Article

Squirrel Search Optimization with Deep Transfer Learning-Enabled Crop Classification Model on Hyperspectral Remote Sensing Imagery

Manar Ahmed Hamza, Fadwa Alrowais, Jaber S. Alzahrani, Hany Mahgoub, Nermin M. Salem and Radwa Marzouk

Abstract: With recent advances in remote sensing image acquisition and the increasing availability of fine spectral and spatial information, hyperspectral remote sensing images (HSI) have received considerable attention in several application areas such as agriculture, environment, forestry, and mineral mapping, etc. HSIs have become an essential method for distinguishing crop classes and accomplishing growth information monitoring for precision agriculture, depending upon the fine spectral response to the crop attributes. The recent advances in computer vision (CV) and deep learning (DL) models allow for the effective identification and classification of different crop types on HSIs. This article introduces a novel squirrel search optimization with a deep transfer learning-enabled crop classification (SSODTL-CC) model on HSIs. The proposed SSODTL-CC model intends to identify the crop type in HSIs properly. To accomplish this, the proposed SSODTL-CC model initially derives a MobileNet with an Adam optimizer for the feature extraction process. In addition, an SSO algorithm with a bidirectional long-short term memory (BiLSTM) model is employed for crop type classification. To demonstrate the better performance of the SSODTL-CC model, a wide-ranging experimental analysis is performed on two benchmark datasets, namely dataset-1 (WHU-Hi-LongKou) and dataset-2 (WHU-Hi-HanChuan). The comparative analysis pointed out the better outcomes of the SSODTL-CC model over other models with a maximum of 99.23% and 97.15% on test datasets 1 and 2, respectively.

Keywords: hyperspectral remoting sensing; crop mapping; image classification; deep transfer learning; hyperparameter optimization

1. Introduction

Due to advancements in remote sensing image acquisition mechanisms and the increasing availability of rich spatial and spectral data by means of various sensors, hyperspectral imaging has become more prominent [1]. Especially, hyperspectral remote sensing image (HSI) classification has become a major source for real-time application in fields such as...
mineral mapping, agriculture, environment, and forestry, etc. [2,3]. Usually, the HIS is taken at a large number of contiguous narrow spectral wavelengths for the improved analysis of the earth object. Since the spectral resolution could be in nm, the hyperspectral sensor offers significant facility in data analysis [4] for many humanitarian tasks, including precision agriculture for improved farming practices, discrimination amongst vegetation classes for better treatment, etc. [5]. The current study emphasizes using and analyzing HSI in the agriculture area. Conventional techniques, such as statistical-based analyses and field surveys, are time-consuming [6]. Cutting-edge remote sensing technologies involving HSI provide an appropriate solution and might fill the gap with solutions such as crop classification. In the HSI framework, the classification has the common objective of automatically labeling the pixel (spectral pattern or signature) into a predetermined class [7]. The classification is implemented either by utilizing the transformed feature or the original feature. An HSI has numerous features and is hard to adapt to a single convolutional kernel size. When the number of model layers is increased, many useful features are lost [8–10].

The authors of [11] proposed a rotation-invariant local binary pattern-based weighted generalized closest neighbor (RILBP-WGCN) approach for an HSI classifier. The presented RILBP is an improved texture-based classifier paradigm, which employs LBP filters to any designated bands to generate a wide sketch of spatial texture data. Similarly, the presented WGCN approach effectually maintained the spatial uniformity amongst the adjacent pixel employing a local weight method and point-to-set distances. Meng et al. [12] concentrated on a DL-based crop mapping, utilizing one-shot hyperspectral satellite imagery, whereas three CNN techniques, such as 1D-CNN, 2D-CNN, and 3D-CNN, were executed for end-to-end crop mapping. Furthermore, a manifold learning-based visualized method, i.e., t-distributed stochastic neighbor embedding (t-SNE), was established for demonstrating the discriminative capability of deep semantic feature extracting by the distinct CNN approaches.

In [13], a hybrid model was established for estimating the chlorophyll content from the crops utilizing HIS segmentation with active learning, which contains two important stages. First, it can utilize a sparse multinomial logistic regression (SMLR) method for learning the class posterior probability distribution with quadratic programming or joint probability distributions. Second, it can utilize the data developed from the preceding step for segmenting the HSI utilizing a Markov random field segment. Farooq et al. [14] examine patch-based weed identification utilizing HSI. A CNN was estimated and correlated to a histogram of oriented gradients (HoG) for this solution. Appropriate patch sizes were examined. The restriction of RGB imagery was established. In [15], a deep one-class crop (DOCC) structure that contains a DOCC extracting element and an OCC extraction loss element was presented for large-scale OCC mapping. The DOCC structure takes only the instances of one target class as input for extracting the crop of interest by positive and unlabeled learning and automatically extracts the feature for OCC mapping.

In [16], a low altitude UAV hyperspectral remote sensing platform was created for collecting higher spatial resolution remote sensing images of degraded grassland. The GDIF-3D-CNN classifier method was utilized for classifying the pure pixel and every pixel data set, whose accuracy and performance were enhanced by optimizing the eight parameters of the method. Wei et al. [17] present a fine classifier approach dependent upon multi-feature fusion and DL. During this case, the morphological profiles, GLCM texture, and endmember abundance features were leveraged to exploit the spatial data of HIS. Next, the spatial data were fused with original spectral data to generate a classifier outcome by utilizing a DNN with a conditional random field (DNN + CRF) method. In detail, the DNN is a deep detection method that extracts depth features and mines the potential data.

For smaller samples and higher-dimension HSIs, it becomes very complex to learn wide-ranging image features; subsequently, it becomes hard to precisely recognize complex HSI. The UAV-borne HSIs have rich spatial data, and the spatial resolution reaches centimeter level; however, the higher spatial resolution causes serious spatial heterogeneity and spectral variability. Nowadays, the deep learning (DL) method is extensively employed in
image processing because of its effective feature learning abilities [9]. Currently, the most common DL-based network framework is the convolution neural network (CNN). CNN has the features of parameter sharing, equivariant mapping, and sparse interaction, which reduce the training parameter size and complexity of the network. Such features permit the algorithm to generate a certain degree of invariance in scaling, shifting, and distortion and also create fault tolerance and stronger robustness [10]. Consequently, CNN has been extensively employed in HSI classification.

This article introduces a novel squirrel search optimization with a deep transfer learning-enabled crop classification (SSODTL-CC) model on HSIs. The proposed SSODTL-CC model initially derives a MobileNet with an Adam optimizer for the feature extraction process. The utilization of the Adam optimizer allows for effectual adjustment of the hyperparameters of the MobileNet model. In addition, a bidirectional long-short term memory (BiLSTM) method is employed for crop type classification. To enhance the classifier efficiency of the BiLSTM model, the SSO algorithm is employed for hyperparameter optimization, which shows the novelty of the work. To demonstrate the better performance of the SSODTL-CC model, a wide-ranging experimental analysis is performed on a benchmark dataset.

2. Materials and Methods

In this article, a new SSODTL-CC model has been developed to identify the crop type in HSIs properly. To do so, the proposed SSODTL-CC model performed feature extraction using MobileNet with an Adam optimizer. In addition, the BiLSTM model received feature vectors and performed crop type classification. To enhance the classifier efficiency of the BiLSTM model, the SSO algorithm was employed for hyperparameter optimization. Figure 1 illustrates the block diagram of the SSODTL-CC technique.

![Block diagram of the SSODTL-CC technique.](image)
2.1. Data Collection

In this section, the experimental validation of the proposed model is performed against two datasets [18], namely dataset-1 (WHU-Hi-LongKou) and dataset-2 (WHU-Hi-HanChuan). The dataset-1 comprises a total of 9000 samples with nine class labels, holding 1000 samples under each class. In addition, dataset-2 comprises a total of 16,000 samples with 16 class labels, holding 1000 samples under each class. Figure 2 shows the sample HSIs from various classes, such as water spinach, soybean, strawberry, corn, sesame, and broad-leaf soybean.

![Sample images: (a) water spinach, (b) soybean, (c) strawberry, (d) corn, (e) sesame, and (f) broad-leaf soybean.](image)

**Figure 2.** Sample images: (a) water spinach, (b) soybean, (c) strawberry, (d) corn, (e) sesame, and (f) broad-leaf soybean.

2.2. Feature Extraction: MobileNet Model

During the feature extraction process, the HSIs were passed into the MobileNet model to generate feature vectors. MobileNet is a CNN-based technique that is extensively applied in classifier procedures. The most important benefit of utilizing the presented method is that the model needs moderately low computation work in comparison with the CNN, which makes it appropriate to operate with a mobile device and a computer that operates with lower computational capabilities. The presented method is a fundamental architecture that combines convolution layers that are applied to efficiently distinguish details according to two controllable attributes that change between parameter precision and potential. The presented method is valuable in diminishing the size of the system.

The MobileNet structure is very effective with the least amount of attributes, namely Palmprint detection. This concerns a depth-wise convolution. The fundamental architecture is dependent on discrete abstracted layers, i.e., a module of dissimilar convolution layers that seem to be the quantal structure that measures a typical in-depth complication [19]. The resolution multiplier variable $\omega$ is added to minimize the measurement of the input dataset and inner layer representation with the analogous variable.
The feature vector map of size $F_m \times F_m$, and the filter is of size $F_s \times F_s$. The input variable is embodied by $p$, and the output variable is denoted by $q$. For the basic abstract layer of the structure, the whole computation work is considered as variable $c_e$, and it could be evaluated as follows:

$$c_e = F_s \cdot F_s \cdot \omega \cdot \alpha F_m \cdot \alpha F_m + \omega \cdot \rho \cdot \alpha F_m \cdot \alpha F_m$$

(1)

The $\omega$ multiplier value can be considered within one to $n$. The variable resolution multiplier is known as $\alpha$. The computational effort is recognized as the variable $\text{cost}_e$ and is evaluated by the following equation:

$$\text{cost}_e = F_s \cdot F_s \cdot \omega \cdot \rho \cdot F_m \cdot F_m$$

(2)

The proposed approach incorporates the pointwise and depth-wise convolutions that are circumscribed by the reduction variable known as the variable $d$, which is evaluated in the following:

$$d = \frac{F_s \cdot F_s \cdot \omega \cdot \alpha F_m \cdot \alpha F_m + \omega \cdot \rho \cdot \alpha F_m \cdot \alpha F_m}{F_s \cdot F_s \cdot \omega \cdot \rho \cdot F_m F_m}$$

(3)

The two hyper characteristics, resolution and width multipliers, enable changing the optimal window size for accurate prediction based on the context. The third values suggest that it contains three input channels. The principle under the MobileNet structure replaced the complicated convolutional layer, which comprises a convolutional layer with $3 \times 3$ buffers for the input dataset, along with a pointwise convolutional layer of size $1 \times 1$ that combines the filtered variable to construct an element.

2.3. Crop Type Classification: BiLSTM Model

At the time of image classification, the extracted feature vectors are fed into the BiLSTM model. The BiLSTM approach receives the feature vector as input and executes the detection method. The LSTM signifies a different RNN method, which solves the problem of gradient vanishing of RNN by offering a threshold method and memory unit [21]. However, $x$ denotes the network input at different times, $y$ refers to the network outcome, $h$ stands for the hidden layer (HL), $u$ refers to the weighted input to HLs, $w$ demonstrates
the weighted input of the previous node HL to the existing node HL, and \( v \) signifies the weighted input in HL to the output layer.

During the actual implementation of the LSTM technique, the LSTM unit was upgraded at time \( t \) as:

\[
    i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \tag{7}
\]

\[
    f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f) \tag{8}
\]

\[
    \tilde{c}_t = \tanh(W_c h_{t-1} + U_c x_t + b_c) \tag{9}
\]

\[
    c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{10}
\]

\[
    o_t = \sigma(W_o h_{t-1} + U_o x_t + b_o) \tag{11}
\]

\[
    h_t = o_{t-1} \odot \tanh(c_t) \tag{12}
\]

At this point, \( \odot \) stands for the equal product of elements, and \( \sigma \) denotes the sigmoid function. \( x_t \) signifies the input vector at time \( t \). \( h_t \) refers to the HL vector named as the output vector and the storage of all the data at time \( t \) and the preceding time. \( b_i, b_f, b_c, b_o \) demonstrates the offset vector. \( W_i, W_f, W_c, W_o \) implies the weight of various gates to the HL vector \( h_t \). \( U_i, U_f, U_c, U_o \) stands for the weighted input vector. \( x_t \) stands for the input, forgotten, unit, and output gates, correspondingly. Utilizing the 3-gates infrastructure, the LSTM permits the recurrent network to maintain the useful data of the task from the memory units at the time of the trained method, therefore evading the problem of the RNN disappearing but reaching an extensive range of data.

In addition to processing the series data, the BLSTM presents more backward estimate procedures, for instance, different normal LSTM cases. This process employs the subsequent data of sequences. At last, the forward and reverse estimations are executed. The values were resultant of the output layer simultaneously; thus, as the outcome, all of the sequence data are reached in \( 2 \times 2 \) directions, which is utilized to complete a variety of natural language processing tasks.

### 2.4. Hyperparameter Tuning: SSO Algorithm

For enhancing the classifier efficiency of the BiLSTM model, the SSO algorithm is employed for hyperparameter optimization. The SSO technique is proposed by the foraging behavior of a flying squirrel; subsequently, an effectual method employed small animals for migration. According to the food foraging hierarchy of squirrels [22], the optimum SSO algorithm is iteratively developed in an arithmetical model. There are important characteristics in SSA, that is, population sizes \( NP \), maximal value of iteration \( Iter_{\text{max}} \), the predator existence possibility \( P_{\text{dep}} \), decision variables value \( n \), gliding constants \( G_c \), scaling factors \( s_f \), upper and lower limits to decision variable \( FS_{U} \) and \( FS_{L} \). They are given in the following:

The position of the squirrel is randomly loaded from the searching space:

\[
    FS_{i,j} = FS_{L} + rand() \ast (FS_{U} - FS_{L}), \ i = 1, 2, \ldots, \ NP, \ j = 1, 2, \ldots, \ n \tag{13}
\]

However, \( rand() \) denotes an arbitrary value in [0, 1]. The fitness measure \( f = (f_1 f_2, f_{NP}) \) of a squirrel position was processed by replacing the decision variable with FF:

\[
    f_i = f_i(FS_{i,1}, FS_{i,2}, \ldots, FS_{i,n}), \ i = 1, 2, \ldots, \ NP \tag{14}
\]

Next, the quality of food sources is evaluated by the fitness measure of a squirrel position as follows:

\[
    [\text{sorted}_f, \text{sorte_index}] = \text{sort}(f) \tag{15}
\]

In addition, the organization of food sources was processed, which comprised hickory trees, normal trees, and oak trees (acorn nuts). The optimal food source (lower fitness) was assumed to be the hickory nut tree \( (FS_{nt}) \), the successive food sources that exist are denoted as acorn nut trees \( (FS_{nt}) \), and the rest are called normal trees \( (FS_{nt}) \):
\[ FS_{ht} = FS(sorte - index(1)) \]  
\[ FS_{at}(1 : 3) = FS(sorte - index(2 : 4)) \]  
\[ FS_{nt}(1 : NP - 4) = FS(sorte - index(5 : NP)) \]  

The three states that denote the dynamic gliding approach of squirrels are described in the following.

Scenario 1. The squirrel resides in an acorn nut tree and jumps to a hickory nut tree. A novel location can be given as follows:

\[ FS_{\text{new}} = \begin{cases} 
FS_{\text{old}} + d_g G_c (FS_{\text{ht}} - FS_{\text{at}}) & \text{if } R \geq P_{dp} \\
\text{random location} & \text{otherwise}
\end{cases} \]  

Now \( d_g \) indicates the gliding distance, \( R \) denotes a function that proceeds the measured value of a uniform distribution value within 0 and 1, and \( G_c \) denotes a gliding constant.

Scenario 2. The squirrel resides in a normal tree and moves to acorn nut trees for gathering needed food. A novel location can be determined by:

\[ FS_{\text{nt}} = \begin{cases} 
FS_{\text{old}} + d_g G_c (FS_{\text{at}} - FS_{\text{nt}}) & \text{if } R \geq P_{dp} \\
\text{random location} & \text{otherwise}
\end{cases} \]

Here, \( R_2 \) indicates a function that provides a measure of uniform distribution value in \([0, 1]\).

Scenario 3. Squirrels on normal trees go to hickory nut trees once they meet the routine objectives. Now, a novel position of squirrel can be determined by:

\[ S_{\text{nt}} = \begin{cases} 
FS_{\text{old}} + d_g G_c (FS_{\text{ht}} - FS_{\text{nt}}) & \text{if } R \geq P_{dp} \\
\text{random location} & \text{otherwise}
\end{cases} \]

where \( R_3 \) shows a function that suggests the measure of uniform distribution amongst \([0, 1]\). Hence, this measure is a maximum that invokes high perturbation. For achieving an appropriate method, a scaling factor (sf) is employed as a divisor of \( d_g \).

The foraging nature of flying squirrels depends on the season, which varies frequently. Therefore, the seasonal observation must be implemented; thus, the trapping is removed in the local optimal result. The seasonal constant \( S_c \) and minimal value can be given as:

\[ S_c^t = \sqrt{\sum_{k=1}^{n} (FS_{at,k} - FS_{ht,k})^2}, \quad t = 1, 2, 3 \]  
\[ S_{c\text{min}} = \frac{10E - 6}{365\text{iter} / (\text{Iter max} / 2.5)} \]  

For \( S_c^t < S_{c\text{min}} \), the winter becomes the highest, the squirrel loses its exploring ability, and the method of searching for food sources and locations changes:

\[ FS_{\text{nt}} = FS_L + \text{Lévy}(n) \times (FS_U - FS_L) \]  

Now the Lévy distribution is employed to improve the global search to an enhanced method:

\[ \text{Lévy}(x) = 0.01 \times \frac{r_a \times \sigma}{|r_b|^{1/\beta}} \]  
\[ \sigma = \left( \frac{\Gamma(1 + \beta) \times \sin(\pi \beta / 2)}{\Gamma((1 + \beta) / 2 \times \beta \times 2^{(\beta - 1) / 2})} \right)^{1/\beta} \]

This approach stops when the maximal constraint is fulfilled. If not, the nature of creating a novel location and approving the seasonal observation need to be repeatedly followed.
3. Experimental Validation

3.1. Result Analysis of SSODTL-CC Model

This section investigates the performance of the proposed model on test images. Figure 3 showcases the sample classification results obtained by the SSODTL-CC model. The figure implies that the proposed model has obtained effective classification results. In addition, some of the misclassified regions by the SSODTL-CC model are marked in blue circles.

\[ \sigma = \left( \frac{\gamma(1+\beta) \times \sin \left(\frac{\pi \beta}{2}\right)}{\gamma \left(\frac{1+\beta}{2}\right) \times \beta \times 2^{\left(\frac{\beta-1}{2}\right)}}\right)^{1/\beta} \]

Figure 3. Sample classification result of the SSODTL-CC technique under dataset-1: (a) input image, (b) class labels, and (c) classification output.

Figure 4 inspects the confusion matrices created by the SSODTL-CC model on the classification of nine classes under dataset-1. The figure reports that the SSODTL-CC model has categorized all the classes under different sets of datasets. For the entire dataset, the SSODTL-CC model recognized 956 samples under corn, 975 samples under cotton, 971 samples under sesame, 971 samples under broad-leaf soybean, 964 samples under narrow-leaf soybean, 949 samples under rice, 965 samples under water, 958 samples under roads and houses, and 967 samples under mixed weed. Similarly, the SSODTL-CC model has categorized the class labels proficiently on 70% of the training samples and 30% of the testing samples on dataset-1.
Figure 4 inspects the confusion matrices created by the SSODTL-CC model on the classification of nine classes under dataset-1. The figure reports that the SSODTL-CC model has categorized all the classes under different sets of datasets. For the entire dataset, the SSODTL-CC model recognized 956 samples under corn, 975 samples under cotton, 971 samples under sesame, 971 samples under broad-leaf soybean, 964 samples under narrow-leaf soybean, 949 samples under rice, 965 samples under water, 958 samples under roads and houses, and 967 samples under mixed weed. Similarly, the SSODTL-CC model has categorized the class labels proficiently on 70% of the training samples and 30% of the testing samples on dataset-1.

Figure 4. Confusion matrix of the SSODTL-CC technique under dataset-1. (a) Entire dataset-1. (b) 70% of Training dataset-1 and (c) 30% of Testing dataset-1.

Table 1 reports detailed crop classification outcomes of the SSODTL-CC model on all of dataset-1. The experimental values indicated that the SSODTL-CC model gained effectual outcomes under every individual class. For instance, in the corn class, the SSODTL-CC model offered accuracy, precision, and recall of 99.24%, 97.55%, and 95.60%, respectively. Similarly, on the mixed weed class, the SSODTL-CC model reached accuracy, precision, and recall of 99.27%, 96.70%, and 96.70%, respectively. Overall, the SSODTL-CC model showed a maximum average accuracy, precision, and recall of 99.20%, 96.43%, and 96.40%, respectively.

Table 2 depicts a brief crop classification outcome of the SSODTL-CC approach on 70% of training dataset-1. The experimental values stated that the SSODTL-CC method gained effectual outcomes under every individual class. For instance, in the corn class, the SSODTL-CC model offered accuracy, precision, and recall of 99.19%, 97.04%, and 95.49%, respectively. In addition, in the mixed weed class, the SSODTL-CC system obtained accuracy,
prec_n, and reca_l of 99.27%, 96.67%, and 96.67%, respectively. Overall, the SSODTL-CC model demonstrated maximum average acc_u, prec_n, and reca_l of 99.19%, 96.38%, and 96.35%, correspondingly.

Table 1. Result analysis of the SSODTL-CC technique with distinct classes under all of dataset-1.

| Class Labels          | Entire Dataset Samples |
|-----------------------|------------------------|
|                       | Accuracy | Precision | Recall | Kappa Score |
| Corn                  | 99.24    | 97.55     | 95.60  | -           |
| Cotton                | 99.22    | 95.59     | 97.50  | -           |
| Sesame                | 98.97    | 93.82     | 97.10  | -           |
| Broad-leaf soybean    | 99.16    | 95.38     | 97.10  | -           |
| Narrow-leaf soybean   | 99.39    | 98.07     | 96.40  | -           |
| Rice                  | 99.26    | 98.34     | 94.90  | -           |
| Water                 | 99.13    | 95.73     | 96.50  | -           |
| Roads and Houses      | 99.17    | 96.67     | 95.80  | -           |
| Mixed weed            | 99.27    | 96.70     | 96.70  | -           |
| **Average**           | **99.20**| **96.43** | **96.40**| **95.95**  |

Table 2. Result analysis of the SSODTL-CC technique with distinct classes under 70% of training dataset-1.

| Class Labels          | Training Samples (70%) |
|-----------------------|------------------------|
|                       | Accuracy | Precision | Recall | Kappa Score |
| Corn                  | 99.19    | 97.04     | 95.49  | -           |
| Cotton                | 99.21    | 95.63     | 97.27  | -           |
| Sesame                | 99.02    | 94.01     | 97.46  | -           |
| Broad-leaf soybean    | 99.06    | 95.45     | 96.38  | -           |
| Narrow-leaf soybean   | 99.33    | 97.96     | 96.01  | -           |
| Rice                  | 99.32    | 98.57     | 95.43  | -           |
| Water                 | 99.03    | 95.07     | 96.29  | -           |
| Roads and Houses      | 99.27    | 97.00     | 96.14  | -           |
| Mixed weed            | 99.27    | 96.67     | 96.67  | -           |
| **Average**           | **99.19**| **96.38** | **96.35**| **95.89**  |

Table 3 defines the detailed crop classification outcomes of the SSODTL-CC model on 30% of testing dataset-1. The experimental values indicated that the SSODTL-CC model gained effectual outcomes under every individual class. For instance, in the corn class, the SSODTL-CC approach presented acc_u, prec_n, and reca_l of 99.37%, 98.68%, and 95.85%, correspondingly. Furthermore, in the mixed weed class, the SSODTL-CC methodology reached acc_u, prec_n, and reca_l of 99.26%, 96.76%, and 96.76%, respectively. Overall, the SSODTL-CC model portrayed enhanced average acc_u, prec_n, and reca_l of 99.23%, 96.54%, and 96.53%, correspondingly.

Figure 5 illustrates the confusion matrices created by the SSODTL-CC approach on the classification of sixteen classes under dataset-2. The figure reveals that the SSODTL-CC model categorized all the classes under different sets of datasets. On the entire dataset, the SSODTL-CC model recognized 783 samples under class 1, 757 samples under class 2, 766 samples under class 3, 728 samples under class 4, 721 samples under class 5, 774 samples under class 6, 764 samples under class 7, 788 samples under class 8, 779 samples under
class 9, 779 samples under class 10, 733 samples under class 11, 806 samples under class 12, 771 samples under class 13, 829 samples under class 14, 733 samples under class 15, and 821 samples under class 16. Similarly, the SSODTL-CC approach categorized the class labels proficiently on 70% of the training samples and 30% of the testing samples on dataset-2.

**Table 3.** Result analysis of the SSODTL-CC technique with distinct classes under 30% of testing dataset-1.

| Class Labels          | Accuracy | Precision | Recall | Kappa Score |
|-----------------------|----------|-----------|--------|-------------|
| Corn                  | 99.37    | 98.68     | 95.85  | -           |
| Cotton                | 99.26    | 95.5      | 98.02  | -           |
| Sesame                | 98.85    | 93.33     | 96.22  | -           |
| Broad-leaf soybean    | 99.37    | 95.21     | 98.93  | -           |
| Narrow-leaf soybean   | 99.52    | 98.31     | 97.32  | -           |
| Rice                  | 99.11    | 97.74     | 93.53  | -           |
| Water                 | 99.37    | 97.32     | 96.99  | -           |
| Roads and Houses      | 98.93    | 95.99     | 95.11  | -           |
| Mixed weed            | 99.26    | 96.76     | 96.76  | -           |
| **Average**           | **99.23**| **96.54** | **96.53**| **96.08**  |

Table 4 demonstrates the detailed crop classification outcomes of the SSODTL-CC model on all of dataset-2. The experimental values exposed that the SSODTL-CC model gained effectual outcomes under every individual class. For instance, in class 1, the SSODTL-CC algorithm obtained \( \text{accu}_1, \text{prec}_1, \) and \( \text{reca}_1 \) of 97.39%, 79.57%, and 78.30% correspondingly. In addition, in class 16, the SSODTL-CC model gained \( \text{accu}_1, \text{prec}_1, \) and \( \text{reca}_1 \) of 97.17%, 74.98%, and 82.10%, correspondingly. Overall, the SSODTL-CC model outperformed higher average \( \text{accu}_1, \text{prec}_1, \) and \( \text{reca}_1 \) of 97.13%, 74.98%, and 82.10%, respectively.

Table 5 reports a brief crop classification outcome of the SSODTL-CC model on 70% of training dataset-2. The experimental values exposed that the SSODTL-CC model gained effectual outcomes under every individual class. For instance, in class 1, the SSODTL-CC model offered \( \text{accu}_1, \text{prec}_1, \) and \( \text{reca}_1 \) of 97.43%, 79.74%, and 79.29%, correspondingly. In addition, in class 16, the SSODTL-CC model reached \( \text{accu}_1, \text{prec}_1, \) and \( \text{reca}_1 \) of 97.21%, 74.08%, and 81.65%, correspondingly. Overall, the SSODTL-CC methodology exhibited maximal average \( \text{accu}_1, \text{prec}_1, \) and \( \text{reca}_1 \) of 97.13%, 77.08%, and 77.05%, correspondingly.

Table 6 defines the detailed crop classification outcome of the SSODTL-CC technique on 30% of testing dataset-2. The experimental values indicated that the SSODTL-CC algorithm gained effectual outcomes under every individual class. For sample, in class 1, the SSODTL-CC model offered \( \text{accu}_1, \text{prec}_1, \) and \( \text{reca}_1 \) of 97.29%, 76.80%, and 82.99%, correspondingly. In the same way, in class 16, the SSODTL-CC system reached \( \text{accu}_1, \text{prec}_1, \) and \( \text{reca}_1 \) of 97.06%, 76.80%, and 82.99%, respectively. Overall, the SSODTL-CC approach showed maximal average \( \text{accu}_1, \text{prec}_1, \) and \( \text{reca}_1 \) of 97.15%, 77.25%, and 77.09%, correspondingly.

### 3.2. Discussion

To ensure the improved crop classification results of the SSODTL-CC model, a comparison study with recent models on two datasets is given in Table 7 [22,23]. Figure 6 investigates a comparative classification outcome of the SSODTL-CC model with existing models on dataset-1. The results indicated that the SVM model gained an ineffectual outcome with the least \( \text{accu}_1 \) of 95.98%. In line with this, the FNEA-OO model certainly accomplished increased performance with an \( \text{accu}_1 \) of 97.07%. In addition, the SRFMC, CNN, and CNN-CRF models depicted closer \( \text{accu}_1 \) values of 98.20%, 98.08%, and...
98.80%, respectively. However, the SSODTL-CC model demonstrated superior performance with an accu of 99.23%.

Figure 5. Confusion matrix of the SSODTL-CC technique under dataset-2. (a) Entire dataset-2. (b) 70% of Training dataset-2, and (c) 30% of Testing dataset-2.

Figure 7 examines a comparative classification outcome of the SSODTL-CC model with existing approaches on dataset-2. The outcomes indicated that the SVM model gained an ineffectual outcome with the least accu of 77.34%. Likewise, the FNEA-OO model certainly accomplished an increased performance with an accu of 86.49%. Then, the SVRFMC, CNN, and CNN-CRF models depicted closer accu values of 86.95%, 87.72%, and 94.67%, correspondingly. At last, the SSODTL-CC methodology demonstrated superior performance with an accu of 97.15%.

From these results and discussions, it is evident that the SSODTL-CC model has the capability of attaining improved crop classification outcomes on HSIs.
Table 4. Result analysis of the SSODTL-CC technique with distinct classes under all of dataset-2.

| Class Labels | Accuracy | Precision | Recall | Kappa Score |
|--------------|----------|-----------|--------|-------------|
| Class-1      | 97.39    | 79.57     | 78.30  | -           |
| Class-2      | 97.24    | 79.18     | 75.70  | -           |
| Class-3      | 97.14    | 77.37     | 76.60  | -           |
| Class-4      | 96.89    | 76.39     | 72.80  | -           |
| Class-5      | 96.63    | 73.42     | 72.10  | -           |
| Class-6      | 97.25    | 78.34     | 77.40  | -           |
| Class-7      | 97.07    | 76.63     | 76.40  | -           |
| Class-8      | 97.31    | 78.25     | 78.80  | -           |
| Class-9      | 97.41    | 80.06     | 77.90  | -           |
| Class-10     | 96.85    | 73.35     | 77.90  | -           |
| Class-11     | 96.87    | 75.80     | 73.30  | -           |
| Class-12     | 97.36    | 77.95     | 80.60  | -           |
| Class-13     | 97.39    | 80.40     | 77.10  | -           |
| Class-14     | 97.36    | 76.69     | 82.90  | -           |
| Class-15     | 96.84    | 75.41     | 73.30  | -           |
| Class-16     | 97.17    | 74.98     | 82.10  | -           |
| **Average**  | 97.13    | 77.11     | 77.08  | 75.55       |

Table 5. Result analysis of the SSODTL-CC technique with distinct classes under 70% of training dataset-2.

| Class Labels | Accuracy | Precision | Recall | Kappa Score |
|--------------|----------|-----------|--------|-------------|
| Class-1      | 97.43    | 79.74     | 79.29  | -           |
| Class-2      | 97.31    | 78.98     | 78.42  | -           |
| Class-3      | 97.04    | 77.56     | 76.28  | -           |
| Class-4      | 96.77    | 75.11     | 72.77  | -           |
| Class-5      | 96.63    | 72.73     | 73.35  | -           |
| Class-6      | 97.29    | 78.6      | 76.89  | -           |
| Class-7      | 97.11    | 77.94     | 75.28  | -           |
| Class-8      | 97.29    | 78.02     | 77.91  | -           |
| Class-9      | 97.35    | 79.91     | 77.78  | -           |
| Class-10     | 96.69    | 72        | 75.65  | -           |
| Class-11     | 96.94    | 76.06     | 74.21  | -           |
| Class-12     | 97.4     | 78.79     | 80.03  | -           |
| Class-13     | 97.36    | 80.84     | 77.25  | -           |
| Class-14     | 97.37    | 76.72     | 82.98  | -           |
| Class-15     | 96.89    | 75.86     | 73.02  | -           |
| Class-16     | 97.21    | 74.08     | 81.65  | -           |
| **Average**  | 97.13    | 77.06     | 77.05  | 75.5        |
Table 6. Result analysis of the SSODTL-CC model with distinct classes under 30% of testing dataset-2.

| Class Labels | Testing Samples (30%) |
|--------------|-----------------------|
|              | Accuracy | Precision | Recall | Kappa Score |
| Class-1      | 97.29    | 79.15     | 75.93  | -           |
| Class-2      | 97.06    | 79.76     | 69.07  | -           |
| Class-3      | 97.38    | 76.90     | 77.45  | -           |
| Class-4      | 97.19    | 79.63     | 72.88  | -           |
| Class-5      | 96.63    | 75.18     | 69.21  | -           |
| Class-6      | 97.15    | 77.78     | 78.53  | -           |
| Class-7      | 96.98    | 73.82     | 79.05  | -           |
| Class-8      | 97.33    | 78.75     | 80.77  | -           |
| Class-9      | 97.54    | 80.43     | 78.20  | -           |
| Class-10     | 97.23    | 76.26     | 82.90  | -           |
| Class-11     | 96.71    | 75.17     | 71.19  | -           |
| Class-12     | 97.27    | 76.09     | 81.94  | -           |
| Class-13     | 97.48    | 79.26     | 76.70  | -           |
| Class-14     | 97.33    | 76.62     | 82.72  | -           |
| Class-15     | 96.71    | 74.43     | 73.94  | -           |
| Class-16     | 97.06    | 76.80     | 82.99  | -           |
| Average      | 97.15    | 77.25     | 77.09  | 75.64       |

Table 7. Comparative analysis of the SSODTL-CC technique with recent algorithms in terms of accuracy.

| Methods   | Dataset-1 | Dataset-2 |
|-----------|-----------|-----------|
| SVM       | 95.98     | 77.34     |
| FNEA-OO   | 97.07     | 86.49     |
| SVRFMC    | 98.20     | 86.95     |
| CNN       | 98.08     | 87.72     |
| CNN-CRF   | 98.80     | 94.67     |
| SSODTL-CC | 99.23     | 97.15     |

Figure 6. Comparative analysis of the SSODTL-CC technique under dataset-1.
4. Conclusions

In this article, a new SSODTL-CC model was developed to properly identify the crop type in HSIs. To do so, the proposed SSODTL-CC model performed feature extraction using MobileNet with an Adam optimizer. In addition, the BiLSTM model received feature vectors and performed crop type classification. To enhance the classifier efficiency of the BiLSTM model, the SSO algorithm was employed for hyperparameter optimization. To demonstrate the better performance of the SSODTL-CC model, a wide-ranging experimental analysis was performed on two benchmark datasets, namely dataset-1 (WHU-Hi-LongKou) and dataset-2 (WHU-Hi-HanChuan). The comparative analysis pointed out the better outcomes of the SSODTL-CC model over the recent approaches, with a maximum of 99.23% and 97.15% on test datasets 1 and 2, respectively. Therefore, the SSODTL-CC model can be utilized for effective crop type classification on HSIs. In the future, the classification performance of the SSODTL-CC model can be enhanced by the design of hybrid DL models.

Author Contributions: Conceptualization, M.A.H.; Data curation, F.A.; Formal analysis, F.A. and J.S.A.; Investigation, J.S.A.; Methodology, M.A.H.; Project administration, H.M.; Resources, H.M.; Software, N.M.S.; Supervision, N.M.S.; Validation, R.M.; Visualization, R.M.; Writing—original draft, M.A.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by King Khalid University, grant number RGP 2/46/43, Princess Nourah bint Abdulrahman University, grant number PNURSP2022R77 and Umm al-Qura University, grant number 22UQU4340237DSR24.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

Acknowledgments: The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Large Groups Project under grant number (46/43). Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R77), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The...
authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant Code: 22UQU4340237DSR19.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bhosle, K.; Musande, V. Evaluation of deep learning CNN model for land use land cover classification and crop identification using hyperspectral remote sensing images. *J. Indian Soc. Remote Sens.* 2019, 47, 1949–1958. [CrossRef]

2. Wang, C.; Chen, Q.; Fan, H.; Yao, C.; Sun, X.; Chan, J.; Deng, J. Evaluating satellite hyperspectral (Orbita) and multispectral (Landsat 8 and Sentinel-2) imagery for identifying cotton acreage. *Int. J. Remote Sens.* 2021, 42, 4042–4063. [CrossRef]

3. Wei, L.; Yu, M.; Liang, Y.; Yuan, Z.; Huang, C.; Li, R.; Yu, Y. Precise crop classification using spectral-spatial-location fusion based on conditional random fields for UAV-borne hyperspectral remote sensing imagery. *Remote Sens.* 2019, 11, 2011. [CrossRef]

4. Wei, L.; Yu, M.; Zhong, Y.; Zhao, J.; Liang, Y.; Hu, X. Spatial–spectral fusion based on conditional random fields for the fine classification of crops in UAV-borne hyperspectral remote sensing imagery. *Remote Sens.* 2019, 11, 780. [CrossRef]

5. Uddin, M.P.; Mamun, M.A.; Hossain, M.A. PCA-based feature reduction for hyperspectral remote sensing image classification. *ITEE Tech. Rev.* 2021, 38, 377–396. [CrossRef]

6. Papp, L.; Van Leeuwen, B.; Szilassi, P.; Tobak, Z.; Szatmári, J.; Árvai, M.; Mészáros, J.; Pásztor, L. Monitoring invasive plant species using hyperspectral remote sensing data. *Land* 2021, 10, 29. [CrossRef]

7. Singh, P.; Pandey, P.C.; Petropoulos, G.P.; Pavlides, A.; Srivastava, P.K.; Koutsinas, N.; Deng, K.A.K.; Bao, Y. Hyperspectral remote sensing in precision agriculture: Present status, challenges, and future trends. In *Hyperspectral Remote Sensing*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 121–146.

8. Zhong, Y.; Wang, X.; Wang, S.; Zhang, L. Advances in spaceborne hyperspectral remote sensing in China. *Geo-Spat. Inf. Sci.* 2021, 24, 95–120. [CrossRef]

9. Fangi, E.; D’Amico, G.; Francini, S.; Giannetti, F.; Lasserre, B.; Marchetti, M.; Chiriči, G. The new hyperspectral satellite PRISMA: Imagery for forest types discrimination. *Sensors* 2021, 21, 118. [CrossRef] [PubMed]

10. Lassalle, G. Monitoring natural and anthropogenic plant stressors by hyperspectral remote sensing: Recommendations and guidelines based on a meta-review. *Sci. Total Environ.* 2021, 788, 147758. [CrossRef] [PubMed]

11. Sharma, M.; Biswas, M. Classification of hyperspectral remote sensing image via rotation-invariant local binary pattern-based weighted generalized closest neighbor. *J. Supercomput.* 2021, 77, 5528–5561. [CrossRef]

12. Meng, S.; Wang, X.; Hu, X.; Luo, C.; Zhong, Y. Deep learning-based crop mapping in the cloudy season using one-shot hyperspectral satellite imagery. *Comput. Electron. Agric.* 2021, 186, 106188. [CrossRef]

13. Nandibewoor, A.; Hegadi, R. A novel SMLR-PSO model to estimate the chlorophyll content in the crops using hyperspectral satellite images. *Clust. Comput.* 2019, 22, 443–450. [CrossRef]

14. Farooq, A.; Hu, J.; Jia, X. Weed classification in hyperspectral remote sensing images via deep convolutional neural network. In Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 3816–3819.

15. Lei, L.; Wang, X.; Zhong, Y.; Zhao, H.; Hu, X.; Luo, C. DOCC: Deep one-class crop classification via positive and unlabeled learning for multi-modal satellite imagery. *Int. J. Appl. Earth Obs. Geoinf.* 2021, 105, 102598. [CrossRef]

16. Pi, W.; Du, J.; Bi, Y.; Gao, X.; Zhu, X. 3D-CNN based UAV hyperspectral imagery for grassland degradation indicator ground object classification research. *Ecol. Inform.* 2021, 62, 101278. [CrossRef]

17. Wei, L.; Wang, K.; Lu, Q.; Liang, Y.; Li, H.; Wang, Z.; Wang, R.; Cao, L. Crops fine classification in airborne hyperspectral imagery based on multi-feature fusion and deep learning. *Remote Sens.* 2021, 13, 2917. [CrossRef]

18. Zhong, Y.; Hu, X.; Luo, C.; Wang, X.; Zhao, J.; Zhang, L. WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF. *Remote Sens. Environ.* 2020, 250, 112012. [CrossRef]

19. Wang, W.; Li, Y.; Zou, T.; Wang, X.; You, J.; Luo, Y. A novel image classification approach via dense-MobileNet models. *Mob. Inf. Syst.* 2020, 2020, 7602384. [CrossRef]

20. Bock, S.; Weiß, M. A proof of local convergence for the Adam optimizer. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 30 September 2019; pp. 1–8.

21. Hameed, Z.; Garcia-Zapirain, B. Sentiment classification using a single-layered BiLSTM model. *IEEE Access* 2020, 8, 73992–74001. [CrossRef]

22. Jain, M.; Singh, V.; Rani, A. A novel nature-inspired algorithm for optimization: Squirrel search algorithm. *Swarm Evol. Comput.* 2019, 44, 148–175. [CrossRef]

23. Zhong, Y.; Wang, X.; Xu, Y.; Wang, S.; Jia, T.; Hu, X.; Zhao, J.; Wei, L.; Zhang, L. Mini-UAV-borne hyperspectral remote sensing: From observation and processing to applications. *IEEE Geosci. Remote Sens. Mag.* 2018, 6, 46–62. [CrossRef]