IJWR

Transfer Learning for Crop Classification in Data-Scarce Regions Using Satellite Imagery

Erfan Shakouri, Saeed Zare, Masoume Azimzadeh, Parvin Ahmadi*

ICT Research Institute, Tehran, Iran; erfanshakouri.aielec@gmail.com, sd.zare73@gmail.com, azim_ma@itrc.ac.ir, p.ahmadi@itrc.ac.ir

ABSTRACT

Satellite imagery provides valuable data to address the growing demand for agricultural production. However, analyzing such vast amounts of data requires advanced artificial intelligence methods, such as deep learning. The primary challenge lies in the scarcity of labeled training data, as its preparation is both costly and time-consuming. To address this issue, this study integrates remote sensing data, deep neural networks, and transfer learning techniques to estimate the cultivated area of strategic crops in Iran. Given the diverse climates and topographies across Iran's provinces, in addition to Sentinel-1 and Sentinel-2 satellite data, MODIS sensor and SRTM elevation data were also utilized. To compensate for data limitations, transfer learning was employed to enhance model performance in data-deficient regions (Kermanshah and Markazi). This approach resulted in an approximate 10% improvement in Cohen's Kappa coefficient. Furthermore, the study investigated the minimum data required for fine-tuning the models. The results demonstrated that even with a reduction of over 60% in the target province's training data, transfer learning still achieved model performance comparable to scenarios where it was not applied.

Keywords— Crop Classification, Transformers, CNN, Transfer Learning, Satellite Image Time Series, Data-scarce.

1. Introduction

The goal of achieving "Zero Hunger" as one of the Sustainable Development Goals (SDGs) faces serious threats amid growing challenges from climate change and population growth [1]. According to the UN Food and Agriculture Organization (FAO) report, over 720 million people worldwide suffer from hunger. Accurate and timely assessment of agricultural production is critical to addressing this crisis [1]. Satellite mapping serves as a powerful tool for monitoring agricultural output and assessing food security at a global scale. This technology enables precise identification of crop types, growth monitoring, and yield prediction through the analysis of time-series satellite imagery [2] and [3]. Updated and accurate crop distribution maps help decision-makers identify food-insecure regions and plan necessary interventions to address food crises [4]. Furthermore, this information is essential for developing sustainable agricultural policies, optimizing water resource management, and mitigating climate change impacts on crop production [5].

Crop classification acts as the fundamental step in estimating cultivated areas and serves as the primary stage for yield prediction [3]. Deep learning models significantly contribute to agricultural crop classification. Consequently, current research has shifted focus toward advanced deep learning algorithms, as these models can achieve high accuracy in solving crop classification problems. Notable examples include Transformer and CNN models. For instance, Transformers (among the most powerful deep learning architectures) are particularly suitable for processing time-series satellite imagery due to their ability to model long-term dependencies in data. These models can effectively capture seasonal vegetation changes and variations in crop growth patterns [5]. Recent literature has emphasized the effectiveness of hierarchical fusion across radar, optical, and elevation data for improving model generalization [6]. Our framework is consistent with these findings, as it similarly employs multi-source

doi http://dx.doi.org/10.22133/ijwr.2025.537225.1301

Citation E. Shakouri, S. Zare, M. Azimzadeh, P. Ahmadi, "Transfer Learning for Crop Classification in Data-Scarce Regions Using Satellite Imagery", *International Journal of Web Research*, vol.9, no.1,pp.25-37, 2026, doi: http://dx.doi.org/10.22133/ijwr.2025.537225.1301.

**Coressponding Author*

Article History: Received: 26 July 2025; Revised: 6 December 2025; Accepted: 20 December 2025.

data (Sentinel-1, Sentinel-2, MODIS, and SRTM) to address the spatial and climatic diversity of Iran's agricultural regions.

One of the main challenges in using deep neural networks for agricultural crop classification is the scarcity of high-quality labeled data. These data, known as ground truth data, are essential for training deep learning models [1]. In many regions, particularly in developing countries, access to high-resolution, large-scale labeled datasets is limited, making it difficult to train high-performance deep learning models [3]. The similarity in crop growth patterns across different regions of the world enables the use of models trained in one area to predict crops in another. This approach, known as transfer learning, has gained increasing attention in deep learning applications [7]. Transfer learning, in essence, involves leveraging knowledge acquired from one problem to solve another. In the context of agricultural crop classification, this means that a model trained on a large dataset can be adapted to improve performance on smaller datasets [8].

As a developing country, Iran faces challenges due to the lack of labeled agricultural data. Acquiring such data in certain regions is difficult and costly due to accessibility constraints, high expenses, and inadequate infrastructure. To address this challenge, this study employs a transfer learning approach. Specifically, a deep learning model is first trained on readily available labeled data (e.g., from advanced agricultural regions). The pre-trained model is then fine-tuned using limited target-region data. This way, knowledge from the source region is transferred to the target region, enhancing model performance in the target area.

Given the phenological diversity of crops across different climates, this study incorporates not only Sentinel-1 and Sentinel-2 satellite data but also SRTM and Terra data to account for geometric features, enabling models to better comprehend climatic variations. A key challenge in transfer learning is determining the minimum required training data for model fine-tuning to achieve acceptable performance, which is thoroughly investigated in this paper. Furthermore, to evaluate transfer learning's efficacy in reducing labeled data dependency, the method was implemented on various convolutional and transformer-based models, with comparative analysis of the results.

The paper is structured as follows: Section 2 reviews prior research on transfer learning, examining the advantages and limitations of each approach. Section 3 details the study area and dataset characteristics. Section 4 comprehensively describes the research methodology, including employed models and experimental scenarios. Section 5 presents and analyzes experimental results, while Section 6 provides deeper interpretation and discusses the underlying reasons for these outcomes. Finally, Section 7 summarizes key research findings and examines their significance.

2. Related Work

Deep learning models have significantly enhanced the accuracy of agricultural crop classification systems, which play a vital role in crop yield estimation and food security assurance. Depending on the input data type, various deep learning architectures can be employed for crop classification tasks.

One prominent approach utilizes pixel-based data, where either all pixels of a field or their averaged values can be processed. Among these, CNNs have emerged as particularly effective models. CNNs are widely applied in remote sensing applications, including land cover classification.

CNN models operate by applying convolutional operations across temporal, spectral, and spatial dimensions, or combinations thereof. Notably, the integration of temporal and spectral dimensions has proven particularly effective, leading to improved performance in agricultural crop classification [9]. Consequently, convolutional operations can be applied to either all individual pixels or their averaged values along the relevant temporal and spectral dimensions.

Processing all pixels significantly increases computational load and processing time, despite many pixels within a given field exhibiting similar spectral behavior. Conversely, using only a single averaged pixel substantially reduces the training dataset volume. An optimal solution involves considering a representative subset of pixels from each field, necessitating pixel-block data structures. This approach employs limited pixel samples per block to maintain computational efficiency while preserving sufficient training data volume for effective network training. For such pixel-block data, the PSE+TAE architecture was proposed in [10]. The model comprises two key components:

1. The Pixel-Set Encoder (PSE): A multilayer perceptron (MLP) architecture that computes learned statistical descriptors from the spectral distribution of pixel-block observations.

2. The Temporal Attention Encoder (TAE): A transformer-based network that processes the spatially-enhanced features extracted by the PSE while incorporating temporal information.

This integrated architecture combines the PSE's spatial feature extraction with the TAE's temporal processing capabilities, ultimately constructing an optimized spatio-temporal classification model.

Deep learning models require massive amounts of labeled data for training, yet many countries (particularly developing nations) face limited access to such data, with their acquisition being prohibitively expensive.

To address data scarcity in agricultural crop classification, the transfer learning method was proposed in reference [4] [6][11][12]. This method utilizes a TransformerEncoder model pre-trained on the large-scale BreizhCrops dataset, subsequently fine-tuned on the smaller Vojvodina dataset, effectively transferring knowledge from data-rich to data-scarce regions. The study evaluates three transfer learning strategies:

- Feature Extraction (FE): Only unfreezes the output layer
- Partial Fine-Tuning (PFT): Freezes the input layer while training subsequent layers
- Full Fine-Tuning (FFT): Unfreezes all layers

Results demonstrate PFT's superior performance, indicating that preserving generic features from BreizhCrops in early layers while learning Vojvodina-specific patterns in deeper layers significantly enhances model accuracy. In Reference [2] further is examined how the number of frozen layers impacts transfer learning efficacy.

Recent studies have explored both inductive and sample-free transfer learning strategies for crop classification [13][14]. Building on these approaches, our study focuses on a multi-source data fusion framework combined with partial fine-tuning to tackle the challenges posed by limited labeled data.

These findings confirm the robustness of transfer learning techniques under data-scarce and temporally varying agricultural conditions, which aligns closely with the objectives of this study.

To address the data scarcity issue, in [7], high-resolution drone imagery was employed as input data for agricultural crop classification using a transfer learning approach. Pre-trained CNNs such as VGG16 and GoogLeNet, initially trained on ImageNet, were utilized. Leveraging transfer learning techniques, these models achieved classification accuracies of 90% in Mozambique and 83% in Malawi for diverse crop types. The results demonstrate that the number of frozen layers during transfer learning significantly impacts model performance. Similarly, studies [1], [3], and [8] have investigated the influence of transfer learning on the performance of crop field classification models.

Previous research has demonstrated that deep learning models achieve strong performance when trained on large-scale data. However, these models often underperform in data-scarce regions, necessitating an investigation into the impact of transfer learning to address this limitation. Nevertheless, existing studies have not comprehensively evaluated the effect of transfer learning on both CNN-based and transformer-based models using a unified benchmark dataset. Furthermore, the precise amount of training data required to achieve satisfactory performance remains underexplored. To bridge these gaps, this paper conducts a systematic analysis of transfer learning's influence on deep learning model performance under data-deficient conditions and quantitatively determines the minimum data volume required to attain acceptable accuracy thresholds.

3. Materials

3.1 Study Area

This study focuses on the provinces of Kermanshah and Markazi as target regions. Given the high costs associated with data collection in these target provinces, pre-training data from Hamedan, Ilam, Qazvin, and Kurdistan provinces (referred to as source provinces) were utilized. The significant landscape diversity across these regions stemming from variations in topography, climate, and agricultural practices has increased the complexity of the study.

To enhance classification accuracy, satellite data providing geometric information were employed. Care was taken to select source provinces that closely share climatic and topographic similarities with the target provinces while offering more accessible and cost-effective data collection opportunities (Figure 1).

3.2 Ground truth data

Iran cultivates a diverse range of agricultural products, with certain crops classified as strategic crops (including wheat, barley, sugar beet, canola, alfalfa, and tomato). Accurate statistical data on these crops is essential for precise agricultural policymaking. In this study, these crops constitute the main classification categories. Complementing the primary classes, others categories include watermelon, potato, rangeland, fallow land, vegetables, garlic, others, orchards, plowed fields, uncultivated land, peas, and corn. Each agricultural field is divided into 9×9 pixel-blocks, which serve as individual samples for analysis. The dataset utilized in this study comprises:

- 44,561 samples of main crops
- 43,042 samples of other crops

Table 1 displays the distribution of main and other crops across both data groups. Furthermore, the source-to-target province data ratio exhibits variability ranging from 2:1 to 7:1.

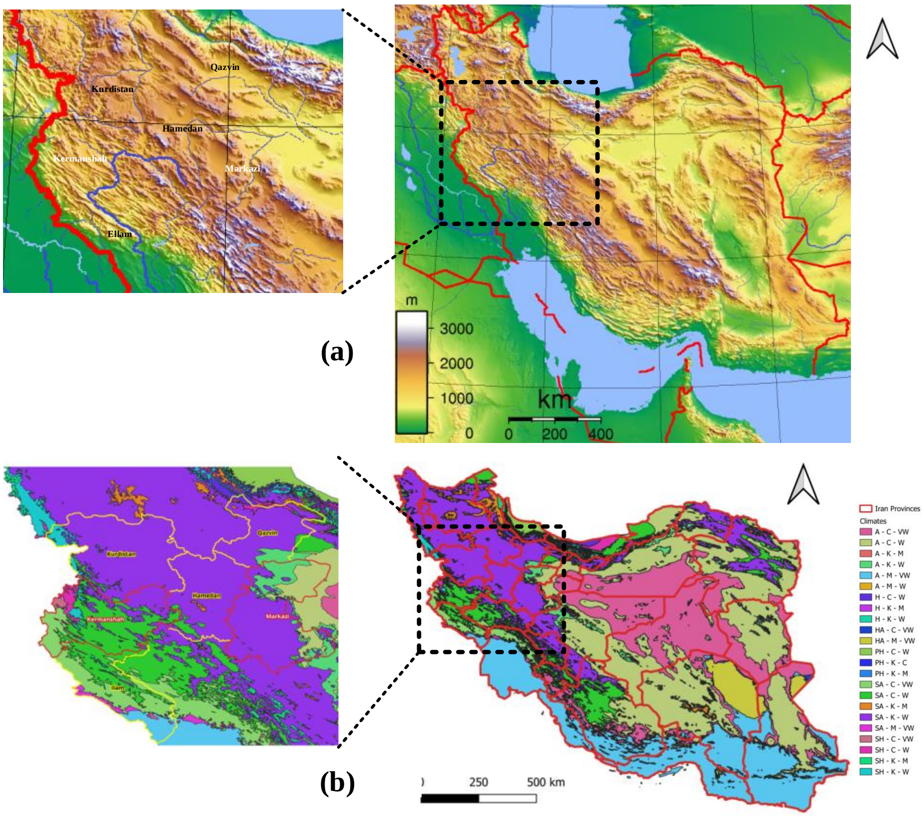

Figure. 1. Geometric information of Iran, a) topographic map b) climatic map

Table 1. Data distribution according to main and other crops

Class	***Label***	***Number of Sample***		***Pretrain-to-Retrain Ratio***
		Retrain	***Pretrain***	
Main	Wheat and barley	8202	31110	3.79
	Alfalfa	793	3394	4.28
	Sugar beet	55	419	7.62
	Rapeseed	156	383	2.46
	Tomato	12	37	3.08
Others	Barren	3233	1144	3.94
	Plow	2757	12840	
	Fallow	1204	5832	
	Garden	600	1707	
	Vegetable	411	1017	
	Pasture	190	3962	
	Other	161	1238	
	Garlic	61	497	
	Corn	57	555	
	Peas	13	1382	
	Watermelon	21	0	
	Potato	0	4160	

3.3 Data Path and Satellite Image Time Series

This section describes the data preparation methodology for different models, as illustrated in Figure 2.

The ground truth data was collected via GPS, with polygons delineated using Google satellite imagery. These polygons are the main sources for determining the boundaries of the desired lands so that information such as satellite data can be collected for them.

This study utilizes satellite data from Sentinel-1 (radar) and Sentinel-2 (optical) with a spatial resolution of 10 meters. The Sentinel-2 imagery provides valuable information about vegetation cover and soil type, with a 5-day revisit frequency [5] and [15]. However, these images are not usable under cloudy conditions. Conversely, Sentinel-1 radar imagery remains reliable in all weather conditions and provides complementary data to enhance result accuracy [16]. For this analysis, specific bands (as indicated in Figure 2) from both satellites were employed.

Iran exhibits diverse climatic conditions. To account for this variability, land surface temperature (LST) data (day/night) from the Terra satellite (MODIS sensor) were utilized [17] ,[18].

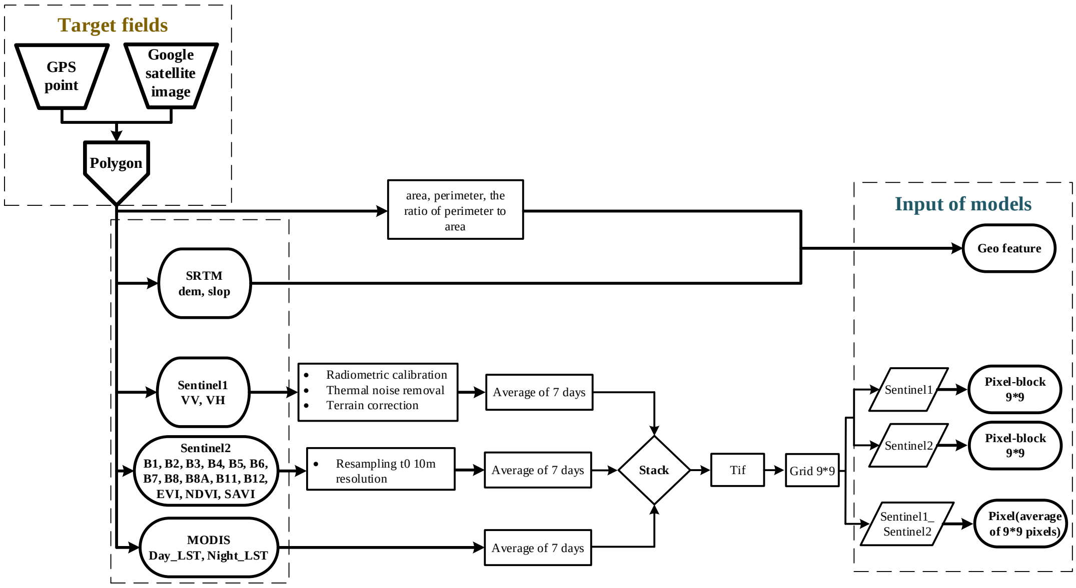

Figure. 2. Data Characteristics and Preparation Methodology for Each Model

Furthermore, different regions of Iran feature varied topographic characteristics. To incorporate these variations, terrain parameters including slope and elevation derived from SRTM (Shuttle Radar Topography Mission) data were employed [19].

Given that the maximum update frequency for the examined satellite data is less than 7 days, all satellite data were acquired within a 7-day window and subsequently averaged to represent the 7-day period. These data were collected over a one-year timeframe (from November 8, 2022 to November 21, 2023) across 55 distinct 7-day periods to comprehensively cover the entire crop growth cycle.

The multi-source satellite data, after preprocessing and integration, are converted into TIFF format images. These images are then partitioned into 9×9 pixel-blocks (gridding). Any block segments falling outside the designated field polygons are masked and excluded from computations.

Missing data values are filled with -1. At this stage, the dataset is prepared and requires model-specific adaptations to match each algorithm's input requirements.

In the subsequent processing stage, for models requiring pixel-block inputs, the Sentinel-1 and Sentinel-2 data are first separated (note that Terra satellite data has been integrated with these datasets), and dedicated pixel-blocks are created for each satellite's data. For models designed with single-pixel input requirements, the mean values of pixels within each block are computed and utilized as model inputs.

Since geometric features for a given polygon remain relatively constant throughout the year, these parameters are extracted once and reused for all temporal analyses.

4. Methodology

In this study, we employ transfer learning techniques to address the challenge of data scarcity in provinces where acquiring training data is costly and difficult. The methodology section first describes various models based on both pixel and pixel-block input architectures. We then detail transfer learning techniques, including optimal parameter freezing strategies to achieve high performance. To determine the minimum required data volume for target provinces, models are fine-tuned using varying amounts of provincial data. Finally, multiple evaluation metrics are presented to compare model performance across different scenarios.

4.1 Models

The previous section detailed the data preparation pipeline for various models. This section now categorizes and describes the models based on their input type, dividing them into pixel-based and pixel-block-based models.

Pixel-based models

The models considered for pixel-based input utilize CNN architectures, specifically including 2D-CNN and CNN Autoencoder, each of which is described in detail below.

a) CNN2D

Accurate crop classification requires simultaneous utilization of spectral and temporal features in remote sensing data. To address this challenge, in [9] a hybrid deep learning model based on 2D-CNN architecture that concurrently incorporates both spectral and temporal guidance to enhance classification accuracy proposed. This architecture employs two-dimensional convolutional filters across both dimensions, enabling extraction of spatio-temporal features essential for agricultural pattern recognition.

b) Autoencoder CNN

In [20], a hybrid architecture was used to classify agricultural products, which combines the strengths of several deep learning models. CNNs have proven highly effective in classification tasks due to their ability to extract temporal and spectral features, while autoencoders are recognized for their capacity to derive meaningful representations through unsupervised learning. By integrating these architectures, the proposed model in [20], enhances both feature extraction and classification performance. This hybrid approach not only increases accuracy, but also has high robustness in the face of changes in crop pattern extraction, making it suitable for real-world applications in the agricultural field.

Pixel-block-based models

The models considered for pixel-block input are based on the Pixel-Set Encoder - Temporal Attention Encoder (PSE-TAE) deep learning architecture proposed in [10]. This architecture consists of two main parts: PSE for considering spatial features and TAE for considering temporal features, which are described below.

PSE: In deep learning applications, CNNs are typically employed for extracting textural features from high-resolution imagery. However, these discriminative features become less discernible in medium-to-low resolution images. To address this limitation, the PSE module is utilized. This encoder computes statistical descriptors from randomly sampled pixel sets through Multilayer Perceptron (MLP) networks, effectively extracting spectral characteristics for each agricultural parcel.

TAE: The TAE model is built upon a self-attention mechanism that provides comprehensive data representation by analyzing relationships across different positions in a temporal sequence. The architecture employs a positional encoder utilizing sinusoidal and cosinusoidal functions to inject positional information into the PSE-derived inputs before processing through the TAE module. A multi-head attention mechanism then enables simultaneous focus on diverse temporal positions, allowing the model to better capture complex inter-temporal relationships. Finally, the TAE outputs are fed into an MLP network for classification.

The model discussed in [10]. exclusively utilizes Sentinel-2 satellite data. However, given the limitations posed by adverse weather conditions (e.g., clouds and fog), incorporating Sentinel-1 data becomes essential. In [21], in order to solve this problem, both Sentinel-1 and Sentinel-2 satellites were used and, considering these two satellites, several different methods of entering data into the model were adopted. In this study, TAE fusion and late fusion methods are used.

As mentioned, Iran has different topographic and climatic conditions, this issue becomes even more important in transfer learning. Therefore, it is necessary to use TERRA and SRTM as well. Therefore, in this article, this information is added to the aforementioned combinations to finally introduce the three desired combinations. Each combination is explained below.

a) PSE fusion

In this fusion approach, the two selected bands from Terra are first combined with Sentinel-2 bands. The resulting composite data is then fed as input to the primary PSE module. Simultaneously, the Sentinel-1 bands are processed through a separate PSE module. These two parallel PSE streams are subsequently merged, with geometric features incorporated before being passed to the TAE module for temporal analysis.

b) TAE fusion

In this configuration (similar to PSE fusion), two separate PSE modules are implemented. However, unlike the previous approach, these modules remain distinct and each feeds into an independent TAE module. The outputs from both TAE modules are then merged and passed to the MLP classifier for final crop classification.

c) Late fusion

In the final fusion approach, fusion occurs after the MLP layer, where each branch independently generates its prediction. Then, the average of these predictions is considered as the final output of the model.

4.2 Evaluation Metrics

This study employs standard evaluation metrics, including Overall Accuracy (OA) and Cohen's Kappa coefficient [22], to assess model performance. OA quantifies the ratio of correct predictions to total predictions [5]. However, it may yield misleading results under class imbalance. Cohen's Kappa coefficient serves as a more robust measure by evaluating agreement between ground truth and predicted labels while accounting for class distribution [23].

The Cohen's Kappa metric is calculated using a confusion matrix and is primarily used for multi-class classification algorithms. As shown in Equation (1), it consists of two parameters: Pr(a) and Pr(e). Pr(a) represents the observed agreement, obtained by dividing the number of true positives (TP) by the total number of instances. Pr(e) denotes the chance agreement, which is calculated separately for each class and then summed. The chance agreement for each class is derived by multiplying the actual probability (the proportion of instances in that class in the ground truth) by the predicted probability (the proportion of instances in that class in the predictions).

$$K = \frac{P_{r(a)} - P_{r(e)}}{1 - P_{r(e)}} \qquad (1)$$

4.3 Evaluation Scenarios

To evaluate the performance of the models and the impact of data volume, we have designed various scenarios. These scenarios aim to investigate the effect of transfer learning for areas where data collection is expensive and to determine the minimum data required to achieve acceptable performance of pixel-based and pixel-block-based models. These scenarios are shown in Figure 3 and described below.

Training and Evaluation without Using Transfer Learning (Scenario 1)

In this scenario, only the target provinces' datasets are used for training and evaluation. This scenario helps us examine the effect of limited data on different deep learning models.

Implementation of Transfer Learning (Scenario 2)

In this scenario, pre-training is first performed using data from the reference provinces. Then, by applying transfer learning techniques, each model is frozen at two levels and retrained using data from the target provinces. At level 1, the initial parameters of the models are frozen, while at level 2, most parameters are frozen and only the final parameters remain trainable. The number of frozen parameters for each model at each level is shown in Table 2. The purpose of this scenario is to investigate the effect of transfer learning for regions suffering from data scarcity and to determine the appropriate number of parameters to freeze in order to achieve optimal performance.

Minimum Required Dataset using Transfer Learning (Scenario 3)

In this scenario, pre-training is first performed using data from the reference provinces. Then, by gradually reducing the target province's data (by 10% at each step) and retraining the model, we aim to determine the minimum data volume required for model training while maintaining acceptable performance. For this purpose, we use the model that demonstrated superior performance compared to other models in the previous scenario. Additionally, to further investigate the impact of transfer learning, models are trained and evaluated without employing transfer learning, using only the progressively reduced target province data at each stage.

4.4 Experimental Setup

To ensure robust evaluation and prevent overfitting, we adopted the following experimental protocol. The dataset was split into training (80%), validation (10%), and test (10%) sets, ensuring stratification by both crop type and geographic region to maintain representativeness. A 5-fold cross-validation strategy was employed on the training set to further verify model stability. During training, we monitored the performance on the validation set and implemented early stopping with a patience of 15 epochs based on the validation loss. The optimization process was guided by tracking both overall accuracy and Cohen's Kappa coefficient on the validation set, with the final performance reported on the held-out test set.

All models were trained for 100 epochs with a batch size of 256. We used the NAdam optimizer [Citation, if any] with a fixed learning rate of 0.001. For the pixel-block-based models (PSE, TAE, Late Fusion), 40 pixels were randomly sampled from each agricultural field to form the input. The detailed architectures were as follows:

- ***PSE-TAE Models:*** The Pixel-Set Encoder (PSE) utilized MLP layers of [17, 32, 64] and [135, 128]. The Temporal Attention Encoder (TAE) employed 4 attention heads with a key dimension of 32, followed by MLP layers [512, 128, 128] and a dropout rate of 0.2. The final classifier consisted of a dense layer outputting 17 units, corresponding to the crop classes.
- ***Autoencoder CNN:*** This model took input of shape (55, 26). The encoder comprised two 1D convolutional layers (64 and 32 filters) with batch normalization, followed by global average pooling. The bottleneck and decoder mirrored this structure in reverse. The final output layer had 17 units.
- ***2D-CNN (TempCNN2D):*** The input shape was (55, 26, 1). The architecture consisted of three consecutive blocks of Conv2D, BatchNormalization, Dropout, and MaxPooling2D layers with 64, 128, and 128 filters, respectively. The features were then flattened and passed through dense layers (512, 128) before the final 17-unit classification layer.

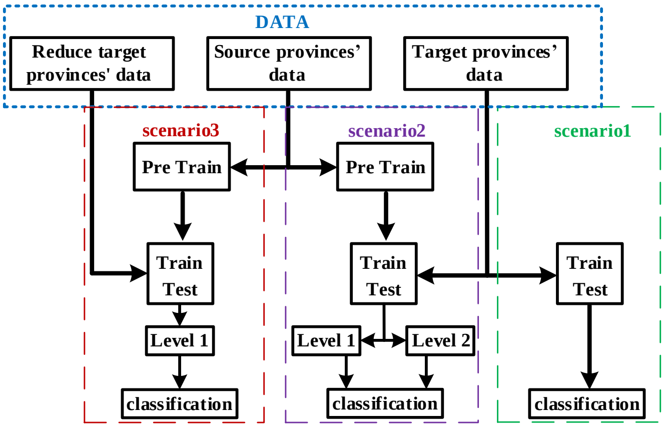

Figure. 3. Evaluation scenarios

Table 2. Number of frozen parameters for each model at each level

	Level 1		*Level 2*		
	Trainable	*Non-Trainable*	*Trainable*	*Non-Trainable*	*Total*
Pse fusion	777997	40672	19413	799256	818669
Tae fusion	299733	40672	19413	320992	340405
Late fusion	302762	40672	22442	320992	343434
cnn_2d	289680	2432	67984	224128	292112
autoencoder_cnn	74090	11808	11850	74048	85898

5. Results

This section consists of three main parts. The first part examines the results of training and evaluating the target models using the target provinces' datasets. Subsequently, transfer learning was employed to enhance performance, with the corresponding results detailed in the second part. The final part analyzes the outcomes obtained by progressively reducing the retraining dataset while maintaining the use of transfer learning.

5.1 Training and Evaluation without Transfer Learning

The classification results obtained without using transfer learning are presented in Table 3. The accuracy of pixel-based models is higher than that of pixel-block-based models, a superiority that is also reflected in Cohen's Kappa coefficient. The lowest accuracy and Cohen's Kappa coefficient belong to the late fusion model, while the highest accuracy and Cohen's Kappa coefficient are achieved by the 2D-CNN model, with values of 93.75 and 88.63, respectively.

5.2 Implementation of Transfer Learning

To enhance classification performance, we employed transfer learning, with the results presented in Table 4.

The findings from this scenario demonstrate that freezing the model parameters at Level 1 yields superior outcomes. The highest accuracy achieved at Level 2 was 96.37% (TAE fusion model), which only surpassed the Autoencoder_CNN model when compared to Level 1 accuracies. Across both levels, pixel-block-based models consistently outperformed pixel-based models. This distinction was particularly evident at Level 1, where all pixel-block-based models achieved higher accuracy than their pixel-based counterparts. The optimal performance was attained by the TAE fusion model at Level 1, achieving near 98% accuracy. Notably, all other pixel-block-based models also exceeded 97% accuracy. The Cohen's Kappa coefficient mirrored the accuracy trends across all models - higher accuracy models correspondingly demonstrated higher Cohen's Kappa coefficient values.

Given the PSE fusion model's good performance, particularly at level 1, we employed this model with level 1 to determine the minimum required data using transfer learning.

5.3 Minimum data requirements using transfer learning

In the final scenario, the target province datasets were progressively reduced by 10% at each stage. Training and evaluation were then conducted both with and without transfer learning, with the resulting accuracy and Cohen's Kappa coefficient metrics shown in Figures 4 and 5, respectively. As mentioned, the PSE fusion model was used in this scenario, employing Level 1 transfer learning. The results show that both accuracy and Cohen's Kappa coefficient metrics followed similar trends across all models.

When transfer learning was not used, accuracy decreased exponentially, from 92.7% with all dataset to 79% when 70% of data was removed (a 13.7% decrease). In contrast, when transfer learning was applied, accuracy declined only from 97.6% to 91.9% (just 5.7% decrease) under the same 70% data reduction.

Notably, while all dataset without transfer learning achieved 92.7% accuracy, using transfer learning with 60% less data still yielded higher accuracy (93.5%). This clearly demonstrates the significant impact of transfer learning.

5.4 Model Complexity and Efficiency

We analyzed the computational complexity of the models based on their parameter counts (Table 2). The PSE fusion model was the most complex with 818,669 parameters, followed by the TAE fusion (340,405) and Late fusion (343,434) models. In contrast, the pixel-based models were more lightweight, with the 2D-CNN having 292,112 parameters and the Autoencoder CNN being the most efficient with only 85,898 parameters. This higher parameter count directly correlated with increased computational demand and longer training times observed for the pixel-block-based models compared to their pixel-based counterparts.

Table 3. Model accuracy and Cohen's Kappa coefficient without transfer learning

	Pse fusion	*Tae fusion*	*Late fusion*	*cnn_2d*	*autoencoder_cnn*
Accuracy	92.71	92.94	91.52	93.75	93.41
Cohens's Kappa coefficient	86.89	87.28	84.71	88.63	88

Table 4. Model accuracy and Cohen's Kappa coefficient with transfer learning

	Pse fusion		*Tae fusion*		*Late fusion*		*cnn_2d*		*autoencoder_cnn*	
	Level 1	***Level 2***	***Level 1***	***Level 2***	***Level 1***	***Level 2***	***Level 1***	***Level 2***	***Level 1***	***Level 2***
Accuracy	97.6	96.34	97.86	96.37	97.34	93.24	96.4	94.5	94.84	93.78
Cohen's Kappa coefficient	95.66	93.42	96.16	93.46	95.21	87.85	93.5	90	90.86	88.76

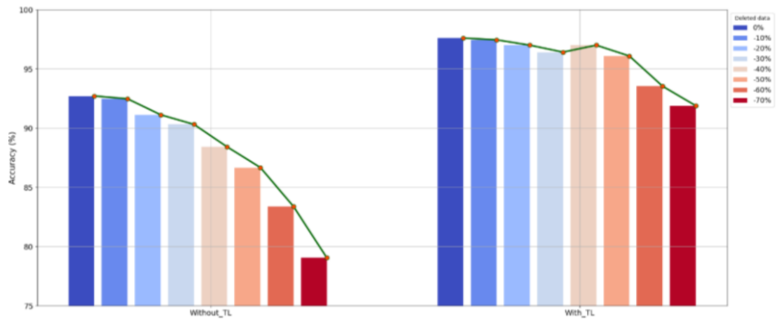

Figure. 4. Accuracy of PSE fusion model with and without transfer learning under target province dataset reduction

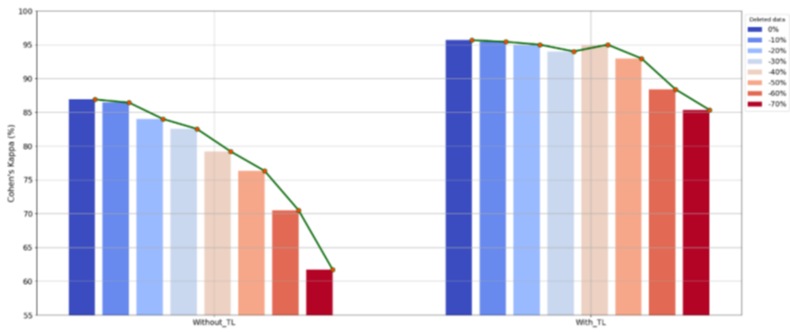

Figure. 5. Cohen's Kappa coefficient of PSE fusion model with and without transfer learning under target province dataset reduction

6. Discussion

Agricultural crop classification in Iran presents significant challenges, including the scarcity of labeled datasets, the high costs associated with data collection in certain regions, and the considerable topographic and climatic variations across different areas. To address these issues, this research utilized transfer learning and incorporated geometric information to improve the performance of classification models. Additionally, we examined the minimum required training data to maintain acceptable model accuracy while minimizing sampling efforts in high-cost regions.

The accuracy and Cohen's Kappa coefficient metrics for models with and without transfer learning are presented in Tables 3 and 4. To further evaluate the impact of transfer learning in target provinces, these assessment metrics are collectively illustrated in Figures 6 and 7 for the specified scenarios.

According to Figure 6, when transfer learning is not employed, the accuracy of pixel-block-based models is lower than that of pixel-based models. However, with the application of transfer learning, particularly at Level 1, the accuracy of pixel-block-based models shows a significant improvement, generally surpassing that of pixel-based models. For instance, at Level 1, the accuracy of pixel-block-based models increases by an average of 5.2% compared to the scenario without transfer learning, while the increase for pixel-based models is only 2%.

A key factor contributing to this disparity is the number of model parameters. As shown in Table 2, pixel-block models have a substantially higher parameter count compared to pixel-based models, necessitating larger training datasets for effective learning. Without transfer learning, pixel-block-based models fail to train adequately, resulting in lower accuracy. Transfer learning addresses this by enabling these models to be pre-trained on sufficient data from source provinces and fine-tuned on target province data, leading to a marked accuracy improvement.

Model performance is superior at Level 1 compared to Level 2. One reason for this is the dataset size requirement: at Level 2, a greater number of components are frozen, demanding an exceptionally large pre-training dataset to properly train the frozen parameters. Given the limited volume of source province data used for pre-training, freezing fewer components (Level 1) yields better results.

For future work, the volume of pre-training data can be increased, and based on the extent of this increase, a greater number of layers can be frozen to evaluate the system's performance. Additionally, in this study, layer freezing was implemented at two levels; expanding the number of freezing levels could help identify the optimal number of layers to freeze.

The discrepancy between accuracy metrics and Cohen's Kappa coefficient, when comparing scenarios with and without transfer learning while reducing the volume of target province datasets, is illustrated in Figure 8. As expected, both metrics exhibit similar trends. According to Figure 8, it can

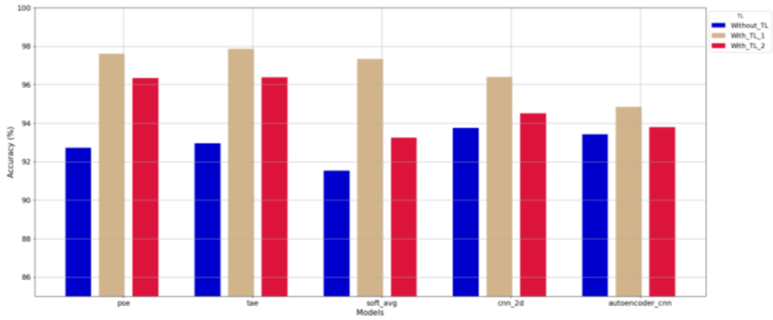

Figure. 6. Accuracy of models with and without transfer learning

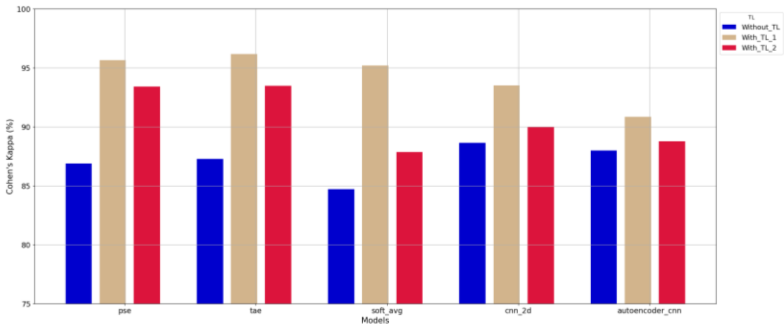

Figure. 7. Cohen's Kappa coefficient of models with and without transfer learning

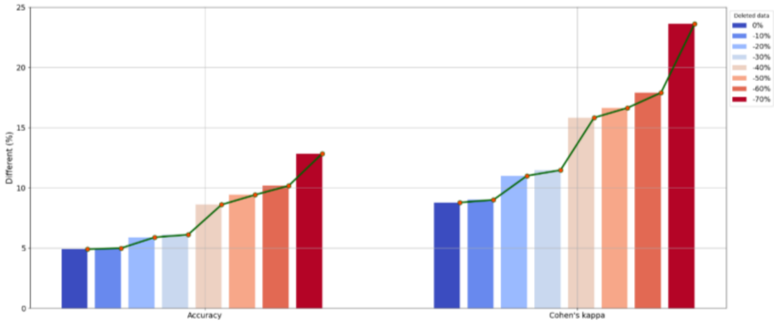

Figure. 8. Differences in accuracy and Cohen's Kappa coefficient metrics based on transfer learning usage and reduction of target province dataset volume

be concluded that the impact of transfer learning is more pronounced when the target province datasets undergo more significant reductions.

When utilizing the complete dataset from target provinces with transfer learning, Cohen's Kappa coefficient shows an 8.7% improvement. However, with a 70% reduction in target province data, this metric demonstrates a 23.6% enhancement compared to scenarios without transfer learning.

These findings are consistent with recent studies emphasizing the robustness of transfer learning under severe data scarcity [11], [24], where comparable performance levels were achieved even with substantial reductions in labeled data availability.

7. Conclusions

In this study, we addressed the challenge of limited training data availability in certain regions of Iran where data collection is costly and difficult by employing various deep learning models combined with transfer learning techniques. To comprehensively evaluate the impact of transfer learning, we utilized CNN-based models for pixel-based data analysis and transformer-based networks for pixel-block data analysis. Sentinel-1 and Sentinel-2 satellite data served as primary model inputs, supplemented with additional geospatial features including land surface temperature, elevation, and slope to account for Iran's diverse climatic and topographic conditions.

This study designed various scenarios to examine the impact of transfer learning on agricultural crop classification. The results demonstrated that employing transfer learning enhanced classification performance across all scenarios. This improvement was particularly evident in pixel-block-based models, as their higher parameter count benefited significantly from pre-training with source province datasets, which effectively optimized their weight initialization. Furthermore, the effect of transfer learning was evaluated under progressive reduction of target province training data, revealing its exceptional efficacy in data-scarce regions. For instance, with a 70% reduction in training data, Cohen's Kappa coefficient decreased by 25% without transfer learning, whereas the decrease was only 10% when transfer learning was applied. These findings strongly confirm that transfer learning serves as an effective solution for improving model accuracy in data-limited regions.

Future studies may leverage meta-learning and time-series transformer variants such as SITSMamba [25], [26] to further enhance model adaptability.

Declaration

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contributions

ES: Simulating, training and testing the models, Writing the main sections of the paper (Introduction, Proposed dataset and Methods, Results, Discussion)

SZ: Simulating, training and testing the models, Writing the main sections of the paper (Proposed dataset and Methods, Results)

MA: Literature review and gathering references; Writing Abstract, Introduction and Conclusion; Final editing of the paper for language and structure

PA: Data analysis and preparation of tables and figures; Reviewing and editing the paper for technical and scientific accuracy, Contribution in writing specialized sections (Proposed dataset and Methods, Results, Discussion)

Conflict of interest

The authors have no competing interests that might be perceived to influence the results and discussion reported in this paper.

Acknowledgements

This research has been supported by Iran Telecommunication Research Center (ITRC), and we are thankful to them for their

encouragements and support to conduct this research.

References

[1] Y. Hu *et al.*, "An Interannual Transfer Learning Approach for Crop Classification in the Hetao Irrigation District, China," *Remote Sensing*, vol. 14, p. 1208, 03/01 2022.

[2] H.-W. Jo, A. Koukos, V. Sitokonstantinou, W.-K. Lee, and C. Kontoes, "Towards global crop maps with transfer learning," *arXiv preprint arXiv:2211.04755*, 2022.

[3] K. K. Gadiraju and R. R. Vatsavai, "Remote Sensing Based Crop Type Classification Via Deep Transfer Learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 4699-4712, 2023.

[4] O. Antonijevic, S. Jelić, B. Bajat, and M. Kilibarda, "Transfer learning approach based on satellite image time series for the crop classification problem," *Journal of Big Data*, vol. 10, 04/29 2023.

[5] M. Račič, K. Oštir, A. Zupanc, and L. Čehovin Zajc, "Multi-Year Time Series Transfer Learning: Application of Early Crop Classification," *Remote Sensing*, vol. 16, p. 270, 01/10 2024.

[6] V. Barrière, M. Claverie,M. Schneider, G. Lemoine, R. Andrimont, “Boosting Crop Classification by Hierarchically Fusing Satellite Modalities Across Seasons and Regions,” Remote Sensing of Environment, vol. 305, p.114110., 2024.

[7] A. Nowakowski *et al.*, "Crop type mapping by using transfer learning," *International Journal of Applied Earth Observation and Geoinformation*, vol. 98, p. 102313, 06/01 2021.

[8] P. Hao, L. Di, C. Zhang, and G. Liying, "Transfer Learning for Crop classification with Cropland Data Layer data (CDL) as training samples ",*Science of The Total Environment*, vol. 733, p. 138869, 04/01 2020.

[9] C. Pelletier, G. I. Webb, and F. Petitjean, "Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series," *Remote Sensing*, vol. 11, no. 5, p. 523, 2019.

[10] V. S. F. Garnot, L. Landrieu, S. Giordano, and N. Chehata, "Satellite image time series classification with pixel-set encoders and temporal self-attention," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12325-12334.

[11] J. Mai, Q. Feng, Sh. Fu, R. Wang, “Enhancing Crop Type Mapping in Data-Scarce Regions Through Transfer Learning: A Case Study of the Hexi Corridor,” Remote Sensing, vol. 17, no. 9, p. 1494, 2025.

[12] H. Hoppe et al., “Transferability of Machine Learning Models for Crop Classification,” Remote Sensing, vol. 16, no. 9, p. 1493, 2024.

[13] X. Guo et al., “Fine Classification of Crops Based on an Inductive Transfer Learning Approach,” GIScience & Remote Sensing, 61, no. 1, p. 2319939, 2024.

[14] H. Xue, Y. Fan, G. Dong, S. He, Y. Lian, and W. Luan, "Crop classification in the middle reaches of the Hei River based on model transfer," Scientific Reports, vol. 14, no. 1, Art. 28964, Nov. 22, 2024.

[15] I. Gallo, L. Ranghetti, N. Landro, R. La Grassa, and M. Boschetti, "In-season and dynamic crop mapping using 3D convolution neural networks and sentinel-2 time series," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 195, pp. 335-352, 2023/01/01/ 2023.

[16] J. Yao, J. Wu, C. Xiao, Z. Zhang, and J .Li, "The Classification Method Study of Crops Remote Sensing with Deep Learning, Machine Learning, and Google Earth Engine," *Remote Sensing*, vol. 14, no. 12, p. 2758, 2022.

[17] O. E. Adeyeri *et al.*, "Land surface dynamics and meteorological forcings modulate land surface temperature characteristics," *Sustainable Cities and Society*, vol. 101, p. 105072, 2024/02/01/ 2024.

[18] M. O. Román *et al.*, "Continuity between NASA MODIS Collection 6.1 and VIIRS Collection 2 land products," *Remote Sensing of Environment*, vol. 302, p. 113963, 2024/03/01/ 2024.

[19] J. Liu, K. Yang, A. Tariq, L. Lu, W. Soufan, and A. El Sabagh, "Interaction of climate, topography and soil properties with cropland and cropping pattern using remote sensing data and machine learning methods," *The Egyptian Journal of Remote Sensing and Space Sciences*, vol. 26, no. 3, pp. 415-426, 2023/12/01/ 2023.

[20] Aashish, A. Thakkar, S. Yadav, S. Saini, and K. Lata, "CNN and Autoencoders based Hybrid Deep Learning Model for Crop Disease Detection ",in *2022 International Conference on Connected Systems & Intelligence (CSI)*, 2022, pp. 1-7.

[21] S. Ofori-Ampofo, C. Pelletier, and S. Lang, "Crop Type Mapping from Optical and Radar Time Series Using Attention-Based Deep Learning," *Remote Sensing*, vol. 13 , no. 22, p. 4668, 2021.

[22] X. Zhou, J. Wang, B. Shan, and Y. He, "Early-Season Crop Classification Based on Local Window Attention Transformer with Time-Series RCM and Sentinel-1," *Remote Sensing*, vol. 16, no. 8, p. 1376, 2024.

[23] D. J. Gallardo-Romero, O. E. Apolo-Apolo, J. Martínez-Guanter, and M. Pérez-Ruiz, "Multilayer Data and Artificial Intelligence for the Delineation of Homogeneous Management Zones in Maize Cultivation," *Remote Sensing*, vol. 15, no. 12, p. 3131, 2023.

[24] A. Mirzaei, H. Bagheri, and I. Khosravi, "Enhancing crop classification accuracy by synthetic SAR-optical data generation using deep learning," *arXiv preprint* arXiv:2402.02121, 2024.

[25] J. Reuss, J. Macdonald, S. Becker, K. Schultka, L. Richter, and M. Körner, "Meta-learning for few-shot time series crop type classification: A benchmark on the EuroCropsML dataset," arXiv preprint arXiv:2504.11022, Apr. 2025. [Online]. Available: https://doi.org/10.48550/arXiv.2504.11022.

[26] X. Qin, X. Su, and L. Zhang, "SITSMamba for crop classification based on satellite image time series," arXiv preprint arXiv:2409.09673, Sep. 15, 2024. [Online]. Available: https://arxiv.org/abs/2409.09673.

Erfan Shakouri is a Ph.D. candidate in Electrical and Electronics Engineering at Shahed University, Tehran, Iran, where his research focuses on in-memory computing, with a particular emphasis on the implementation of artificial intelligence algorithms in

memory. He is also a researcher at the AI Laboratory of the ICT Research Institute, where he works on the development and evaluation of AI models. His research interests include image processing and time-series modeling for agricultural monitoring and crop area estimation, as well as the application of transfer learning techniques in data-scarce environments.

Saeed Zare is a researcher specializing in remote sensing and machine learning. Earned a Bachelor's degree in Geomatics from Tabriz University and a Master's degree in Remote Sensing from Khajeh Nasir al-Din Tusi University. Research interests include classification, machine learning, remote sensing, crop type mapping, and time-series analysis. Current work focuses on developing deep learning models for geospatial data analysis, particularly in agricultural monitoring and land cover classification.

Masoumeh Azimzadeh is a researcher and software engineering specialist with extensive expertise in intelligent systems, data analytics, and Persian language processing. She received her B.S. in Software Engineering from Kharazmi University, Tehran, and her M.S. in Software Engineering from Islamic Azad University, South Tehran Branch. She earned her Ph.D. in Software Engineering from Islamic Azad University, Science and Research Branch, where her research focused on IoT service placement in fog computing environments. She is currently a researcher at the Iran Telecommunication Research Center (ITRC), where she has also held managerial positions. Her professional and research activities include leading and contributing to numerous national projects in areas such as native search engines, big data analytics, artificial intelligence, smart agriculture, and digital banking.

Parvin Ahmadi received the degree of B.Sc. and M.Sc. both in Electrical Engineering from Isfahan University of Technology and Ph.D. from Sharif University of Technology, Tehran, Iran. Her research interests are mainly in the areas of signal and image processing, computer vision and pattern recognition. She is now an Assistant Professor at Artificial Intelligence Center, Iran ICT Research Institute.