Deep Transfer Learning based Fusion Model for Environmental Remote Sensing Image Classification Model

Anwer Mustafa Hila, Fahd N. Al-Wesabi, Khalid J Alzahrani, Mesfer Al Duhayyim, Manar Ahmed Hamza, Mohammed Rizwanullah and Vicente García Díaz

*Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam Bin Abdulaziz University, AlKhajr, Saudi Arabia; **Department of Computer Science, King Khalid University, Muhayel Aseer, Saudi Arabia & Faculty of Computer and IT, Sana’a University, Sana’a, Yemen; **Department of Clinical Laboratories Sciences, College of Applied Medical Sciences, Taif University, Taif, Saudi Arabia; **Department of Natural and Applied Sciences, College of Community - AlFaj, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia; **Department of Computer Science, School of Computer Science Engineering, University of Oviedo, Oviedo, Spain

ABSTRACT
Remote-sensing images comprise massive amount of spatial and semantic data that can be employed for several applications. Presently, deep learning (DL) models for RS image processing become a familiar research area. Due to the advancements of recent satellite imaging sensors, the issue of huge amount of data processing becomes a challenging problem. To accomplish this, deep transfer learning (DTL) models are developed to resolve the semantic gap among various datasets. This study develops a new DTL-based fusion model for environmental remote-sensing image classification, called DTLF-ERSIC technique. The proposed technique focuses on the design of fusion model to combine multiple feature vectors and thereby attains maximum classification performance. The DTLF-ERSIC technique incorporates the entropy-based fusion of three feature extraction techniques, namely, Discrete Local Binary Pattern (DLBP), Residual Network (ResNet50), and EfficientNet models. Besides, a rain optimization algorithm (ROA) with fuzzy-rule-based classifier (FRC) is applied to predict the class labels of the test RS images and shows the novelty of the work. A comprehensive experimental analysis of the DTLF-ERSIC technique takes place on benchmark dataset and examined the results in terms of different performance measures. The simulation results reported the supremacy of the DTLF-ERSIC technique over the recent state-of-art techniques.

INTRODUCTION
Satellite images of earth are created using an imaging satellite that might be functioned by the enterprises/governments. This image is taken from remote-sensing (RS) technology and, in general, RS could be represented as the procedure of analyzing and collecting data regarding an event/region, entity without being in physical interaction with it (Zhang et al., 2019). RS information is deliberated as an important source of data for several applications, like land use classification, particularly while incorporated with artificial intelligence technologies. A variety of RS images are now available, for example, synthetic aperture radar (SAR), multi/hyper-spectral, and very high resolution (VHR) images. Therefore, various methods for the study of this type of image are proposed in the previous year including RS image detection, identification, classification, and compression. Land cover classification with RS image plays a significant part in various applications like urban planning, land resource management, environmental protection, and precision agriculture (Dai et al., 2021).

In recent times, the spread of deep learning (DL) approaches has resulted in outstanding performances on RS image classification (Lv et al., 2020). Because of the great accomplishment attained by convolutional neural network (CNN) in the computer vision community, a significant amount of CNN-based methods have been presented for RS classification. This model gives better classification performances when compared to another conventional model (Hagag et al., 2015). But the CNN feature is employed in an insufficient manner through this research, it is due to that the CNN feature only extracted from the fully connected layers is used by the rich hierarchical data from discarded convolution layer. One general solution is to resize the original scene image into a predefined size that inevitably loses efficient discriminative data (Byju et al., 2020). Simultaneously, in the DCNN applications for scene classification of RS image, a huge amount of tagged information is generally needed for training; however, the object is very difficult and different in the RS images, therefore the image using individual labels is smaller when artificial annotation is laborious and time-consuming.
In recent times, it was demonstrated that DL approaches could adaptively learn image features, that is, appropriate for certain scene classification tasks, and attain an effective classification accuracy when compared to conventional scene classification methods (Vakalopoulou et al., 2015). But there exist two main challenges that severely affect the usage of DL methods (Xu et al., 2021): (1) The DL approach requires huge training data for training the module and is also time-consuming. Various studies have illustrated that the present pre-trained CNN methods pre-trained on huge datasets like ImageNet could be transmitted to other detection tasks. Nogueira et al., (2017) have demonstrated that the pre-trained CNN employed as feature-extractor attained greater results when compared to fully training a novel CNN for three RS scenes dataset.

This study develops a new DTL-based fusion model for environmental RS image classification called DTLF-ERSIC technique. The DTLF-ERSIC technique designs an entropy-based fusion of three feature extraction techniques, namely, Discrete Local Binary Pattern (DLBP), Residual Network (ResNet50), and EfficientNet models. Moreover, a rain optimization algorithm (ROA) with fuzzy rule-based classifier (FRC) is applied for the classification of RS images. In order to ensure the betterment of the DTLF-ERSIC technique, an extensive simulation process takes place on benchmark dataset.

The upcoming sections of the study are planned as follows. Section 2 provides a comprehensive-related works; section 3 introduces the proposed model; section 4 offers the experimental validation; and section 5 draws the conclusion.

**Literature review**

Alkhelaiwi et al. (Alkhelaiwi et al., 2021) proposed an effective method that utilizes PPDL-based methods for addressing privacy concerns based on the data from satellite images while employing public DL model. In this work, they presented a partially homomorphic encryption system (a Paillier system), which allows processing of secured data without disclosure of the basic information. This approach attains an effective result while employed to a customized CNN and present TL approaches. Also, the presented encryption system enables to train CNN models on encrypted data directly that require low computation complexity. Tong et al. (Tong et al., 2020) proposed a system for applying deep models attained from labelled land cover datasets to categorize unlabelled HRRS images. The primary objective is based on DNN to present the contextual data contained in distinct kinds of land cover and proposed a pseudo-labelling and sample election system to improve the transferability of deep model. In order to attain pixelwise land cover classifications using the target image, they based mainly on the fine-tuned CNN and developed a hybrid classification by integrating hierarchical segmentation and patchwise classification.

In Lu et al. (X. Lu et al., 2019), a bi-directional adaptive feature fusion approach is explored for handling the RS scene classification. The DL and SIFT features are merged for getting discriminative image presentations. The fused features could define the scene efficiently by using DL features as well as conquer the rotation and scale variability using SIFT features. With global CNN and SIFT features, the presented approach attains an improved scene classification accuracy. Shawky et al. (Shawky et al., 2020) proposed an efficient classification method called CNN-MLP for utilizing the benefit of these two methods: MLP and CNN. The feature is made by the pre-trained CNN without FC layer. Because of constraint training images for each class, the presented method employs data augmentation technique for expanding the training image. Next, an MLP is applied for classifying the last feature map to the determined class.

In Yang et al. (Yang et al., 2018), a classification method DCNN_MSFF is presented on the basis of DCNN and MSFF models. In the first stage, the RS image is converted to attain an amount of distinct scale ones for augmentation. In the second stage, they are inputted to the DCNN for extracting features, and the distinct scale features of the convolution and averagely the FC layers are pooled/encoded. Finally, the processed feature is fused, and the MKSVM model is employed for classifying the scene. Bera and Shrivastava (Bera & Shrivastava, 2020) proposed a spatial feature extraction method with DCNN model for HSI classification. Since the optimizer plays a significant part in learning procedure of DCNN method, they have introduced the effects of seven distinct optimizers on DCNN method in the applications of HSI classification. The seven distinct optimizers employed in this work consist of Adagrad, SGD, RMSprop, Adadelta, AdaMax, Nadam, and Adam.

Liu et al. (Liu et al., 2020) employed unsupervised TL method for the CNN training. They convert similarity learning to deep ordinal classification through various pre-trained CNNs over large-scale labelled day-to-day image set, that together define image similarity and provide pseudo-label for the classification. The presented approach ends up with a novel lightweight method named SBS-CNN model, that is, trained from scratch with fully unlabelled RS image, and the resultant features are compact. A new loss function named weighted Wasserstein ordinal losses are made for considering the ordinal
relationships between classes, therefore efficiently navigate parameters update at the time of training. Li et al. (Li et al., 2020) proposed an HSRRS image scene classification model with TL-DeCNN method in a few shot HSRRS scene sample. Particularly, three common DeCNNs of ResNet50, InceptionV3, and VGG19, trained on the ImageNet2015, the weight of their convolution layers for the TL-DeCNN are transmitted correspondingly.

In Pu et al. (Pu et al., 2021), a new cross attention and graph convolution integration algorithms were introduced. In the initial step, PCA model is employed for reducing the dimension of hyperspectral image to acquire lower dimension features, that is, highly expressive. In the next step, the models use a cross (vertical direction and horizontal direction) attention algorithm for allocating weight. Finally, the produced deep features and the relationships among the deep features are applied for completing the predictions of hyperspectral data. In Liang et al. (Liang et al., 2018), a novel RS image classification algorithm stimulated by DL method is presented, that is, depending on Stacked Denoising AE. In the initial phase, the deep network models are constructed by the stacked layer of Denoising AE. In the second phase, using noised input, the unsupervised Greedy layerwise training algorithms are employed for training every layer sequentially for further robust expressing, features are attained in supervised learning with BPNN, and the entire networks are enhanced using error BP.

The proposed model

In this study, an effective DTLF-ERSIC technique is derived to detect and classify the RS with maximal detection rate. The DTLF-ERSIC technique encompasses different subprocesses namely data augmentation, fusion-based feature extraction, FRC-based classification, and ROA-based parameter optimization. Figure 1 showcases the overall working process of proposed DTLF-ERSIC model. The detailed working of these modules is offered in the following subsections.

Data augmentation

The augmentation methods could be employed for generating different versions of image, that is, improved in model training when generalizing it. Various transformations on the training set are examined in Perez & Wang, (2017). To every input image, there is a generation of additional five images, that is, rotated employs the image transformation equations that rotate all the pixels \((x, y)\) by multiplying transformation matrix using predefined rotate angle \(\theta\) in Eq. (1). This image is vertically flipped, zoomed, and horizontally flipped, also it can be implemented. E.g., for a horizontal flip, the pixel located at coordinate \((x, y)\) would be located at coordinate \((-x, y)\) in the novel image by Eq. (2). In such equations, each \(x\) value is switched to its negative counterpart, whereas the \(y\) value remains unchanged. To the dataset of size \(N\), the presented approach makes dataset with size \(5N\).

Figure 1. Overall process of DTLF-ERSIC model.
Fusion of feature vectors

The DTLF-ERSIC technique incorporates the entropy-based fusion of three feature extraction techniques, namely, DLBP, ResNet50, and EfficientNet models.

\[
R = \begin{bmatrix}
\cos\theta & -\sin\theta \\
\sin\theta & \cos\theta
\end{bmatrix}
\]

\[
y = -f(x)
\]

Overview of DLBP model

LBP operators provide descriptors for each image with the gray level of every pixel. In its standard version, a pixel is deliberated by eight neighbours, therefore making a square of \(3 \times 3\) pixels. After that, for every eight adjacent, their relation with the central pixels was measured: when their grey level is exceeding that of the central pixels, they are substituted with, or else with 0. The resultant binary pattern is later transformed to a decimal value. The LBP operators are measured by:

\[
LBP(x, y_c) = \sum_{p \in P} s(q_t - t_c)
\]
where $P$ denotes a structure that describes the pixel closer to the central one, $i$, and $i$ are correspondingly the grey level of the central pixels and its $p$th neighbors, and $q(z)$ indicates a quantization function determined by:

$$q(z) = \begin{cases} 1, & z \geq 0 \\ 0, & otherwise \end{cases}$$

It is noteworthy that the number of neighbors and the radius are parameters, that is, adapted at will and are not static (Nanni et al., 2020). In DLBP models, it is an adapted version of LBP method. The concept is to discover, per pixel patch, the optimal threshold divides by the pixel inside it. This threshold is attained by minimizing a residual error measured by:

$$\varepsilon(\tau) = \frac{1}{N} \left\{ \sum_{i \in [r(\tau) \leq \tau]} (I(r_i) - \mu_0)^2 + \sum_{i \in [r(\tau) \leq \tau]} (I(r_i) - \mu_1)^2 \right\}$$

(5)

$$\mu_0 = \frac{1}{N_0} \sum_{i \in [r(\tau) \leq \tau]} I(r_i)$$

(6)

$$\mu_1 = \frac{1}{N_1} \sum_{i \in [r(\tau) \leq \tau]} I(r_i)$$

(7)

The residual error $\varepsilon(\tau)$ is later formed as the intraclss variance $\sigma_W^2$. The optimal threshold is the one that maximizes the variances among the classes $\sigma_B^2$ of the pixel in the patch measured by:

$$\text{Optimal threshold} = \sigma - \sigma_W^2$$

(8)

When this threshold is established, the weights of every pixel patch on the last histogram is evaluated as follows:

$$\omega = \sqrt{\frac{\sigma_B^2(f^y)}{\sigma^2 + C}}$$

(9)

where $C$ denotes a constant which serve to manage this case in which $\sigma^2$ is closer to 0 and might create the weight of the vote to vary. Kobayashi’s work is located at 0.012. The features are extracted by (radius, neighbor) = (1, 8), (2, 8), (3, 8).

**ResNet50 model**

ResNet employs the residual block to resolve the gradient disappearance and degradation challenges present in standard CNN. The residual blocks deepen the network depth and enhance the efficiency of the system. Especially, ResNet networks have made great achievements in the ImageNet (Z. Lu et al., 2020) classification competition. The residual block in ResNet implements the residual by including the input and output of the residual blocks. The equation of the residual function can be expressed by:

$$y = F(x, W_{opt}) + x$$

(10)

where $x$ denotes the input of residual block; $W$ indicates the weight of residual block; $y$ signifies the output of the residual block. ResNet network includes various residual blocks where the convolutional kernel size of the convolutional layers is distinct. The classical framework of RestNet consists of ResNet50, ResNet101, and RestNet18. Feature extracted by the ResNet residual network is located in FC layers for image classification. The FRC method is mainly employed for classification.

**EfficientNet model**

An EfficientNet (Bansal et al., 2021) has a kind of CNN that utilizes a unique scaling technique, which evenly scales all dimensions, for instance, width, depth, and resolution adopting compound coefficients. The scaling has been important concerned to one of the data science experts throughout the world. The scaling of network was carried out mostly across three dimensions like width, depth, and resolution. But it has been monitored that scaling initially does improve accuracy; however, the spike in accuracy saturates slowly with scaling. Conversely, standard ConvNets could not carry out scaling “efficiently”. An EfficientNets simplify these issues of scaling by something called Compound Scaling. During this approach, a Compound Coefficient has been utilized for scaling the network uniformly through the three dimensions.

**Entropy-based fusion process**

Data fusion is employed in various ML methods and CV applications. Feature fusion is an important task that integrates several feature vectors. The proposed method is depending on feature fusion using entropy. The three vectors are determined by:

$$f_{\text{ResNet1x8n}} = \{\text{ResNet1x1}, \text{ResNet1x11x2}, \text{ResNet1x11x3}, \ldots, \text{ResNet1x11x8} \}$$

(11)

$$f_{\text{EfficientNet191x8m}} = \{\text{EfficientNet1x1}, \text{EfficientNet1x2}, \text{EfficientNet1x3}, \ldots, \text{EfficientNet1x8} \}$$

(12)

$$f_{\text{LBP1x8p}} = \{\text{LBP1x1}, \text{LBP1x2}, \text{LBP1x3}, \ldots, \text{LBP1x8} \}$$

(13)

Furthermore, extracted feature is fused in an individual vector.

$$\text{Fused feature vector}_{1x8} = \sum_{i=1}^{3} f_{\text{ResNet1x8n}}, f_{\text{EfficientNet1x8m}}, f_{\text{LBP1x8p}}$$

(14)

whereas $f$ denotes fused vector (1x1186). The entropy is used on feature vector for the election of optimum feature according to the score (Saba et al., 2020). The FS
method is arithmetically described in Eqs. (11)–(14). Entropy is applied for selecting 1186 score-based features from 7835 features.

\[
B_{He} = -NH_{E} \sum_{i=1}^{n} p(f_i) \quad (15)
\]

\[
F_{select} = B_{He}(\max(f_i, 1186)) \quad (16)
\]

In Eqs. (15) & (16), p denotes features likelihood and He signifies entropy. The lastly elected feature is provided to the classifier for differentiating the healthy and glioma image.

**Optimal FRC-based classification process**

The fused feature vectors are fed into the RFC model to estimate the proper class labels. Fuzzy classifier belongs to rule-based methods, has considerable benefits based on their performances, in addition to their subsequent and design analyses. An exclusive advantage of fuzzy classifiers is the interpretability of classification rules. Assume that \( x = (x_1, x_2, \ldots, x_D) \in R^D \) is a \( D \) dimension feature space and \( C = \{c_1, c_2, \ldots, c_m\} \) represents a group of class labels. Fuzzy classifiers are provided with a base of production rule of the succeeding forms:

\[ R_k : IF S_1 \times x_1 = A_{1i} AND S_2 \times x_2 = A_{2j} AND \ldots AND S_D \times x_D = A_{Dn} \] \( \text{THEN} \) class \( i \), \( i = 1, \ldots, R \),

whereas \( A_{ki} \) denotes the fuzzy term which describes the \( k \)th feature in the \( i \)th fuzzy rule \( (k = 1, \ldots, D) \), \( R \) denotes the amount of fuzzy rules and \( S = (s_1, s_2, \ldots, s_D) \) indicates the binary feature vector (Hodashinsky et al., 2019), in which \( s_k \) represents the presence \( (s_k = 1) \) or absence \( (s_k = 0) \) of a feature in the classifiers. On a provided datasets \( \{(x_p; c_p), p = 1, 2, \ldots, Z\} \) the class label is determined by:

\[
\text{class} = c_t, t = \arg \max_{j=1,2,\ldots,m} \beta_j
\]

\[
\beta_j(x_p) = \frac{1}{\sum_{R_j} \prod_{k=1}^{D} \mu_{A_{kj}}(x_{pk})}
\]

\( \mu_{A_{ki}}(x_{pk}) \) denotes the symmetric membership function for the fuzzy \( A_{ki} \) at the point \( x_{pk} \). The amount of classification rates is determined as a ratio among the amount of properly allocated class labels and the overall quantity of objects to be categorized. 

\[
E(\theta, S) = \sum_{p=1}^{Z} \left\{ \begin{array}{ll} 
1, & \text{if} \ f_p = \arg \max_{j=1,2,\ldots,m} f_j(x_p; \theta, S) \\
0, & \text{otherwise}
\end{array} \right.
\]

whereas \( f(x_p; \theta, S) \) denotes the output of fuzzy classifiers using the parameter \( \theta \) and feature \( S \) at the point \( x_p \). The method generates the primary rule base for the fuzzy classifiers which contain one rule of all classes. The rule is made according to the extreme value in the training samples \( T = \{(x_p; c_p), p = 1, 2, \ldots, Z\} \), Where \( m \) represents the number of classes; \( D \) denotes the number of features.

To fine tune the membership function of the FRC model, the ROA is applied and thereby boosts the classifier performance. At this point, the rain performance has been inspired as it can be determined from the traditional section. All the solutions to issue are demonstrated as raindrops. According to this problem, any points in the answer space were obvious in an arbitrary fashion as raindrop falls on earth's surface. When the primary answer population has been created, the radius of all droplets is assigned from an arbitrary approach to limited extent. Besides, all droplets confirm the neighborhood dependent upon size. The single droplet that is not yet connected validates the termination restriction of the place that is covered. For resolving these problems in dimension space, all the droplets are comprised of \( n \) variable. Therefore, the primary step the maximal as well as minimal restrict of the variables are estimated as the restrictions were calculated as a radius of the droplet. Then, two endpoints of variable were sampled and it can be frequent still attaining the end variable. Next, the cost of initial droplet has been updated by moving in downward way. It can be carried out to all the droplets, and the cost as well as place of all droplets are allocated. The radius of droplet is altered from two fashions:

If the two droplets with radius \( r_1 \) and \( r_2 \), afterward it can be nearly one another with classic area and it links to develop huge droplet of radius \( R \):

\[
R = (r_1^n + r_2^n)^{1/n}
\]

where \( n \) refers the number of variables to every droplet. If the droplet with radius \( r1 \) is not shifted dependent upon soil feature which is demonstrated as \( \alpha \), water has been detected by the soil.

\[
R = (\alpha r_1^n)^{1/n}
\]

Noticeably, \( \alpha \) demonstrates the volume of droplet which is absorbed from every iteration in \([0–100]\) percent. Also, it defines the minimum values to droplets radius \( r_{min} \), in which droplet with less radius of that \( r_{min} \) is reduced.
As noted previously, the population value is decreased after any iterations and maximal droplet is utilized with massive domain of analysis (Pustokhina et al., 2021). With improving the analysis procedure, the local exploring potential of drops is maximizing proportionally to diameter of droplets. Therefore, with maximized number of rounds, weaker droplets obtain vanished or connect to robust drop with maximal domain of examination, and primary population is decreased intensively and define the correct answer(s). It can be assumed that there are some differences amongst the recently projected optimized technique from ROA and presently executed search method utilized Rain Fall Algorithm (RFA) that are combined as:

- In ROA, the primary population amount is altered then every iteration due to connection of neighboring drop. It improves the searching ability of technique and decreases the optimized cost considerably.
- Once the size of droplet has been modified, the connection of neighboring droplet or adsorption with soil has been implemented. This efficiency alters the exploring potential of all droplets and classifies the droplets.
- In RFA and different searching techniques, all the populations are comprised of neighbor points and the droplet was improved one step from arbitrary fashion. Also, all the populations recognize the optimum path to lesser point. Once the path has been initiated, it can be moved in downward way iteratively by step and the cost function has been decreased from single iteration.

According to the approximation and idealization of the technique, the rain technique has been demonstrated. In depth, tuned parameters of these techniques as primary raindrop number (population number), fundamental raindrop radius, etc. Then, the value has been assigned to every droplet according to the cost function. Afterward, all the droplets are moved downward. Therefore, near droplets were related to one another that effectively improve outcomes. Afterward, it can be appropriate for identifying extrema points of an objective function.

Performance validation

This section reports the RS results analysis of the proposed DTLF-ERSIC technique in terms of diverse dimensions. Two benchmark datasets, namely, The University of California (UC) Merced Land-Use and AID dataset are used [http://weegee.vision.ucmerced.edu/datasets/landuse.html, https://captain-whu.github.io/AID/]. The UCM dataset includes 21 class labels with 2100 images. In addition, the AID dataset has 10 K images with 30 class labels. Figures 2–3 demonstrate the sample test images from the applied two datasets.

Figure 4 depicts the confusion matrix of the DTLF-ERSIC technique on the applied UCM dataset. The figure portrayed that the DTLF-ERSIC technique has accomplished maximum classification performance on the applied dataset.

Table 1 offers the detailed classification results analysis of the DTLF-ERSIC technique on the applied UCM dataset. The table values denoted that the DTLF-ERSIC technique has classified all the images effectively into distinct class labels. For instance, the DTLF-ERSIC manner has classified the images into “agricultural” class with the sens. of 0.950, spec. of 0.999, and acc. of 0.997. Also, the DTLF-ERSIC approach has classified the images into “buildings” class with the sens. of 0.970, spec. of 0.998, and acc. of 0.997.

Besides, the DTLF-ERSIC algorithm has classified the images into “harbor” class with the sens. of 0.980, spec. of 0.996, and acc. of 0.995. Additionally, the DTLF-ERSIC manner has classified the images into “river” class with the sens. of 0.980, spec. of 1.000, and acc. of 0.999. Moreover, the DTLF-ERSIC methodology has classified the images into “tennis court” class with the sens. of 0.960, spec. of 0.997, and acc. of 0.995.

Figure 5 illustrates the average results analysis of the DTLF-ERSIC technique on the applied UCM dataset. The figure described that the DTLF-ERSIC technique has resulted in a maximum average sens. of 0.967, spec. of 0.998, precision of 0.968, acc. of 0.997, and F-score of 0.967.

Figure 6 illustrates the confusion matrix of the DTLF-ERSIC approach on the applied AID dataset. The figure exhibited that the DTLF-ERSIC methodology has accomplished maximal classification efficiency on the applied dataset.

Table 2 provides the detailed classification outcomes analysis of the DTLF-ERSIC approach on the applied AID dataset. The table values referred that the DTLF-ERSIC manner has classified all the images efficiently into different class labels. For instance, the DTLF-ERSIC algorithm has classified the images into “airport” class with the sens. of 0.967, spec. of 1.000, and acc. of 0.999. In addition, the DTLF-ERSIC manner has classified the images into “church” class with the sens. of 0.978, spec. of 0.999, and acc. of 0.998. Followed by, the DTLF-ERSIC manner has classified the images into “park” class with the sens. of 0.994, spec. of 0.999, and acc. of 0.984. Furthermore, the DTLF-ERSIC methodology has classified the images into “school” class with the sens. of 0.994, spec. of 0.999, and acc. of 0.999. Likewise, the DTLF-ERSIC methodology has classified the images into “viaduct” class with the sens. of 0.983, spec. of 0.999, and acc. of 0.999.
Figure 3. AID Sample Images.

Figure 4. Confusion Matrix of UCM Dataset.
Table 1. Results analysis of Proposed DTLF-ERSIC Model of UCM Dataset.

| No. of Classes       | Sensitivity | Specificity | Precision | Accuracy | F-Score |
|----------------------|-------------|-------------|-----------|----------|---------|
| Agricultural         | 0.950       | 0.999       | 0.979     | 0.997    | 0.965   |
| Airplane             | 0.920       | 1.000       | 1.000     | 0.996    | 0.938   |
| Baseball             | 0.960       | 1.000       | 1.000     | 0.998    | 0.980   |
| Diamond              |             |             |           |          |         |
| Beach                | 0.960       | 0.998       | 0.960     | 0.996    | 0.960   |
| Buildings            | 0.970       | 0.998       | 0.960     | 0.997    | 0.965   |
| Chaparral            | 1.000       | 0.999       | 0.971     | 0.999    | 0.965   |
| Dense                | 0.940       | 1.000       | 1.000     | 0.979    | 0.969   |
| Residential          |             |             |           |          |         |
| Forest               | 0.980       | 0.998       | 0.952     | 0.977    | 0.966   |
| Freeway              | 0.980       | 1.000       | 0.990     | 0.999    | 0.985   |
| Golf Course          | 0.940       | 0.997       | 0.931     | 0.994    | 0.935   |
| Harbor               | 0.980       | 0.996       | 0.925     | 0.995    | 0.952   |
| Intersection         | 0.990       | 0.998       | 0.961     | 0.998    | 0.975   |
| Medium               | 0.970       | 1.000       | 1.000     | 0.999    | 0.985   |
| Residential          |             |             |           |          |         |
| Residential          |             |             |           |          |         |
| Mobile Home          | 0.950       | 0.999       | 0.969     | 0.960    |         |
| Park                 | 0.960       | 1.000       | 0.990     | 0.979    | 0.960   |
| Parking Lot          | 0.970       | 0.998       | 0.951     | 0.996    | 0.960   |
| River                | 0.980       | 1.000       | 1.000     | 0.999    | 0.990   |
| Runway               | 0.990       | 0.995       | 0.908     | 0.995    | 0.947   |
| Sparse               | 0.960       | 0.997       | 0.941     | 0.995    | 0.951   |
| Residential          |             |             |           |          |         |
| Storage Tanks        |             |             |           |          |         |
| tennis court         |             |             |           |          |         |
| Average              | 0.967       | 0.998       | 0.968     | 0.997    | 0.967   |

Figure 5. Average analysis of DTLF-ERSIC model with different measures under UCM dataset.

Figure 7 depicts the average outcomes analysis of the DTLF-ERSIC approach on the applied AID dataset. The figure stated that the DTLF-ERSIC algorithm has resulted in a maximal average sens. of 0.975, spec. of 0.999, precision of 0.976, acc. of 0.998, and F-score of 0.975.

A brief comparative results analysis of the DTLF-ERSIC manner with existing techniques on the UCM dataset takes place in Table 3 and Figure 8. The results exhibited the FBA and TS-Fusion methods have obtained lower performance with the lower acc. of 0.974 and 0.980. Followed by, the IV3-CapsNet and Bi-MobileNetv2 techniques have gained moderately closer acc. of 0.991 and 0.993 respectively. Furthermore, the MVFLN +VGG technique has resulted in a competitive acc. of 0.995. However, the proposed DTLF-ERSIC technique has accomplished superior outcomes with a maximum acc. of 0.997.

A detailed comparative outcomes analysis of the DTLF-ERSIC manner with recent algorithms on the AID dataset takes place in Table 4 and Figure 9. The outcomes outperformed that the TS-Fusion and FT-VGGNet-16 systems have achieved minimal performance with the lesser acc. of 0.833 and 0.905. Simultaneously, the IV3-CapsNet...
and FDP-RN methods have obtained moderately closer acc. of 0.926 and 0.955 correspondingly. Concurrently, the CNN-MLP manner has resulted in a competitive acc. of 0.974. However, the proposed DTLF-ERSIC algorithm has accomplished maximum results with improved acc. of 0.998.
By observing into the aforementioned tables and figures, it can be clear that the DTLF-ERSIC technique has been found to be an effective tool for RS image classification.

Conclusion

In this study, an effective DTLF-ERSIC technique is derived to detect and classify the RS with maximum detection rate. The DTLF-ERSIC technique encompasses different subprocesses namely data augmentation, fusion-based feature extraction, FRC-based classification, and ROA-based parameter optimization. The DTLF-ERSIC technique incorporates the entropy-based fusion of three feature extraction techniques namely DLBP, ResNet50, and EfficientNet models. In addition, the ROA is applied to effectively tune the membership function of the FRC classifier to boost the classification performance. In order to ensure the betterment of the DTLF-ERSIC technique, an extensive simulation process takes place on benchmark dataset. The simulation results reported the supremacy of the DTLF-ERSIC technique over the recent state of art techniques. In future, advanced DL models with hyper-parameter optimization processes can be employed for the effective extraction of features in the RS image classification process.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work under grant number (RGP 2/209/42).

This work was supported by Taif University Researchers Supporting Program (project number: TURSP-2020/128), Taif University, Saudi Arabia.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the King Khalid University [RGP 2/209/42]; Taif University [TURSP-2020/128].

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated during the current study.

References

Alkhelaiwi, M., Boulila, W., Ahmad, J., Koubaa, A., & Driss, M., et al. (2021). An efficient approach based on privacy-preserving deep learning for satellite image classification. Remote Sensing, 13(11), 2221. https://doi.org/10.3390/rs13112221

Bansal, P., Kumar, R., & Kumar, S., et al. (2021). Disease detection in apple leaves using deep convolutional neural network. Agriculture, 11(7), 617. https://doi.org/10.3390/agriculture11070617

Bera, S., & Shrivastava, V. K. (2020). Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification. International Journal of Remote Sensing, 41(7), 2664–2683. https://doi.org/10.1080/01431161.2019.1694725

Byju, A. P., Sumbul, G., Demir, B., & Bruzzzone, L., et al. (2020). Remote-sensing image scene classification with deep neural networks in JPEG 2000 compressed domain. IEEE Transactions on Geoscience and Remote Sensing, 59(4), 3458–3472. https://doi.org/10.1109/TGRS.2020.3007523

Dai, L., Liu, G., Huang, L., Gang, X., Zhao, X., Junjin, R. et al. (2021). Feature transfer method for infrared and visible image fusion via fuzzy lifting scheme]. Infrared Physics & Technology, 114(1), 103621. http://weegsee.vision.ucmerced.edu/datasets/landuse.htmlhttps://captain-whu.github.io/AID/

Hagag, A., Amin, M., & Abd El-Samie, F. E., et al. (2015). Multispectral image compression with band ordering and wavelet transforms. Signal, Image and Video Processing, 9(4), 769–778. https://doi.org/10.1007/s11760-015-0516-4

Hodashinsky, I., Sarin, K., Shelupanov, A., & Slezkin, A., et al. (2019). Feature selection based on swallow swarm optimization for fuzzy classification. Symmetry, 11(11), 1423. https://doi.org/10.3390/sym11111423

Li, W., Wang, Z., Wang, Y., Wu, J., Wang, J., Jia, Y., & Gui, G., et al. (2020). Classification of high-spatial-resolution remote sensing scenes method using transfer learning and deep convolutional neural network. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 1986–1995. https://doi.org/10.1109/JSTARS.2020.2988477

Liang, P., Shi, W., & Zhang, X., et al. (2018). Remote sensing image classification based on stacked denoising autoencoder. Remote Sensing, 10(2), 16. https://doi.org/10.3390/rs10010016

Liu, Y., Ding, L., Chen, C., & Liu, Y., et al. (2020). Similarity-based unsupervised deep transfer learning for remote sensing image retrieval. IEEE Transactions on Geoscience and Remote Sensing, 58(11), 7872–7889. https://doi.org/10.1109/TGRS.2020.2984703

Lu, X., Ji, W., Li, X., & Zheng, X., et al. (2019). Bidirectional adaptive feature fusion for remote sensing scene classification. Neurocomputing, 328, 135–146. https://doi.org/10.1016/j.neucom.2018.03.076

Lu, Z., Bai, Y., Chen, Y., Su, C., Lu, S., Zhan, T., Wang, S., et al. (2020). The classification of gliomas based on a pyramid dilated convolution resnet model. Pattern Recognition Letters, 133, 173–179. https://doi.org/10.1016/j.patrec.2020.03.007

Lv, Z., Liu, T., & Benediktsson, J. A., et al. (2020). Object-oriented key point vector distance for binary land cover change detection using VHR remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 58(9), 6524–6533. https://doi.org/10.1109/TGRS.2020.2977248

Ma, H., Liu, Y., Ren, Y., Wang, D., Yu, L., & Yu, J., et al. (2020). Improved CNN classification method for groups of buildings damaged by earthquake, based on high resolution remote sensing images. Remote Sensing, 12(2), 260. https://doi.org/10.3390/rs12020260
Nanni, L., De Luca, E., Facin, M. L., & Maguolo, G., et al. (2020). Deep learning and handcrafted features for virus image classification. Journal of Imaging, 6(12), 143. https://doi.org/10.3390/jimaging6120143

Nogueira, K., Penatti, O. A., & Dos Santos, J. A., et al. (2017). Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recognition, 61, 539–556. https://doi.org/10.1016/j.patcog.2016.07.001

Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning https://arxiv.org/abs/1712.04621. arXiv preprint arXiv:1712.04621.

Pu, S., Wu, Y., Sun, X., & Sun, X., et al. (2021). Hyperspectral image classification with localized graph convolutional filtering. Remote Sensing, 13(3), 526. https://doi.org/10.3390/rs13030526

Pustokhina, I. V., Pustokhin, D. A., Nguyen, P. T., Elhoseny, M., & Shankar, K., et al. (2021). Multi-objective rain optimization algorithm with WELM model for customer churn prediction in telecommunication sector. In Complex & Intelligent Systems. Springer. (pp. 1–13).

Saba, T., Mohamed, A. S., El-Affendi, M., Amin, J., & Sharif, M., et al. (2020). Brain tumor detection using fusion of hand crafted and deep learning features. Cognitive Systems R Essearch, 59, 221–230. https://doi.org/10.1016/j.cogsys.2019.09.007

Shawky, O. A., Hagag, A., El-Dahshan, E. S. A., & Ismail, M. A., et al. (2020). Remote sensing image scene classification using CNN-MLP with data augmentation. Optik, 221, 165356. https://doi.org/10.1016/j.ijleo.2020.165356

Tong, X. Y., Xia, G. S., Lu, Q., Shen, H., Li, S., You, S., & Zhang, L., et al. (2020). Land-cover classification with high-resolution remote sensing images using transferable deep models. Remote Sensing of Environment, 237, 111322. https://doi.org/10.1016/j.rse.2019.111322

Vakalopoulou, M., Karantzalos, K., Komodakis, N., & Paragios, N., et al. (2015, July). Building detection in very high resolution multispectral data with deep learning features. In 2015 IEEE international geoscience and remote sensing symposium (IGARSS) Italy (pp. 1873–1876). IEEE.

Xu, Z., Liu, G., Xiao, G., Tang, L., & Li, Y., et al. (2021). JCa2Co: A joint cascade convolution coding network based on fuzzy regional characteristics for infrared and visible image fusion. IET Comput Vis, (7). https://doi.org/10.1049/cvi.12046

Yang, Z., Mu, X. D., & Zhao, F. A., et al. (2018). Scene classification of remote sensing image based on deep network and multi-scale features fusion. Optik, 171, 287–293. https://doi.org/10.1016/j.ijleo.2018.06.024

Zhang, Z., Jiang, R., Mei, S., Zhang, S., & Zhang, Y., et al. (2019). Rotation-invariant feature learning for object detection in VHR optical remote sensing images by double-net. IEEE Access, 8, 20818–20827. https://doi.org/10.1109/ACCESS.2019.2960931