Ghosh 2023).

The optimal number of trees in a random forest (RF) model is a topic of ongoing debate. Some researchers advocate for using the largest number of trees that computational resources allow (Probst and Boulesteix 2018), while others argue that performance may plateau or even degrade when the number of trees becomes excessively large (Oshiro et al. 2012). Theoretical analyses indicate that the expected error rate can exhibit non-monotonic behaviour as the number of trees increases, although this observation may not hold for other performance metrics (Probst and Boulesteix 2018). Empirical studies have identified a threshold beyond which additional trees yield diminishing returns in performance improvement (Oshiro et al. 2012). Non-deterministic strategies for optimizing the number of trees have shown promise, achieving high accuracy while also reducing computational costs (Senagi and Jouandeau 2018). These approaches demonstrate significant improvements in accuracy, tree count, and execution time compared to deterministic methods (Senagi and Jouandeau 2018). For example, Wald et al. (2013) reported that a random forest with 100 trees and 200 selected features is optimal for bioinformatics research, whereas Ramadhan et al. (2017) found that 300 is optimal for gender classification based on voice frequency. Ultimately, the optimal number of trees depends on the specific dataset and the computational resources available.

In this study, the RF algorithm was implemented on the Google Earth Engine platform. Training was conducted using pixel-based spectral band values derived from median composite satellite imagery. Field data labelled as "potato fields" and synthetically generated data representing "other fields" and "others" were used for both training and validation. A five-fold cross-validation was performed, with 80% of the data used for training and 20% for validation in each fold. Spectral bands from Sentinel-2 and Landsat 8 imagery were utilized as input features. Final classification was carried out following a sensitivity analysis on the number of trees. A value of 100 was selected for the number of trees and applied consistently across all classifications. Subsequently, the relative importance of each spectral band was assessed through feature importance scores. To evaluate the reliability of the classification results, performance metrics including overall accuracy, user's accuracy, producer's accuracy, *F*-score, and the kappa coefficient were calculated. Finally, surface areas for each class were estimated, and classification maps were generated.

Gradient Tree Boost (GTB) Classification

Gradient tree boost (GTB), also referred to as gradient boosted decision trees, represents a supervised machine learning approach based on ensemble learning principles. The method involves the construction of multiple shallow decision trees, known as weak learners, which are added sequentially to form a composite predictive model (Friedman 2001). Each successive tree is trained to approximate the residuals or errors produced by the preceding ensemble, thereby incrementally enhancing overall classification performance. Unlike bagging-based methods such as random forests, boosting algorithms build trees in a stage-wise fashion, where each subsequent model aims to minimize the residual errors of the ensemble accumulated so far. This process enables GTB to effectively reduce both model bias and variance over successive iterations.

In classification tasks, GTB minimizes a chosen loss function—commonly the logistic loss for binary outcomes—using gradient descent techniques in function space (Dumont et al. 2009; Hastie et al. 2009). At each iteration, a new decision tree is fitted to the negative gradients (pseudo-residuals) of the loss function computed with respect to current predictions. The algorithm incorporates regularization mechanisms such as shrinkage (i.e. learning rate adjustment), tree pruning, and subsampling strategies to mitigate overfitting and stabilize learning. Advanced implementations, including XGBoost, introduce additional enhancements such as second-order gradient optimization and sparsity-aware splitting to improve computational efficiency (Chen and Guestrin 2016). The XGBoost framework has gained popularity for its scalability and efficiency in implementing gradient boosting algorithms (Ayyadevara 2018).

In this study, the gradient tree boost (GTB) algorithm was implemented following the same procedure used for the random forest algorithm, as mentioned in the last paragraph of the "Random Forest (RF) Classification" section.

Performance Evaluation

The performance of the machine learning algorithms was evaluated using k-fold cross-validation. For this purpose, metrics such as overall accuracy, consumer's accuracy, producer's accuracy, kappa coefficient, and *F*-score were calculated from the error matrix, and their means and standard deviations were assessed. After the final classification, feature importances along with the same metrics were again calculated based on the error matrix to evaluate the results. Finally, the total surface area of the potato fields was calculated and compared with the official data from the Turkish Statistical Institute (TÜİK).

K-fold cross-validation is a resampling technique used to assess the generalizability of a model by partitioning the dataset into k equal parts; each subset is used once as a validation set while the others serve as training data (Soper 2021). This approach reduces bias and variance in model evaluation and provides a more stable estimate of predictive performance.

Feature importance indicates the relative contribution of each input variable (spectral bands in this case) to the prediction process within tree-based models (Breiman 2001). In ensemble algorithms like random forest or gradient boosting, it helps in identifying which features most significantly influence the model's outcome.

An error matrix, also known as a confusion matrix, is a fundamental tool for evaluating classification performance by comparing predicted and actual class labels (Foody 2002). It enables the derivation of various accuracy metrics, such as overall accuracy, user's (consumer's) and producer's accuracy.

Overall accuracy measures the proportion of correctly classified instances across all classes. Consumer's (user's) accuracy indicates the reliability of a given class in the classified map, while producer's accuracy reflects the classifier's ability to identify that class correctly (Congalton 1991). The kappa coefficient accounts for agreement occurring by chance, providing a more robust measure than overall accuracy alone (Cohen 1960). The F-score combines precision and recall into a single metric, particularly useful in imbalanced datasets (Sokolova and Lapalme 2009).

Processing

All processes conducted in this study were carried out on the Google Earth Engine (GEE) platform, with the exception of the final mapping steps, which were performed using ArcGIS Pro. The entire processing workflow was executed through coding within the GEE environment. Due to GEE's user-allocated memory limitation, the workflow was divided into four separate code scripts. These were as follows: random forest (RF) classification using Sentinel-2 imagery, RF classification using Landsat 8 imagery, gradient tree boost (GTB) classification using Sentinel-2 imagery, and GTB classification using Landsat 8 imagery. Each of these code scripts followed the same methodological steps as described in the following paragraphs.

Initially, the study area polygon and the sample data in point format were uploaded to the GEE workspace. Subsequently, satellite imagery was filtered for clouds and scaled appropriately in the study area limit. Images from June, July, and August of 2020 were queried, and multiple images were combined into a single composite by calculating the median pixel values across bands. This process produced one median composite image representing the summer season of 2020.

The spectral band values corresponding to the pixel locations in the composite image were extracted to generate the sample dataset to be used for machine learning. A fivefold k-fold cross-validation procedure was applied with 100 number of trees. The k-fold cross-validation split the data into training and validation sets by 80% and 20%, respectively. Reserving 20% of all available data for validation means allocating one-fifth of the data. Therefore, using 5 folds allows the entire dataset to be thoroughly and representatively covered for both training and validation without leaving out any portion. The overall accuracy, consumer's accuracy, producer's accuracy, kappa coefficient, and *F*-score for each fold, as well as the averages and the standard deviations, were calculated and reported.

Following the cross-validation, the final machine learning model was trained with 100 numbers of trees and all the available sample data. After model training, the relative feature importances of the spectral bands were calculated. The trained model was then used to classify the satellite imagery, and classification accuracy was assessed using all the sample data. To evaluate the classification performance, the error matrix, overall accuracy, consumer's accuracy, producer's accuracy, kappa coefficient, and *F*-score were computed.

Finally, the surface area of each class was calculated, and the resulting classified raster images were exported from GEE and mapped using ArcGIS Pro.

The overall methodological process is outlined in flowchart format in Fig. 5.

Results

This study presents a multi-sensor approach using random forest and gradient tree boost algorithms on Sentinel-2 and Landsat 8 data to accurately map potato fields in a semi-arid region of Konya, Türkiye, relying on minimal ground-truth data.

K-fold cross-validation was performed using 5 folds, and for each fold, the overall accuracy, kappa coefficient, user's accuracy, producer's accuracy, and *F*-score values

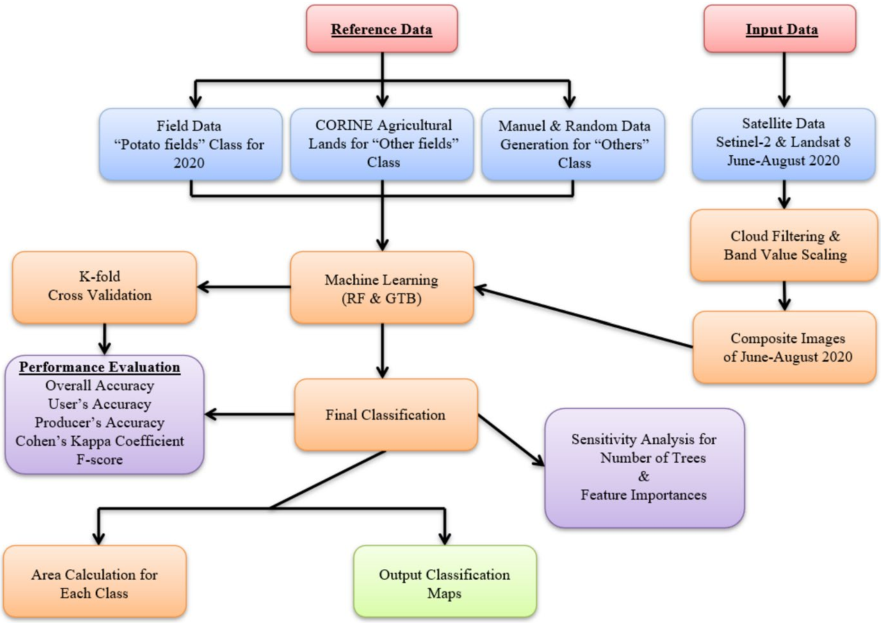

Fig. 5 Methodological flow chart

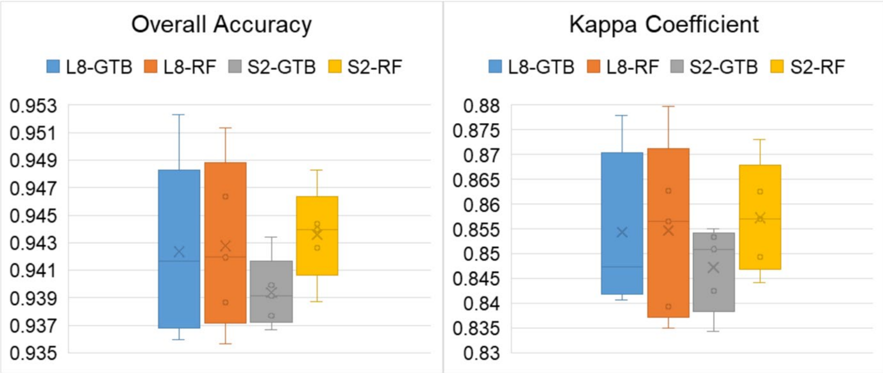

Fig. 6 Box plot for overall accuracy and kappa coefficient values for k-fold cross validation

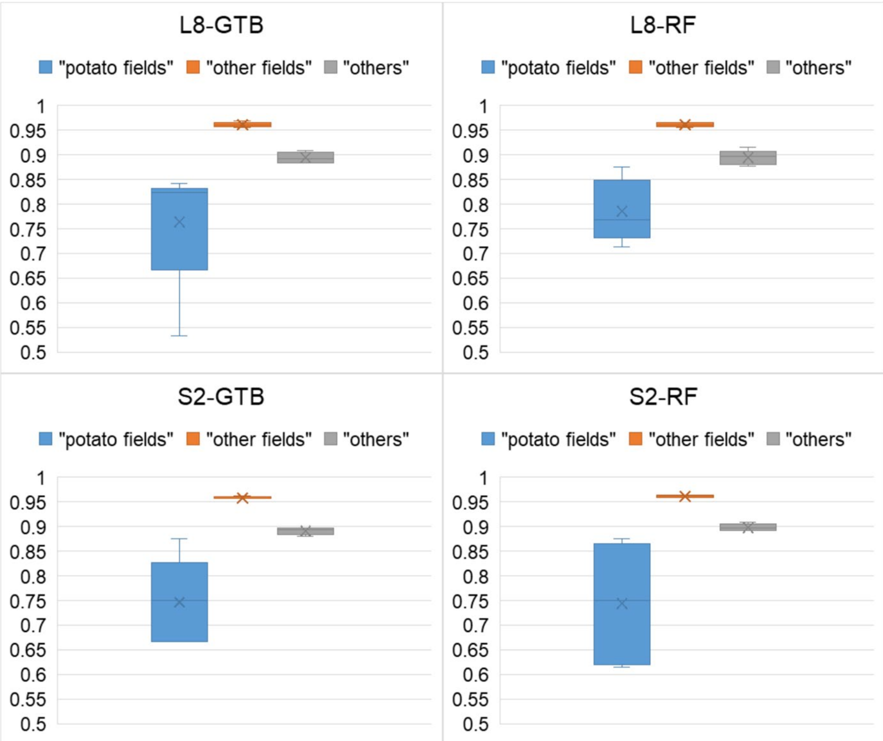

Fig. 7 Box plot for k-fold cross validation F-score values

were calculated. The resulting values from these calculations were visualized using box-whisker plots. The overall accuracy and kappa coefficient plots are presented in Fig. 6, F-score plots in Fig. 7, user's accuracy plots in Fig. 8, and producer's accuracy plots in Fig. 9.

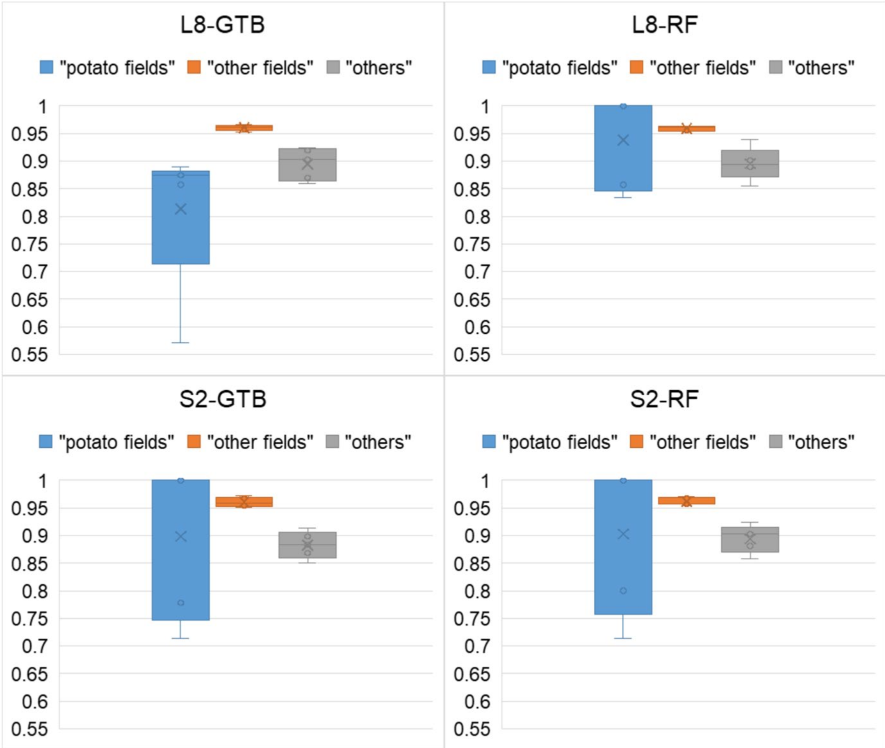

Fig. 8 Box plot for k-fold cross validation user's accuracies

Figure 6 presents the distribution of overall accuracy (OA) and kappa coefficient values derived from five-fold cross-validation for all four classification scenarios (S2-RF, L8-RF, S2-GTB, L8-GTB). The box-whisker plots summarize the central tendency and variability of classification performance across the folds. Among the models, S2-RF achieves the highest median and mean OA, indicating strong classification accuracy, with moderate variability across folds. In contrast, S2-GTB exhibits the lowest OA but also the least variability, suggesting consistent yet limited performance. Both L8-RF and L8-GTB show comparable OA values, with L8-RF performing slightly better overall. The kappa coefficient values mirror the trends observed in OA, consistently scoring about 0.1 points lower, which is expected due to the metric's adjustment for chance agreement. Again, S2-RF yields the highest kappa values, reinforcing its robust predictive capability, while S2-GTB records the lowest. This alignment between OA and kappa across all models suggests a stable classification behaviour with respect to both observed accuracy and chance-corrected agreement.

Figure 7 displays the distribution of *F*-score values derived from five-fold cross-validation for each classification scenario. The *F*-score, representing the harmonic mean of precision and recall, provides a comprehensive measure of

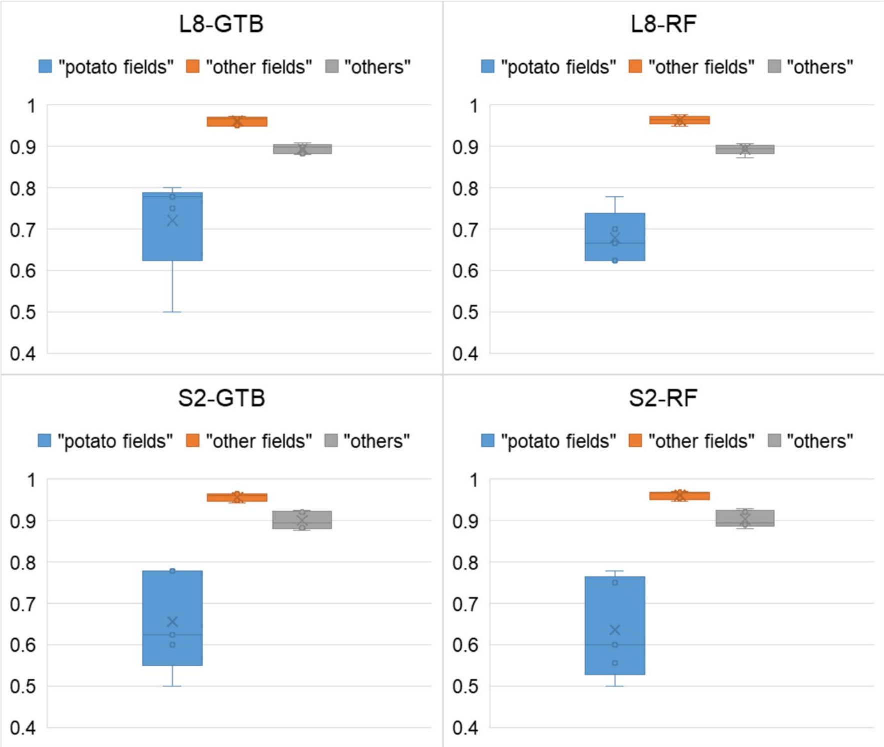

Fig. 9 Box plot for k-fold cross validation producer's accuracies

the models' classification performance, particularly in balancing correct identification and minimizing false positives. For the "potato fields" class, the L8-RF model achieves the highest median *F*-score with the least overall variation, indicating strong performance in both precision and recall across folds. On the other hand, S2-RF yields the lowest median *F*-score and the highest variability among folds. The wider variability suggests a higher sensitivity to the specific partitioning of training and validation data.

Figure 8 illustrates the user's accuracy (consumer's accuracy) for each classification scenario, as obtained from five-fold cross-validation. This metric reflects the likelihood that a pixel labelled as a particular class in the map truly belongs to that class on the ground, thus representing classification reliability from the map user's perspective. Among all models, L8-RF demonstrates the highest median user's accuracy for "potato fields" class. It is followed by S2-RF and S2-GTB, which are quite similar and show a slightly lower median with a wider interquartile range across folds. In contrast, L8-GTB results in the lowest median and moderate variability in user's accuracy, implying a greater likelihood of commission errors. Overall, the plots reveal that models employing the random forest algorithm outperform gradient tree boosting in minimizing mislabelling, particularly when applied to Landsat 8 imagery, which exhibits superior user-oriented reliability.

Figure 9 presents the producer's accuracy values obtained from five-fold cross-validation for each classification scenario. This metric quantifies the probability that real-world class instances are correctly detected by the classifier, thereby indicating the model's ability to avoid omission errors. Among the four models, L8 classifications achieve the highest median producer's accuracy, highlighting its strong capability in capturing ground truth classes effectively. L8-RF yields lower median accuracy but with lower variation than L8-GTB. GTB and RF using S2 image show quite similar results.

In the final classification stage, feature importance values for the spectral bands of the satellite images used by the trained models are calculated and presented in Fig. 10. Subsequently, the error matrices obtained from the classification are provided in Table 1. Based on these error matrices, overall accuracy, kappa coefficient, *F*-score, and user's/producer's accuracy metrics were calculated and visualized in Fig. 11. After classification, the surface areas of the classes—including the "potato fields"—were calculated and presented in Table 2.

In this study, a sensitivity analysis was conducted to assess how model performance is affected by the number of trees. For this purpose, the number of trees in the ML models was set from 10 to 300 in increments of 10, and the overall accuracy obtained from each training run was reported. According to the results given in Fig. 10, L8-RF and S2-RF produced very similar values, while the overall accuracy of the GTB models showed slightly more variation, with S2-GTB yielding somewhat higher values. However, the key observation is that the RF models exhibited no significant improvement beyond 100 trees. Although using more than 100 trees may

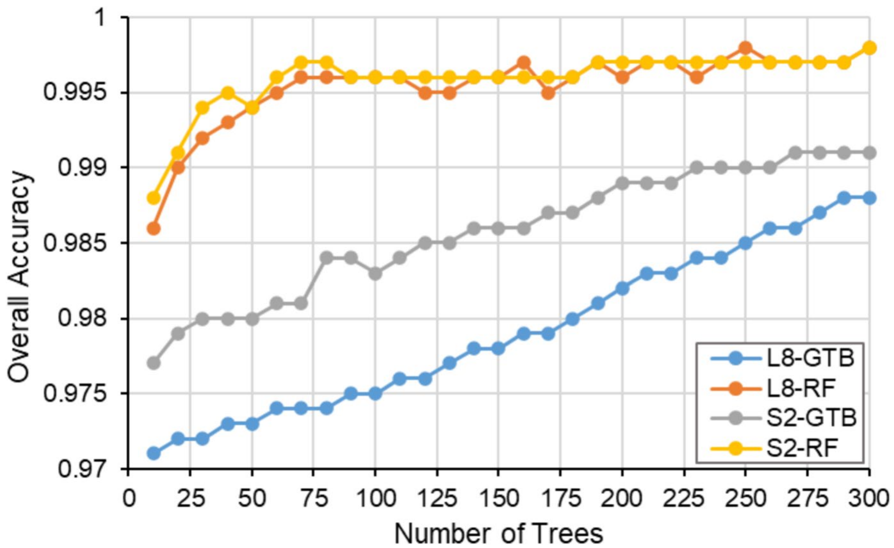

Fig. 10 Sensitivity analysis for number of trees: overall accuracy changes by number of trees per each model-satellite combination

Table 1 Error matrices for each final model-satellite classification: “Pot. F.” for “potato fields”, “Oth. F.” for “other fields”, “Oth.” for “others” classes

L8-GTB				**L8-RF**			
	“Pot. F.”	“Oth. F.”	“Oth.”		“Pot. F.”	“Oth. F.”	“Oth.”
“Pot. F.”	36	8	0	“Pot. F.”	43	1	0
“Oth. F.”	1	3598	50	“Oth. F.”	0	3641	8
“Oth.”	0	64	1225	“Oth.”	0	10	1279
S2-GTB				**S2-RF**			
	“Pot. F.”	“Oth. F.”	“Oth.”		“Pot. F.”	“Oth. F.”	“Oth.”
“Pot. F.”	38	6	0	“Pot. F.”	43	1	0
“Oth. F.”	2	3604	43	“Oth. F.”	0	3640	9
“Oth.”	0	32	1257	“Oth.”	0	11	1278

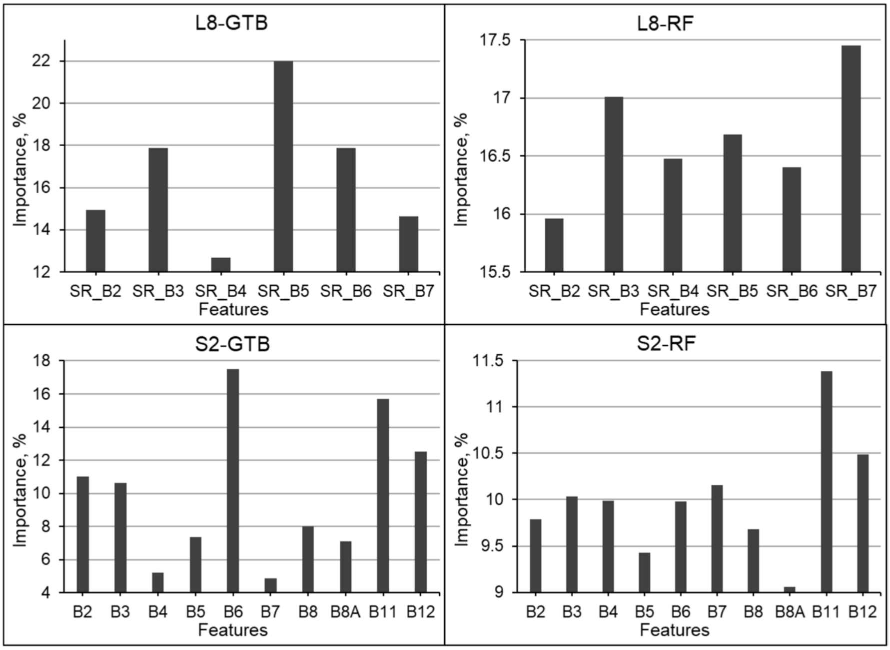

Fig. 11 Feature importances for each final model-satellite training

Table 2 Calculated surface areas per each model-satellite classification: “Pot. F.” for “potato fields”, “Oth. F.” for “other fields”, “Oth.” for “others” classes

Surface Areas, sq km			
	“Pot. F.”	“Oth. F.”	“Oth.”
L8-GTB	8.80	734.01	1345.50
L8-RF	8.77	744.41	1335.13
S2-GTB	8.55	650.19	1429.56
S2-RF	7.81	647.66	1432.84

result in slightly higher accuracy, the gain is not substantial enough to justify the increased computational effort and processing time. In contrast, GTB showed steadily increasing accuracy values, but it is likely that performance gains plateau beyond 250–300 trees. Therefore, to ensure comparability across models and align with common practice in the literature, all models were ultimately run with 100 trees.

Figure 11 displays the relative feature importance values of the spectral bands utilized in the final classification models. These values quantify the contribution of each band to the overall decision-making process within the ensemble classifier. In the L8-GTB classification, the bands with the highest importance values are, respectively, near infrared, shortwave infrared 1, and green. In the L8-RF classification, shortwave infrared 2, green, and shortwave infrared 1 ranked highest in importance. For the S2-GTB model, the top contributing bands are red edge 2, shortwave infrared 1, and shortwave infrared 2, respectively. In the S2-RF classification, shortwave infrared 1, shortwave infrared 2, and red edge 3 were the most important bands in descending order. The range of feature importance values in the RF models was notably narrower compared to the GTB models, indicating that RF utilizes satellite image bands more evenly during the decision-making process. Overall, infrared bands were found to be significantly more influential than visible wavelength bands.

Table 1 presents the confusion matrices derived from the final classification results, offering a detailed summary of classification performance across the three land cover classes: “potato fields”, “other fields”, and “others”. The diagonal elements of the matrices represent correctly classified pixels, while the off-diagonal entries indicate misclassifications between classes. For each model–satellite variation, overall accuracy, kappa coefficient, *F*-score, user’s accuracy, and producer’s accuracy values were calculated based on the corresponding error matrices.

Figure 12 presents a comprehensive graphical summary of the classification performance metrics—including overall accuracy (OA), kappa coefficient, *F*-score, user’s accuracy, and producer’s accuracy—derived from the confusion matrix reported in Table 1. The results demonstrate that the RF model using L8 and S2 images consistently performed significantly better than GTB for “potato fields” classifications. RF models for L8 and S2 yielded quite similar performance metrics. Conversely, GTB models for L8 and S2 yielded lower and more class-wise varying performance metrics than RF models. All evaluated models achieved high classification accuracies. The random forest (RF) model trained on Landsat 8 (L8) imagery consistently outperformed other model-satellite variations, achieving the highest OA value as 0.9962 and kappa coefficient as 0.9904. The Sentinel-2 (S2) RF classifier also exhibited strong performance, with metrics closely trailing those of the L8-RF model. Notably, the minimal discrepancy between user’s and producer’s accuracy in the S2/L8-RF scenario indicates balanced reliability in both omission and commission error rates. Gradient tree boosting (GTB) classifiers showed greater variability and the lowest accuracy metrics, though they still maintained acceptable classification performance. These findings underscore the robustness and stability of the random forest algorithm for land cover classification in the given dataset and conditions.

Table 2 summarizes the calculated surface areas (km^2) for each classified land cover class derived from the final classification outputs, including “potato fields”,

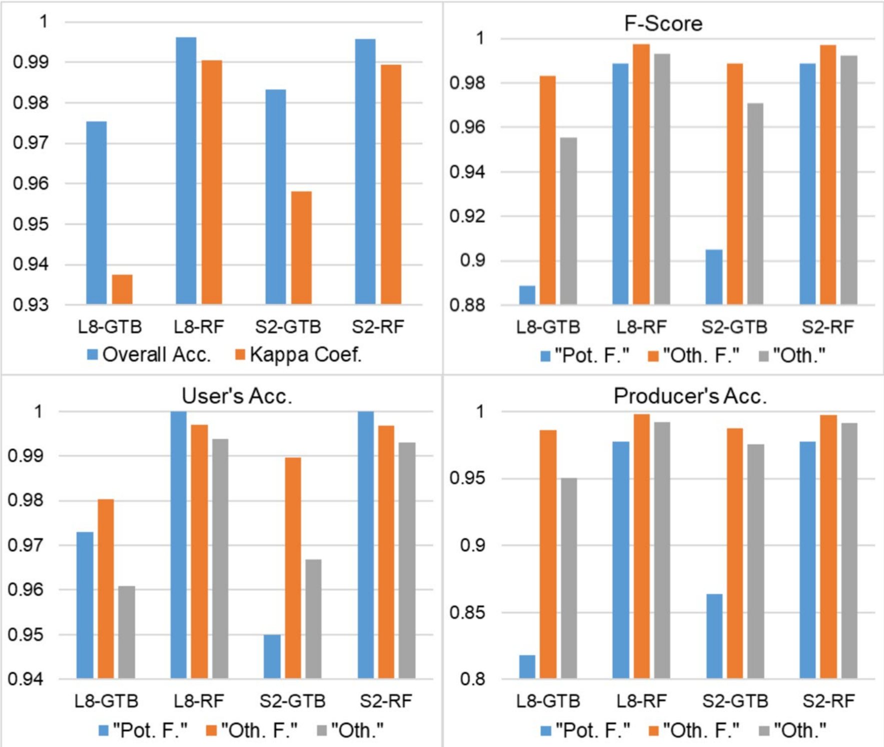

Fig. 12 Overall accuracy, kappa coefficient, *F*-score, and user's/producer's accuracy metrics for each model-satellite final classification

"other fields", and "others" class categories. The average surface area for "potato fields" is 8.48 km^2. Except for the S2-RF estimation, the other three estimations yielded similar values. The official agricultural statistics published by the Turkish Statistical Institute (TÜİK 2024) underscores 15 km^2 for potato fields in Çumra district by 2020. The difference between estimated areas and the official number differs from 41 to 48%.

The result maps of the classification are presented in Fig. 13. These maps visually illustrate the distribution of "potato fields", "other fields", and "others" classes across the study area. Sentinel-2 results displayed more fragmented field boundaries compared to Landsat 8 results.

Discussion

This study aimed to evaluate the effectiveness of two widely used machine learning algorithms—random forest (RF) and gradient tree boosting (GTB)—for crop classification using multispectral satellite imagery from Landsat 8 (L8) and Sentinel-2

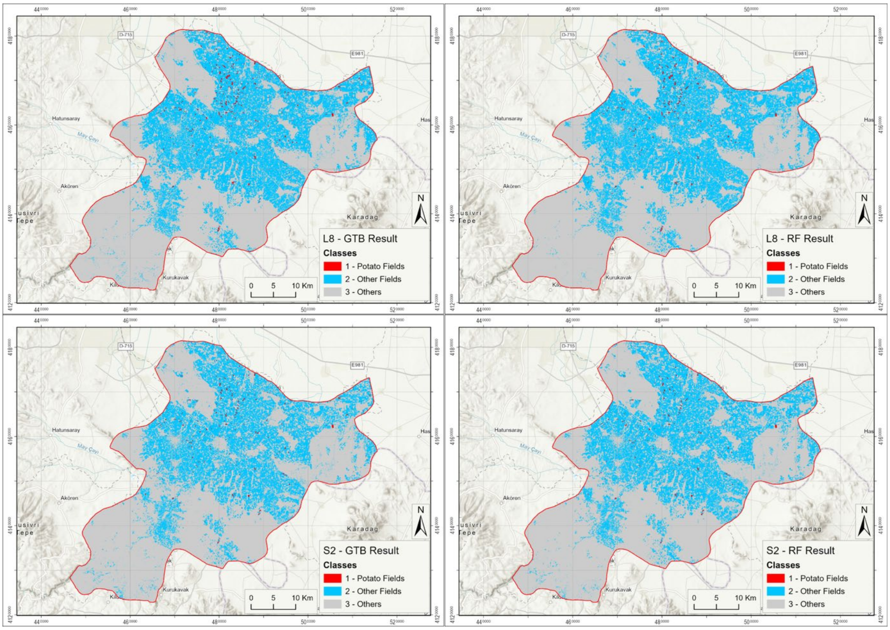

Fig. 13 The classification result maps for each model-satellite final classification

(S2). The classification focused specifically on identifying potato fields within the Konya Plain (Çumra district), utilizing both spatial and spectral features derived from optical data during the growing season. The model performances were validated through various classification metrics and fivefold cross-validation procedures.

The number of field data points used was 44, which is relatively small compared to other machine learning studies. However, it is important to note that the field data used is of very high quality. In practice, collecting or providing large quantities of field data is often challenging due to various constraints. The balance between data quantity and quality in machine learning algorithms is an ongoing area of research. While some studies suggest that larger datasets lead to better performance (Hyontai 2018), others emphasize that data quality is more critical than sheer quantity (Kariluoto et al. 2021). The quality and quantity of training data significantly impact classification accuracy, with at least 30 high-quality ground truth polygons per crop type typically yielding an average accuracy of 85% (Das et al. 2021).

In machine learning applications, the performance and accuracy of prediction algorithms generally improve with larger training datasets. However, in real-world field studies, obtaining an ideal quantity of ground truth data is often challenging. In this study, real field data were utilized, with the number of potato-labelled samples limited to 44 for the entire study area. The primary novelty of this work lies in evaluating whether high-accuracy predictions can be achieved despite this limited ground truth availability. To facilitate more effective model training, the “others” class was subdivided into two categories: “other fields” and “other”. The “other fields” class represents agricultural areas not labelled as “potato fields”, while the “other” class encompasses all remaining non-agricultural land covers. This division was intended to enable the machine learning algorithm to better differentiate between potato fields and other crop types, as the undivided “others” class predominantly included non-crop areas. Given that 44 labelled potato samples alone were insufficient for robust model training, a significantly larger number of samples labelled as “other fields” and “other” were generated and incorporated to provide broader land cover context. This strategy was expected to enhance model performance by improving class separability during training. The effectiveness of this approach was validated through prediction performance metrics, which demonstrated high classification accuracy, confirming the robustness of the methodology.

The composite satellite images used for training and prediction were created by calculating the pixel-based medians of images acquired in June, July, and August of 2020. This timing was chosen to capture the potato plant leaves at their mature stage. In the field data, planting occurred on April 14–15 for all fields except six. Of these six fields, three were planted 2 weeks earlier than the majority, while the remaining three were planted 6 weeks later. It is possible that the fields planted 2 weeks earlier were harvested sooner than the majority, and the fields planted 6 weeks later may not have reached full maturity. Consequently, these six fields in the dataset might introduce a small margin of error due to variations in plant development stages.

Some of the automatically and randomly generated point data labelled as “other fields” and “others” may have fallen within areas that were planted with potatoes in reality but were not included in the field data. This may have resulted in points that were actually potatoes being incorrectly labelled in the training dataset used for the algorithm. Such mislabelling could introduce an error margin that might not be detected.

The results indicate that among the four classification scenarios (S2-RF, S2-GTB, L8-RF, L8-GTB), the S2-RF model consistently outperformed others, achieving the highest overall accuracy and kappa values (Fig. 6). Notably, although the S2-GTB configuration showed the lowest variance across folds, it also resulted in the lowest accuracy, suggesting a trade-off between model stability and predictive performance. In contrast, L8-based models showed moderate but consistent performance, with L8-RF slightly outperforming L8-GTB. These findings align with previous literature emphasizing the utility of Sentinel-2’s higher spatial and spectral resolution in agricultural applications (Immitzer et al. 2016; Pelletier et al. 2019).

The *F*-score values (Fig. 7) indicate that L8-RF is emerging as the best performing model in terms of precision and recall. Interestingly, the user’s and producer’s accuracy plots (Figs. 8 and 9) show that while the L8-GTB model achieved the highest producer’s accuracy with high standard deviation, L8-RF had the second highest median producer’s accuracy with more stable variation. Additionally, L8-RF outperformed other model-satellite options in terms of user’s accuracy. This suggests that the S2-RF model stood out for both identifying potato fields with low omission and commission error. From a general perspective, the RF model provided more accurate classification than GTB. In terms of satellite images, the differences in results can be attributed to temporal and spatial resolution disparities between the sensors, and potentially to differing spectral sensitivities that affect model generalization.

The number of trees was set to 100 when applying the random forest (RF) and gradient tree boost (GTB) algorithm, as detailed in the methodology section. Although the literature suggests that this parameter has minimal impact, it may be beneficial to perform a sensitivity analysis using different values. Normally, increasing the number of trees leads to better performance. Yet, the computational need would also increase while setting a higher number of trees. In general, 100 trees are becoming a standard for the ML classifications. A sensitivity analysis was performed to evaluate the impact of the number of trees on model performance. Results showed that RF models (L8-RF and S2-RF) stabilized after 100 trees, while GTB models (especially S2-GTB) continued to improve but with diminishing returns beyond 250–300 trees (Fig. 10). Considering computational efficiency and consistency with standard practice, all models were finalized using 100 trees.

Feature importance analysis (Fig. 11) provides further insight into model behaviour, revealing that bands related to red-edge and shortwave infrared (SWIR) wavelengths contributed more significantly to model predictions, particularly in Sentinel-2 imagery. This highlights the critical role of longer-wavelength bands in discriminating between crop types, as these bands are highly sensitive to vegetation structure, biomass, and water content—attributes that are particularly relevant for distinguishing crops such as potatoes. This is consistent with the physiological characteristics of vegetation and the known spectral separability of potato fields during peak phenological stages (Clevers and Kooistra 2012). This result is also consistent across both Landsat 8 and Sentinel-2 models, reinforcing the importance of NIR and SWIR spectral domains in agricultural land cover classification tasks. The outcome underlines the physiological relevance of these bands for vegetation mapping and affirms their utility in enhancing model accuracy. Additionally, with less variance between the feature importances, RF used the features more evenly than the GTB.

Confusion matrices and associated classification metrics for the final classification (Table 1 and Fig. 12) confirm the previously discussed trends. The RF models achieved the highest classification consistency, indicated by superior kappa values, while also maintaining high and stable *F*-scores and user's/producer's accuracies across classes.

The estimated surface areas for each land cover class have an average of 8.48 km^2 for potato fields (Table 2). While three models produced similar estimates, the S2-RF result deviated more significantly. Compared to the official 2020 TÜİK figure of 15 km^2, the estimates show a 41–48% difference. However, the reliability of official agricultural data in Türkiye is widely questioned, limiting its suitability for scientific validation.

Various data sources, including multispectral bands, vegetation indices, and temporal information, contribute to improved classification results (Saxena et al. 2021). For this study, pixel data associated with potatoes were utilized from 10 bands of Sentinel-2 and 6 bands of Landsat 8, with spatial resolutions of 20 m and 30 m, respectively. When considering the number of bands and resolution together with accuracy values, it can be concluded that the algorithm trained with higher-resolution data and more bands performs better.

Overall, the study demonstrates that Sentinel-2 imagery combined with the random forest algorithm yields robust classification results for potato fields in arid and semi-arid agricultural landscapes. Additionally, the fivefold cross-validation framework provided a reliable mechanism to evaluate model generalizability and avoid overfitting, an essential requirement for operational agricultural monitoring systems.

Deploying the satellite- and machine-learning–based model for potato field mapping presents three key challenges: limited data availability due to cloud cover, temporal gaps, and sparse ground truth; high computational demands for processing multispectral time-series imagery; and reduced generalizability across crops owing to spectral and phenological variability. Mitigation strategies include multi-sensor data fusion (to overcome cloud-induced gaps), cloud-based scalable GPU infrastructures for computation, and domain-adaptation techniques (e.g., multi-crop training and feature alignment) to enhance transferability. Despite the promising results, several factors, including variability in planting and harvesting times and potential mislabelling of non-potato data points, may introduce some margin of error.

The RF model, as implemented in this study, demonstrates strong computational scalability for applications over larger areas or for different crop types, primarily due to its execution within the Google Earth Engine (GEE) platform. The use of GEE's cloud-based, parallel processing architecture allows the model to efficiently handle massive datasets, making the extension of this analysis to a national scale computationally practical. The framework is also readily adaptable for classifying other crops, provided that new high-quality training data is supplied and the temporal compositing window is adjusted to match the unique phenological cycle of the target crop. The primary challenge in scaling for new crop types is therefore not computational, but rather the acquisition of sufficient and accurate ground-truth data for each new class.

ML ability to process multi-temporal satellite data enables timely crop mapping and early anomaly detection, supporting applications such as policy development, subsidy allocation, and supply chain optimization. The model can be integrated into Geographic Information Systems (GIS) and existing agricultural decision-support platforms to enhance operational use. Furthermore, its compatibility with IoT-based smart farming technologies such as soil sensors, weather stations, and UAVs offers potential for real-time crop health monitoring and adaptive management. Given its flexibility in handling various spectral and temporal inputs, the methodology is transferable to other major crops (e.g., wheat, maize, rice) and can be adapted to diverse agroecological regions. With appropriate calibration, the model presents a scalable solution for global agricultural monitoring across different climatic conditions.

While the application of machine learning and high-resolution satellite imagery offers significant benefits for agricultural monitoring and food security, it is also important to acknowledge the ethical considerations associated with this technology. The ability to monitor land use at a detailed scale raises potential concerns regarding data privacy, particularly for individual landowners. The methods developed in this study are intended for regional-scale agricultural planning and resource management by public agencies. However, it is crucial that the use and dissemination of such high-resolution classification maps are governed by clear ethical guidelines and data protection policies to prevent misuse and protect the privacy of individuals. Future applications should prioritize aggregated and anonymized data to ensure that the societal benefits of precision agriculture do not compromise fundamental privacy rights. Responsible implementation and transparent policies are key to harnessing the full potential of this technology for the public good.

Conclusions

This study presented a comprehensive evaluation of machine learning–based potato fields classification using satellite imagery in the Konya Plain. Using multispectral composites from Sentinel-2 and Landsat 8 during the peak growing season, two ensemble learning algorithms—random forest (RF) and gradient tree boosting (GTB)—were tested and compared.

The methodology involved data preprocessing, feature extraction, k-fold cross validation, sensitivity analysis for tree number, and performance validation through confusion matrices. The results revealed that spectral bands associated with vegetation structure, especially red-edge and SWIR bands, played a significant role in classification success. Sentinel-2 images generally outperformed Landsat 8 due to their finer spatial resolution and richer spectral content. Among the models, Sentinel-2 with random forest was found to be the best performing for potato field detection.

The key finding is that Sentinel-2 imagery coupled with random forest with limited field data provides a reliable and scalable solution for potato field classification.

The integration of the random forest (RF) algorithm with Sentinel-2 imagery provides a viable method for effective and accurate crop type mapping, with significant implications for agricultural management and planning in Türkiye and similar agro-ecological zones worldwide. This approach has the potential to be extended to other crops and regions, thereby supporting the broader application of remote sensing and machine learning in precision agriculture.

Future work may focus on investigating the performance of integrating temporal features, vegetation indices, SAR data, or deep learning architectures to enhance performance in heterogeneous agricultural environments and multi-year analyses.

Acknowledgements The authors would like to thank Konya Şeker Industry and Trade Inc. staff for their collaboration.

Author Contribution Conceptualization, B.K., Ç.S., and S.Y.Y.; methodology, E.T., B.K., and Ç.S.; software, E.T., B.K., and Ç.S.; validation, G.E.A., E.Y., E.A., and M.C.C.; formal analysis, E.T., G.E.A., E.Y., and E.A.; investigation, B.K., Ç.S., M.C.C., and S.S.Y.; resources, E.T., B.K., Ç.S., and M.C.C.; data curation, E.T.,G.E.A., E.Y., and E.A.; writing—original draft preparation, E.T., B.K., Ç.S., and S.S.Y.; writing—review and editing, G.E.A., E.Y., E.A., and M.C.C.; visualization, E.T., G.E.A., E.Y., and E.A.; supervision, B.K., Ç.S., and S.S.Y.; project administration, B.K., Ç.S., and S.S.Y. All authors have read and agreed to the published version of the manuscript.

Funding No funding was received for conducting this study.

Data Availability The raw data supporting the conclusion of this article will be made available by the authors on request.

Declarations

Conflicts of Interest The authors declare no competing interests.

References

Abatzoglou JT, Dobrowski SZ, Parks SA, Hegewisch KC (2018) Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. Sci Data 5:170191. https://doi.org/10.1038/sdata.2017.191

Abdelhamid AA, Alhussan AA, Qenawy AST, Osman AM, Elshewey AM, Eed M (2025) Potato harvesting prediction using an improved ResNet-59 model. Potato Res 68(2):1049–1068. https://doi.org/10.1007/s11540-024-09773-6

Akar Ö, Güngör O (2015) Integrating multiple texture methods and NDVI to the random forest classification algorithm to detect tea and hazelnut plantation areas in northeast Turkey. Int J Remote Sens 36(2):442–464. https://doi.org/10.1080/01431161.2014.995276

Ayyadevara VK (2018) Gradient boosting machine. In: Ayyadevara VK (ed) Pro machine learning algorithms. Apress, Berkeley, pp 117–134. https://doi.org/10.1007/978-1-4842-3564-5_6

Belgiu M, Drăguţ L (2016) Random forest in remote sensing: a review of applications and future directions. ISPRS J Photogramm Remote Sens 114:24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bertellini P, D'Addese G, Franchini G, Parisi S, Scribano C, Zanirato D, Bertogna M (2023) Binary classification of agricultural crops using Sentinel satellite data and machine learning techniques. FedCSIS 2023 18th Conference on Computer Science and Intelligence Systems (IEEE), 17–20 September 2023; Warsaw University of Technology. Warsaw, Poland, pp 859–864

Bourriz M, Hajj H, Laamrani A, Elbouanani N, Abdelali HA, Bourzeix F, El-Battay A, Amazirh A, Chenbouni A (2025) Integration of hyperspectral imaging and AI techniques for crop type mapping: present status, trends and challenges. Remote Sens 17(9):1574. https://doi.org/10.3390/rs17091574

Breiman L (2001) Random forests. Mach Learn 45:5–32

Chen T, Guestrin C (2016) XGboost: a scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. August 2016, pp 785–794

Clevers JG, Kooistra L (2012) Using hyperspectral remote sensing data for retrieving canopy chlorophyll and nitrogen content. IEEE J Sel Top Appl Earth Observ Remote Sens 5(2):574–583. https://doi.org/10.1109/JSTARS.2011.2176468

Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Meas 20(1):37–46. https://doi.org/10.1177/001316446002000104

Congalton RG (1991) A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens Environ 37(1):35–46. https://doi.org/10.1016/0034-4257(91)90048-B

Das A, Kumar M, Maity S, Bhattacharya BK (2021) Impact of training data quality on machine learning based crop classification using time series C-band SAR data. In: International India Geoscience and Remote Sensing Symposium (InGARSS), 6-10 December 2020; Virtual Symposium, pp. 352-355. https://doi.org/10.1109/InGARSS51564.2021.9791961

de Crisóstomo Castro Filho H, de Abílio Carvalho Júnior O, de Ferreira Carvalho OL, de Pozzobon Bem P, dos Santos de Moura R, de Olino Albuquerque A, Rosa Silva C, Guimaraes Ferreira PH, Guimaraes RF, Trancoso Gomes RA (2020) Rice crop detection using LSTM, Bi-LSTM, and machine learning models from Sentinel-1 time series. Remote Sens 12(16):2655. https://doi.org/10.3390/rs12162655

Dimov D, Löw F, Ibrakhimov M, Stulina G, Conrad C (2017) SAR and optical time series for crop classification. In: International Geoscience and Remote Sensing Symposium (IGARSS), Forth Worth, Texas, USA, 23-28 July 2017, pp. 811-814. https://doi.org/10.1109/IGARSS.2017.8127076

Dumont M, Marée R, Wehenkel L, Geurts P (2009) Fast multi-class image annotation with random subwindows and multiple output randomized trees. In: International Conference on Computer Vision Theory and Applications (VISAPP), Lisboa, Portugal, 5-8 February 2009, pp. 196-203

Eed M, Alhussan AA, Qenawy AT, Osman AM, Elshewey AM, Arnous R (2025) Potato consumption forecasting based on a hybrid stacked deep learning model. Potato Res 68:809–833. https://doi.org/10.1007/s11540-024-09764-7

Elghazel H, Aussem A, Perraud F (2011) Trading-off diversity and accuracy for optimal ensemble tree selection in random forests. In: Okun O, Valentini G, Re M (eds) Ensembles in machine learning applications, studies in computational intelligence vol 373. Springer, Berlin, Heidelberg. pp 169–179. https://doi.org/10.1007/978-3-642-22910-7_10.

El-Kenawy ESM, Alhussan AA, Khodadadi N, Mirjalili S, Eid MM (2025) Predicting potato crop yield with machine learning and deep learning for sustainable agriculture. Potato Res 68:759–792. https://doi.org/10.1007/s11540-024-09753-w
El-Kenawy ESM, Eid MM, Abualigah L (2024a) Machine learning in public health forecasting and monitoring the zika virus. Metaheuristic Optim Rev 1(2):1–11. https://doi.org/10.54216/MOR.010201
El-Kenawy ESM, Khodadadi N, Mirjalili S, Abdelhamid A, Eid MM, Ibrahim A (2024b) Greylag goose optimization: nature-inspired optimization algorithm. Expert Syst Appl 238:122147. https://doi.org/10.1016/j.eswa.2023.122147
Evrendilek F, Gulbeyaz O (2011) Boosted decision tree classifications of land cover over Turkey integrating MODIS, climate and topographic data. Int J Remote Sens 32(12):3461–3483. https://doi.org/10.1080/01431161003749469
FAOSTAT (2024) Crops and livestock products. https://www.fao.org/faostat/en/#data/QCL. Accessed 25 July 2024
Foody GM (2002) Status of land cover classification accuracy assessment. Remote Sens Environ 80(1):185–201. https://doi.org/10.1016/S0034-4257(01)00295-4
Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 29(5):1189–1232. https://doi.org/10.1214/aos/1013203451
Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, Husak G, Rowland J, Harrison L, Hoell A, Michaelsen J (2015) The climate hazards infrared precipitation with stations-a new environmental record for monitoring extremes. Sci Data 2:150066. https://doi.org/10.1038/sdata.2015.66
Garg U, Kukreti V, Pundir RS, Manchanda M, Gupta N (2023) Prediction of turkey forest fire using random forest regressor. 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), 14–16 March 2023; Dehradun. Uttarakhand, India, pp 1–7
Gilbertson JK, Kemp J, Van Niekerk A (2017) Effect of pan-sharpening multi-temporal Landsat 8 imagery for crop type differentiation using different classification techniques. Comput Electron Agric 134:151–159. https://doi.org/10.1016/j.compag.2016.12.006
Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning. Springer, New York. https://doi.org/10.1007/978-0-387-84858-7
Husain SO, Meghana A, Soujanya R, Sunil G, Muvva VBR (2024) A Multi-Module Spatial Semantic Network with Gradient Boosting for Efficient Crop Classification. In: 2024 International Conference on Distributed Systems, Computer Networks and Cybersecurity (ICDSCNC), Bengaluru, India, 20-21 September 2024, pp. 1-5
Hyontai SUG (2018) Performance of machine learning algorithms and diversity in data. In: 22nd International Conference on Circuits, Systems, Communications and Computers, 14–17 July, Mallorca, Spain, p 04019
Immitzer M, Vuolo F, Atzberger C (2016) First experience with Sentinel-2 data for crop and tree species classifications in central Europe. Remote Sens 8(3):166. https://doi.org/10.3390/rs8030166
Iqbal MA, Talukder KH (2020) Detection of potato disease using image segmentation and machine learning. In: 2020 international conference on wireless communications signal processing and networking (WiSPNET), 4–6 August 2020; Chennai, India, pp 43–47
Iwańska M, Oleksy A, Dacko M, Skowera B, Oleksiak T, Wójcik-Gront E (2018) Use of classification and regression trees (CART) for analyzing determinants of winter wheat yield variation among fields in Poland. Biomed Lett 55(2):197–214. https://doi.org/10.2478/bile-2018-0013
Kariluoto A, Kultanen J, Soininen J, Pärnänen A, Abrahamsson P (2021) Quality of data in machine learning. In: 21st International Conference on Software Quality, Reliability and Security Companion (QRS-C), 6–10 December 2021; Hainan Island, China, pp 216–221
Khan AH, Zafar Z, Shahzad M, Berns K, Fraz MM (2023) Crop type classification using multi-temporal Sentinel-2 satellite imagery: a deep semantic segmentation approach. In: ICRAI 2023 5th International Conference on Robotics and Automation in Industry (IEEE), 3–5 March 2023; Peshawar, Pakistan, pp 1–6
Liaw A, Wiener M (2002) Classification and regression by random forest. R News 2(3):18–22
Lizarazo I, Rodriguez JL, Cristancho O, Olaya F, Duarte M, Prieto F (2023) Identification of symptoms related to potato Verticillium wilt from UAV-based multispectral imagery using an ensemble of gradient boosting machines. Smart Agric Technol 3:100138. https://doi.org/10.1016/j.atech.2022.100138
Millard K, Richardson M (2015) On the importance of training data sample selection in random forest image classification: a case study in peatland ecosystem mapping. Remote Sens 7(7):8489–8515. https://doi.org/10.3390/rs70708489

Monsalve-Tellez JM, Torres-León JL, Garcés-Gómez YA (2022) Evaluation of SAR and optical image fusion methods in oil palm crop cover classification using the random forest algorithm. Agriculture 12(7):955. https://doi.org/10.3390/agriculture12070955

Neetu RSS (2019) Exploring machine learning classification algorithms for crop classification using Sentinel 2 data. Int Arch of the Photogramm Remote Sens Spat Inf Sci 42:573–578. https://doi.org/10.5194/isprs-archives-XLII-3-W6-573-2019

Ok AO, Akar O, Gungor O (2012) Evaluation of random forest method for agricultural crop classification. Eur J Remote Sens 45(1):421–432. https://doi.org/10.5721/EuJRS20124535

Oshiro TM, Perez PS, Baranauskas JA (2012) How many trees in a random forest? In: Machine learning and data mining in pattern recognition: 8th International Conference, 13–20 July 2012; Berlin, Germany, pp 154–168

Pelletier C, Webb GI, Petitjean F (2019) Temporal convolutional neural network for the classification of satellite image time series. Remote Sens 11(5):523. https://doi.org/10.3390/rs11050523

Probst P, Boulesteix AL (2018) To tune or not to tune the number of trees in random forest. J Mach Learn Res 18(181):1–18

Radwan M, Alhussan AA, Ibrahim A, Tawfeek SM (2024) Potato leaf disease classification using optimized machine learning models and feature selection techniques. Potato Res. https://doi.org/10.1007/s11540-024-09763-8

Ramadhan MM, Sitanggang IS, Nasution FR, Ghifari A (2017) Parameter tuning in random forest based on grid search method for gender classification based on voice frequency. DEStech Trans Comput Sci Eng 10. https://doi.org/10.12783/dtcse/cece2017/14611

Rodriguez J, Lizarazo I, Prieto F, Angulo-Morales V (2021) Assessment of potato late blight from UAV-based multispectral imagery. Comput Electron Agric 184:106061. https://doi.org/10.1016/j.compag.2021.106061

Savitha C, Talari R (2025) Evaluating the performance of random forest, support vector machine, gradient tree boost, and CART for improved crop-type monitoring using greenest pixel composite in Google Earth Engine. Environ Monit Assess 197:437. https://doi.org/10.1007/s10661-025-13880-3

Saxena V, Dwivedi RK, Kumar A (2021) Analysis of machine learning algorithms for crop mapping on satellite image data. In: 10th International Conference on System Modeling & Advancement in Research Trends (SMART), 10–11 December 2021; Moradabad, India, pp 23–28

Senagi K, Jouandeau N (2018) A non-deterministic strategy for searching optimal number of trees hyperparameter in random forest. In: Federated Conference on Computer Science and Information Systems (FedCSIS), 9–12 September 2018; Poznań, Poland, pp 73–80

Sharma V, Ghosh SK (2023) Evaluating the potential of 8 band Planetscope dataset for crop classification using random forest and gradient tree boosting by Google Earth Engine. Int Arch Photogramm Remote Sens Spatial Inf Sci 48:325–330. https://doi.org/10.5194/isprs-archives-XLVIII-M-1-2023-325-2023

Shukla G, Garg RD, Srivastava HS, Garg PK (2018) Performance analysis of different predictive models for crop classification across an aridic to ustic area of Indian states. Geocarto Int 33(3):240–259. https://doi.org/10.1080/10106049.2016.1240721

Şimşek FF (2025) Comparison of agricultural crop type classifications with different machine learning algorithms (RF-SVM-ANN-XGBoost) by generating ground truth data from farmer declaration parcels. Int J Eng Geosci 10(2):207–220. https://doi.org/10.26833/ijeg.1552141

Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. Inf Process Manag 45(4):427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Soper DS (2021) Greed is good: rapid hyperparameter optimization and model selection using greedy k-fold cross validation. Electronics 10(16):1973. https://doi.org/10.3390/electronics10161973

Tatsumi K, Yamashiki Y, Torres MAC, Taipe CLR (2015) Crop classification of upland fields using random forest of time-series Landsat 7 ETM+ data. Comput Electron Agr 115:171–179. https://doi.org/10.1016/j.compag.2015.05.001

Thapa S, Thapa S (2019) Scope of value-addition in potato. Int J Hortic Agric Food Sci 3:132–465. https://doi.org/10.22161/ijhaf.3.3.4

Towfek SK, Khodadadi N (2023) Deep convolutional neural network and metaheuristic optimization for disease detection in plant leaves. J Intell Syst Internet Things 10(1):66–75. https://doi.org/10.54216/JISIoT.100105

TÜİK (2024) Agricultural statistics summary. https://www.tuik.gov.tr/. Accessed 25 July 2024

Uribeetxebarria A, Carrasco A, Aranguren M, Legorburu J, Castellón A, Aizpurua A (2023) Classifying potato yield in North Spain from Sentinel-2 data based on random forest and multitemporal

auxiliary data. In: Stafford JV (Ed) Precision agriculture ’23. Wageningen Academic Publishers, Leiden, pp 329–336. https://doi.org/10.3920/978-90-8686-947-3_40
Venkatanaresh M, Kullayamma I (2022) A new approach for crop type mapping in satellite images using hybrid deep capsule auto encoder. Knowl-Based Syst 256:109881. https://doi.org/10.1016/j.knosys.2022.109881
Wald R, Khoshgoftaar T, Dittman DJ, Napolitano A (2013) Random forest with 200 selected features: an optimal model for bioinformatics research. 12th International Conference on Machine Learning and Applications, 4–7 December 2013; Miami. Florida, USA, pp 154–160
Wang X, Fang S, Yang Y, Du J, Wu H (2023) A new method for crop type mapping at the regional scale using multi-source and multi-temporal Sentinel imagery. Remote Sens 15(9):2466. https://doi.org/10.3390/rs15092466
Watts JD, Lawrence RL (2008) Merging random forest classification with an object-oriented approach for analysis of agricultural lands. Int Arch Photogramm Remote Sens Spat Inf Sci 37(B7)
Yamaç SS (2018) Estimation of long-term reference evapotranspiration using limited weather data in sugar beet plantation area from Middle Anatolian, Turkey. 1. International agricultural structure and irrigation congress special issue, Süleyman Demirel Uni J of Agric Fac, 160–165
Yang J, Xu J, Zhai DL (2021) Integrating phenological and geographical information with artificial intelligence algorithm to map rubber plantations in Xishuangbanna. Remote Sens 13(14):2793. https://doi.org/10.3390/rs13142793
Yassen MA, Abdel-Fattah MG, Ismail I, El Kenawy ESM, Moustafa HE (2024) An AI-based system for predicting renewable energy power output using advanced optimization algorithms. J Artif Intell Metaheuristics 8(1):1–8. https://doi.org/10.54216/JAIM.080101
Yavuz D, Kara M, Suheri S (2012) Comparison of different irrigation methods in terms of water use and yield in potato farming. J Selcuk Univ Nat Appl Sci 2:1–12
Zhang Q, Wang G, Wang G, Song W, Wei X, Hu Y (2023) Identifying winter wheat using Landsat data based on deep learning algorithms in the North China Plain. Remote Sens 15(21):5121. https://doi.org/10.3390/rs15215121
Zhao J, Tian G, Qiu C, Gu B, Zheng K, Liu Q (2022) Weed detection in potato fields based on improved YOLOv4: optimal speed and accuracy of weed detection in potato fields. Electronics 11(22):3709. https://doi.org/10.3390/electronics11223709

Authors and Affiliations

Emre Tokluoğlu[1] · Bedri Kurtuluş[1,3] · Çağdaş Sağır[2,3,9] · Günseli Erdem Altın[4] · Elifnur Yurdakul[5] · Ersin Ateş[6] · Mustafa Can Canoğlu[7] · Sevim Seda Yamaç[8]

✉ Sevim Seda Yamaç
sevimseda.yamac@erbakan.edu.tr; sevimseda10@gmail.com

1 Department of Geological Engineering, Muğla Sıtkı Koçman University, Muğla 48000, Türkiye

2 Department of Geological Engineering, Faculty of Engineering, Middle East Technical University, Ankara 06800, Türkiye

3 Ekofly Engineering Consultancy R&D Ltd., Middle East Technical University Technopolis ETKİM Block, Ankara 06800, Türkiye

4 Department of Civil Engineering, Faculty of Engineering, Yeditepe University, Istanbul 34755,

Türkiye

[5] Department of Earth and Geo-Environmental Sciences, University of Bari Aldo Moro, Bari 70125, Italy

[6] Geographic Information Systems Program, Vocational School of Ayaş, Ankara University, Ankara 06710, Türkiye

[7] Department of Environmental Engineering, Faculty of Engineering and Architecture, Sinop University, Sinop 57000, Türkiye

[8] Department of Biosystems Engineering, Eregli Faculty of Agriculture, Necmettin Erbakan University, Konya 42310, Türkiye

[9] Hydrogeology Division, Fugro Sial Geosciences Consultancy and Engineering Ltd. Co., Ankara 06510, Türkiye