RESEARCH ARTICLE

Multi-Label Scene Classification on Remote Sensing Imagery Using Modified Dingo Optimizer With Deep Learning

MAHMOUD RAGAB[ID]

Information Technology Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
Department of Mathematics, Faculty of Science, Al-Azhar University, Nasr City, Cairo 11884, Egypt
Center of Excellence in Smart Environment Research, King Abdulaziz University, Jeddah 21589, Saudi Arabia
Center of Research Excellence in Artificial Intelligence and Data Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia

e-mail: mragab@kau.edu.sa

ABSTRACT Multi-label scene classification on remote sensing imagery (RSI) includes the classification of images into multiple categories or labels, where each image belongs to more than one class or scene. This is a common task in RS and computer vision, especially for applications like urban planning, land cover classification, and environmental monitoring. By leveraging the power of deep learning (DL), this model extracts high-level features from the imagery, facilitating efficient and accurate scene classification, which is indispensable for applications including environmental analysis, land use monitoring, and disaster management. This study introduces a new Multi-Label Scene Classification on Remote Sensing Imagery using Modified Dingo Optimizer with Deep Learning (MSCRSI-MDODL) technique. The MSCRSI-MDODL technique targeted the identification and classification of multiple target classes from the RSI. In the presented MSCRSI-MDODL technique, attention Squeeze and Excitation (SE) with DenseNet model, named improved DenseNet model is applied for the extraction of features. Besides, MDO algorithm can be employed for the optimal hyperparameter tuning of the improved Densenet model. For scene classification process, the MSCRSI-MDODL technique makes use of stacked dilated convolutional autoencoders (SDCAE) model. The simulation analysis of the MSCRSI-MDODL model is tested on benchmark RSI datasets. The comprehensive result analysis portrayed the higher performance of the MSCRSI-MDODL technique over other existing techniques for RSI classification.

INDEX TERMS Remote sensing images, deep learning, scene classification, hyperparameter tuning, computer vision.

I. INTRODUCTION

Remote sensing (RS) plays a pivotal role in earth observation because it aids in discovering as well as identifying acts based on physical characteristics [1]. At present, Remote sensing images (RSI) have gained a high reputation in government and business fields for several reasons like weather, agriculture, forestry and biodiversity monitoring to surface variations. In recent research, RS employed

The associate editor coordinating the review of this manuscript and approving it for publication was Jon Atli Benediktsson[ID].

Deep Learning (DL) method to satellite images to remove effective information [2]. Efforts are still being made to extract many discriminatory features for the classification of satellite images in recent days. The traditional techniques are mainly focused on handcrafted features like texture and color features [3]. The mid-level models are proposed to develop further representation to aid high-level statistical techniques. The scene classification images are an effective and highly challenging task in real-time applications like urban planning from High-Spatial Resolution (HSR), geospatial object detection, geographic image retrieval, natural hazard

detection and environmental monitoring [4]. Scene classification is an essential step in several real-time applications of RS.

In terms of Machine Learning (ML), there are huge kinds of applications that arise from RSI which are proposed by employing single-label classification [5]. The main aim is to allocate a semantic/ single label set to an image. Moreover, usually, the real-time RSIs are more difficult and consist of more than a single semantic class in a particular image. So, single-label identification remains inadequate to define the attendance of complex areas which can carry out semantically complex content. To simplify the highly accurate image of RSI, the training of RSI must be taken through the Multi-Label classification (MLC) task, where an assumed image can be connected with many semantic concepts that are occupied by a predetermined set of labels [6]. With this method, the identification issue becomes more complicated when compared to single-label classification that was detailed in a recent important comparative research and investigated huge types of MLC models. The research comes with two kind of main challenges which can able to limit the performance of MLC techniques such as high-dimensional label spaces and the presence of difficult label correlations [7].

From Deep Learning (DL) researchers, these sorts of challenges are increasing huge extents of attention, especially on models that are proficient in learning long-range needs automatically (for example, with the usage of self-attention devices in vision transformer (ViT) base network framework) as well as controlling high dimensional label space [8]. Their inner devices and framework like hierarchical design and characteristics in Convolutional Neural Networks (CNNs), with non-linearity and local connectivity, are proficient in encoding data enormously [9]. However, their flexible plan provides data to be removed and discriminative representations must be trained from noisy data in an end manner. It helps to gain more exact classification performance in less-constrained situations when related to common MLC models. Then, the current achievement of this sort of technique can be able to connect with their capability to influence a massive range of labelled data for absorbing important knowledge [10].

This study introduces a new Multi-Label Scene Classification on Remote Sensing Imagery using the Modified Dingo Optimizer with Deep Learning (MSCRSI-MDODL) technique. In the presented MSCRSI-MDODL technique, attention Squeeze and Excitation (SE) with the DenseNet model, named the improved DenseNet model is applied for the extraction of features. Besides, the MDO algorithm can be employed for the optimal hyperparameter tuning of the improved Densenet model. For the scene classification process, the MSCRSI-MDODL technique makes use of the stacked dilated convolutional autoencoders (SDCAE) model. The simulation analysis of the MSCRSI-MDODL model is tested on benchmark RSI datasets.

II. RELATED WORKS

In [11], a multistage self-guided separation network (MGSNet) is developed. Numerous feature concerns between dissimilar network branches are extended through contrastive regularization (CR). In addition to that, a self-guided system is designed to discover the mutual feature among intraclass samples and then enhance feature representation constancy. Xu et al. [12] designed a new network technique called Lie Group Regional Influence Network (LGRIN). Initially, by mapping numerous space samples are attained and then features are removed. Next, multidilation pooling is combined into the CNNs framework. Similarly, the image regional influence system model is mainly proposed in order to monitor the attention of the identification approach by employing the regional level control of decomposition. In [13], an adaptive learning method for transporting a CNN-based technique is designed. Initially, an adaptive transform is mainly employed to alter RS image size. And then, an adaptive transferring technique is mainly projected for the purpose of classification. Lastly, in grouping with the label smoothing model, an adaptive label is offered in order to create easy labels depending on the images of the identification method forecasts for every type.

Liang et al. [14] designed an original dual-stream design which associates object-based location as well as global-based visual features. At first, the research study removed the appearance visual features from the entire scene image which is based on CNN. Next, the designed method identifies ground objects and build a chart in order to train spatial location feature depending on the GCN. Yang et al. [15] introduced a Semantic-Aware Graph Network (SAGN) model which consists of a Scene Decision Module (SDM), a dynamic graph feature update module, an adaptive semantic analysis module (ASAM) and a Dense Feature Pyramid Network (DFPN). Instead of transferring single-label into multi-label problems, the SAGN can elaborate appropriate techniques by employing various semantics unseen in HRRS images for the purpose of scene classification. Deng et al. [16] developed a common design that combined CNNs as well as Vision Transformer (ViT) (CTNet) by including two modules. First, compressed image patches are sent to a pre-trained ViT technique in order to mine semantic features. Then, the pre-trained CNN is transported in order to remove local structural feature in the C-stream.

In [17], a multi-level fusion Swin Transformer (MFST) combines an adaptive feature compression (AFC) as well as a multi-level feature merging (MFM) method is proposed. The MFM part restricts the semantic gap in multi-level feature through the patch inclusion in lower-level feature map as well as adjacent networks in the top-down path. By adaptive channel reduction, the AFC model generates multi-level feature which have tiny sizes and clear semantic data. Liu et al. [18] proposed a multi-level label-aware (MLLA), which is a semi-supervised act identification design. At first, a multi-level prototype awareness method is designed. At last,

based on this design, a multi-level pseudo-label generation technique is projected in order to allocate multi-level pseudo-labels to the unlabeled data.

III. THE PROPOSED MODEL

In this study, we have designed an automated scene classification using the MSCRSI-MDODL technique on the RSI. The MSCRSI-MDODL technique targeted the identification and classification of multiple target classes from the RSI. In the presented MSCRSI-MDODL technique, three phases of operations are mainly involved namely improved MDO-based hyperparameter tuning, DenseNet model for feature extraction, and SDCAE-based classification. Fig. 1 demonstrates the entire process of MSCRSI-MDODL approach.
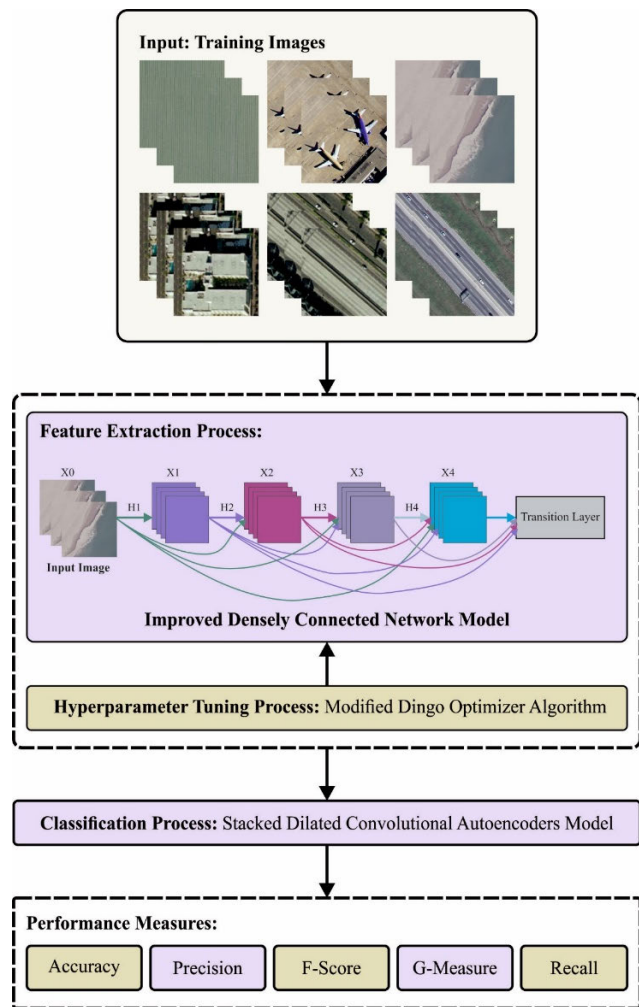

FIGURE 1. Overall process of MSCRSI-MDODL approach.

A. FEATURE EXTRACTION USING IMPROVED DENSENET MODEL

Primarily, the improved DenseNet model is applied for the extraction of features. DenseNet is the formation of dense connections which ensure maximum flow of data between layers. DenseNet is composed of 3 transition layers,

3×3 max-pooling layer and 7×7 Conv layers with a stride of 2, a classification layer, and 4 dense blocks [19]. Among these, dense block implements 1×1 and 3×3 Conv. The transition layer has 2×2 *avg* pooling layers and 1×1 Conv layer. The classification layer has FC layer and 7×7 global *avg* pooling layers.

DenseNet has $\frac{L(L+1)}{2}$ direct connection with L layers different from classical CNN with L layers. Note that x_0 is the input image, the CNN has 5 layers, $x_i (i = 0, 1, 2, \cdots 5)$ is the mapping features of ith layer, $H_i()$ is non-linear conversion that includes activation function, batch normalization, convolution, pooling, etc. Therefore, the output of 5th layer is a non-linear conversion of output feature map of each prior layer,

$$x_5 = H_5 ([x_0, x_1, \cdots x_4]) \tag{1}$$

Attention mechanism is considered a resource allocation model and split into different classes, namely, multistage attention, channel attention, pixel attention, etc. The key concept of Squeeze and Excitation (SE) block is to learn the weight features based on the loss thus mapping feature has greater weight. Squeeze is a compressing feature with a spatial dimension and transforms 2D feature channels into real numbers. Excitation is the same as Recurrent Neural Network (RNN), which learns the weight to model the relationship between the feature channels and generate weight for the feature channel through parameters. where $U \in R^{H \times W \times C}$, for any transformation F_{tr} mapping $X(X \in R^{H' \times W' \times C'})$ input into the mapping feature U, we create a respective SE block to implement feature *re*-calibration. First, the feature U passed over the squeeze operation F_{sq} that compresses U into $1 \times 1 \times C$ feature. Next, the feature from F_{sq} is excited through the excitation operation F_{ex}. Lastly, the re-calibration feature $\tilde{X}(\tilde{X} \in R^{H' \times W' \times C'})$ achieved F_{scale}, in which F_{scale} implies the weight of excitation output is successively weighted to the prior feature channels through multiplication.

B. HYPERPARAMETER TUNING USING MDO ALGORITHM

The MDO algorithm is exploited to tune the hyperparameter values of the improved DenseNet architecture. The MDO is the new biologically inspired global optimization algorithm based on the hunting tactics of dingoes [20]. This model includes persecuting individuals, scavenging behavior, and grouping techniques. The dingo dog is at risk of extinction in Australia. During classification, the MDO algorithm is used to calculate the weighted value for approximating the loss function so as to optimize the performance. Meanwhile, the computation of the loss function is an indispensable process for enhancing the classifier accuracy. As well, it assists in reducing the classification error rate while detecting the attack class. Therefore, the computation of the loss function can be enhanced by the MDO approach. The central benefits of using the MDO algorithm are high search efficacy, low computation burden, and high convergence rate when compared to the other optimization algorithms. Also, the surviving probability of dingoes is considered.

The input $\vec{\chi}$ is a feature matrix, and the output is optimum value of $\vec{x_b}$. Now, initialize the input parameters including the probability of hunting P, the persecution attack, the set of population $\vec{\chi}$, and the probability of the group. The three different rules namely group attack, scavenger, and persecution attack procedures are calculated until the maximal amount of iterations is reached. By using the subsequent equation, the new position of search agent $x_k(iter \rightarrow +1)$ can be assessed during rule 1 formation:

$$\vec{x_k}(iter + 1) = \vartheta_1 * \sum_{j=1}^{n_P} \frac{|\vec{\delta_j}(iter) - \vec{x_j}(iter)|}{n_p} - \vec{X_b}(iter) \quad (2)$$

In Eq. (2), the random number is indicated as ϑ_1, the random number within the range of $[2, \frac{P_{size}}{2}]$ can be denoted as n_p, the total population size is represented by P_{size}, the subset of search agents is specified as $\vec{\delta_j}$ (iter), the current searching agent can be implied as $\vec{x_k}(iter \rightarrow +1)$, and the fittest search agent of the prior iteration is x_b (iter). By using the following equation, the existing position updating is performed while executing rule 2:

$$\vec{x_k}(iter \rightarrow +1) = \vec{x_b}(iter) + \vartheta_1 * \exp^{\vartheta_2} * (\vec{x_{s_1}}(iter) - \vec{x_k}(iter)) \quad (3)$$

In Eq. (3), the arbitrary values from 1 to the maximum size of search agents are s_1. The uniformly generated random number within $[-1, 1]$ is ϑ_2,

By using the following model, the position updating is performed While executing rule 3:

$$E\vec{x_k}(iter + l) = \frac{1}{2} * \{\exp^{\vartheta_2} * \vec{x_{s_1}} \rightarrow (iter) - (-1)^\varepsilon * \vec{x_k}(iter) \rightarrow\} \quad (4)$$

In Eq. (4), a randomly generated binary number is ε. Next, by using the following equation, the search agent with lower survival rate can be estimated:

$$\S(k) = \frac{M_{fit} - fit(k)}{M_{fit} - N_{fit}} \quad (5)$$

In Eq. (5), the existing fitness value of *the* k^{th} search agent is $fit(k)$, and the worst and the best fitness values of the present generation are M_{fit} and N_{fit}, correspondingly. By using the following model, the fitness value of the search agent can be evaluated:

$$\vec{x_k^{new}}(iter) = \vec{x_b}(iter) + \frac{1}{2} * \{\exp^{\vartheta_2} * \vec{x_{s_1}}(iter) - (-1)^\varepsilon * \vec{x_k}(iter)\} \quad (6)$$

If the existing iteration is higher than the prior iteration, the last position updating is performed using Eq. (7):

$$\vec{x_b}(iter) \rightarrow = \vec{x_k^{new}}(iter) - \nabla * \left(\frac{1}{2} - \aleph\right) \quad (7)$$

Here, uniformly distributed pseudo-random integers within $[-2, 2]$ and $[0, 1]$ are ∇, and $\aleph$.

The MDO method derives FF to accomplish high efficiency of classification. It defines a positive integer to signify the superior result of the solution candidate. The decline of the classifier error rate is assumed as an FF.

$$fitness(x_i) = ClassifierErrorRate(x_i)$$
$$= \frac{number\ of\ misclassified\ samples}{Total\ number\ of\ samples} * 100 \quad (8)$$

C. IMAGE CLASSIFICATION USING SDCAE MODEL

In this work, the SDCAE model is utilized for the image classification process. Dilated convolutional autoencoder (DCAE) has a similar design to classical AE [21]. An activating method used to transform the input into convolution layer:

$$i^c = s(A * E^c + y^c). \quad (9)$$

In Eq. (9), E^c and y^c are matrices with bias related to c^{th} mapping features i^c, and the 2D inputs converted from the numerical vector is A.

In the proposed model, $(*)$ is a ReLU activation function, AF input vector. Subsequently, the feature mapping of the hidden layer is transported into the reconstruction via inverted convolution:

$$A = s\left(\sum_c i^c * \check{E}^c + \check{y}\right), \quad (10)$$

In Eq. (10), A is a set of mapping features with a similar design as input A; also, i refers to the set of mapping features with a similar structure as input A. The $\dot{E}$ & E weight matrices have similar initial values. Fig. 2 illustrates the infrastructure of DCAE.
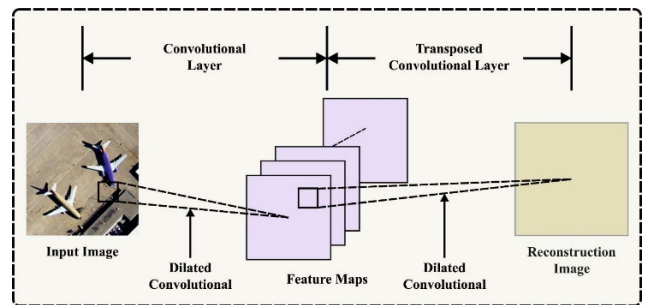

FIGURE 2. Structure of DCAE.

The training purpose is DCAE is to minimize the variance amongst the input vectors A and the reconstructed B. MSE is the cost function:

$$M(A, B) = \frac{1}{x} \sum_{j}^{x} (A_j - B_j)^2. \quad (11)$$

The DCAEs are used for creating the DNN by layering many DCAEs. The hidden layer output of the prior DCAE is applied as an input of subsequent DCAE. Stacked DCAEs are a demanding layer-wise unsupervised training process. Dilated convolution has a wide range of visual fields without losing data. Due to these advantages, it is suitable for data

extraction. Firstly, dilated convolution is used to expand the activation function of the layer, which allows them to learn global features. In comparison with max-pooling, Dilated convolution safeguards the input data from the lost data. Next, the DCAEs' training model does not need a labelled dataset. Lastly, DCAEs have fewer parameters compared to FCN. Thus, DCAE outperforms other unsupervised DL algorithms for efficiency and effectiveness.

IV. RESULTS AND DISCUSSION

In this section, the scene classification results of the MSCRSI-MDODL method are tested on the LandUse dataset [22]. It holds 2100 samples with 21 classes as defined in Table 1.

TABLE 1. Details on database.

Labels	Class	No. of Samples
L0	Agricultural	100
L1	Airplane	100
L2	Baseballdiamond	100
L3	Beach	100
L4	Buildings	100
L5	Chaparral	100
L6	Denseresidential	100
L7	Forest	100
L8	Freeway	100
L9	Golgcourse	100
L10	Harbor	100
L11	Intersection	100
L12	Mediumresidential	100
L13	Mobilehomepart	100
L14	Overpass	100
L15	Parkinglot	100
L16	River	100
L17	Runway	100
L18	Sparseresidential	100
L19	Storagetanks	100
L20	Tenniscourt	100
Total Number of Samples		2100

Fig. 3 shows the confusion matrix formed by the MSCRSI-MDODL technique under 70% of the TR phase. The simulated values reported the efficient recognition of all 21 classes.

In Table 2, the scene classification results of the MSCRSI-MDODL technique are tested on 70% of the TR Phase. The simulated values indicate that the MSCRSI-MDODL method reaches effectual outcomes with 21 classes. It is also noticed that the MSCRSI-MDODL system gains an average $accu_y$ of 99.84%, $prec_n$ of 98.41%, $sens_y$ of 98.37%, $spec_y$ of 99.92%, and $G_{measure}$ of 98.38%.

Fig. 4 shows the confusion matrix produced by the MSCRSI-MDODL methodology on 30% of the TS phase.
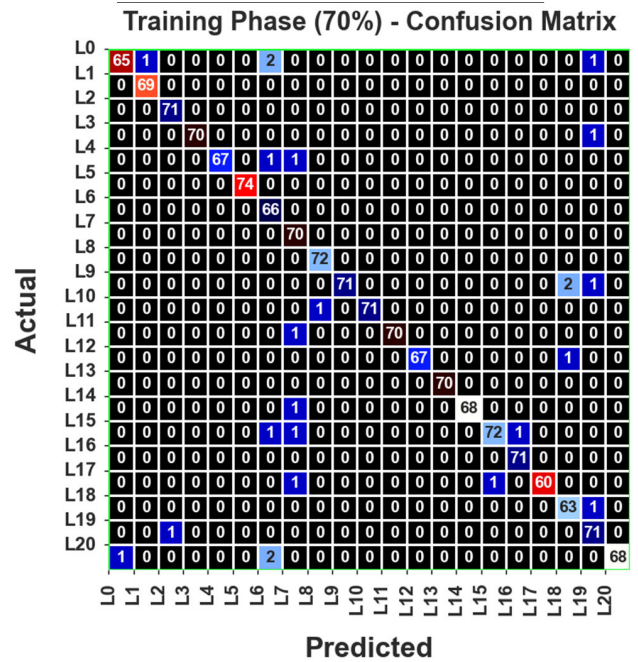

FIGURE 3. Confusion matrix of MSCRSI-MDODL approach on 70% of TR phase.

TABLE 2. Scene classification outcome of MSCRSI-MDODL approach on 70% of TR phase.

Class Labels	$Accu_y$	$Prec_n$	$Sens_y$	$Spec_y$	$G_{measure}$
TR Phase (70%)					
L0	99.66	98.48	94.20	99.93	96.32
L1	99.93	98.57	100.00	99.93	99.28
L2	99.93	98.61	100.00	99.93	99.30
L3	99.93	100.00	98.59	100.00	99.29
L4	99.86	100.00	97.10	100.00	98.54
L5	100.00	100.00	100.00	100.00	100.00
L6	99.59	91.67	100.00	99.57	95.74
L7	99.66	93.33	100.00	99.64	96.61
L8	99.93	98.63	100.00	99.93	99.31
L9	99.80	100.00	95.95	100.00	97.95
L10	99.93	100.00	98.61	100.00	99.30
L11	99.93	100.00	98.59	100.00	99.29
L12	99.93	100.00	98.53	100.00	99.26
L13	100.00	100.00	100.00	100.00	100.00
L14	99.93	100.00	98.55	100.00	99.27
L15	99.73	98.63	96.00	99.93	97.31
L16	99.93	98.61	100.00	99.93	99.30
L17	99.86	100.00	96.77	100.00	98.37
L18	99.73	95.45	98.44	99.79	96.93
L19	99.66	94.67	98.61	99.71	96.62
L20	99.80	100.00	95.77	100.00	97.86
Average	99.84	98.41	98.37	99.92	98.38

The simulated values reported the efficient recognition of all 21 classes.

In Table 3, the scene classification analysis of the MSCRSI-MDODL technique is confirmed in 30% of the TS Phase. The simulated values exhibit that the MSCRSI-MDODL system obtains effectual outcomes with 21 classes. Additionally, it is observed that the MSCRSI-MDODL model
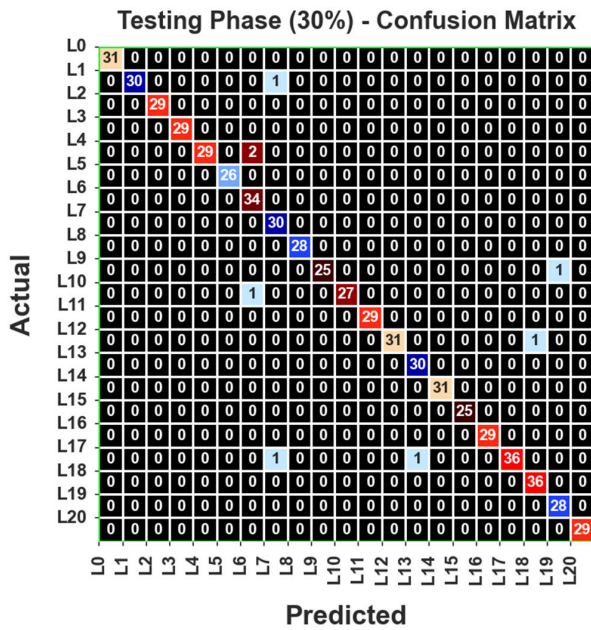
FIGURE 4. Confusion matrix of MSCRSI-MDODL approach on 30% of TS phase.

TABLE 3. Scene classification outcome of MSCRSI-MDODL approach on 30% of TS phase.

Class Labels	$Accu_y$	$Prec_n$	$Sens_y$	$Spec_y$	$G_{measure}$
TS Phase (30%)					
L0	100.00	100.00	100.00	100.00	100.00
L1	99.84	100.00	96.77	100.00	98.37
L2	100.00	100.00	100.00	100.00	100.00
L3	100.00	100.00	100.00	100.00	100.00
L4	99.68	100.00	93.55	100.00	96.72
L5	100.00	100.00	100.00	100.00	100.00
L6	99.52	91.89	100.00	99.50	95.86
L7	99.68	93.75	100.00	99.67	96.82
L8	100.00	100.00	100.00	100.00	100.00
L9	99.84	100.00	96.15	100.00	98.06
L10	99.84	100.00	96.43	100.00	98.20
L11	100.00	100.00	100.00	100.00	100.00
L12	99.84	100.00	96.88	100.00	98.43
L13	99.84	96.77	100.00	99.83	98.37
L14	100.00	100.00	100.00	100.00	100.00
L15	100.00	100.00	100.00	100.00	100.00
L16	100.00	100.00	100.00	100.00	100.00
L17	99.68	100.00	94.74	100.00	97.33
L18	99.84	97.30	100.00	99.83	98.64
L19	99.84	96.55	100.00	99.83	98.26
L20	100.00	100.00	100.00	100.00	100.00
Average	99.88	98.87	98.79	99.94	98.81

achieves an average $accu_y$ of 99.88%, $prec_n$ of 98.87%, $sens_y$ of 98.79%, $spec_y$ of 99.94%, and $G_{measure}$ of 98.81% correspondingly.

To estimate the performance of the MSCRSI-MDODL technique, TR and TS $accu_y$ curves are determined, as shown in Fig. 5. The TR and TS $accu_y$ curves demonstrate the performance of the MSCRSI-MDODL system over various epochs. The figure offers important details about the learning

tasks and generalization abilities of the MSCRSI-MDODL model. With a rise in epoch count, it is evidenced that the TR and TS $accu_y$ curves have improved. It is observed that the MSCRSI-MDODL methodology achieves enriched testing accuracy that can have the potential to identify the patterns in the TR and TS data.
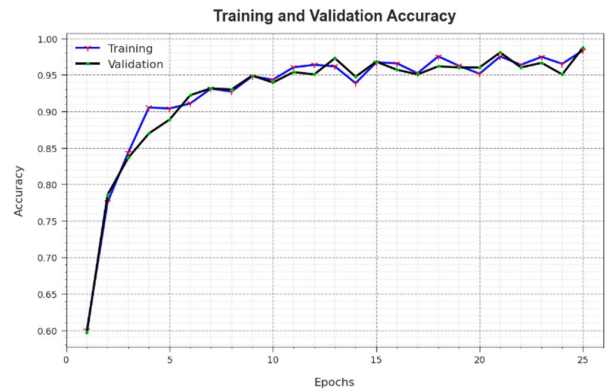

FIGURE 5. $Accu_y$ curve of the MSCRSI-MDODL approach.

FIGURE 6. Loss curve of the MSCRSI-MDODL approach.

Fig. 6 illustrates the overall TR and TS loss values of the MSCRSI-MDODL system over epochs. The TR loss described as the model loss gets diminished over epochs. Primarily, the loss values are reduced as the model adjusts the weight to decrease the predicted error on the TR and TS data. The loss curves exhibit the level to which the model fits the training data. It is noticed that the TR and TS loss is progressively reduced and represents that the MSCRSI-MDODL algorithm efficaciously learns the patterns shown in the TR and TS data. It is also perceived that the MSCRSI-MDODL method modifies the parameters to minimize the difference among the predicted and actual training labels.

The PR performance of the MSCRSI-MDODL system is represented by plotting precision against recall as shown in Fig. 7. The simulated values confirm that the MSCRSI-MDODL method becomes improved PR values with each 21 class. The figure exhibits that the model learns to recognize different 21 class labels. The MSCRSI-MDODL technique attains enriched outcomes in the recognition of positive samples with diminished false positives.
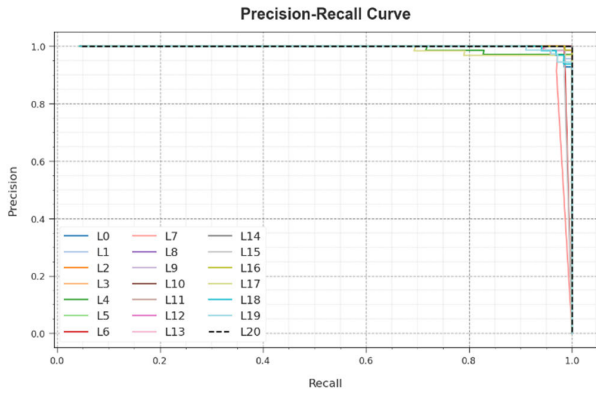
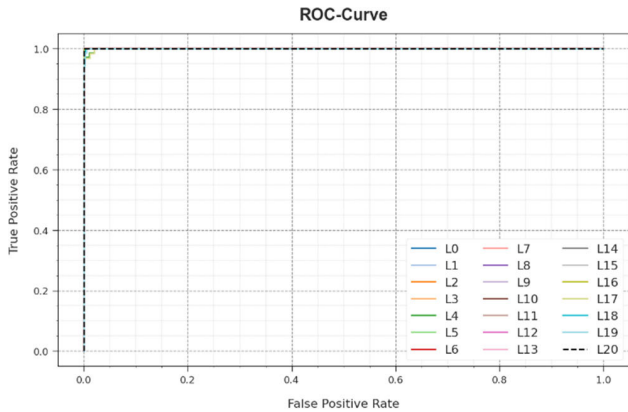**FIGURE 7. PR curve of the MSCRSI-MDODL approach.**

FIGURE 8. ROC curve of the MSCRSI-MDODL approach.

The ROC analysis offered by the MSCRSI-MDODL system is shown in Fig. 8, which has proficiency in the differentiation of the 21 class labels. The figure revealed valuable perceptions of the trade-off among the TPR and FPR rates over different classification thresholds and modifying counts of epochs. It accurately gives the predicted performance of the MSCRSI-MDODL methodology on the classification of various 21 classes.

TABLE 4. Comparative outcome of MSCRSI-MDODL approach with other systems.

Methods	$Accu_y$	$Prec_n$	$Sens_y$	$Spec_y$
MSCRSI-MDODL	99.88	98.87	98.79	99.94
SIDTL-AIC	99.60	95.89	95.86	95.85
SIDTLD+SSA	94.05	98.16	97.39	97.10
DL-C-PTRN	96.23	97.48	95.44	95.35
DL-MOPSO	96.54	96.50	94.45	97.37
DL-AlexNet	96.40	94.00	94.62	96.59
DL-VGG-VD-19	97.80	98.03	95.68	96.27
DL-CaffeNet	94.65	94.39	96.48	95.40

In Table 4 and Fig. 9, the experimental validation of the MSCRSI-MDODL technique with recent models is made [23]. The results highlighted that the SIDTLD+SSA
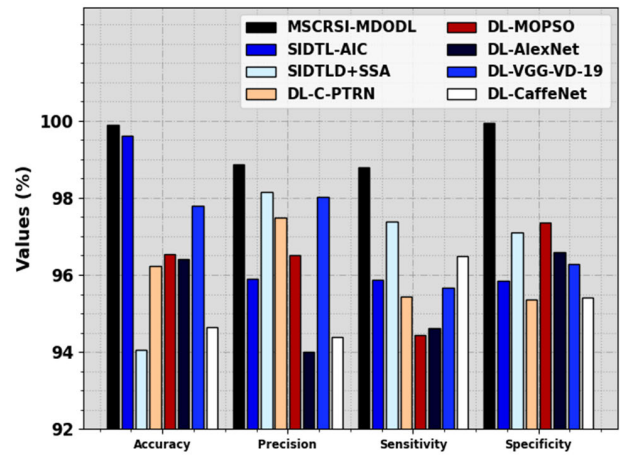

FIGURE 9. Comparative outcome of MSCRSI-MDODL approach with other systems.

and DL-CaffeNet models have obtained worse performance. Along with that, the DL-C-PTRN, DL-MOPSO, DL-AlexNet, and DL-VGG-VD-19 models have reported slightly improved results. Meanwhile, the SIDTL-AIC model reaches considerably boosted performance with $accu_y$, $prec_n$, $sens_y$, and $spec_y$ values of 99.60%, 95.89%, 95.86%, and 95.85% respectively. However, the MSCRSI-MDODL technique reaches higher $accu_y$, $prec_n$, $sens_y$, and $spec_y$ values of 99.88%, 98.87%, 98.79%, and 99.94% respectively. Therefore, the MSCRSI-MDODL technique ensured better performance over other models on the scene classification process.
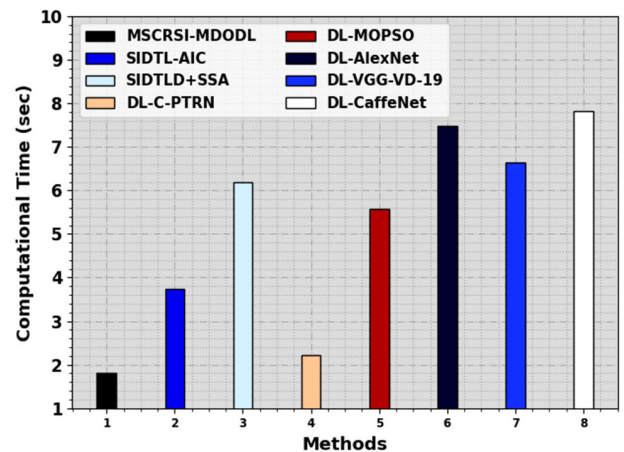

FIGURE 10. CT outcome of MSCRSI-MDODL approach with other systems.

Finally, the computation time (CT) analysis of the MSCRSI-MDODL system is shown in Table 5 and Fig. 10. The extensive simulated values demonstrated that the MSCRSI-MDODL technique outcomes in a reducing CT value of 1.80s. While, the SIDTL-AIC, SIDTLD+SSA, DL-C-PTRN, DL-MOPSO, DL-AlexNet, DL-VGG-VD-19, and DL-CaffeNet methods acquire increased CT values.

TABLE 5. CT outcome of MSCRSI-MDODL approach with other systems.

Methods	Computational Time (sec)
MSCRSI-MDODL	1.80
SIDTL-AIC	3.75
SIDTLD+SSA	6.18
DL-C-PTRN	2.22
DL-MOPSO	5.58
DL-AlexNet	7.48
DL-VGG-VD-19	6.63
DL-CaffeNet	7.83

V. CONCLUSION

In this study, we have designed an automated scene classification using the MSCRSI-MDODL technique on the RSI. The MSCRSI-MDODL technique targeted the identification and classification of multiple target classes from the RSI. In the presented MSCRSI-MDODL technique, three phases of operations are mainly involved namely improved DenseNet model for feature extraction, MDO-based hyperparameter tuning, and SDCAE-based classification. In this work, the improved DenseNet model is applied for the extraction of features. Besides, the MDO algorithm can be employed for the optimal hyperparameter tuning of the improved Densenet model. For the scene classification process, the MSCRSI-MDODL technique makes use of the SDCAE model. The simulation analysis of the MSCRSI-MDODL technique is tested on benchmark RSI datasets. The comprehensive result analysis portrayed the higher performance of the MSCRSI-MDODL model over other existing techniques for RSI classification.

REFERENCES

[1] W. Miao, J. Geng, and W. Jiang, "Multigranularity decoupling network with pseudolabel selection for remote sensing image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5603813.

[2] C. Xu, G. Zhu, and J. Shu, "A combination of lie group machine learning and deep learning for remote sensing scene classification using multi-layer heterogeneous feature extraction and fusion," *Remote Sens.*, vol. 14, no. 6, p. 1445, Mar. 2022.

[3] T. Tian, L. Li, W. Chen, and H. Zhou, "SEMSDNet: A multiscale dense network with attention for remote sensing scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 5501–5514, 2021.

[4] J. Li, M. Gong, H. Liu, Y. Zhang, M. Zhang, and Y. Wu, "Multiform ensemble self-supervised learning for few-shot remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4500416.

[5] F. Peng, W. Lu, W. Tan, K. Qi, X. Zhang, and Q. Zhu, "Multi-output network combining GNN and CNN for remote sensing scene classification," *Remote Sens.*, vol. 14, no. 6, p. 1478, Mar. 2022.

[6] X. Wang, S. Wang, C. Ning, and H. Zhou, "Enhanced feature pyramid network with deep semantic embedding for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7918–7932, Sep. 2021.

[7] S.-C. Hung, H.-C. Wu, and M.-H. Tseng, "Remote sensing scene classification and explanation using RSSCNet and LIME," *Appl. Sci.*, vol. 10, no. 18, p. 6151, Sep. 2020.

[8] Z. Li, B. Hou, X. Guo, S. Ma, Y. Cui, S. Wang, and L. Jiao, "Contrastive learning based on multiscale hard features for remote-sensing image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023.

[9] J. Shen, T. Yu, H. Yang, R. Wang, and Q. Wang, "An attention cascade global–local network for remote sensing scene classification," *Remote Sens.*, vol. 14, no. 9, p. 2042, Apr. 2022.

[10] Y. Lu, M. Gong, Z. Hu, W. Zhao, Z. Guan, and M. Zhang, "Energy-based CNNs pruning for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 3000214.

[11] J. Wang, W. Li, M. Zhang, R. Tao, and J. Chanussot, "Remote-sensing scene classification via multistage self-guided separation network," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5615312.

[12] C. Xu, G. Zhu, and J. Shu, "A lightweight and robust lie group-convolutional neural networks joint representation for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5501415.

[13] W. Wang, Y. Chen, and P. Ghamisi, "Transferring CNN with adaptive learning for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5533918.

[14] J. Liang, Y. Deng, and D. Zeng, "A deep neural network combined CNN and GCN for remote sensing scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4325–4338, 2020.

[15] Y. Yang, X. Tang, Y.-M. Cheung, X. Zhang, and L. Jiao, "SAGN: Semantic-aware graph network for remote sensing scene classification," *IEEE Trans. Image Process.*, vol. 32, pp. 1011–1025, 2023.

[16] P. Deng, K. Xu, and H. Huang, "When CNNs meet vision transformer: A joint framework for remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[17] G. Wang, N. Zhang, W. Liu, H. Chen, and Y. Xie, "MFST: A multi-level fusion network for remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[18] Q. Liu, M. He, Y. Kuang, L. Wu, J. Yue, and L. Fang, "A multi-level label-aware semi-supervised framework for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5616112.

[19] K. Wang, P. Jiang, J. Meng, and X. Jiang, "Attention-based DenseNet for pneumonia classification," *IRBM*, vol. 43, no. 5, pp. 479–485, Oct. 2022.

[20] A. Mahalingam, G. Perumal, G. Subburayalu, M. Albathan, A. Altameem, R. S. Almakki, A. Hussain, and Q. Abbas, "ROAST-IoT: A novel range-optimized attention convolutional scattered technique for intrusion detection in IoT networks," *Sensors*, vol. 23, no. 19, p. 8044, Sep. 2023.

[21] P. Nirmala, T. Manimegalai, J. R. Arunkumar, S. Vimala, G. V. Rajkumar, and R. Raju, "A mechanism for detecting the intruder in the network through a stacking dilated CNN model," *Wireless Commun. Mobile Comput.*, vol. 2022, pp. 1–13, Apr. 2022.

[22] Accessed: Oct. 16, 2023. [Online]. Available: http://weegee.vision.ucmerced.edu/datasets/landuse.html

[23] S. S. Alotaibi, H. A. Mengash, N. Negm, R. Marzouk, A. M. Hilal, M. A. Shamseldin, A. Motwakel, I. Yaseen, M. Rizwanullah, and A. S. Zamani, "Swarm intelligence with deep transfer learning driven aerial image classification model on UAV networks," *Appl. Sci.*, vol. 12, no. 13, p. 6488, Jun. 2022.

MAHMOUD RAGAB received the Ph.D. degree from the Faculty of Mathematics and Natural Sciences, Christian-Albrechts-University (CAU), Kiel, Schleswig-Holstein, Germany. He is a Professor of data science with the Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia; and the Mathematics Department, Faculty of Science, Al-Azhar University, Cairo, Egypt. He was with different research groups at many universities, such as the Combinatorial Optimization and Graph Algorithms Group (COGA), Faculty II Mathematics and Natural Sciences, Berlin University of Technology, Berlin, Germany; The British University in Egypt (BUE); and the Automation, Integrated Communication Systems Group, Ilmenau University of Technology (TU Ilmenau), Thüringen, Germany. He is also a Researcher with various centers such as: the Centre for Artificial Intelligence in Precision Medicines, University of Oxford; the Center of Research Excellence in Artificial Intelligence and Data Science; and the Center of Excellence in Smart Environment Research, King Abdulaziz University. His research focuses on AI algorithms, deep learning, sorting, optimization, mathematical modeling, data science, neural networks, time series analysis, and computation.

• • •