Gazi Üniversitesi
Fen Bilimleri Dergisi
PART C: TASARIM VE TEKNOLOJİ

Gazi University
Journal of Science
PART C: DESIGN AND TECHNOLOGY

GU J Sci, Part C, 14(X): XX-XX (2026)

A Hybrid Framework Based on Transfer Learning for Seasonal Classification of Satellite Images

Eyyüp YILDIZ[1*], Özge ASLAN YILDIZ[1]

[1] *Erzincan Binali Yıldırım University, Faculty of Engineering-Architecture, Department of Computer Engineering, Erzincan, Turkey*

Article Info

Research article
Received: 25/05/2026
Revision: 07/07/2026
Accepted: 21/07/2026

Keywords

Transfer learning, pretrained models, machine learning classification, satellite image classification.

Makale Bilgisi

Araştırma makalesi
Başvuru: 25/05/2026
Düzeltme: 07/07/2026
Kabul: 21/07/2026

Anahtar Kelimeler

Transfer öğrenme, Ön eğitilmiş modeller, Makine öğrenmesi sınıflandırması, Uydu görüntüsü sınıflandırması.

Graphical/Tabular Abstract (Grafik Özet)

This study proposes a hybrid transfer learning framework for seasonal classification of Sentinel-2 satellite images from 81 provinces of Türkiye. Eight pretrained deep models and seven classifiers were compared. The fine-tuned hybrid model combining ConvNeXt, Vision Transformer, and Swin Transformer achieved the best performance. / Bu çalışma, Türkiye'nin 81 ilinden elde edilen Sentinel-2 uydu görüntülerinin mevsimsel sınıflandırılması için transfer öğrenmeye dayalı hibrit bir çerçeve önermektedir. Sekiz ön eğitilmiş model ve yedi sınıflandırıcı karşılaştırılmıştır. ConvNeXt, Vision Transformer ve Swin Transformer'ı birleştiren ince ayar yapılmış hibrit model, en iyi performansı elde etmiştir.

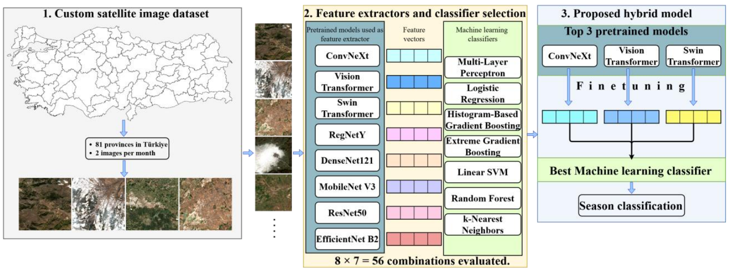


***Figure A:** Overview of the proposed hybrid transfer learning framework / **Şekil A:** Önerilen hibrit transfer öğrenme çerçevesinin genel görünümü.*

Highlights (Önemli noktalar)

- *Eight ImageNet-pretrained deep learning models and seven machine learning classifiers were compared. / Sekiz ImageNet ön eğitilmiş derin öğrenme modeli ve yedi makine öğrenmesi sınıflandırıcısı karşılaştırılmıştır.*
- *For A hybrid transfer learning-based classification framework, ConvNeXt, Vision Transformer, and Swin Transformer were selected as feature extractors and Multi-Layer Perceptron has become the classification component. / Hibrit bir transfer öğrenimi tabanlı sınıflandırma çerçevesi için, ConvNeXt, Vision Transformer ve Swin Transformer özellik çıkarıcı olarak seçilmiştir ve Çok Katmanlı Algılayıcı, sınıflandırma bileşenini oluşturmuştur.*

***Aim (Amaç):** The aim of this study is to develop an effective hybrid transfer learning framework for seasonal classification of satellite images by combining pretrained deep feature extractors with machine learning classifiers. / Bu çalışmanın amacı, ön eğitilmiş derin özellik çıkarıcıları makine öğrenmesi sınıflandırıcılarıyla birleştirerek uydu görüntülerinin mevsimsel sınıflandırılması için etkili bir hibrit transfer öğrenme çerçevesi geliştirmektir.*

***Originality (Özgünlük):** a Türkiye-wide multi-scale satellite image dataset was created, and eight pre-trained deep learning models were evaluated using seven machine learning classifiers. / Türkiye genelinde çok ölçekli bir uydu görüntüsü veri seti oluşturulmuştur ve önceden eğitilmiş sekiz derin öğrenme modelini yedi makine öğrenmesi sınıflandırıcısı ile değerlendirilmiştir.*

***Results (Bulgular):** The quantitative results showed that the ConvNeXt and Multi-Layer Perceptron combination achieved the best individual performance. The hybrid framework improved the performance to 0.725 accuracy and 0.724 F1-score. After fine-tuning, the proposed hybrid model achieved the highest performance, with 0.740 accuracy and 0.742 F1-score. / Nicel sonuçlar, ConvNeXt ve Çok Katmanlı Algılayıcı kombinasyonunun en iyi bireysel performansı elde ettiğini göstermiştir. Hibrit model başarıyı 0,725 doğruluk ve 0,724 F1-skoruna yükseltmiştir. İnce ayar sonrasında önerilen hibrit model 0,740 doğruluk ve 0,742 F1-skoru ile en yüksek performansı elde etmiştir.*

***Conclusion (Sonuç):** The proposed hybrid framework improves classification performance by combining complementary features from convolutional and transformer-based architectures. / Önerilen hibrit çerçeve, konvolüsyonel ve transformer tabanlı mimarilerden elde edilen tamamlayıcı özellikleri birleştirerek sınıflandırma performansını artırmaktadır.*

*Corresponding author, e-mail: eyyup.yildiz@erzincan.edu.tr
DOI: 10.29109/gujsc.1958797

GU J Sci, Part C, 14(X): XX-XX (2026)

Gazi Üniversitesi
Fen Bilimleri Dergisi
PART C: TASARIM VE TEKNOLOJİ

Gazi University
Journal of Science
PART C: DESIGN AND TECHNOLOGY

http://dergipark.gov.tr/gujsc

A Hybrid Framework Based on Transfer Learning for Seasonal Classification of Satellite Images

Eyyüp YILDIZ[1*], Özge ASLAN YILDIZ[1]

[1]*Erzincan Binali Yıldırım University, Faculty of Engineering-Architecture, Department of Computer Engineering, Erzincan, Turkey*

Article Info

Research article
Received: 25/05/2026
Revision: 07/07/2026
Accepted: 21/07/2026

Keywords

Transfer learning, pretrained models, machine learning classification, satellite image classification.

Abstract

Seasonal classification from satellite imagery is a crucial remote sensing task for monitoring vegetation dynamics, agricultural processes, environmental change, and climate-related spatial patterns. However, limited labeled data, regional variability, and the high computational cost of training large-scale networks from scratch make developing generalizable models for this task challenging. This study proposes a hybrid framework based on transfer learning for seasonal classification using monthly satellite imagery from all 81 provinces of Türkiye. Eight ImageNet pretrained deep learning models were evaluated as feature extractors, and the extracted representations were evaluated using seven machine learning classifiers. The selection of feature extractors and classifiers was performed using 10-fold cross-validation repeated five times. Among the feature extractor models, the combination of ConvNeXt and Multi-Layer Perceptron produced the best result with a mean accuracy of 0.698 and a mean F1-score of 0.696. As a result of the comparative analysis, ConvNeXt, Vision Transformer, and Swin Transformer were selected as the three most successful feature extractors, and Multi-Layer Perceptron was selected as the final classifier. The hybrid model, constructed by concatenating the features of these three selected feature extractors, achieved a mean accuracy of 0.725 and a mean F1-score of 0.724. After fine-tuning, the proposed hybrid model achieved the highest performance with a mean accuracy of 0.740 and a mean F1-score of 0.742. The results demonstrate that combining pretrained representations and fine-tuning improved seasonal classification in satellite imagery. The proposed framework offers an effective and computationally practical approach for seasonal classification, providing a promising foundation for future environmental monitoring and agricultural remote sensing applications.

Uydu Görüntülerinin Mevsimsel Sınıflandırılması için Transfer Öğrenmeye Dayalı Hibrit Çerçeve

Makale Bilgisi

Araştırma makalesi
Başvuru: 25/05/2026
Düzeltme: 07/07/2026
Kabul: 21/07/2026

Anahtar Kelimeler

Transfer öğrenme, Ön eğitilmiş modeller, Makine öğrenmesi sınıflandırması, Uydu görüntüsü sınıflandırması.

Öz

Uydu görüntülerinden mevsimsel sınıflandırma, bitki örtüsü dinamiklerini, tarımsal süreçleri, çevresel değişimi ve iklimle ilgili mekânsal desenleri izlemek için önemli bir uzaktan algılama görevidir. Bununla birlikte, sınırlı sayıda etiketli veri, bölgesel değişkenlik ve büyük ölçekli ağların sıfırdan eğitilmesinin yüksek hesaplama maliyeti, bu görev için genelleştirilebilir modeller oluşturmayı zorlaştırmaktadır. Bu çalışma, Türkiye'nin 81 ilinin tamamından elde edilen aylık uydu görüntülerini kullanarak mevsimsel sınıflandırma için transfer öğrenmeye dayalı hibrit bir çerçeve önermektedir. Sekiz adet ImageNet ön eğitilmiş derin öğrenme modeli özellik çıkarıcı olarak değerlendirilmiş ve çıkarılan temsiller yedi makine öğrenmesi sınıflandırıcısı kullanılarak sınıflandırılmıştır. Özellik çıkarıcıların ve sınıflandırıcıların seçimi, beş kez tekrarlanan 10 katlı çapraz doğrulama kullanılarak gerçekleştirilmiştir. Özellik çıkarıcı modeller arasında, ConvNeXt ve çok katmanlı algılayıcı kombinasyonu, ortalama 0,698 doğruluk ve ortalama 0,696 F1 puanı ile en iyi sonucu vermiştir. Karşılaştırmalı analiz sonucunda, ConvNeXt, Vision Transformer ve Swin Transformer en başarılı üç özellik çıkarıcı olarak seçilmiş ve çok katmanlı algılayıcı nihai sınıflandırıcı olarak belirlenmiştir. Seçilen bu üç özellik çıkarıcının özelliklerinin birleştirilmesiyle oluşturulan hibrit model, ortalama 0,725 doğruluk ve ortalama 0,724 F1 puanı elde etmiştir. İnce ayardan sonra, önerilen hibrit model, ortalama 0,740 doğruluk ve ortalama 0,742 F1 puanı ile en yüksek performansı göstermiştir. Sonuçlar, önceden eğitilmiş temsillerin birleştirilmesi ve ince ayarın, uydu görüntülerinde mevsimsel sınıflandırmayı iyileştirdiğini göstermektedir. Önerilen çerçeve, mevsimsel sınıflandırma için etkili ve hesaplama açısından pratik bir yaklaşım sunarak, gelecekteki çevresel izleme ve tarımsal uzaktan algılama uygulamaları için umut vadeden bir temel sağlamaktadır.

*Corresponding author, e-mail: eyyup.yildiz@erzincan.edu.tr
DOI: 10.29109/gujsc.1958797

1. INTRODUCTION (GİRİŞ)

Advances in sensor technology have increased the variety and resolution of satellite imaging processes, enabling the generation of large amounts of data. This has allowed for the investigation of terrestrial, atmospheric, and marine environments with high accuracy and a broad perspective. Increased access to high-resolution satellite imagery has also increased the applicability of artificial intelligence methods to satellite imagery [1]. Deep learning models have demonstrated outstanding performance in computer vision in recent years and have become a standard approach in hierarchical feature extraction [2]. Deep learning models' strong representational power has made it possible for earth observation systems to integrate scene classification from remote sensing data [3]. Consequently, it is crucial to employ deep learning architectures in reliable and scalable information extraction from satellite data [4], [5]. Today, deep learning-based remote sensing models demonstrate superior performance in problems such as land cover classification [6], [7] and object detection [8], [9]. Specifically, the classification of seasonal transition phases from satellite images stands out as a important research area for agricultural product estimation, water resource management and monitoring of global climate change.

Training deep learning models from scratch is a challenging process that requires high computational costs and massive amounts of labeled data [10]. Convolutional Neural Networks (CNNs) encounter optimization issues including overfitting and poor generalization when trained on small datasets, despite their powerful feature extraction capabilities. However, it has been shown that CNN layers trained on image data exhibit hierarchical learning behavior. Accordingly, the initial layers in CNNs learn low-level general features similar to edge detection, texture, or color filters, while the layers closer to the output layer learn more complex and high-level features specific to the target dataset being trained. Thus, transfer learning techniques have emerged as a result of the general-to-specific learning strategy across layers of CNNs, with the goal of reducing computational costs and overcoming limited data [11].

In remote sensing, acquiring large, high-quality labeled datasets for specific geographic regions or narrow time windows presents a significant constraint. The seasonal classification from satellite imagery relies on the temporal tracking of phases representing the life cycle of plants. In particular, the dynamics of winter and snow cover in the Eastern Anatolia Region span more than six months, while early spring phenology is concurrently observed along the southern and western coasts. This results in considerable spatial and temporal heterogeneity when analyzing Türkiye's complex topographic and climatic structure. Therefore, it is challenging to build a classification model that can be adapted and generalized across the country due to geographical and climatic variability.

This study proposes a hybrid framework for seasonal classification across the provinces of Türkiye based on satellite imagery. The proposed hybrid model utilizes pretrained CNNs to extract features from satellite images, followed by machine learning methods for classification. In this context, ResNet50, DenseNet121, ConvNeXt, MobileNet V3, EfficientNet B2, RegNetY, Vision Transformer, and Swin Transformer architectures, trained on the ImageNet dataset, are used as feature extractors. The obtained multidimensional features are classified using linear support vector machine, logistic regression, random forest, multi-layer perceptron, histogram-based gradient boosting, k-nearest neighbors, and extreme gradient boosting classifiers. A total of 56 distinct hybrid models were used by evaluating all combinations of eight pretrained deep learning architectures and seven machine learning classifiers. Next, the three best-performing pretrained models were fine-tuned. This allowed the general features learned by the pretrained CNNs to be quickly adapted to our dataset. Accordingly, the individual performances of the pretrained models and their fine-tuned versions were compared to quantitatively to investigate the effect of domain-specific adaptation in transfer learning. Finally, the fine-tuned feature representations were concatenated and classified using the selected machine learning model. The main contributions of this study can be summarized as follows:

- A new hybrid transfer learning framework was developed to combine the most effective representations of different pretrained architectures.
- Seven machine learning algorithms and eight pretrained deep learning architectures were systematically compared for seasonal classification.
- By evaluating pretrained and fine-tuned models, the quantitative impact of domain-specific adaptation to performance was examined.

The paper is organized as follows: Section 2 reviews related work on transfer learning and hybrid models on satellite images. Section 3 introduces the proposed hybrid model. Section 4 presents the experimental results in detail. Finally, Section 5 concludes the paper.

2. RELATED WORKS (İLGİLİ ÇALIŞMALAR)

We divided the related studies into three categories: transfer learning-based studies, hybrid studies, and season classification studies.

Transfer learning-based studies. Thirumaladevi et al. (2023) proposed a transfer learning-based model for remote sensing image scene classification. They tested two different strategies using pretrained AlexNet, VGG16, and VGG19 models. In the first strategy, features obtained from the fully connected layers of the CNNs were given as input to the support vector machine classifier. In the second strategy, AlexNet, VGG16, and VGG19 were fine-tuned by modifying the last layers of the networks to model the dataset [12]. Kunwar and Ferdush (2024) utilized transfer learning for land use and land cover classification on the EuroSAT dataset. They compared the performance of pretrained Vision Transformer, ResNet50, and VGG16 models with and without data augmentation. Quantitative results showed that the Vision Transformer with data augmentation achieved the highest results. Additionally, it was highlighted that ResNet 50 achieved higher accuracy than VGG16 over the same training iterations [13]. Roy et al. (2025) proposed a model that integrates transfer learning with an encryption approach for real-time and secure classification of satellite images at the edge. They compared the performance of five pretrained models on the EuroSAT dataset. In addition, they utilized The Brakerski–Fan–Vercauteren encryption method to protect the confidentiality of the pretrained models during cloud-based storage and real-time processing. The results showed that transfer learning-based models demonstrated high performance and were memory-efficient [14]. Farag et al. (2026) proposed a novel transfer learning-based model for classifying very similar settlements and barren lands from Landsat-8 satellite images. This study compares the performance of four different segmentation models based on the output feature maps of ResNet 50 [15].

Hybrid studies. Sudiana et al. (2023) proposed a hybrid model for detecting burned areas on Sentinel-1 and Sentinel-2 datasets. The hybrid model is constructed by combining a CNN as a feature extractor and a Random Forest method as a classifier. They evaluated the model's performance on five different test scenarios consisting of different combinations of Sentinel-1 and Sentinel-2 datasets [16]. Kwenda et al. (2024) proposed a novel hybrid model combining CNNs with machine learning algorithms for segmenting forested areas from satellite images. The study used aerial and satellite images from the DeepGlobe dataset. Pretrained VGG16 and ResNet50 models were chosen as feature extractors. In the study, the features extracted from VGG16 and ResNet50 were concatenated and fed into five different machine learning classifiers [17]. Mashraqi et al. (2025) proposed a hybrid deep model for classifying semi-arid lands. In the study, 10 different pretrained CNNs were used for feature extraction. In addition, the ant colony algorithm was integrated into the model to reduce the dimensionality of the features obtained from pretrained models and is used to select the most distinctive features. Random forest algorithm was utilized for classification [18]. Sudiana et al. (2025) developed a hybrid CNN-random forest model for mapping rice fields using satellite imagery. In the study, CNN was used for feature extraction, and the random forest was used as a classifier. The results showed that the proposed hybrid model achieved higher overall accuracy than the CNN-1D and traditional random forest models. The study emphasized that using deep networks as feature extractors and integrating the random forest algorithm instead of fully connected layers both reduced the computational load and increased the generalization capacity of the model, even with limited data samples [19].

Season classification studies. Seasonal classification task from satellite imagery has not yet been extensively studied in the literature. Therefore, there are limited number of similar studies. Cheng and Zhou (2011) proposed an automatic season classification method for outdoor images by combining normalized HSV color histogram features with skin exposure information extracted from images containing people. The method uses face detection and adaptive skin detection to identify relevant skin regions and then integrates color and skin cues through a Bayesian model to classify images into spring, summer, autumn, and winter categories. Experiments conducted on 1,000 outdoor photos showed an average accuracy of 77.5%, indicating that visual appearance cues can be effective for season recognition [20]. Volokitin et al. (2016) investigated the effectiveness of features extracted from CNNs for predicting the time of the year in an outdoor scene. The VGG16 model pretrained on ImageNet reported approximately 68.6% accuracy at the seasonal level and approximately 50.2% accuracy at the monthly level after short fine-tuning [21].

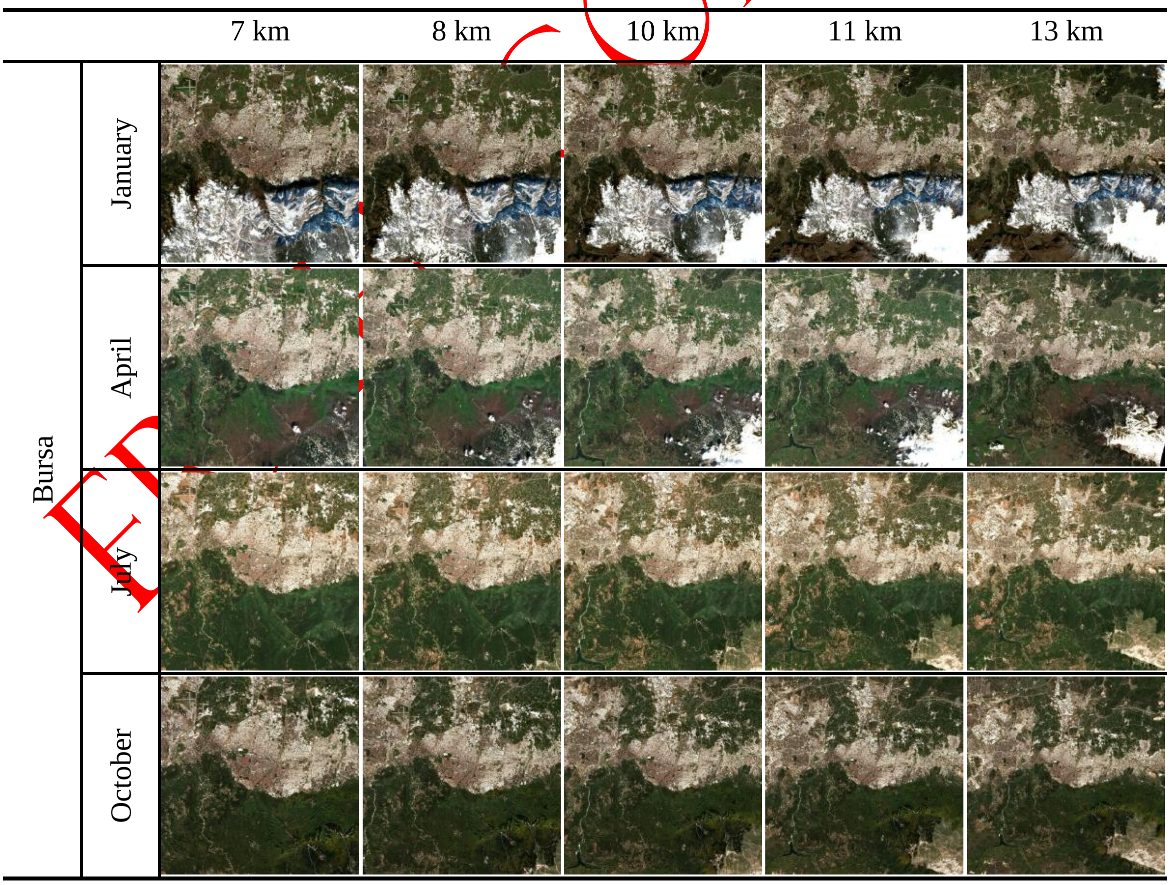
3.MATERIALS AND METHODS (MATERYAL VE METOD)

3.1. Material (Materyal)

In this study, a custom dataset for season classification was created from RGB images obtained from all 81 provinces of Türkiye using Sentinel-2 satellite imagery. Images were obtained via the Google Earth Engine API covering the period from January 2023 to December 2024. Image capturing for each province was performed monthly, considering the province's central coordinates. To capture multi-scale spatial contexts, each province was represented across five distinct spatial scales corresponding to 7, 8, 10, 11, and 13 km. Two images were acquired for each month, resulting in a total of 9720 images across 12 months and five spatial scales. To improve data quality, images with cloud cover below 20% were included in the dataset. After this filtering step, 8868 images remained for the experiments. Details of the distribution of images according to season and scale are given in Table 1.

Table 1. Distribution of images in the dataset according to season and spatial scale (Veri kümesindeki görüntülerin mevsime ve mekansal ölçeğe göre dağılımı)

Season	7 km	8 km	10 km	11 km	13 km	Total
Winter	423	423	424	424	424	2118
Spring	397	398	397	398	398	1988
Summer	483	482	483	482	483	2413
Autumn	470	470	470	470	469	2349
Total	1773	1773	1774	1774	1774	8868



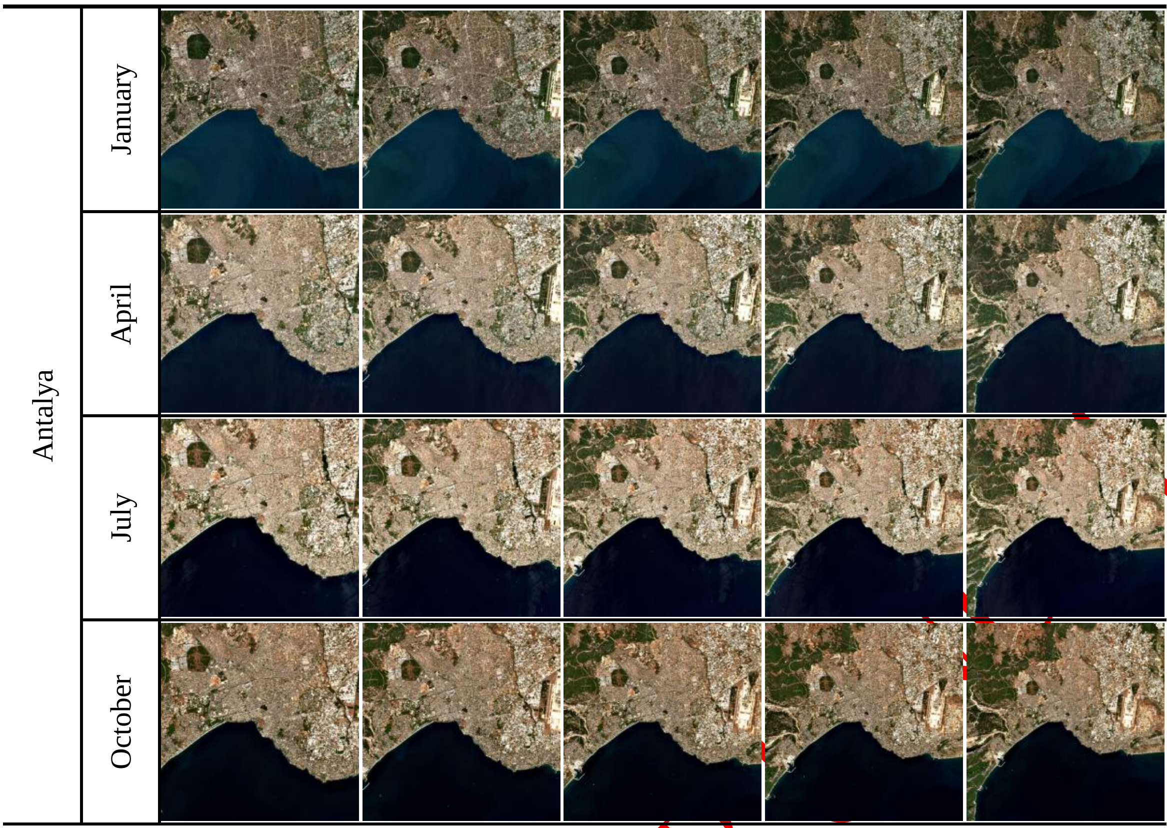


Figure 1. Sample images from Bursa, which exhibit clear seasonal transitions, and Antalya, where seasonal transitions are less visually distinct (Mevsim geçişlerinin belirgin olduğu Bursa ve mevsim geçişlerinin görsel olarak daha az belirgin olduğu Antalya'dan örnek görüntüler)

Figure 1 shows sample images from Bursa, which exhibits clear seasonal transitions, and Antalya, where seasonal transitions are less visually distinct. Images were preprocessed to accommodate the input structure of the pretrained models. All RGB satellite images were resized to 224 × 224 pixels and normalized with ImageNet dataset's mean and standard deviation values. For the fine-tuning stage, random rescaling and cropping within the scale range of 0.85–1.0 and aspect ratios of 0.95–1.05, horizontal flipping with a 50% probability, vertical flipping with a 50% probability, random rotation of ±10 degrees, and slight color distortions in brightness, contrast, saturation, and hue values with a 40% probability were applied as data augmentation.

3.2. Method (Metot)

In the proposed hybrid framework, eight ImageNet-pretrained deep learning architectures and seven machine learning classifiers were employed to optimize feature extraction and classification for satellite image-based seasonal classification.

3.2.1. Pretrained models (Öneğitimli modeller)

The pretrained models, summarized in Table 2, were used only for feature extraction. The final classification layers of pretrained models were removed, and fixed-length feature vectors were obtained from the last representation layer or by applying global average pooling to the final convolutional feature map. For each satellite image sample, features were extracted independently from five spatial scales and then aggregated by mean fusion.

3.2.2. Classification models (Sınıflandırma modelleri)

The feature vectors extracted from the pretrained models were classified using seven machine learning algorithms. Table 3 summarizes the classifiers used in this study. These classifiers were selected to compare linear, nonlinear, distance-based, neural network-based, and ensemble-based decision mechanisms within the same extracted feature space.

Table 2. Brief Explanations of the pretrained models utilized in this study (Çalışmada kullanılan öneğitimli modellerin kısa açıklamaları)

Pretrained model	Explanation and role in this study
ResNet50	Addresses the degradation and vanishing-gradient using residual connections for high-dimensional features [22], [23].
DenseNet121	Enables feature reuse through dense inter-layer connections. Each layer is connected to all preceding layers, which allows for efficient feature reuse, improved gradient flow, and reduced redundancy in feature learning [24], [25].
ConvNeXt	Captures strong local and mid-level spatial patterns with a modern design. It redesigns the vanilla ResNet structure using depthwise separable convolutions and layer normalization thereby improving computational efficiency and training stability [26].
MobileNet V3	Provides compact and computationally efficient feature representations. It is designed for mobile and memory-constrained devices using neural architecture search [27], [28].
EfficientNet B2	Balances network depth, width, and input resolution for efficient feature extraction. It is a scaled variant of EfficientNet that uses mobile inverted bottleneck convolutions and squeeze-and-excitation blocks [29].
Vision Transformer	It divides an input feature map into fixed-size patches and models long-range dependencies through self-attention [30].
Swin Transformer	It uses shifted window-based self-attention to efficiently model visual features at different scales while reducing the computational complexity [31].
RegNetY	Represents optimized convolutional network design with squeeze-and-excitation blocks. It is derived from a systematically designed network design space, where structural parameters such as depth and width are progressively optimized [32].

Table 3. Brief Explanations of the machine learning classifiers utilized in this study (Çalışmada kullanılan makine öğrenimi sınıflandırıcılarının kısa açıklamaları)

Classifier	Explanation and role in this study
Linear SVM	Finds a maximum-margin linear decision boundary. It aims to find the optimal separating hyperplane between classes by maximizing the decision margin.
Logistic Regression	It estimates class probabilities by applying a logistic function to a linear combination of input features.
Random Forest	It constructs multiple decision trees using bootstrap samples and random feature subsets, enabling robust classification by capturing nonlinear relationships.
Multi-Layer Perceptron	Learns nonlinear transformations from extracted feature vectors. It learns nonlinear transformations through hidden layers, allowing more flexible decision boundaries.
Histogram-Based Gradient Boosting	Efficient boosting-based ensemble method that discretizes continuous features into histogram bins and sequentially builds learners to capture nonlinear feature interactions.
k-Nearest Neighbors	Instance-based classifier that assigns a sample to the majority class among its closest training samples in the feature space.
Extreme Gradient Boosting	A powerful gradient-boosted decision tree algorithm that sequentially adds trees to minimize residual errors.

3.2.3. Proposed hybrid transfer learning based framework (Hibrit transfer öğrenme tabanlı önerilen çerçeve)

The proposed hybrid framework is designed to exploit the representation ability of pretrained deep models while avoiding computational cost. Rather than using pretrained models directly, the proposed hybrid framework first employs them as feature extractors and then uses machine learning algorithms to perform the final classification.

As illustrated in Figure 2, the proposed hybrid framework consists of two main stages: selection of the most effective deep feature representations and classifier and representation fusion using fine-tuning. In the first stage, different ImageNet-pretrained models described in Section 3.2.1 are used to extract high-level representations from satellite images, and then these representations are evaluated using seven machine learning classifiers described in Section 3.2.2. Based on the initial evaluation, the three most effective feature extractors and the most effective machine learning classifier are selected. In the second stage, the selected pretrained models are fine-tuned, and their updated feature representations are concatenated for final classification.

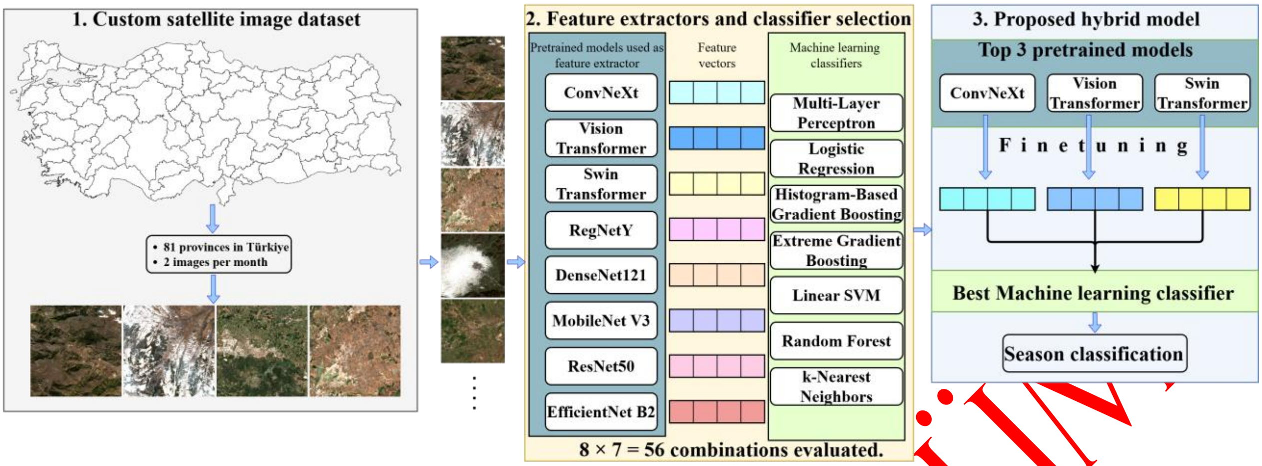


Figure 2. Overview of the proposed hybrid transfer learning framework. First, a multi-scale satellite image dataset was constructed from all 81 provinces of Türkiye. Second, eight pretrained deep learning models and seven machine learning classifiers were evaluated to determine the most effective feature extractor–classifier combinations. Finally, the best classifier and the top three pretrained feature extractors were selected and combined for seasonal classification (Önerilen hibrit transfer öğrenme çerçevesinin genel görünümü. İlk olarak, Türkiye'nin 81 ilinin tamamından çok ölçekli bir uydu görüntüsü veri seti oluşturuldu. İkinci olarak, en etkili özellik çıkarıcı-sınıflandırıcı kombinasyonlarını belirlemek için sekiz önceden eğitilmiş derin öğrenme modeli ve yedi makine öğrenmesi sınıflandırıcısı değerlendirildi. Son olarak, en iyi sınıflandırıcı ve en iyi üç önceden eğitilmiş özellik çıkarıcısı seçilerek mevsimsel sınıflandırma için birleştirildi)

Selection of the most effective deep feature representations. Pretrained models with different structural characteristics are employed to obtain latent feature representations from the satellite images. The use of pretrained models is motivated by the fact that the earlier layers of deep neural networks generally learn transferable visual patterns such as edges, textures, shapes, and local structures, which can be useful across different image domains [21], [22]. Therefore, instead of learning visual representations from scratch, the proposed framework transfers the pretrained visual knowledge to the satellite image classification task.

High-quality latent representations are obtained from satellite images using ImageNet pretrained deep models. After latent representation extraction, the obtained latent representations are classified using machine learning classifiers, each of which is trained separately. For each pretrained model, the representational quality of its extracted feature set is assessed based on the classification performance. Accordingly, the three pretrained models that generate the most discriminative feature representations are selected for further fine-tuning, while the machine learning classifier with the highest overall classification performance is selected as the final classifier. This evaluation enables a comprehensive comparison of both the representation capacity of the pretrained models and the decision-making ability of the machine learning classifiers.

Representation fusion using fine-tuning. The three selected pretrained models are fine-tuned independently for three epochs to adapt the ImageNet-pretrained representations to the specific visual characteristics of our satellite image dataset. Although ImageNet pretraining provides strong generic visual features, satellite images differ in terms of spatial patterns, land-cover structures, seasonal variations, and spectral-textural characteristics. Therefore, fine-tuning allows the selected models to adjust their feature extraction process to the target domain. After fine-tuning, for each input image, three updated feature vectors are concatenated. The concatenated feature vectors enable each satellite image to be represented from three complementary feature perspectives, thereby increasing the descriptive capacity of the final feature space. The selected machine learning classifier is then trained using the concatenated feature representations to make the final prediction.

The proposed hybrid transfer learning-based framework is particularly appropriate for satellite image classification problems where the dataset size is limited compared with the amount of data typically required to train deep neural networks. Since pretrained models already encode visual patterns, machine learning classifiers can operate on informative feature vectors. Moreover, the fine-tuning adapts the selected representations to the target satellite image domain and combines their complementary strengths. Overall, the proposed method provides a hybrid classification pipeline that integrates the advantages of transfer learning, deep feature extraction, and machine learning-based classification.

4.RESULTS (BULGULAR)

4.1. Implementation Details (Uygulama Detayları)

The proposed hybrid framework was implemented and evaluated on a system equipped with an Intel i7-10700KF CPU, 32 GB DDR4 RAM, and an NVIDIA TITAN X Pascal GPU. For frozen feature extraction, the final classification layers of the ImageNet-pretrained networks were removed, and the feature vectors were obtained from the last representation layer of each network or by applying global average pooling to the final convolutional feature map. For each satellite image sample, latent features were extracted independently from the five spatial scales and aggregated using mean fusion. Thus, one fixed-length feature vector was obtained for each satellite image. The feature extraction layers and resulting dimensionalities are reported in Table 4.

Table 4. The feature extraction layers and output sizes of pretrained models (Ön eğitilmiş modellerin özellik çıkarım katmanları ve çıktı boyutları)

Pretrained model	Feature extraction layer	Feature dimension
ResNet50	Output of layer4 followed by adaptive global average pooling	2048
DenseNet121	Output of features block after ReLU and adaptive global average pooling	1024
ConvNeXt	Output of features block followed by adaptive global average pooling	768
MobileNet V3	Output of features block followed by adaptive global average pooling	960
EfficientNet B2	Output of features block followed by adaptive global average pooling	1408
RegNetY	Penultimate representation before the fully connected layer	1512
Vision Transformer	Final encoder representation after replacing classification head with identity	768
Swin Transformer	Normalized final feature map averaged over spatial dimensions	768

We first evaluated the performance of each model–classifier pair to select the best machine learning classifier and the top three pretrained feature extractor models for our proposed hybrid framework. Second, we analyzed the fine-tuning effect of the proposed hybrid framework.

Table 5. Hyperparameter values of the machine learning classifiers and the proposed hybrid framework (Makine öğrenimi sınıflandırıcılarının ve önerilen hibrit çerçevenin hiperparametre değerleri)

Model	Main hyperparameters
Linear SVM	Class weight = balanced, maximum iterations = 500
Logistic Regression	Class weight = balanced, maximum iterations = 500
Random Forest	Number of estimators = 100, class weight = balanced
Multi-Layer Perceptron	One hidden layer with 128 neurons, maximum iterations = 500
Histogram-Based Gradient Boosting	Maximum iterations = 100, L2 regularization coefficient = 0.1
k-Nearest Neighbors	Number of neighbors = 5
Extreme Gradient Boosting	Number of estimators = 100, maximum depth = 6, learning rate = 0.1
Fine-tuning	3 epochs of training, batch size = 16, AdamW optimizer with default values, backbone learning rate =0.00001, head learning rate = 0.001.

The hyperparameters of the machine learning classifiers and the proposed hybrid framework are summarized in Table 5. Fine-tuning was applied to the three selected pretrained models for three epochs with a batch size of 16. During the first epoch, the backbone was frozen and only the classification head was trained. The AdamW optimizer was used with a learning rate of 0.00001 for the backbone parameters and 0.001 for the classification head. Cross-entropy loss was used as the optimization objective during fine-tuning.

The performance of each model–classifier pair was assessed using accuracy and F1-score. To improve statistical robustness and mitigate variance in performance estimates, all pretrained model-classifier combinations were evaluated using a 10-fold group-based cross-validation repeated five times (42, 52, 62, 72, and 82 were used as random seeds). In every fold, all samples from the same province were assigned to either the training or test set, preventing spatial leakage between splits. All reported results correspond to the average performance over five repetitions of 10-fold group-based cross-validation.

4.1. Quantitative Results (Nicel Sonuçlar)

Performance comparison of pretrained models and machine learning classifiers. Table 6 presents the comparative analysis of pretrained models used as deep feature extractors with machine learning classifiers for the seasonal classification task in terms of accuracy and F1-score. The results demonstrate that the choice of pretrained feature extractor has a significant impact on classification performance. Among all combinations, ConvNeXt combined with the Multi-Layer Perceptron achieved the best results. The modern design structure of ConvNeXt can provide highly discriminative representations for capturing seasonal patterns. The competitive results of Vision Transformer and Swin Transformer indicate that transformer-based representations are also effective for seasonal satellite image classification due to their ability to model long-range spatial dependencies.

The multi-layer perceptron consistently achieved the strongest results across pretrained models. Therefore, it can be inferred that the extracted deep features are not always optimally separable by simple distance-based or linear decision boundaries. The multi-layer perceptron classifier benefits from its ability to learn non-linear relationships among high-level deep features, which is particularly important for distinguishing visually similar seasons such as spring and summer or autumn and winter. In tree-based classifiers, histogram-based gradient boosting and extreme gradient boosting generally outperformed random forest, but they often remained below the multi-layer perceptron-based and logistic regression-based models. The lowest performances were generally obtained with k-nearest neighbors. Although pretrained embeddings are semantically meaningful, the class boundaries may not correspond to local neighborhood structures under vanilla distance metrics. The accuracy and F1-score values are very close for most model–classifier combinations. The pretrained models and machine learning classifier pairs not only achieved high accuracy by favoring only the dominant classes, but their performance was also relatively consistent across seasonal categories. To identify the feature extractors in the proposed hybrid framework, the independent representation and generalization capabilities of 8 pretrained models and 7 machine learning classifiers were analyzed. Figure 3 presents the average classification performance of the utilized pretrained models and machine learning classifiers in terms of mean F1-score and mean accuracy. The first row compares the pretrained models averaged over all classifiers, while the second row compares the classifiers averaged over all pretrained models. ConvNeXt, Vision Transformer, and Swin Transformer models were the top three models, achieving the best average results. These results suggest that modern convolutional architectures and transformer-based models provide stronger transferable representations than the other evaluated architectures for capturing spatial heterogeneity and phenological transitions in satellite imagery. Consequently, we selected ConvNeXt, Vision Transformer, and Swin Transformer as the top three pretrained feature extractors for the proposed hybrid framework. Consistent with the results in Table 6, the multi-layer perceptron outperformed all other classifiers on average, as shown in Figure 3. Thus, we selected multi-layer perceptron as a classifier model for the proposed hybrid framework.

Table 6. Performance comparison of all pretrained model and machine learning classifier combinations averaged for 10-fold cross-validation with five repetitions (Tüm ön eğitilmiş model ve makine öğrenimi sınıflandırıcı kombinasyonlarının performans karşılaştırması, beş tekrarlı 10 katlı çapraz doğrulama için ortalama alınarak yapılmıştır)

Pretrained model	Classifier	Accuracy	F1 score
ConvNeXt	Multi-Layer Perceptron	0.698±0.005	0.696±0.007
Vision Transformer	Multi-Layer Perceptron	0.675±0.007	0.673±0.004
Swin Transformer	Multi-Layer Perceptron	0.671±0.005	0.668±0.005
ConvNeXt	Logistic Regression	0.660±0.005	0.658±0.005
DenseNet121	Multi-Layer Perceptron	0.657±0.009	0.654±0.008
Swin Transformer	Logistic Regression	0.644±0.004	0.641±0.002
MobileNet V3	Multi-Layer Perceptron	0.637±0.003	0.635±0.002
Vision Transformer	Logistic Regression	0.637±0.008	0.635±0.008
ConvNeXt	Histogram-Based Gradient Boosting	0.635±0.010	0.633±0.009
RegNetY	Multi-Layer Perceptron	0.629±0.006	0.626±0.005
RegNetY	Histogram-Based Gradient Boosting	0.623±0.004	0.620±0.003
Vision Transformer	Histogram-Based Gradient Boosting	0.617±0.010	0.616±0.010
RegNetY	Logistic Regression	0.614±0.005	0.611±0.006
RegNetY	Extreme Gradient Boosting	0.614±0.005	0.611±0.004
ConvNeXt	Extreme Gradient Boosting	0.613±0.006	0.610±0.006
DenseNet121	Histogram-Based Gradient Boosting	0.614±0.007	0.610±0.005
Vision Transformer	Extreme Gradient Boosting	0.609±0.004	0.607±0.003
Swin Transformer	Histogram-Based Gradient Boosting	0.609±0.007	0.607±0.007
ResNet50	Multi-Layer Perceptron	0.612±0.007	0.606±0.007
MobileNet V3	Logistic Regression	0.608±0.003	0.605±0.003
ResNet50	Histogram-Based Gradient Boosting	0.609±0.010	0.605±0.012
ResNet50	Logistic Regression	0.609±0.005	0.604±0.004
MobileNet V3	Histogram-Based Gradient Boosting	0.602±0.006	0.601±0.005
DenseNet121	Logistic Regression	0.603±0.008	0.600±0.008
DenseNet121	Extreme Gradient Boosting	0.601±0.008	0.597±0.008
ResNet50	Extreme Gradient Boosting	0.597±0.009	0.594±0.010
Swin Transformer	Extreme Gradient Boosting	0.594±0.007	0.592±0.007
MobileNet V3	Extreme Gradient Boosting	0.591±0.004	0.590±0.005
ConvNeXt	Linear SVM	0.592±0.004	0.588±0.004
ConvNeXt	Random Forest	0.585±0.015	0.581±0.014
EfficientNet B2	Multi-Layer Perceptron	0.584±0.005	0.580±0.005
Swin Transformer	Linear SVM	0.583±0.007	0.578±0.007
RegNetY	Random Forest	0.580±0.007	0.574±0.008
Vision Transformer	Random Forest	0.573±0.017	0.570±0.017
Vision Transformer	Linear SVM	0.572±0.014	0.569±0.014
ConvNeXt	k-Nearest Neighbors	0.564±0.007	0.563±0.007
Swin Transformer	Random Forest	0.563±0.009	0.561±0.009
MobileNet V3	Random Forest	0.564±0.004	0.561±0.004
ResNet50	Random Forest	0.566±0.010	0.560±0.012
EfficientNet B2	Logistic Regression	0.563±0.006	0.560±0.004
Vision Transformer	k-Nearest Neighbors	0.555±0.005	0.556±0.005
DenseNet121	Random Forest	0.562±0.012	0.556±0.012
ResNet50	Linear SVM	0.554±0.005	0.548±0.005
EfficientNet B2	Histogram-Based Gradient Boosting	0.546±0.012	0.543±0.013
RegNetY	Linear SVM	0.547±0.007	0.542±0.007
DenseNet121	Linear SVM	0.545±0.004	0.540±0.005
MobileNet V3	Linear SVM	0.542±0.009	0.537±0.008
EfficientNet B2	Extreme Gradient Boosting	0.532±0.006	0.531±0.007
Swin Transformer	k-Nearest Neighbors	0.522±0.004	0.520±0.005
EfficientNet B2	Linear SVM	0.510±0.012	0.506±0.010
EfficientNet B2	Random Forest	0.507±0.008	0.502±0.007
DenseNet121	k-Nearest Neighbors	0.499±0.007	0.496±0.007
RegNetY	k-Nearest Neighbors	0.481±0.006	0.476±0.005
MobileNet V3	k-Nearest Neighbors	0.472±0.005	0.468±0.005
ResNet50	k-Nearest Neighbors	0.455±0.006	0.447±0.005
EfficientNet B2	k-Nearest Neighbors	0.421±0.006	0.419±0.006

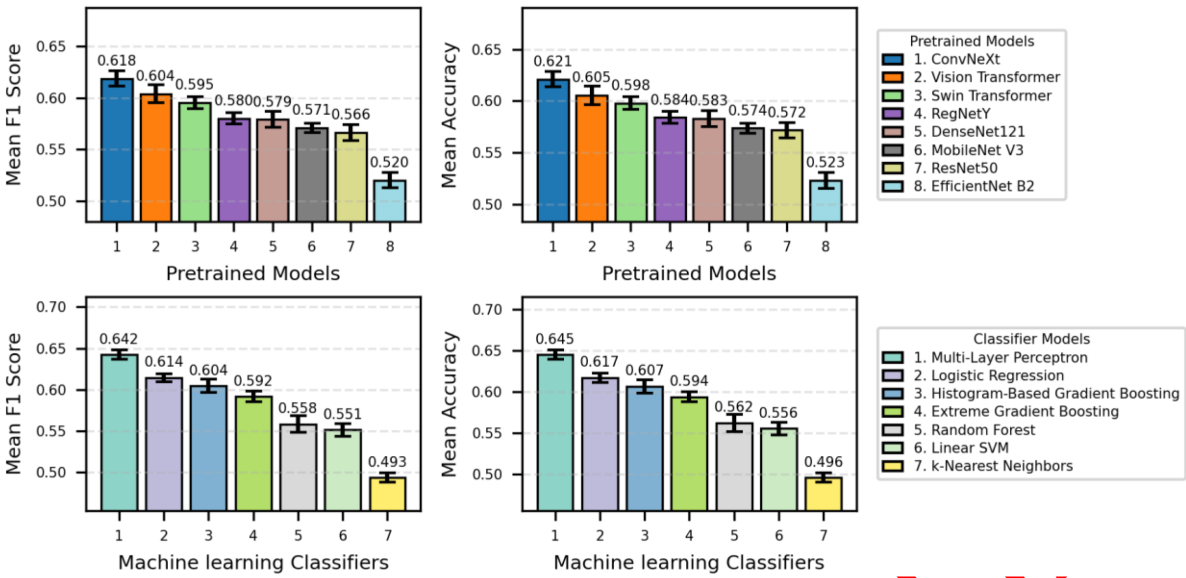


Figure 3. Average performance comparison of all pretrained model and machine learning classifier combinations for 10-fold cross-validation with five repetitions (10 katlı çapraz doğrulama yöntemiyle beş tekrar kullanılarak yapılan tüm ön eğitilmiş modellerin ve makine öğrenimi sınıflandırıcı kombinasyonlarının ortalama performans karşılaştırması)

Table 7 compares the performance of the three best pretrained models with the proposed hybrid framework before and after fine-tuning. Features extracted from ConvNeXt, Vision Transformer, and Swin Transformer were combined to construct the proposed hybrid representation, and the resulting concatenated feature representation was classified using the Multi-Layer Perceptron. We first trained only the multi-layer perceptron with extracted features and achieved a mean accuracy of 0.725 and a mean F1-score of 0.724 under the repeated 10-fold cross-validation. This improvement shows that the selected pretrained architectures provide complementary representations. While ConvNeXt captures strong local and mid-level spatial patterns, Vision Transformers contribute global contextual information, and Swin Transformers provide hierarchical attention-based representations. Therefore, combining these features allows the classifier to exploit a richer and more discriminative feature space than any single architecture alone. After fine-tuning for three epochs, our hybrid framework achieved 0.740 accuracy and 0.742 F1-score. This indicates that, although features pre-trained on ImageNet are effective for satellite image classification, brief fine-tuning helps adapt the extracted representations to the specific characteristics of the our custom dataset. The improvement after fine-tuning indicates that the pretrained representations may not fully capture the domain-specific visual patterns of satellite imagery. Fine-tuning enables the selected feature extractors to adjust their representations toward these remote sensing-specific patterns, thereby increasing their discriminative capability for seasonal classification.

Table 7. Comparison of the performance of the three best pretrained models and the proposed hybrid model before and after fine-tuning (En iyi üç ön eğitilmiş modelin ve önerilen hibrit modelin ince ayar öncesi ve sonrası performanslarının karşılaştırılması)

Pretrained model	Classifier	Accuracy	F1-score
ConvNeXt	Multi-Layer Perceptron	0.698	0.695
Vision Transformer	Multi-Layer Perceptron	0.674	0.673
Swin Transformer	Multi-Layer Perceptron	0.670	0.668
Hybrid	Multi-Layer Perceptron	0.725	0.724
Hybrid (Fine-tuned)	Multi-Layer Perceptron	0.740	0.742

Hybrid: ConvNeXt, Vision Transformer, Swin Transformer

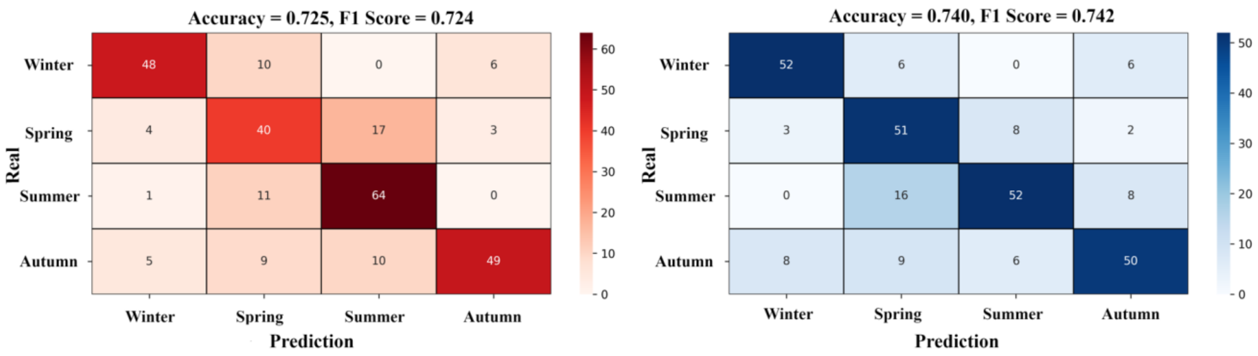


Figure 4. Confusion matrices of the proposed hybrid model before and after fine-tuning (Önerilen hibrit modelin ince ayar öncesi ve sonrası karışıklık matrisleri)

Figure 4 presents the confusion matrices of the proposed hybrid framework before and after fine-tuning. The fine-tuned hybrid model correctly classified 52 of 64 winter samples, 51 of 64 spring samples, 52 of 76 summer samples, and 50 of 73 autumn samples. These results indicate that the proposed hybrid framework maintained relatively balanced performance across all seasonal classes. The most frequent misclassifications occurred between adjacent seasons, particularly spring–summer and summer–autumn, which is expected because neighboring seasons may share similar vegetation cover, illumination conditions, and land-surface characteristics in satellite imagery. Nevertheless, the diagonal dominance of the matrix indicates that the proposed transfer learning-based hybrid framework provides a balanced seasonal classification performance.

5. CONCLUSIONS (SONUÇLAR)

This study proposed a transfer learning-based hybrid framework for seasonal classification from satellite images. A custom dataset was constructed using monthly Sentinel-2 satellite images obtained from 81 provinces of Türkiye. Eight ImageNet-pretrained deep learning models were evaluated as feature extractors. The extracted representations were classified using seven machine learning algorithms to evaluate seasonal classification performance. The quantitative results showed that the selection of pretrained models strongly influences classification performance. Based on the comparative analysis of 56 model–classifier combinations, ConvNeXt, Vision Transformer, and Swin Transformer were selected as the top three feature extractors, while the multi-layer perceptron was selected as the final classifier. Among the models, ConvNeXt achieved the best performance when combined with the multi-layer perceptron classifier. Vision Transformer and Swin Transformer also produced competitive results, indicating that both convolution-based and transformer-based representations are effective for capturing seasonal patterns in satellite imagery. The classifier comparison further demonstrated that the multi-layer perceptron outperformed the other machine learning classifiers. The proposed transfer learning-based hybrid framework, formed by concatenating the feature representations of these three selected models, achieved the best performance after fine-tuning, with an accuracy of 0.740 and an F1-score of 0.742. These results confirm that limited domain-specific fine-tuning can improve the adaptation of ImageNet-pretrained representations. Future work can extend this research by incorporating multispectral data, longer temporal sequences, and remote sensing-specific self-supervised pretraining. In conclusion, the proposed transfer learning-based hybrid framework provides an effective and practical approach for seasonal classification from satellite images and offers a promising basis for future environmental monitoring, agricultural analysis, and climate-related remote sensing applications.

DECLARATION OF ETHICAL STANDARDS (ETİK STANDARTLARIN BEYANI)

The authors of this article declare that the materials and methods they use in their work do not require ethical committee approval and/or legal-specific permission.

Bu makalenin yazarı çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan ederler.

AUTHORS' CONTRIBUTIONS (YAZARLARIN KATKILARI)

Eyyüp YILDIZ: He conceived the idea for the study, conducted the experiments, analysed the results, and wrote the paper.

Çalışma fikrini geliştirdi, deneyleri yürüttü, sonuçları analiz etti ve makaleyi yazdı.

Özge ASLAN YILDIZ: She carried out the methodology of the study, followed the experiments, analyzed the results, and wrote the paper.

Çalışmanın metodolojisini uyguladı, deneyleri yürüttü, sonuçları analiz etti ve makaleyi yazdı.

CONFLICT OF INTEREST (ÇIKAR ÇATIŞMASI)

There is no conflict of interest in this study.

Bu çalışmada herhangi bir çıkar çatışması yoktur.

REFERENCES (KAYNAKLAR)

[1] Krejcar, O., Namazi, H. (2026). AI in remote sensing and satellite image processing-a review. Environmental Earth Sciences. 85(3). 78. https://doi.org/10.1007/s12665-025-12798-w.

[2] LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. Nature. 521(7553), 436-444. https://doi.org/10.1038/nature14539.

[3] Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. IEEE geoscience and remote sensing magazine. 5(4). 8-36. 10.1109/MGRS.2017.2762307.

[4] Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J. (2020). Deep learning in environmental remote sensing: Achievements and challenges. Remote sensing of Environment. 241. 111716. 10.1016/j.rse.2020.111716.

[5] Zhao, T., Wang, S., Ouyang, C., Chen, M., Liu, C., Zhang, J., Yu, L., Wang, F., Xie, Y., Li, J., Wang, F. (2024). Artificial intelligence for geoscience: Progress, challenges, and perspectives. The Innovation. 5(5). 10.1016/j.xinn.2024.100691.

[6] Cheng, G., Han, J., Lu, X. (2017). Remote sensing image scene classification: Benchmark and state of the art. Proceedings of the IEEE. 105(10). 1865-1883. 10.1109/JPROC.2017.2675998.

[7] Moharram, M. A., Sundaram, D. M. (2023). Land use and land cover classification with hyperspectral data: A comprehensive review of methods, challenges and future directions. Neurocomputing. 536. 90-113. https://doi.org/10.1016/j.neucom.2023.03.025.

[8] Gui, S., Song, S., Qin, R., Tang, Y. (2024). Remote sensing object detection in the deep learning era - a review. Remote Sensing. 16(2). 327. https://doi.org/10.3390/rs16020327.

[9] Li, K., Wan, G., Cheng, G., Meng, L., Han, J. (2020). Object detection in optical remote sensing images: A survey and a new benchmark. ISPRS journal of photogrammetry and remote sensing. 159. 296-307. https://doi.org/10.1016/j.isprsjprs.2019.11.023.

[10] Uzun, S., Arslantaş, H. (2024). Determination of Radiation Value by Month Using Artificial Neural Network Model; Ankara, Sivas, Erzurum example. Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji. 12(1). 315-323. 10.29109/gujsc.1420617.

[11] Pires de Lima, R., Marfurt, K. (2019). Convolutional neural network for remote-sensing scene classification: Transfer learning analysis. Remote Sensing. 12(1). 86. https://doi.org/10.3390/rs12010086.

[12] Thirumaladevi, S., Swamy, K. V., Sailaja, M. (2023). Remote sensing image scene classification by transfer learning to augment the accuracy. Measurement: Sensors. 25. 100645. https://doi.org/10.1016/j.measen.2022.100645.

[13] Kunwar, S., Ferdush, J. (2024). Mapping of Land Use and Land Cover (LULC) Using EuroSAT and Transfer Learning. Revue Internationale de Geomatique. 33. 1-13. https://doi.org/10.32604/rig.2023.047627.

[14] Roy, A., Gourisaria, M. K., Chatterjee, R., Jha, A. V., Appasani, B., Bizon, N., Mazare, A. G. (2025). A privacy-preserving, on-board satellite image classification technique incorporating homomorphic encryption and transfer learning. Frontiers in Remote Sensing. 6. 1678882. https://doi.org/10.3389/frsen.2025.1678882.

[15] Farag, N. H., Abdelhafiz, D., Abdrabo, M. A., ELIskandarani, M. A., Hassaan, M. A. (2026). Enhancing the classification of spectrally similar land use/land cover classes using transfer learning in arid regions. Scientific Reports.

https://doi.org/10.1038/s41598-026-38540-5.

[16] Sudiana, D., Lestari, A. I., Riyanto, I., Rizkinia, M., Arief, R., Prabuwono, A. S., Sri Sumantyo, J. T. (2023). A hybrid convolutional neural network and random forest for burned area identification with optical and synthetic aperture radar (SAR) data. Remote Sensing. 15(3). 728. https://doi.org/10.3390/rs15030728.

[17] Kwenda, C., Gwetu, M., Fonou-Dombeu, J. V. (2024). Hybridizing deep neural networks and machine learning models for aerial satellite forest image segmentation. Journal of imaging. 10(6). 132. https://doi.org/10.3390/jimaging10060132.

[18] Mashraqi, A. M., Alshari, E. A., Halawani, H. T., Senan, E. M., Asiri, Y., Alowadhi, B. M. (2025). Hybrid deep learning and optimization-based land use and land cover classification for advancing sustainable agriculture in Najran city, Saudi Arabia. Scientific Reports. 15(1). 41808. https://doi.org/10.1038/s41598-025-25908-2.

[19] Sudiana, D., Putri, S. H., Kushardono, D., Prabuwono, A. S., Sri Sumantyo, J. T., Rizkinia, M. (2025). CNN-random forest hybrid method for phenology-based paddy rice mapping using Sentinel-2 and Landsat-8 satellite images. Computers. 14(8). 336. https://doi.org/10.3390/computers14080336.

[20] Cheng, P., Zhou, J. (2011). Automatic season classification of outdoor photos. Third International Conference on Intelligent Human-Machine Systems and Cybernetics. 1. 46-49. IEEE.

[21] Volokitin, A., Timofte, R., Van Gool, L. (2016). Deep features or not: Temperature and time prediction in outdoor scenes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 63-71.

[22] Bilotta, G., Bibbò, L., Meduri, G. M., Genovese, E., Barrile, V. (2025). Deep learning innovations: Resnet applied to SAR and sentinel-2 imagery. Remote Sensing. 17(12). 1961. https://doi.org/10.3390/rs17121961.

[23] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. 770-778.

[24] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition. 4700-4708.

[25] Kurtoğlu, B., Çiftçi, S. (2024). A comparative study of swin-based enhanced remote sensing image classifications. In Proceedings of the 2024 6th International Conference on Image Processing and Machine Vision. 68-74.

[26] Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., Xie, S. (2022). A convnet for the 2020s. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 11976-11986.

[27] Liu, H. I., Galindo, M., Xie, H., Wong, L. K., Shuai, H. H., Li, Y. H., Cheng, W. H. (2024). Lightweight deep learning for resource-constrained environments: A survey. ACM Computing Surveys. 56(10). 1-42. https://doi.org/10.1145/3657282.

[28] Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q.V. (2019). Searching for mobilenetv3. In Proceedings of the IEEE/CVF international conference on computer vision. 1314-1324.

[29] Tan, M., Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning. 6105-6114. PMLR.

[30] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

[31] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision. 10012-10022.

[32] Radosavovic, I., Kosaraju, R. P., Girshick, R., He, K., Dollár, P. (2020). Designing network design spaces. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 10428-10436.