forests

Article

DBMF: A Novel Method for Tree Species Fusion Classification Based on Multi-Source Images

Xueliang Wang [1,2] and Honge Ren [1,3,*]

1 College of Information and Computer Engineering, Northeast Forestry University, Harbin 150040, China; wangxueliang@qqhru.edu.cn
2 Network Information Center, Qiqihar University, Qiqihar 161006, China
3 Forestry Intelligent Equipment Engineering Research Center, Harbin 150040, China
* Correspondence: nefu_rhe@163.com

Abstract: Multi-source data remote sensing provides innovative technical support for tree species recognition. Tree species recognition is relatively poor despite noteworthy advancements in image fusion methods because the features from multi-source data for each pixel in the same region cannot be deeply exploited. In the present paper, a novel deep learning approach for hyperspectral imagery is proposed to improve accuracy for the classification of tree species. The proposed method, named the double branch multi-source fusion (DBMF) method, could more deeply determine the relationship between multi-source data and provide more effective information. The DBMF method does this by fusing spectral features extracted from a hyperspectral image (HSI) captured by the HJ-1A satellite and spatial features extracted from a multispectral image (MSI) captured by the Sentinel-2 satellite. The network has two branches in the spatial branch to avoid the risk of information loss, of which, sandglass blocks are embedded into a convolutional neural network (CNN) to extract the corresponding spatial neighborhood features from the MSI. Simultaneously, to make the useful spectral feature transfer more effective in the spectral branch, we employed bidirectional long short-term memory (Bi-LSTM) with a triple attention mechanism to extract the spectral features of each pixel in the HSI with low resolution. The feature information is fused to classify the tree species after the addition of a fusion activation function, which could allow the network to obtain more interactive information. Finally, the fusion strategy allows for the prediction of the full classification map of three study areas. Experimental results on a multi-source dataset show that DBMF has a significant advantage over other state-of-the-art frameworks.

Keywords: tree species classification; deep learning fusion method; multi-source images classification

1. Introduction

1.1. Background and Problem

Tree species information is being given an increasing amount of attention as a core indicator in forest resource surveys, as the accurate assessment of the composition of tree species in the forest environment will be an asset for forest ecologists, land managers, and commercial harvesters. It can also be used to study biodiversity patterns, estimate timber stocks, or improve the understanding of forest fire risk [1]. Accurate tree species distribution information is crucial for effective forest resource protection efforts. However, the wide range and large area distribution of forests make it excessively time-consuming and labor-intensive to manually identify the tree species they contain. As such, it is difficult to satisfy the requirements for real-time forest monitoring and protection. There is a demand for an effective classification framework to recognize the distribution information of large-area tree species [2]. Remote sensing provides detailed, spectrally rich, continuous spatial, and multi-temporal information, allowing tree species to be identified on the basis of their spectral and structural characteristics. Although satellite remote sensing sensor

technology is rapidly changing, the high-resolution information of forests requires new classification algorithms to bridge the gap between their needs and the wealth of data information. As such, the present study aimed to improve existing automatic tree species classification techniques [3].

In oversized and hard-to-access areas, remote sensing has been utilized to resolve the deficiencies of traditional field investigation [4]. Within the past four decades, advances in remote sensing technology have made it possible to classify tree species using several image types [5]. Certain remote sensing images limit the ability to recognize tree species due to their relatively low spatial resolution (LR), even when they contain large amounts of information, such as for Sentinel or Landsat data [6]. LiDAR data can provide many structural features of trees [7]. A hyperspectral image (his) is composed of many continuous narrow bands that reveal the spectral details of different tree species [8]. However, as most trees are mixed and marshy plants, it is hard to recognize species in HSIs due to their LR [9]. Conversely, multispectral images (MSIs) provide abundant spatial and contextural information because of their high spatial resolution (HR) [10]. Multispectral images are the most common images used in forest areas [11,12]. Remote sensing images such as LANDSAT MSS, Landsat TM, Spot, and Sentinel-2 have been used for wetland monitoring, land cover mapping, and tree species identification [13–15]. The fusion of MSI ahisHSI information can integrate a variety of relevant features to improve the accuracy of tree species classification. A wide variety of features are reflected in the differences hisng HSI and MSI spectral characteristics [16]. The high spectral resolution and subtle divergence of HSIs are advantageous in tree species identification [17,18]. A single source of spectral information has certain disadvantages that may prevent the accurate classification of tree species in complex areas. Hyperspectral and MS data fusion technology can improve recognition rates. Classification accuracy can also be further improved by the deep learning (DL) of fused spectral and spatial information [19].

1.2. Deep Learning Algorithm

Deep learning is a research subfield in machine learning that aims to build abstract hierarchical models within datasets, employing an incremental approach. Inspired by the deep structure function of the human brain, DL algorithms activate the effective information by many layers of nonlinear transformation operations to compose a learning model to build a mapping relationship between input data and output data [20]. Recently, the DL method has exhibited success for tree species detection [21], crop classification [22], and HSI classification [23]. Deep learning has demonstrated superior results over other commonly used classifiers for trees recognition [24,25]. Franklin et al. [11] classified four tree species by employing random forest (RF) classifier with MSI data captured by a rotating sensor on an unmanned aerial vehicle (UAV) in a broadleaf forest, and the overall accuracy (OA) of classification results reached 78%. Pölönen et al. [26] applied a three-dimensional convolutional neural network (3D-CNN) to recognize three main tree species with HS data from a boreal forest in Finland, and the OA reached 0.96 on the validation dataset by applying the proposed CNN model. Contextual interactions were obtained by exploiting the local spatial–spectral relationship of adjacent pixel vectors in a square window in the contextual deep CNN (CD-CNN) [27]. Xu [28] proposed the long short-term memory (LSTM) model based on frequency band grouping and a multi-scale CNN as spectral and spatial feature extractors, respectively. The dual CNN branch structure proposed by Yang [29] can be used to extract spectral features from low-resolution HSIs and spatial neighborhood feature information from high-resolution MSIs.

There is an abundance of remote sensing data available concerning the classification of tree species, but they have not been deeply mined for potential information [30]. Liu and Wang [31] identified the tree species and estimated stock volume with the VGG16 and UNET model. Because there is no better classification strategy to solve the problem of the accurate classification of regional tree species, it is not yet possible to further improve the fusion classification strategy. In recent years, Sentinel-2 satellite MS data have been applied

in tree species classification technologies because of their low cost and HR [6]. With the DL method, it is possible to improve the existing classification strategy by combining the data of two satellites and extracting feature information for tree species recognition [25]. Although existing tree species classification network designs are relatively sophisticated, they are rarely used with multi-source networks due to their complexity or relatively low accuracy. Firstly, these DL networks can capture finer features, although the training work will be more difficult and very time-consuming and resource-intensive. Secondly, the remaining representation is compressed and connecting identity mappings between thin bottlenecks will inevitably lead to information loss. In addition, due to feature dimensionality reduction, gradient confusion also weakens the ability of gradients to propagate between layers, affecting the training convergence and model performance. Only efficient feature selection and deeper depth mining can eliminate the redundancy to retain the depth features and ensure that the tree species are correctly classified. When the spatial and spectral block are adjusted, this type of classic DL structure can be applied with multi-source networks, yielding state-of-the-art results. Thus, designing more efficient network architectures is essential for yielding efficient models [32].

1.3. Research Objectives

This study aimed to improve the network identification ability for challenging tree species classification tasks by proposing a network for tree species classification wherein CNN and LSTM are fused with multi-source data based on HJ-1A and Sentinel-2 remote sensing images. Sentinel-2 data are used as an MSI, which has HR, and HJ-1A data are used as an HSI, which has LR. To exploit the correlation between the HSI and MSI, we used the spectrum of LR HSI and the corresponding spatial neighborhoods in HR MSI as the input pairs of the network. Features were extracted from the corresponding neighborhoods in the LR HSI and the MSI with two CNN branches. Then, these branches were connected and fed to the fusion activation layer. The final fusion layer output the spectrum classification map.

The main objectives of this work can be summarized as follows:

(1) To design a framework for tree species classification with multi-source data and a DL algorithm.
(2) To assess the performance of the proposed method for tree species classification using HSI data and MSI data.
(3) To analyze the advantages of the proposed model with other models to provide new ideas for the application of the DL method in forestry.

2. Materials and Methods

2.1. Study Area

The experiment was conducted in the Tahe Forestry Bureau (123° to 125° E and 52° to 53° N) located in the center of Daxing'an Mountains, northwest of Chinese Heilongjiang Province, with a border line of 173 km and a total area of 14,420 km^2 (Figure 1). The climate is a cold–temperate continental climate and experiences severe climatic changes, with long dry and cold winters and short hot and humid summers; the annual average temperature is -2.4 °C, and the annual average precipitation is 463.2 mm, which mainly occurs in July and August. The forest covers 81% of the total area, with a storage capacity of 53.4 million m^3. Dominant tree species include birch, larch, spruce, mongolica pine (shortened as mongolica), willow, and poplar [33].
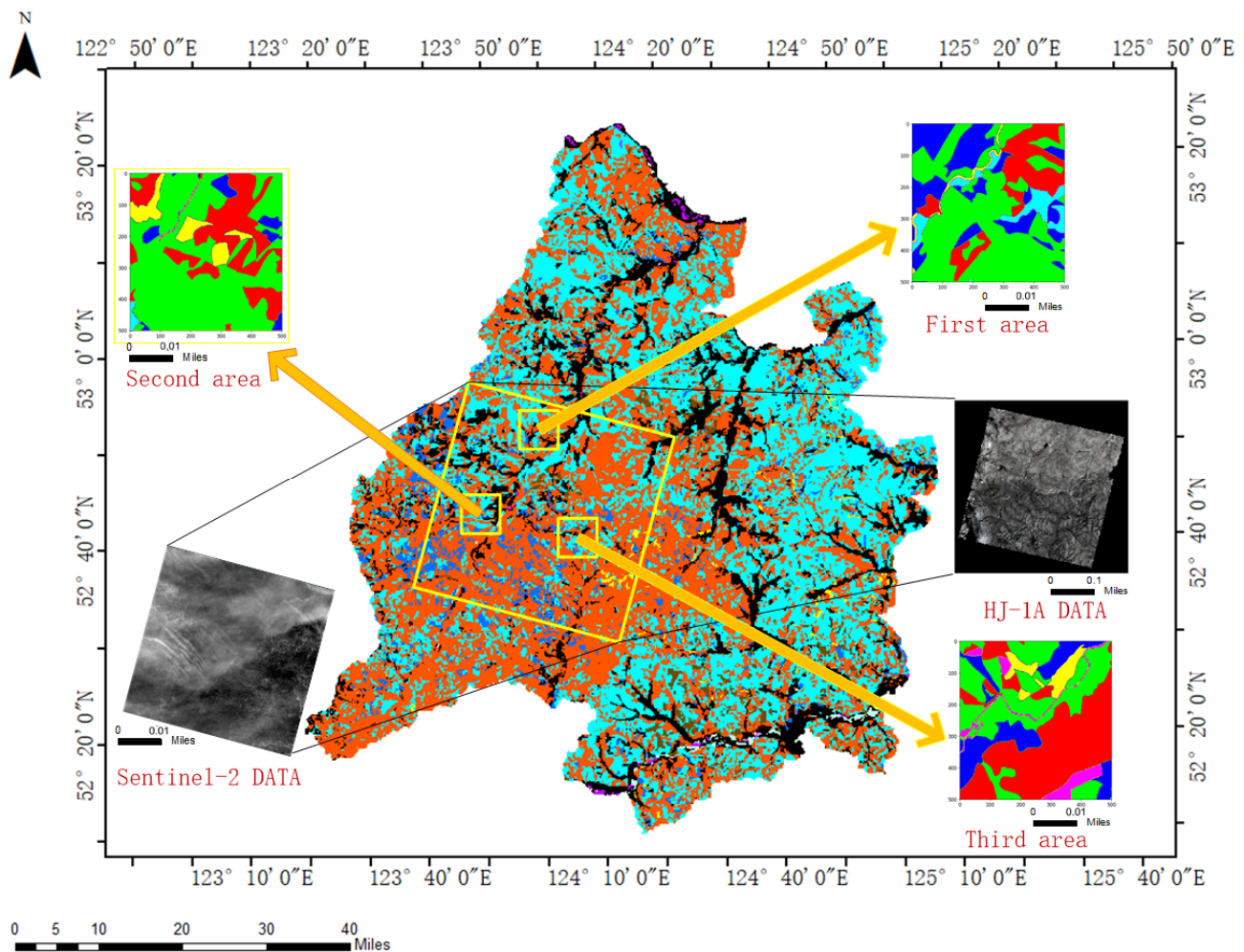
Figure 1. Map of the study area.

2.2. Data

Data captured by HJ-1A and Sentinel-2 were used for tree species classification. The HSI data of HJ-1A and the MSI data of Sentinel-2A were obtained from the official websites of the China Center for Resources Satellite Data and Application and the United States Geological Survey (USGS), respectively, which is presented in Figure 1.

The HJ-1A satellite was equipped with an HS imaging instrument with 115 bands. The images have a 100 m spatial resolution. The Sentinel-2A data has 13 spectral bands [34]. It provides abundant data information for the field of land and coastal remote sensing [35]. To supply the gap of the relatively LR of the HSI, we used the bilinear interpolation method to improve the resolution of the HJ-1A/HSI image (both datasets were taken on 20 August 2016) using ENVI 5.1 software to increase the spatial resolution of the HJ-1A/HSI image to 10 m. The experimental HSI data were resampled by the interpolation algorithm so that it had the same 10×10 m^2 spatial resolution as the MSI on the same ground. The classification of dominant forest species was performed in the study area from second-class data surveyed by the Tahe Forestry Bureau in 2018.

Because of the large area of the forest, the authors attempted to select the area with the most species as the research object, and the size of the three study areas was $500 \times 500 \times 115$ pixels and $500 \times 500 \times 13$ pixels for the HSI and MSI data, respectively. The dominant tree species were birch, larch, spruce, mongolica, willow, and poplar. Table 1 lists the tree species of the present article's three study areas, where the training samples approximately accounted for a third of the total samples.

Table 1. List of 6 tree species samples of the three study areas.

	Birch	**Larch**	**Mongolica**	**Poplar**	**Spruce**	**Willow**
First area	130,124	39,216	57,620	3019	15,330	3492
Second area	150,771	58,829	11,412	2175	17,048	1067
Third area	99,082	82,746	38,114	1013	13,460	15,486

2.3. Classification Models

2.3.1. Bi-LSTM

Long short-term memory is a form of recurrent neural network (RNN) [36], which learns long-term dependency information via feedback connections. To acquire short-term memory and abandon long-term memory, the model captures time-series information and sequence data by cyclic connections on their hidden layers. Recurrent neural networks have gradient vanishing characteristics that destabilize the long-term dependency of the model's learning process. However, LSTM resolves this problem by making the hiding layer store the latest information rather than the previous information.

Long short-term memory is operated as follows:

$$f_t = \sigma\left(W_{hf} \cdot h_{t-1} + W_{xf} \cdot x_t + b_f\right), \tag{1}$$

$$i_t = \sigma(W_{hi} \cdot h_{t-1} + W_{xi} \cdot x_t + b_i), \tag{2}$$

$$\widetilde{C}_t = tanh(W_{hC} \cdot h_{t-1} + W_{xC} \cdot x_t + b_C), \tag{3}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C}_t, \tag{4}$$

$$O_t = \sigma(W_{ho} \cdot h_{t-1} + W_{xo} \cdot x_t + b_o), \tag{5}$$

$$h_t = O_t \cdot tanh(C_t). \tag{6}$$

Long short-term memory is composed of four parts: a forget gate f_t, input gate i_t, output gate O_t, and candidate cell value $\widetilde{C}_t$. Here, x_t is the input; O_t is the output; and b_f, b_i, b_C, and b_o are bias terms. Furthermore, σ denotes the activation function, '·' denotes the matrix multiplication operator, and $W_{\times\times}$ denotes the weight matrix.

According to LSTM, bidirectional LSTM (Bi-LSTM) integrates the input sequence information in both direct and inverted directions. The output at time t includes the forward LSTM output and reverse LSTM output [37].

$$\overrightarrow{h_t} = \varnothing\left(W_{hh}^d \cdot h_{t-1} + W_{xh}^d \cdot x_t + b_h^d\right), \tag{7}$$

$$\overleftarrow{h_t} = \varnothing\left(W_{hh}^i \cdot h_{t-1} + W_{xh}^i \cdot x_t + b_h^i\right), \tag{8}$$

where $\overrightarrow{h_t}$, $\overleftarrow{h_t}$ denote the hidden layer in forward and reverse LSTM directions at time t, respectively. The hidden state h_t is obtained by concatenating $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$, which is then fed to the output layer. The output O_t is calculated as follows:

$$O_t = W_{hp} \cdot h_t + b_p, \tag{9}$$

where W_{hp} denotes the weight parameter and b_p denotes the bias parameter.

2.3.2. Triple-Attention Mechanism

The attention mechanism assigns weighted values so that the spectrum with greater weight is prioritized in the module. The attention module mainly refines the weights of mapped features. Meaningful bands are given more attention by each spectral data point

in the feature map, which is effectively a feature detector. The meaningless bands are weakened.

The triple-attention mechanism is composed of three attention modules: a traditional attention mechanism, average pooling attention mechanism, and max-pooling attention mechanism. The average pooling function is used by the second of these to retain local characteristics, and, similarly, the max-pooling function is used by the last of them to retain local characteristics [38]. The three triple-attention mechanism formula models are as follows:

$$t = \sum_{j=1}^{n} \frac{exp(e_{ij})}{\sum_{m=1}^{n} exp(e_{im})} h_{ij}, \tag{10}$$

$$a = AveragePooling \left\{ \sum_{j=1}^{n} \frac{exp(e_{ij})}{\sum_{m=1}^{n} exp(e_{im})} h_{ij} \right\}, \tag{11}$$

$$m = MaxPooling \left\{ \sum_{j=1}^{n} \frac{exp(e_{ij})}{\sum_{m=1}^{n} exp(e_{im})} h_{ij} \right\}. \tag{12}$$

Multiple weights are assigned to the original features of HS data so that each feature is recognized to the greatest extent possible. Redundant features are removed before the subsequent classification experiment. The triple-attention mechanism formula is calculated as follows:

$$F_{(t,a,m)} = concatencate(t \oplus a \oplus m). \tag{13}$$

The triple-attention mechanism formula provides the new weight, which is assigned to the corresponding features, where t denotes the traditional attention mechanism value, a denotes the average pooling attention mechanism value, m denotes the maximum pooling attention mechanism value, and $F_{(t,a,m)}$ denotes the concatenated features of the three equations. Finally, '$\oplus$' denotes the operation of the feature fusion algorithm. The fusion information concatenated with the output of the triple-attention mechanism provides rich, detailed features for subsequent classification work.

2.3.3. Sandglass Block

The purpose of the sandglass block is to minimize parameters and computational cost by flipping the inverted residuals. The sandglass block uses the thickness of each block to represent the corresponding relative number of channels, which is presented in Figure 2. The residual block reverses building shortcuts between bottlenecks and includes a depth-wise convolution (detached blocks) at both ends of the residual path [32].

To preserve information from the bottom layers when transiting to the top layers and to facilitate gradient propagation across layers, we positioned shortcuts to connect high-dimensional representations. Because deep-wise convolution is relatively lightweight, higher-dimensional features can be applied to encode richer spatial information and generate more expressive representations.

Let $I \in T^{D_f \times D_f \times M}$ be the input tensor and $O \in T^{D_f \times D_f \times M}$ be the output tensor of a building block. The formulation of the proposed building block, without considering the depth-wise convolution or activation layers, is as follows:

$$O = \varnothing_e(\varnothing_r(I)) + I, \tag{14}$$

where $\varnothing_e$ denotes the two pointwise convolutions for channel expansion, and $\varnothing_r$ denotes the channel reduction. This mechanism creates a bottleneck in the middle of the residual path to save parameters and computation costs. The shortcut can be applied to connect representations with many channels instead of bottlenecks. The shortcut delivers information from the input I to the output O so that many gradients propagate across multiple layers. Depth-wise spatial convolutions are adopted to encode spatial information from the end
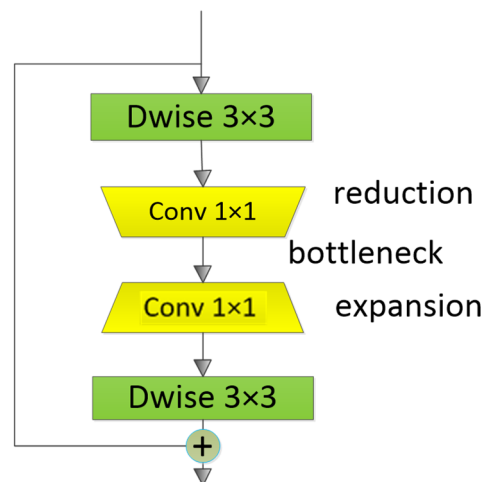
of the residual path to learn expressive spatial contextual information. The block can be formulated as follows:

$$\hat{o}_1 = \varnothing_{1,p}\varnothing_{1,d}(I), \tag{15}$$

$$\hat{o}_2 = \varnothing_{2,p}\varnothing_{2,d}(\hat{o}_1) + \hat{o}_1, \tag{16}$$

$$\hat{o}_i = \varnothing_{i,p}\varnothing_{i,d}(\hat{o}_{i-1}) + \hat{o}_{i-1}, \tag{17}$$

where $\varnothing_{i,p}$ and $\varnothing_{i,d}$ are the i-th point-wise convolutions and depth-wise convolutions, respectively. Both convolutions are conducted into high-dimensional spaces to extract rich features.

Figure 2. Sandglass block with bottleneck structure. Note: The bottleneck falls into the middle of the depth-wise convolutions, which have high-dimensional representations, so that the module encodes expressive spatial information.

2.3.4. Double Branch Multi-Source Fusion Network

Meaningful features are extracted by double branch multi-source fusion (DBMF) from the HSI bands that are exploited via the interpolation algorithm concerning MSI at the same pixel. The expected features of the HSI are extracted so that all the HSI bands can jointly reconstruct the feature map, which minimizes spectral distortion. Compared with other networks, DBMF has lower complexity because it requires less computation.

The proposed DBMF method is illustrated in Figure 3. The spectral information of a pixel is determined by its average reflectance spectrum, while the spatial features are associated with the surrounding pixels. In this framework, the spectral branch extracts spectral features as the spatial branch extracts spatial features; they are then fused for subsequent operations. To minimize the complexity of tree categories and maintain sample imbalance, we fused two single-source datasets by the framework of spatial and spectral branches to complete the classification process.

The HJ-1A HSI data are input to the HS branch, which is resampled by the bilinear interpolation algorithm to make its spatial resolution the same as that of the MSI captured by Sentinel-2. The HS input is composed of different spectral sizes of the HSI paths to fully exploit the structural information of the images. The Bi-LSTM first extracts the HSI data characteristics, then assigns multiple weights to each spectral feature through the triple-attention mechanism. Dense networks then perform residual calculations for numerous shallow features to generate representative residual features.

The spectral branch implementation is described in Table 1. The kernel size of the bidirectional layer is $(1 \times 1 \times 7)$. Feature maps in the shape of $(7 \times 7 \times 47, 24)$ are obtained, then permuted to the triple-attention block to make a size of $(7 \times 7 \times 47, 24)$. A feature map in the shape of $(7 \times 7 \times 1, 60)$ is then obtained, and the BN-Mish-Conv block calculates a $(1 \times 1 \times 47)$ kernel. The BN-Lfu-Conv block consists of a batch normalization (BN) layer

and an active function with an Lfu unit (inspired by Mish [39]), as well as an individual convolution layer.

The MSI data captured by the MS Sentinel-2 satellite are input into the MS branch. The MSI has a much higher spatial resolution than the HSI, and therefore they are subject to very different spatial scales. Each convolutional layer is followed by a BN-Mish-Conv block. The sandglass block is applied to eliminate redundant spatial features, the output of which is a combination of deep and shallow learning CNNs that mine and merge deep spatial characteristics.
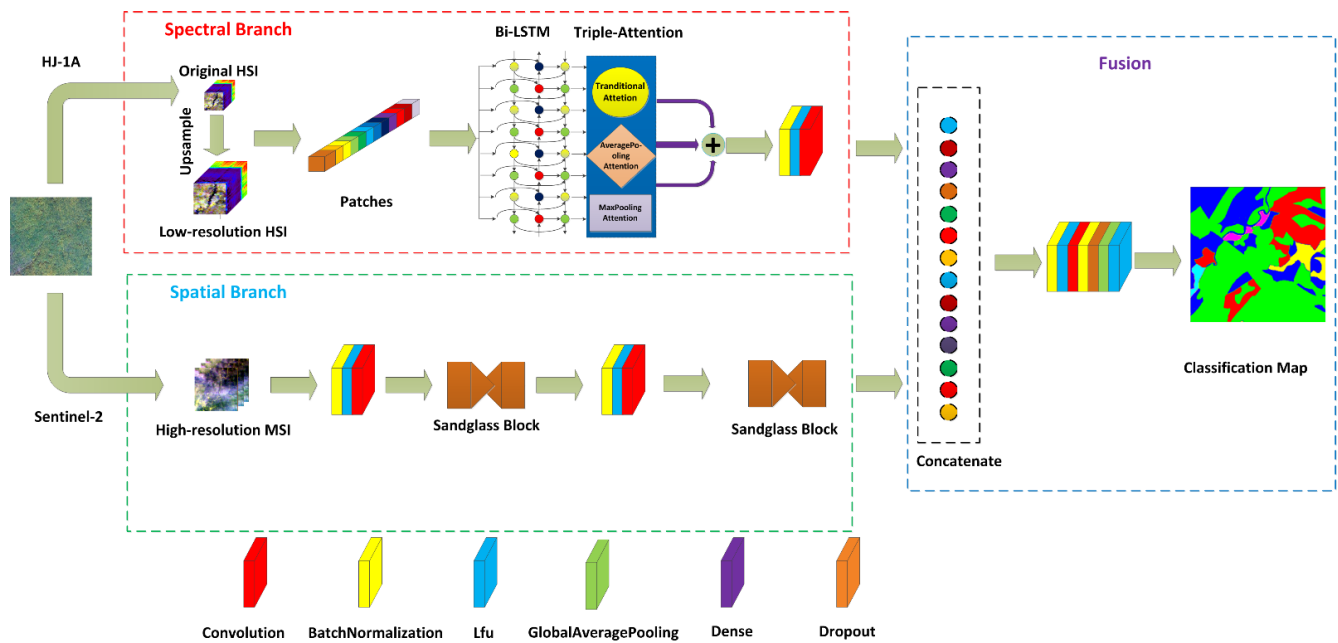

Figure 3. Architecture of the proposed DBMF method.

The red dotted branch of the network employs the expected bands grouped by Bi-LSTM and the triple-attention model to extract spectral characteristics. The green dotted branch of the network employs the convolution and sandglass block to extract spatial characteristics. In the right architecture, the spectral information and spatial information are combined to form a new layer for spectral-spatial classification. The details of the spatial branch are similar with the spectral branch, which is shown in Table 1. The layer is a basic structure composed of a BN and Mish activation function plus a 3D-CNN with a kernel size of $(3 \times 3 \times 1)$ and 24 channels. A sandglass block is applied to the network to change the feature map shape into $(7 \times 7 \times 1, 24)$. These two steps are repeated to finally bring a spatial shape of (1×60) to the next fusion layer.

The spectral and spatial feature maps are obtained by spectral and spatial branches, respectively. The two features are fused together for subsequent operations to account for relevant domains that do not contain the two features. Finally, the full classification result is obtained with the BN-Lfu-Conv block and BN-Drop-Pool-Lfu block calculated in sequence.

In DBMF, concatenation is applied as the activation function. An appropriate activation function can accelerate the counter-propagation and convergence of the network. The complete loss function of the proposed DBMF model is shown to be the following:

$$L = \alpha L_{HS} + \beta L_{MS} + \gamma L_{fusion}, \tag{18}$$

$$L_{fusion} = -\frac{1}{n} \sum_{i=1}^{n} \left[y_i \log\left(\hat{y}_{fusion,i}\right) + (1 - y_i) \log\left(1 - \hat{y}_{fusion,i}\right) \right], \tag{19}$$

$$L_{HS} = -\frac{1}{n} \sum_{i=1}^{n} [y_i \log(\hat{y}_{HS,i}) + (1 - y_i) \log(1 - \hat{y}_{HS,i})], \tag{20}$$

$$L_{MS} = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_{MS,i}) + (1 - y_i)\log(1 - \hat{y}_{MS,i})], \tag{21}$$

where L_{fusion} is the main loss function; L_{HS} and L_{MS} are the two auxiliary loss functions; $\hat{y}_{HS}$, $\hat{y}_{MS}$, and $\hat{y}_{fusion}$ are the corresponding predicted labels for the i-th training sample; and α, β, and γ are the scalar weights. For convenience, the scalar weights were fixed to 1 in the experiments of the present study. Furthermore, y_i is the true label and n is the size of the training set [40]. The model is trained and optimized employing end–end and stochastic gradient descent algorithm, respectively.

2.3.5. Evaluation Indicators

To test the tree species classification accuracy of the proposed method, we used 70% of field samples of the test as validation samples, independent of the training samples. The OA, average accuracy (AA), and Kappa coefficient (kappa) were determined using Equations (22)–(24), respectively.

$$OA = \frac{\sum_{i=1}^{k} C(i,i)}{M}, \tag{22}$$

$$AA = \frac{\sum_{i=1}^{k} OA}{K}, \tag{23}$$

$$kappa = \frac{M \sum_{i=1}^{k} C(i,i) - \sum_{i=1}^{k}(C(i,+)C(+,i))}{M^2 - \sum_{i=1}^{k}(C(i,+)C(+,i))}, \tag{24}$$

where OA represents the proportion of correctly classified samples in the whole test sample, AA denotes the average accuracy of every tree species, and $kappa$ is a statistical measure that reflects the consistency between the ground truth and classified ground maps. For comparison, 30% of the sample was randomly assigned to the training group with the remaining data assigned to the test group.

3. Results

The proposed method for tree species data classification was compared with the traditional classifier support vector machines (SVM) [41], as well as recently developed CNN-type methods, CD-CNN [27], double-branch multi-attention mechanism network (DBMA) [42], and double-branch dual-attention mechanism network (DBDA) [43]. Because SVM does not belong to DL, only the other four algorithms were learned in the training process. All the methods were tested on a single HSI, single MSI, and fused HSI and MSI with data resolution upscaled to the same pixel level. The development environment of the algorithms is listed as follows: (1) Windows 10, (2) Xeon (R) CPU 2.30 GHz processor, (3) 32 GB RAM, (4) NVIDIA Tesla P100 GPU, (5) Python 3.7, and (6) Tensorflow Keras framework. The first subsection demonstrates the training process of the DBMF model.

3.1. Parameter Adjustment

Parameter setting and adjustment are important parts of the DBMF test because they reflect network performance. To ensure fairness in the experimental results, the multi-data fixed resolution was given the same ground truth map, although the two types of data had different resolutions. Table 2 describes each layer of the DBMF method. In the experiment, the proposed method and the other methods operated for comparison were based on the same ground truth map.

Table 2. DBMF architecture.

| HSI Tunnel			MSI Tunnel		
Layer	**Kernel Size**	**Output Shape**	**Layer**	**Kernel Size**	**Output Shape**
Input		$(7 \times 7 \times 100)$	Input		$(7 \times 7 \times 100)$
Bidirectional	$(1 \times 1 \times 7)$	$(7 \times 7 \times 47, 24)$	BN-Lfu-Conv	$(3 \times 3 \times 1)$	$(7 \times 7 \times 1, 24)$
Permute		$(7 \times 7 \times 47, 24)$	Concatenate		$(7 \times 7 \times 1, 24)$
Triple-Attention		$(7 \times 7 \times 47, 24)$	Sandglass Block		$(7 \times 7 \times 1, 24)$
BN-Lfu-Conv	$(1 \times 1 \times 47)$	$(7 \times 7 \times 1, 60)$	BN-Lfu-Conv	$(3 \times 3 \times 1)$	$(7 \times 7 \times 1, 24)$
Pooling	$(7 \times 7 \times 1)$	$(1 \times 1 \times 1, 60)$	Concatenate		$(7 \times 7 \times 1, 24)$
Concatenate		(1×60)	Sandglass Block		$(7 \times 7 \times 1, 24)$
			Concatenate		(1×60)
Layers (Fusion)			Output Shape (Fusion)		
JOINT			(1×60)		
BN-MISH-Conv			$(7 \times 7 \times 1, 60)$		
BN-Dropout-Pooling-Lfu			(1×60)		

Training the DBMF model takes a long amount of time due to its complex branch structure and numerous parameters, and therefore it was repeatedly tested. The input kernels were calculated and characterized by OA and AA. The network model obtained the best performance with the input kernel of (7×7), the details of which will be described in the next section. The Adam strategy was also used in the network with a learning rate set to 0.01 as per the desired training performance and convergence speed of the network [44].

3.2. Training DBMF

During the training of the CNN model, the size of the convolution kernel has a great impact on classification accuracy [45]. If the convolution kernel is set to be larger, the detailed information will be lost, but the associated data will be retained. On the contrary, if the size is set to be smaller, the result is just the opposite. Therefore, the size of the convolution kernel directly affects the results of tree species classification. In this experiment, the most common sizes of convolution kernels are used with n, $n = 1, 3, 5, 7, 9, 11$, and 13. The model was trained with a batch size of 100 samples during 1000 epochs, and the result of which is presented in Figure 4. From Figure 5, it can be seen that the size 7 had the highest accuracy rate. Size 9 converged rapidly in the early stage, which almost reached the peak in 800 cycles, but the accuracy rate did not rise from then on, and therefore the learning ability was limited. However, from the subfigure of the size 7 learning curve, the accuracy still improved in the stage of 800 epochs, the gradient of which did not completely disappear. Although size 1 and size 3 were stable, the accuracy rate was unable to satisfy the given request. The curves of size 11 and size 13 fluctuated considerably, which were not suitable for mass machine learning operations.
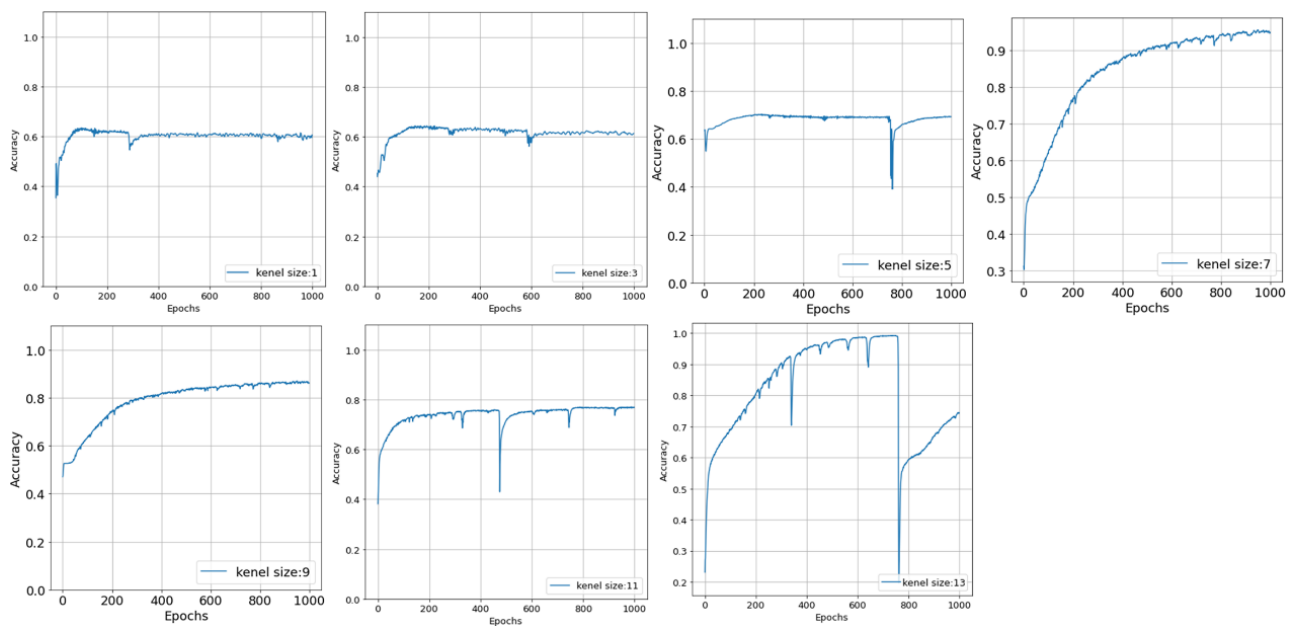
Figure 4. The accuracy curve of DBMF with different kernel sizes. Note: The subfigures are ordered by sizes 1, 3, 5, 7, 9, 11, and 13.
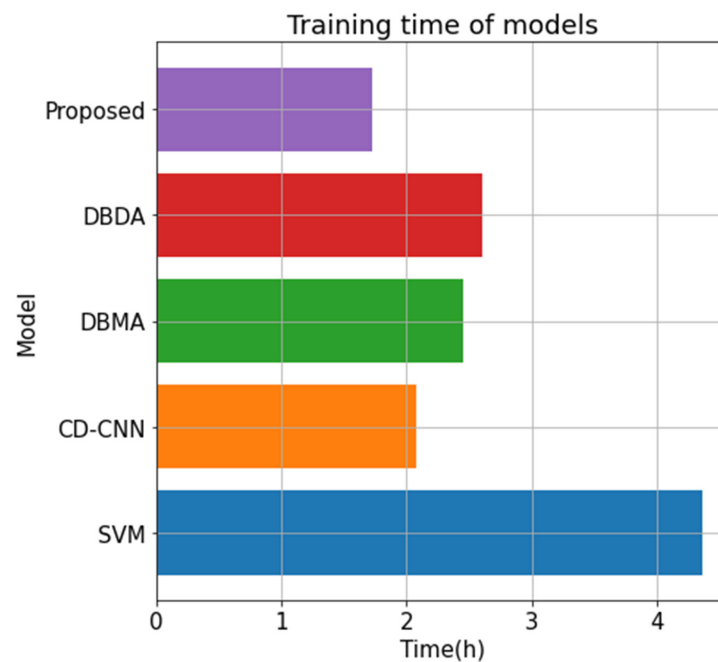

Figure 5. Training time.

The size of the parameter file of the DBMF with all layers was 303 megabytes. The stochastic gradient descent method was also used for training. Figure 5 shows the training time consumed by the tested models, where DBMF took approximately 100 min, which was considerably faster (30 min) than CD-CNN, whose learning speed was the fastest in other compared methods, while SVM took the longest time. It is shown that DBMF was the fastest learning model compared with the others, and the DL capability was better than traditional machine learning in tree species classification.

Figure 6d shows the changes in accuracy and loss rate in the training process of the model. Increasing the number of iterations changed the whole network model so the complete training process can be observed in time and avoid overfitting. As the number of

iterations slowly increased, the accuracy also gradually increased. When the epoch number approached 400, the accuracy tended to be stable. Similarly, as the number of iterations increased, the loss value gradually decreased. When the number of cycles reached about 300, the loss rate tended to be stable, but when it exceeded 400, the loss value began to rise and overfitting occurred. The 400 iterations of the model had the lowest accuracy and stability loss.
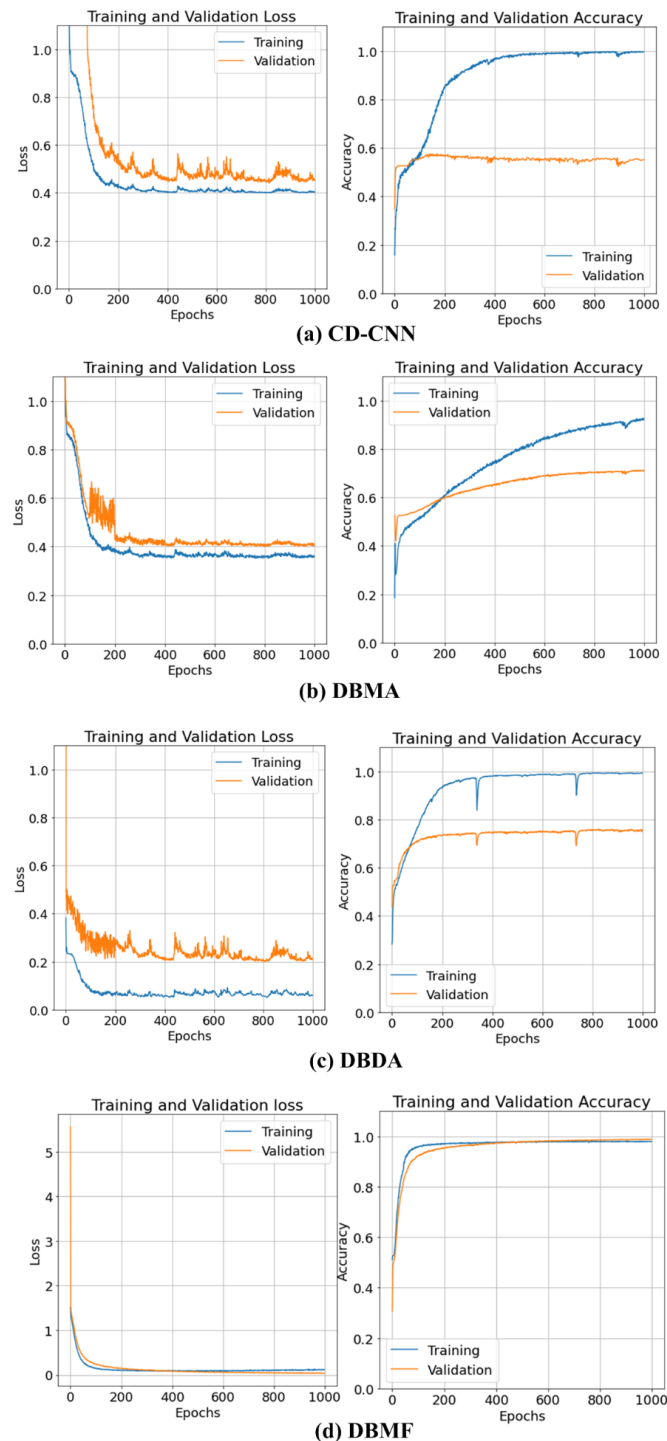
Figure 6. The loss and accuracy of compared models. Note: (**a**) The loss and accuracy of CD-CNN, (**b**) The loss and accuracy of DBMF. (**c**) The loss and accuracy of DBDA. (**d**) The loss and accuracy of DBMF.

Since this paper used the stochastic gradient descent method to pre-train the DBMF model, it can be seen from Figure 6a–c that the training accuracy of these three algorithms for the datasets was general, but the loss rate was still relatively low. After 100 iterations of training, the function curve did not converge, but, rather, it diverged. Therefore, these three methods were not suitable for the presented datasets.

Figure 7 shows that OA changed slowly with the slow increase of the number of samples. They all had one point in common: the more training samples, the higher the accuracy. However, in each stage, the method proposed in this paper was better than the other tested methods, which showed its strong ability to learn spatial and spectral features. The OA of the proposed model reached the peak of 92% when the proportion of training samples accounted for 80%, and then tended to be stable. The accuracy of other algorithms was also stable, but the accuracy was lower than the proposed DBMF method.
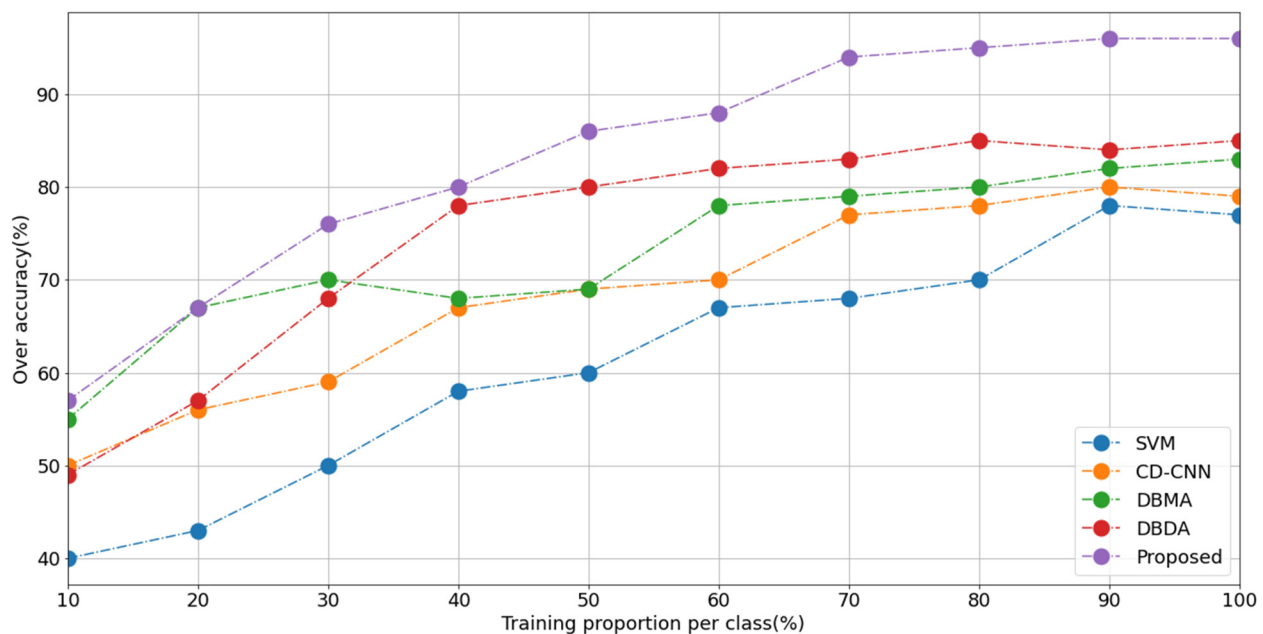
Figure 7. Accuracy with the proportion of training samples.

3.3. Detailed Results of State-of-the-Art Networks

All methods were tested five times, and the results were averaged from the three study areas, of which, the OA, AA, and kappa detailed statistics are listed in Table 3.

Table 3. Classification accuracy (%) comparison of three areas.

NO.	SVM			CD-CNN			DBMA			DBDA			DBMF		
	HJ1	Sen2	HJ + Sen2	HJ1	Sen2	HJ + Sen2	HJ1	Sen2	HJ + Sen2	HJ1	Sen2	HJ + Sen2	HJ1	Sen2	HJ + Sen2
OA	40.67	36.31	42.60	47.64	44.35	48.66	62.13	60.43	63.05	81.67	81.57	82.12	84.32	82.62	90.84
AA	36.39	34.58	36.76	44.38	42.31	46.72	56.79	55.60	58.32	80.18	78.98	82.09	83.60	82.31	90.16
Kappa	31.12	30.89	33.01	43.56	42.08	45.19	50.94	54.27	55.84	77.32	72.15	81.37	83.18	84.91	90.02

Compared with SVM, DBMF improved the recognition rate by 48% (Tables 3 and 4). The SVM produced the worst classification (0.43), whereas DBMF produced the best result (0.91) among the methods tested because it associated spatial and spectral neighborhood information of the same pixel, unlike SVM, which used them independently. The OA of CD-CNN was the worst among the methods, as 2D-CNN did not integrate spatial and spectral information. Compared with DBMA, which extracts two types of features

independently and employs an attention mechanism, DBMF increased the OA by 27%. The DBMF had better classification effects by extracting multiple-scale information on MSIs and HSIs compared with the DBDA method, which has an improved activation function but does not prevent overfitting.

Table 4. The tree classification confusion matrix for tree species classification using ground truth (rows) compared to the SVM model species prediction (columns; numbered labels refer to species numbers shown in row labels).

Species	Species Code	0	1	2	3	4	5	Recall (%)	F-Score (%)
Birch	0	4623	7142	0	0	0	0	39.29	35.72
Larch	1	4280	35,357	0	0	0	0	89.20	70.13
Spruce	2	4334	14,268	0	0	0	0	0	0
Mongolica	3	812	2531	0	0	0	0	0	0
Willow	4	45	1003	0	0	0	0	0	0
Poplar	5	22	884	0	0	0	0	0	0
	Precision	32.7501	57.78	16.66	16.66	16.66	16.66		

As shown in Tables 5–8, the recognition rate of CD-CNN, DBMA, and DBDA models was low, except for the DBMF model. In general, the classification for spruce was difficult for the models, but the recognition rate of DBMF was higher than the other models for spruce, which also improved the OA. The classification performance of the SVM model was the worst, which only recognized larch and birch so that almost no other tree species were recognized. Obviously, there were defects in high-dimensional and large batch data. For instance, CD-CNN incorrectly classified spruce into larch, the percentage of this almost reached 75%, while the recognition rate of spruce was zero. The recognition rate of DBMA for willow was 0.61, which was better than CD-CNN. In addition to the equivalent effect of spruce, the recognition ability of other tree species was slightly stronger than CD-CNN. As shown in Tables 6 and 7, the recognition rate of DBDA for willow was 0.9, which was better than DBMA, but it was not significantly enhanced for spruce. The classification average accuracy of the proposed DBMF for six tree species was found to generally be better than the other compared algorithms.

Table 5. The tree classification confusion matrix for tree species classification using ground truth (rows) compared to the CDCNN model species prediction (columns; numbered labels refer to species numbers shown in row labels).

Species	Species Code	0	1	2	3	4	5	Recall (%)	F-Score (%)
Birch	0	8036	3423	47	259	0	0	68.30	62.98
Larch	1	2972	35,991	158	515	1	0	90.8	77.67
Spruce	2	1860	11,217	5078	447	0	0	27.29	42.24
Mongolica	3	705	1030	140	1468	0	0	43.91	48.15
Willow	4	19	688	5	11	325	0	31.01	47.3
Poplar	5	159	682	11	54	0	0	0	0
	Precision	58.43	67.86	93.36	53.30	99.69	16.66		

Table 6. The tree classification confusion matrix for tree species classification using ground truth (rows) compared to the DBMA model species prediction (columns; numbered labels refer to species numbers shown in row labels).

Species	Species Code	0	1	2	3	4	5	Recall (%)	F-Score (%)
Birch	0	8592	2400	420	341	0	12	73.03	69.55
Larch	1	3140	35,666	317	475	0	39	89.98	80.59
Spruce	2	632	8501	9301	149	0	19	50	64.91
Mongolica	3	562	1267	0	1511	0	3	45.19	51.70
Willow	4	0	701	14	26	307	0	29.29	45.31
Poplar	5	16	336	0	0	0	554	61.14	72.27
	Precision	66.38	72.97	92.52	60.39	99.9	88.35		

Table 7. The tree classification confusion matrix for tree species classification using ground truth (rows) compared to the DBDA model species prediction (columns; numbered labels refer to species numbers shown in row labels).
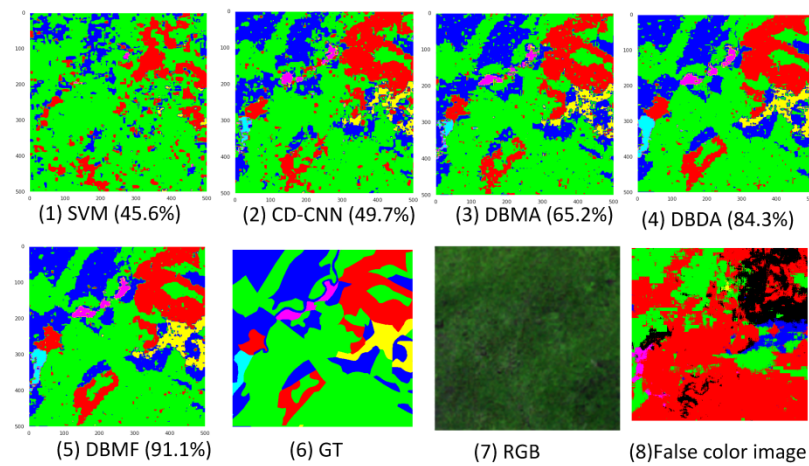
Species	Species Code	0	1	2	3	4	5	Recall (%)	F-Score (%)
Birch	0	10,817	710	18	219	0	1	91.94	79.25
Larch	1	3902	35,259	79	356	2	39	88.95	86.96
Spruce	2	390	4966	13,002	244	0	0	69.89	81.97
Mongolica	3	394	327	10	2612	0	0	78.13	76.22
Willow	4	0	134	10	79	825	0	78.72	88
Poplar	5	30	58	0	0	0	818	90.28	92.74
	Precision	69.63	85.05	99.1	74.41	99.75	95.33		

Table 8. The tree classification confusion matrix for tree species classification using ground truth (rows) compared to the DBMF model species prediction (columns; numbered labels refer to species numbers shown in row labels).
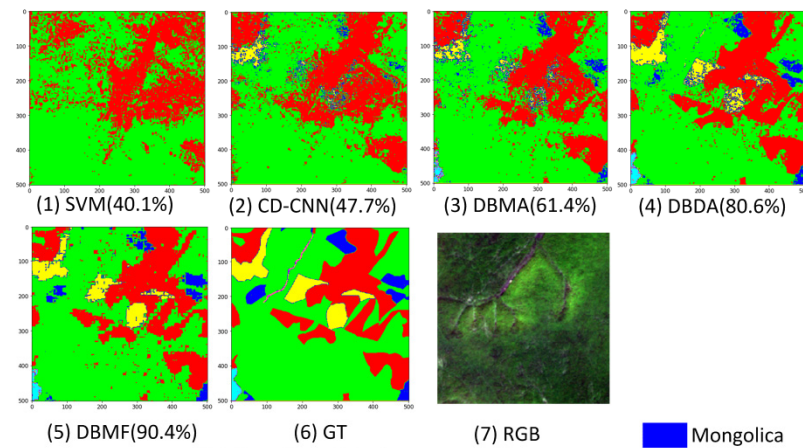
Species	Species Code	0	1	2	3	4	5	Recall (%)	F-Score (%)
Birch	0	10,991	275	109	390	0	1	93.49	83.59
Larch	1	2869	36,993	744	19	10	2	90.82	93.63
Spruce	2	482	795	17,241	19	57	8	92.67	93.37
Mongolica	3	126	131	140	2903	22	20	86.74	86.98
Willow	4	73	23	54	0	898	0	85.87	88.36
Poplar	5	1	36	44	0	0	825	91.05	93.69
	Precision	75.58	96.631	94.08	87.21	91	96.49		

Figure 8 illustrates the tree species classification maps with different methods on the multi-source datasets of three areas. The quality of the DBMF classification map is significantly higher than those of other methods and has the clearest boundaries among them. Furthermore, DBMF readily recognizes tree species while the other methods do not, where the OA reached 95.2%. The accuracies of DBMF for each tree species were 93.5% (birch), 90.8% (larch), 92.5% (mongolica), 86.7% (poplar), 85.9% (spruce), and 91.1% (willow). The DBMF method consistently outperformed other methods in feature mapping, suggesting that the fusion strategy is preferable to operations on single-source data. The DBMA and DBDA methods can simultaneously extract spectral and spatial information, but DBMF presents a better classification effect than both, yielding results closer to the
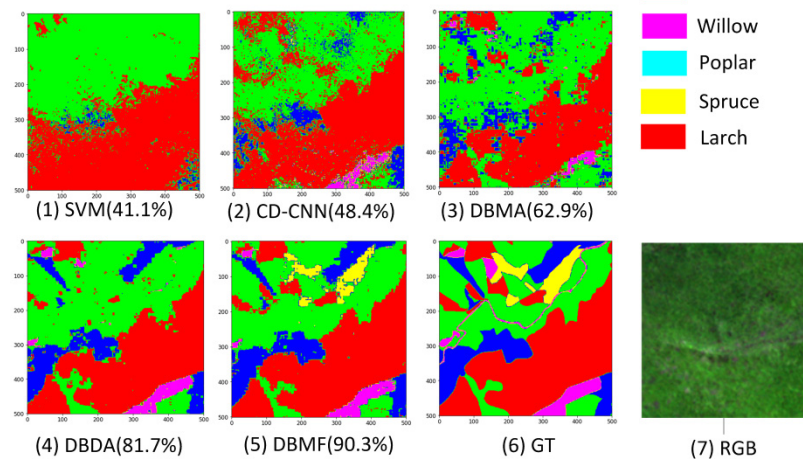
reference map and providing more robust spectral and spatial features with smoother appearances than the other methods.

(1) SVM (45.6%) (2) CD-CNN (49.7%) (3) DBMA (65.2%) (4) DBDA (84.3%)

(5) DBMF (91.1%) (6) GT (7) RGB (8)False color image

(a) Results of 1st area

(1) SVM(40.1%) (2) CD-CNN(47.7%) (3) DBMA(61.4%) (4) DBDA(80.6%)

(5) DBMF(90.4%) (6) GT (7) RGB

(b) Results of 2nd area

Mongolica
Birch
Willow
Poplar
Spruce
Larch

(1) SVM(41.1%) (2) CD-CNN(48.4%) (3) DBMA(62.9%)

(4) DBDA(81.7%) (5) DBMF(90.3%) (6) GT (7) RGB

(c) Results of 3rd area

Figure 8. Full classification maps on tree species with a multi-source dataset of the three study areas obtained by (1) SVM, (2) CD-CNN, (3) DBMA, (4) DBDA, (5) DBMF, (6) ground truth, and (7) wrong method.

4. Discussion

4.1. Influence of Sandglass Block on DBMF

In Section 2.3.3, the sandglass block in DBMF is illustrated. Here, the effect of tree species classification with the proposed datasets is proven. As an important parameter, the best size of image patches for DBMF was 7, and the other conditions were the same as Section 3.1. Next, the performances between 'complete DBDA' and DBDA without sandglass block named 'none' are compared.

As shown in Figure 9, DBMF with sandglass block surpassed itself without the sandglass block. There was an almost 5% OA improvement in the datasets, and the complete one trained faster than that without sandglass block by about 40 min. Since the sandglass block adds shortcut connections between high-dimensional representations and performs the depth-wise convolution in high-dimensional feature space, it can quicken counter-propagation and cause the difference in performance [32].
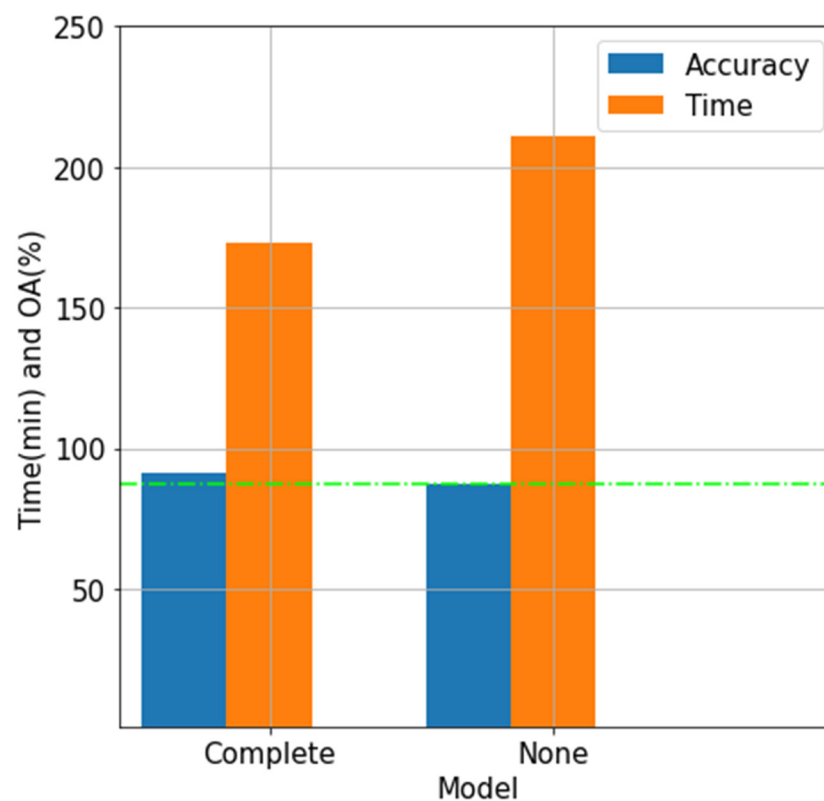
Figure 9. Training time and OA of the DBMF method with sandglass block and none.

4.2. DBMF versus Other Models for the Accuracy of Tree Species Classification

The DBMF model, which applied the Bi-LSTM and triple-attention mechanism, is designed to extract the spectral information that can not only select features through the triple-attention mechanism, but also fully mine a host number of deep features through the residual-dense network and realize the integration of deep features through the Bi-LSTM memory network, effectively reducing redundant features and improving the ability of features fusion. Similarly, the spatial information was extracted through the sandglass block, the effect of which has been conducted in the previous section [38] (Objective 1). As shown in Section 3, the network operation results revealed that DBMF with HJ-1A HS and Sentinel-2 data outperformed all other methods in OA, kappa, and AA. In addition, it performed especially well in recognizing the spruce and larch trees, which would provide commercial value and rarity that is beneficial for the industry. The DBMF method is followed by DBDA, DBMA, and CD-CNN, with SVM exhibiting the worst effect. In the results where fragmentation, rough edges, low accuracy, and severe mixing occurred

among four least species, in contrast, DBMF could recognize the six tree species perfectly. In general, although DBMF deepened the convolution network layer, no gradient degradation and no overfitting occurred in the training process, allowing the best classification results to be obtained (Objective 2). Thus, this method can be used as an effective method for the classification of complex tree species in the northeast.

4.3. Dig Deep Reason for the Results

The SVM method has a statistically significant advantage where the segmentation size does not need to be considered when pixel-based reflectance samples are used. Ghosh and Fassnacht [46,47] have claimed that the support vector machine method could be used to deal with complex classification problems, such as tree species classification, wherein they classified five tree species using SVM with Hyperion data. In the present paper, mongolica pine, spruce, willow, and poplar were almost identified as larch. Therefore, SVM has serious defects in the recognition ability of coniferous forest species. Hartling [48] significantly recognized eight dominant tree species, the OA of which reached 0.82 compared to the SVM classifiers with 0.52. The views that were obtained from the compared DL models are basically consistent with the study conducted by Hartling, as shown in Figure 9. The CD-CNN method exploited the local spatial–spectral relationships neighboring each pixel vector to explore local contextual interactions, however, which hardly recognize any Poplar, in contrast, DBMF have a good effect on the extraction of Poplar. Because CD-CNN is based on 2D-CNN, it has certain shortcomings for complex multi-tree species classification with samples that have small sizes. However, the 3D-CNN method is more lightweight, general, and fast-converging, wherein the convolution operation can retain finer spectral information and provide a good classification model structure for tree classification works where it is difficult to obtain huge samples [17]. Therefore, this method cannot satisfy the request in tree species classification. The DBMA method based on 3D-CNN employs the channel-wise attention and spatial-wise mechanism to enhance features, although the training process is more time-consuming than CD-CNN because of the parameters, whereas DBMF fails to consider the order and relationship of the HS data. Therefore, the two shortcomings above make DBMA more time-consuming and perform worse compared to the DBMF method for tree classification [43]. Compared with DBMA, DBDA adds the Mish activation function to extract the information of some negative parameters while simultaneously increasing the complexity of the algorithm and slowing down the efficiency in the training process. This makes the tree species map more accurate than DBMA in classifying tasks, but at a cost of it performing worse than DBMA. Because the proposed DBMF method can be improved in terms of training speed by employing sandbox so that it does not lose the extraction of effective information ability and the connection with the spatial-spectral feature, it does not pre-process or post-process data and the spatial and spectral information of HSIs so that they can be fully utilized to achieve the desired classification accuracy. It solves the problem of information loss caused by feature downscaling and feature screening of the original spectral–spatial information, and therefore DBMF can be applied in the field of forestry scientific management (Objective 3).

5. Conclusions

In this paper, a DBMF method was developed to improve the feature extraction process for tree species classification. The proposed method effectively combines high-resolution spatial information with spectral information. The spectral information extraction branch employs Bi-LSTM and a triple-attention mechanism; the spatial information extraction branch uses the sandglass module and a 3D-CNN based on the connection between BN and the fusion activation function. The features extracted from the two branches were activated by the fusion function to prevent overfitting. The statistical performance of the proposed method on datasets for six dominant tree species captured by the HJ-1A and Sentinel-2 satellites was found to be better than that of other state-of-the-art methods. As such, the proposed DBMF is suitable for forestry remote sensing application technology.

This study demonstrates the potential of the proposed DBMF model as a powerful classifier for complex landscapes, such as for tree species classification. In the future, we will apply the DBMF framework to other HS datasets, develop a more efficient framework and feature selection processes, and use a wider range of training datasets to extend the model to the complete inventory process.

Author Contributions: X.W. proposed the algorithm and performed the experiment. H.R. supervised the study, analyzed the results, and provided insightful suggestions for the manuscript. X.W. drafted the manuscript. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to still being used in a proceeding project.

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

References

1. Fricker, G.A.; Ventura, J.D.; Wolf, J.A.; North, M.P.; Davis, F.W.; Franklin, J. A convolutional neural network classifier identifies tree species in mixed-conifer forest from hyperspectral imagery. *Remote Sens.* **2019**, *11*, 2326. [CrossRef]
2. Lin, Y.; Hyyppä, J. A comprehensive but efficient framework of proposing and validating feature parameters from airborne LiDAR data for tree species classification. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *46*, 45–55. [CrossRef]
3. Shen, X.; Cao, L. Tree-species classification in subtropical forests using airborne hyperspectral and LiDAR data. *Remote Sens.* **2017**, *9*, 1180. [CrossRef]
4. Dalponte, M.; Bruzzone, L.; Gianelle, D. Tree species classification in the Southern Alpsbased on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. *Remote Sens. Environ.* **2012**, *123*, 258–270. [CrossRef]
5. Cho, M.A.; Mathieu, R.; Asner, G.P.; Naidoo, L.; Van Aardt, J.A.N.; Ramoelo, A.; Debba, P.; Wessels, K.; Main, R.; Smit, I.P.; et al. Mapping tree species composition in South African savannas using an integrated airborne spectral and LiDAR system. *Remote Sens. Environ.* **2012**, *125*, 214–226. [CrossRef]
6. Immitzer, M.; Vuolo, F.; Atzberger, C. First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sens.* **2016**, *8*, 166. [CrossRef]
7. Song, J.H.; Han, S.H.; Yu, K.Y.; Kim, Y.I. Assessing the possibility of land-cover classification using lidar intensity data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2002**, *34*, 259–262.
8. Shang, X.; Chisholm, L.A. Classification of Australian Native Forest Species Using Hyperspectral Remote Sensing and Machine-Learning Classification Algorithms. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2481–2489. [CrossRef]
9. Jiao, L.; Sun, W.; Yang, G.; Ren, G.; Liu, Y. A hierarchical classification framework of satellite multispectral/hyperspectral images for mapping coastal wetlands. *Remote Sens.* **2019**, *11*, 2238. [CrossRef]
10. Gao, L.; Hong, D.; Yao, J.; Zhang, B.; Gamba, P.; Chanussot, J. Spectral superresolution of multispectral imagery with joint sparse and low-rank learning. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 2269–2280. [CrossRef]
11. Franklin, S.E.; Ahmed, O.S. Deciduous tree species classification using object-based analysis and machine learning with unmanned aerial vehicle multispectral data. *Int. J. Remote Sens.* **2018**, *39*, 5236–5245. [CrossRef]
12. Gini, R.; Sona, G.; Ronchetti, G.; Passoni, D.; Pinto, L. Improving tree species classification using UAS multispectral images and texture measures. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 315. [CrossRef]
13. Munyati, C. The potential for integrating Sentinel 2 MSI with SPOT 5 HRG and Landsat 8 OLI imagery for monitoring semi-arid savannah woody cover. *Int. J. Remote Sens.* **2017**, *38*, 4888–4913. [CrossRef]
14. Mohajane, M.; Essahlaoui, A.; Oudija, F.; Hafyani, M.E.; Hmaidi, A.E.; Ouali, A.E.; Randazzo, G.; Teodoro, A.C. Land use/land cover (LULC) using landsat data series (MSS, TM, ETM+ and OLI) in Azrou Forest, in the Central Middle Atlas of Morocco. *Environments* **2018**, *5*, 131. [CrossRef]
15. Hashim, B.M.; Sultan, M.A.; Attyia, M.N.; Al Maliki, A.A.; Al-Ansari, N. Change detection and impact of climate changes to Iraqi southern marshes using Landsat 2 Mss, Landsat 8 Oli and sentinel 2 Msi data and Gis applications. *Appl. Sci.* **2019**, *9*, 2016. [CrossRef]
16. Sun, W.; Du, Q. Graph-regularized fast and robust principal component analysis for hyperspectral band selection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3185–3195. [CrossRef]
17. Zhang, B.; Zhao, L.; Zhang, X. Three-dimensional convolutional neural network model for tree species classification using airborne hyperspectral images. *Remote Sens. Environ.* **2020**, *247*, 111938. [CrossRef]

18. Shi, Y.; Skidmore, A.K.; Wang, T.; Holzwarth, S.; Heiden, U.; Pinnel, N.; Zhu, X.; Heurich, M. Tree species classification using plant functional traits from LiDAR and hyperspectral data. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 207–219. [CrossRef]
19. Xie, Q.; Zhou, M.; Zhao, Q.; Meng, D.; Zuo, W.; Xu, Z. Multispectral and hyperspectral image fusion by MS/HS fusion net. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1585–1594.
20. Deng, L.; Yu, D. Deep learning: Methods and applications. *Found. Trends Signal Process.* **2014**, *7*, 197–387. [CrossRef]
21. Li, W.; Fu, H.; Yu, L.; Cracknell, A. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote Sens.* **2017**, *9*, 22. [CrossRef]
22. SSidike, P.; Sagan, V.; Maimaitijiang, M.; Maimaitiyiming, M.; Shakoor, N.; Burken, J.; Mockler, T.; Fritschi, F.B. dPEN: Deep Progressively Expanded Network for mapping heterogeneous agricultural landscape using WorldView-3 satellite imagery. *Remote Sens. Environ.* **2019**, *221*, 756–772. [CrossRef]
23. Sidike, P.; Asari, V.K.; Sagan, V. Progressively Expanded Neural Network (PEN Net) for hyperspectral image classification: A new neural network paradigm for remote sensing image analysis. *ISPRS J. Photogramm. Remote Sens.* **2018**, *146*, 161–181. [CrossRef]
24. Guan, H.; Yu, Y.; Ji, Z.; Li, J.; Zhang, Q. Deep learning-based tree classification using mobile LiDAR data. *Remote Sens. Lett.* **2015**, *6*, 864–873. [CrossRef]
25. Liao, W.; Van Coillie, F.; Gao, L.; Li, L.; Zhang, B.; Chanussot, J. Deep learning for fusion of APEX hyperspectral and full-waveform LiDAR remote sensing data for tree species mapping. *IEEE Access* **2018**, *6*, 68716–68729. [CrossRef]
26. Pölönen, I.; Annala, L.; Rahkonen, S.; Nevalainen, O.; Honkavaara, E.; Tuominen, S.; Viljanen, N.; Hakala, T. Tree species identification using 3D spectral data and 3D convolutional neural network. In Proceedings of the 2018 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, The Netherlands, 23–26 September 2018; pp. 1–5.
27. Lee, H.; Kwon, H. Going deeper with contextual CNN for hyperspectral image classification. *IEEE Trans. Image Process.* **2017**, *26*, 4843–4855. [CrossRef] [PubMed]
28. Xu, Y.; Zhang, L.; Du, B.; Zhang, F. Spectral–spatial unified networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5893–5909. [CrossRef]
29. Yang, J.; Zhao, Y.Q.; Chan, J.C.W. Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network. *Remote Sens.* **2018**, *10*, 800. [CrossRef]
30. Liu, J.; Wang, X.; Wang, T. Classification of tree species and stock volume estimation in ground forest images using Deep Learning. *Comput. Electron. Agric.* **2019**, *166*, 105012. [CrossRef]
31. Koga, Y.; Miyazaki, H.; Shibasaki, R. A CNN-based method of vehicle detection from aerial images using hard example mining. *Remote Sens.* **2018**, *10*, 124. [CrossRef]
32. Zhou, D.; Hou, Q.; Chen, Y.; Feng, J.; Yan, S. Rethinking bottleneck structure for efficient mobile network design. In *Computer Vision–ECCV 2020: Proceedings of the16th European Conference, Glasgow, UK, 23–28 August 2020*; Part III 16; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; pp. 680–697.
33. Wang, L.; Fan, W.Y. Identification of forest dominant tree species group based on hyperspectral remote sensing data. *J. Northeast. For. Univ.* **2015**, *43*, 134–137.
34. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [CrossRef]
35. Malenovský, Z.; Rott, H.; Cihlar, J.; Schaepman, M.E.; García-Santos, G.; Fernandes, R.; Berger, M. Sentinels for science: Potential of Sentinel-1, -2, and-3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sens. Environ.* **2012**, *120*, 91–101. [CrossRef]
36. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
37. Liu, Q.; Zhou, F.; Hang, R.; Yuan, X. Bidirectional-convolutional LSTM based spectral-spatial feature learning for hyperspectral image classification. *Remote Sens.* **2017**, *9*, 1330. [CrossRef]
38. Cai, W.; Liu, B.; Wei, Z.; Li, M.; Kan, J. TARDB-Net: Triple-attention guided residual dense and BiLSTM networks for hyperspectral image classification. *Multimed. Tools Appl.* **2021**, *80*, 11291–11312. [CrossRef]
39. Misra, D. Mish: A self regularized non-monotonic neural activation function. *arXiv* **2019**, arXiv:1908.08681.
40. Li, H.C.; Hu, W.S.; Li, W.; Li, J.; Du, Q.; Plaza, A. A^{3}CLNN: Spatial, spectral and multiscale attention ConvLSTM neural network for multisource remote sensing data classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, 1–15. [CrossRef]
41. Lu, X.; Yang, D.; Jia, F.; Zhao, Y. Coupled Convolutional Neural Network-Based Detail Injection Method for Hyperspectral and Multispectral Image Fusion. *Appl. Sci.* **2021**, *11*, 288. [CrossRef]
42. Hsu, C.W.; Chang, C.C.; Lin, C.J. A Practical Guide to Support Vector Classification. 2003. Available online: https://www.csie.ntu.edu.tw/~{}cjlin/papers/guide/guide.pdf (accessed on 1 October 2021).
43. Ma, W.; Yang, Q.; Wu, Y.; Zhao, W.; Zhang, X. Double-branch multi-attention mechanism network for hyperspectral image classification. *Remote Sens.* **2019**, *11*, 1307. [CrossRef]
44. Li, R.; Zheng, S.; Duan, C.; Yang, Y.; Wang, X. Classification of hyperspectral image based on double-branch dual-attention mechanism network. *Remote Sens.* **2020**, *12*, 582. [CrossRef]

45. Huang, S.; Tang, J.; Dai, J.; Wang, Y. signal status recognition based on 1dcnn and its feature extraction Mechanism analysis. *Sensors* **2019**, *19*, 2018. [CrossRef]

46. Fassnacht, F.E.; Neumann, C.; Förster, M.; Buddenbaum, H.; Ghosh, A.; Clasen, A.; Joshi, P.K.; Koch, B. Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central European test sites. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2547–2561. [CrossRef]

47. Dalponte, M.; Ørka, H.O.; Gobakken, T.; Gianelle, D.; Næsset, E. Tree species classification in boreal forests with hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2012**, *51*, 2632–2645. [CrossRef]

48. Hartling, S.; Sagan, V.; Sidike, P.; Maimaitijiang, M.; Carron, J. Urban tree species classification using a WorldView-2/3 and LiDAR data fusion approach and deep learning. *Sensors* **2019**, *19*, 1284. [CrossRef] [PubMed]