CLASSIFICATION OF DEGRADED SPECIES IN DESERT GRASSLANDS BASED ON MULTI-FEATURE FUSION AND UNMANNED AERIAL VEHICLE HYPERSPECTRAL

/

基于多特征融合与无人机高光谱的荒漠草原退化物种分类

Tao ZHANG [1)], **Fei HAO** [*2)], **Yuge BI** [1)], **Jianmin DU** [1)], **Weiqiang PI** [3)], **Yanbin ZHANG** [1)],
Xiangbing ZHU [1)], **Xinchao GAO** [1)], **Eerdumutu JIN** [1)]

[1)] Inner Mongolia Agricultural University, Mechanical and Electrical Engineering College, Hohhot / China
[2)] Hohhot Vocational College, Mechanical and Electrical Engineering Department, Hohhot / China
[3)] Huzhou Vocational and Technical College, College of Mechatronics and Automotive Engineering, Huzhou / China
Tel: 17852092475; E-mail: taozhang626@163.com

ABSTRACT

Accurate spatial distribution of grassland degradation indicator species is of great significance for grassland degradation monitoring. In order to realize the intelligent remote sensing grassland degradation monitoring task, this paper collects remote sensing data of three degradation indicator species of desert grassland, namely, constructive species, dominant species, and companion species, through the UAV hyperspectral remote sensing platform, and proposes a multi-feature fusion (MFF) classification model. In addition, vertical convolution, horizontal convolution, and group convolution mechanisms are introduced to further reduce the number of model parameters and effectively improve the computational efficiency of the model. The results show that the overall accuracy and kappa coefficient of the model can reach 91.81% and 0.8473, respectively, and it also has better classification performance and computational efficiency compared to different deep learning classification models. This study provides a new method for high-precision and efficient fine classification study of degradation indicator species in grasslands.

摘要

准确掌握草地退化指示物种的空间分布对草地退化监测有着重要意义。为实现智能化遥感草地退化监测任务，本文通过无人机高光谱遥感平台对荒漠草原建群种、优势种和伴生种三种退化指示物种遥感数据采集，并提出一种多特征融合（MFF）的分类模型。此外，引入垂直卷积、水平卷积和分组卷积机制，进一步减少模型参数量，有效提升模型的计算效率。结果表明，该模型的总体精度和 kappa 系数分别可达 91.81%、0.8473。同时，与不同深度学习分类模型相比，也具有更优的分类性能和计算效率。本研究为草地退化指示物种的高精度、高效率的精细分类提供了一种新方法。

INTRODUCTION

Grassland desertification is one of the ten most serious ecological and environmental problems in the world *(Tang et al., 2016)*. Inner Mongolia Autonomous Region is the province with the largest grassland area in China, with a total grassland area of 87 million hectares, accounting for 73.4% of the land area of the region. However, under the combined influence of global climate change and human factors, more than 90% of the grassland area in the Inner Mongolia Autonomous Region has been severely degraded *(He et al., 2021)*. Grassland degradation will not only lead to ecological problems such as soil erosion, sand and dust storms, a decrease in grassland productivity, and the loss of biodiversity *(Liu et al., 2021; Wang et al., 2020)*, but will also affect the development of local animal husbandry *(Briske et al., 2015)*. Therefore, there is an urgent need for efficient and accurate technology and methods for effective monitoring of desert grasslands, to provide assistance for the restoration and management of desert grasslands.

In recent years, some scholars have applied machine learning to grassland degradation monitoring to classify grassland characteristics. For example, Yang et al. *(2021)* achieved the classification of desert steppe species by constructing a decision tree classification model. However, this machine learning classification method requires manual extraction of a large amount of feature information, which is time-consuming and labor-intensive.

Subsequently, some scholars began to introduce deep learning into the classification task of grassland degradation indicator species to achieve an integrated process from feature extraction to classification. For example, Pi et al. *(2021)* realized the end-to-end classification of grassland degradation indicator species by building a 3D convolutional neural network classification model. However, the constructed model not only has a large number of network parameters but also needs to be improved in classification performance. In the classification of grassland degradation indicator species, the speed of model inference is a necessary factor to improve the efficiency of grassland monitoring. Therefore, it is very important to explore a high efficiency and high precision classification model for grassland degradation indicator species.

Currently, a convolutional neural network (CNN) is a widely used hyperspectral image classification method in hyperspectral image classification *(Liu et al., 2022; Xu et al., 2021)*. CNN is divided into a 1D convolutional neural network (1DCNN), a 2D convolutional neural network (2DCNN), and a 3D convolutional neural network (3DCNN). Since the 1DCNN filter is one-dimensional, it can only extract the spectral features of hyperspectral images *(Hsieh and Kiang, 2020)*; the 2DCNN filter is two-dimensional, so it can only extract the spatial features of hyperspectral images *(Shenming et al., 2022)*; while the 3DCNN filter is three-dimensional, which has one more spectral dimension compared to 2DCNN, so it can extract the spatial-spectral features of hyperspectral images, which usually is better than 1DCNN and 2DCNN in terms of classification performance *(Jung et al., 2022)*. However, 1DCNN, 2DCNN, and 3DCNN ignore the correlation between the pixels to be classified and the neighboring pixels when classifying hyperspectral images.

To address the above problems, this paper uses an unmanned aircraft remote sensing platform to collect data on desert grassland species and proposes a multi-feature fusion (MFF) classification model. The model effectively combines the spatial and spectral features of hyperspectral images, as well as the correlation information between the pixels to be classified and neighboring pixels, which further improves the classification performance of the model. At the same time, vertical convolution, horizontal convolution, and group convolution mechanisms are introduced to further reduce the number of model parameters and effectively improve the computational efficiency of the model. The purpose of this paper is to explore a high precision and high-efficiency method to monitor indicator species of desert grassland degradation, in order to provide help for the management and monitoring of grassland degradation.

MATERIALS AND METHODS
Overview of the study area

The present study area is located on Gegenthala grasslands (111 ° 52′47′ ′ E, 41 ° 46′48′ ′ N) in the Inner Mongolia Autonomous Region, China. The grassland type in this area is *Stipa breviflora* grassland desert, and the vegetation is sparse and low, with a cover of only 15-25% and an average plant height of only 8 cm. The type of climate belongs to a temperate monsoon climate, with an average annual temperature of 3.5 °C and an altitude of 1456 m *(Zhang et al.*, *2022)*. After investigation, there are more than 20 types of vegetation, which can be subdivided into constructive, dominant, and companion species. Here, constructive species are *Stipa breviflora*, dominant species are *Cleistogenes songorica* and *Artemisia frigida*, and companion species are *Neopallasia pectinata*, *Leymus chinensis*, and *Ceratoides latens (Pan et al.*, *2016)*.

Experimental equipment

The instruments used in this study include mainly a hyperspectral imager, a UAV, and a gimbal. Among them, the hyperspectral imager adopts the Pika XC2 hyperspectrometer made by Resonon, USA, with a weight of 2.2 kg. The spectral range of this spectrometer is between 400 and 1000 nm, the maximum number of bands can be set to 447 bands, the spectral resolution is 1.34 nm, and the number of spatial channels is 1600. The Jinan Share HEX-8 eight-rotor UAV was selected, with a professional A3 pro flight control system and dual positioning systems of BeiDou and GPS, which can perform accurate positioning and autonomous route flight. The aircraft has a maximum load of 40 kg and can fly continuously for 30 min at full load with a maximum flight speed of 10 m/s. The gimbal is DJI Ronin-M with a weight of 3.6 kg.

Data collection

Based on local climatic conditions and vegetation growth, data was collected from July 5 to 15, 2019. To reduce the shadow interference, the collection time was between 11:00 and 13:00. At the same time, the weather was ensured to be clear and cloudless and the wind speed was lower than 3.5 m/s. After several flight experiments, the spatial size of the collected hyperspectral images was set at 6062 lines × 1600 samples, and the number of bands was 231; the flight height of the UAV was set at 30 m and the spatial resolution was 2.1 cm at this time.

During the acquisition of 2.5 hm² hyperspectral image data, the UAV flew autonomously according to the planned route and set the bypass overlap rate at 55%. Furthermore, three types of vegetation, namely constructive species, dominant species, and companion species, were surveyed in the UAV flight area by arranging 1 m × 1 m sample boxes; Only one type of vegetation was included in a single sample box; the vegetation coverage, latitude and longitude, and type of vegetation were recorded in the sample boxes.

Data processing

The acquired hyperspectral image data were manually checked to remove distorted and overexposed images. Reflectance correction of hyperspectral image data was performed using Spectral Pro software to obtain the reflectance values of the real features, as shown in Figure 1. From the figure, it can be seen that the reflectance values of the last 40 bands fluctuate greatly due to the influence of noise and should be given to be removed, and finally 191 bands are retained. In addition, in order to reduce the subsequent calculation cost and correlation between the bands, the PCA dimensionality reduction algorithm is applied to reduce the number of bands to 15 bands, and this group of bands retains 96.90% of the features of the hyperspectral image.
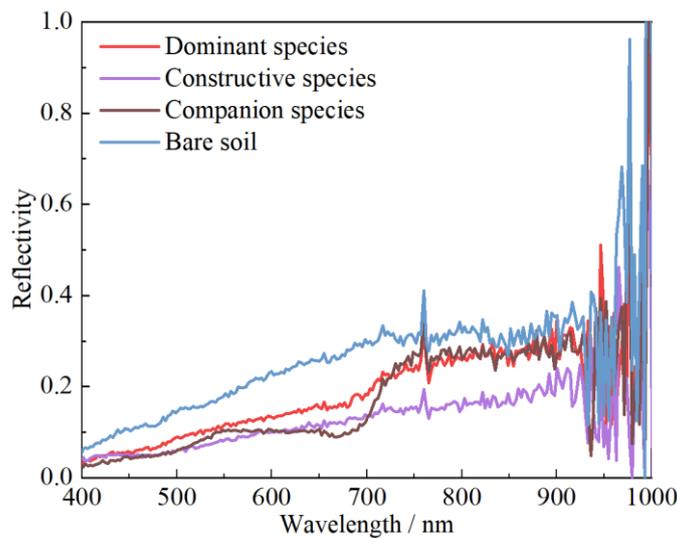

Fig. 1 - Spectral curves of target ground objects

Dataset Production

Taking into account the computational cost, the hyperspectral image data was cropped to a spatial size of 601 lines × 601 samples. Taking into account the data balance problem, the cropped images should contain three feature types: constructive species, dominant species, and companion species. Subsequently, the cropped images were labeled with samples according to the extensive ground survey. Finally, 187482 constructive species samples, 158945 dominant species samples, 3049 companion species samples, and 11725 bare soil samples were obtained, totaling 361201 sample data. This hyperspectral false-color image and ground truth image are shown in Figure 2.
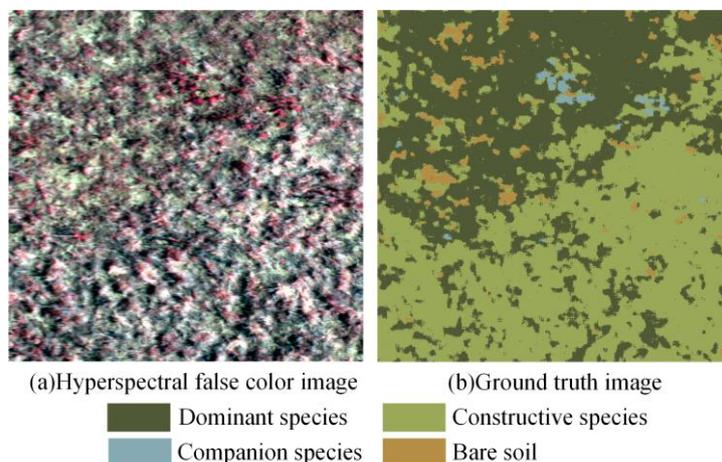

(a)Hyperspectral false color image   (b)Ground truth image

Dominant species   Constructive species

Companion species   Bare soil

Fig. 2 - Hyperspectral images and ground truth images

ALGORITHM PRINCIPLE
Spatial feature extraction

The spatial feature extraction of hyperspectral images is generally performed using convolution kernels of n × n (n > 1). In Inception-v3 *(Szegedy et al., 2016)*, it is noted that n × n convolution kernels can be replaced by vertical convolution (n × 1) and horizontal convolution (1 × n) to effectively reduce the number of model parameters while ensuring classification accuracy. However, the order of vertical and horizontal convolution also has some influence on the classification performance of the model. In this paper, the dual-branch horizontal vertical convolution (DHVC) module is proposed, see Fig. 3.

In the figure, $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$ is the input hyperspectral data, $\mathbf{Y} \in \mathbb{R}^{H \times W \times C'}$ is the output after DHVC spatial feature extraction, H×W is the length and width of the input hyperspectral data, C and C' are the number of bands, and 1×1@C/8 conv means there are C/8 convolution kernels for 1×1, and the rest are the same.

In the DHVC module, the first 1×1 convolution kernel is used to reduce the spectral dimensionality of the input data to reduce the model parameters and computational cost. Subsequently, horizontal convolution (1×3) and vertical convolution (3×1) by two branches are used for spatial feature extraction, and this structure can compensate for the impact on model classification accuracy due to the different order of horizontal and vertical convolution. After the extraction of spatial features from the two branches, the outputs of the two branches are stitched in the channel dimension. At the same time, 1×1 convolution kernels are used to fuse the features across the channels of the spliced data, and the spectral dimension of the data is expanded to C' to obtain the final output $\mathbf{Y} \in \mathbb{R}^{H \times W \times C'}$, which is set to $C' = 2C$ in this paper. for the two 1×1 convolution kernels, the LeakyReLU activation functions are accessed after them. In addition, group convolution is introduced in horizontal and vertical convolution to reduce the number of model parameters, and the groups are set to C/64.
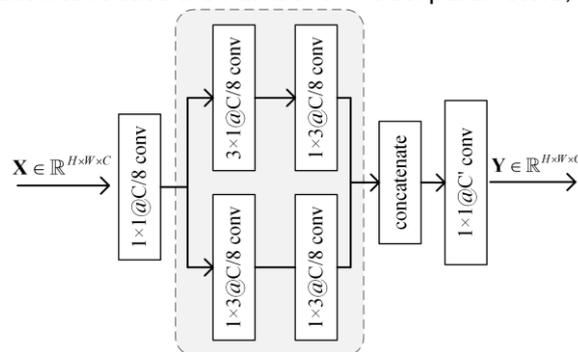

Fig. 3 - Structure of the DHVC module

Spectral feature extraction

The spectral features are extracted on the spectral dimension of individual pixels in the input data block. A 1×1 convolution kernel is equivalent to a linear fitting function, so a 1×1 convolution kernel is used to fit the spectral features. Meanwhile, the LeakyReLU activation function is introduced to activate the fitted spectral features so that they have a nonlinear relationship and can better extract the spectral features. Therefore, the extraction of spectral features can be represented by 1×1@C' conv+LeakyReLU.

Extract of correlation sequence features

In this paper, a correlation sequence feature extraction (CSFE) module is proposed for the extraction of correlation information between the pixels to be classified and the neighboring pixels, see Figure 4.
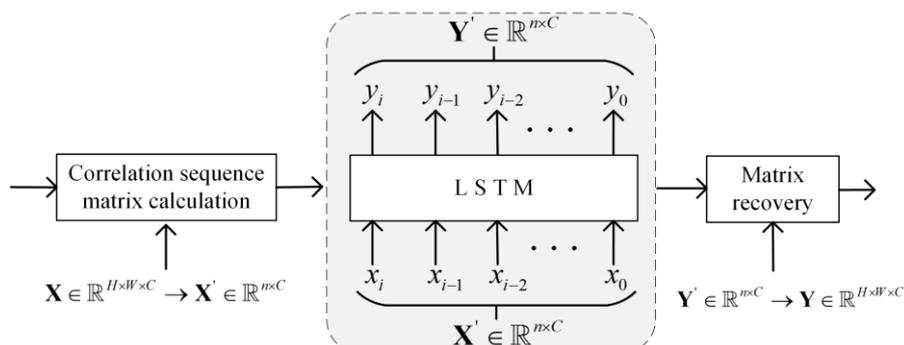

Fig. 4 - Structure of the CSFE module

Assuming that the input hyperspectral data $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$, the correlation size between the pixel to be classified $x_0 \in \mathbb{R}^C$ and the neighboring pixels $x_i \in \mathbb{R}^C, i = 1,2,\ldots,n$ (n = H × W) can be calculated by the Euclidean distance function. Then the correlation sequence matrix $\mathbf{X}' \in \mathbb{R}^{n \times C}$ is obtained by reordering based on the correlation size from smallest to largest. The correlation sequence is brought into a single-layer long short-term memory (LSTM) network as a temporal sequence for correlation sequence feature extraction, and $\mathbf{Y}' \in \mathbb{R}^{n \times C'}$ is obtained. Finally, the pixels in the $\mathbf{Y}'$ matrix are restored to their original positions to obtain $\mathbf{Y} \in \mathbb{R}^{H \times W \times C'}$.

MFF Model Framework

In the MFF model, feature pre-extraction is first performed by a 3×3@64 2D convolution kernel with padding set to 1. Subsequently, it is passed into the joint extraction module of spatial and spectral features and correlated sequence features.

The model is set up with 4 layers of the joint extraction module, and the extracted spatial and spectral features and related sequence features are summed in each layer, and a 2×2 maximum pooling function is used for spatial downsampling between the layers. The spectral feature outputs for each layer are set to 64, 128, 256, and 512, respectively. Finally, a fully connected (FC) layer with 4 neurons is used for the final classification. The structure of this network is shown in Fig. 5, where {64} indicates that the spectral dimension of the module output is 64, and the rest is the same.
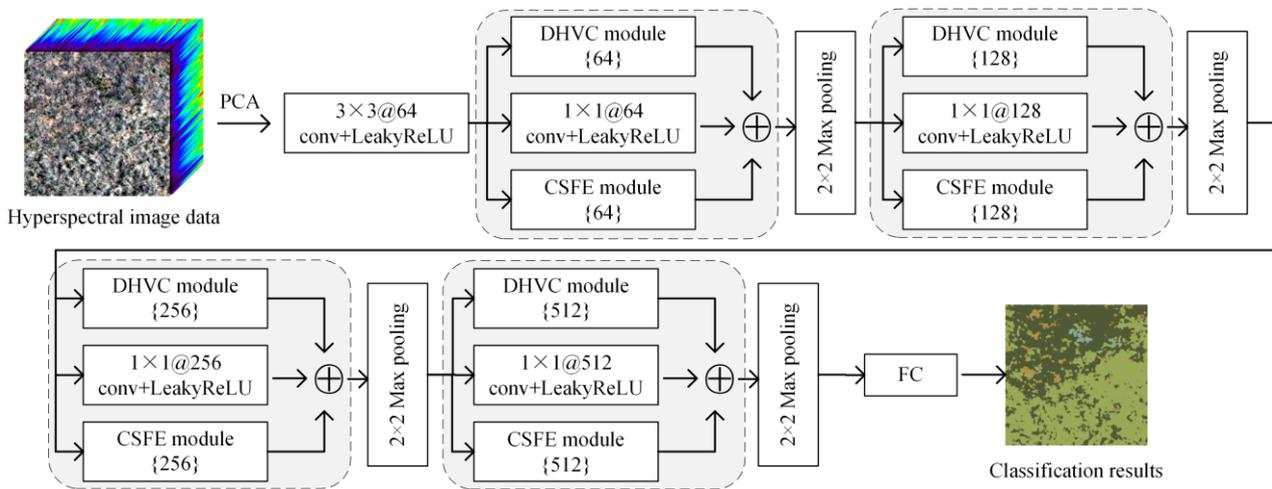

Fig. 5 - Structure of the MFF classification model

EXPERIMENT AND ANALYSIS

This experiment uses the Pytorch framework for model construction. Computer hardware devices are i7-11800H CPU, RTX 3060 graphics card, and 16 GB RAM. The learning rate is set to 0.001, patch to 13, epochs to 10, batch size to 128, and Adam optimization function is selected. In the experiment, 50% of the data is selected as the training set and the rest as the validation set. In addition, the overall accuracy (OA) and the kappa coefficient are selected as the evaluation criteria of the model.

Ablation experiments

To verify the effectiveness of spatial feature extraction, spectral feature extraction and correlated sequence feature extraction of the classification model proposed in this paper, ablation experiments will be analyzed, and the experimental results are shown in Table 1. It can be seen from the table that the classification accuracy of the model extracting only spatial features is the worst, and its OA value is only 87.79%. After adding spectral feature extraction, the OA value of the model is improved by 2.3%, and the kappa is improved by 0.048. The MFF model that simultaneously extracts spatial features, spectral features, and related sequence features have the best classification performance, with an OA of 91.81% and a kappa of 0.8473, indicating the effectiveness of the model proposed in this paper.

Table 1

Analysis of ablation experiments

Spatial features	Spectral features	Correlation sequence features	OA / %	Kappa
√			87.79	0.7663
√	√		90.09	0.8143
√	√	√	91.81	0.8473

RESULTS
Experimental results and analysis

To better evaluate the classification performance of the MFF model proposed in this paper, a total of three classical models, 2DCNN, Resnet18 and Densenet121, were selected for comparison experiments. The experimental results are shown in Table 2, the classification effect is plotted in Fig. 6, and the bold in the table indicates the highest accuracy.

As can be seen in Figure 6, the classification effect map of 2DCNN has the worst performance, the classification effect map of Resnet18 contains too many noise points, and the classification effect map of Densenet121 has too many constructive species identified as dominant species. In general, 2DCNN, Resnet18, and Densenet121 are not good at extracting features from small ground objects such as desert grasslands, while the classification result of the proposed MFF model is closer to the ground truth image and retains the distribution and details of ground objects.
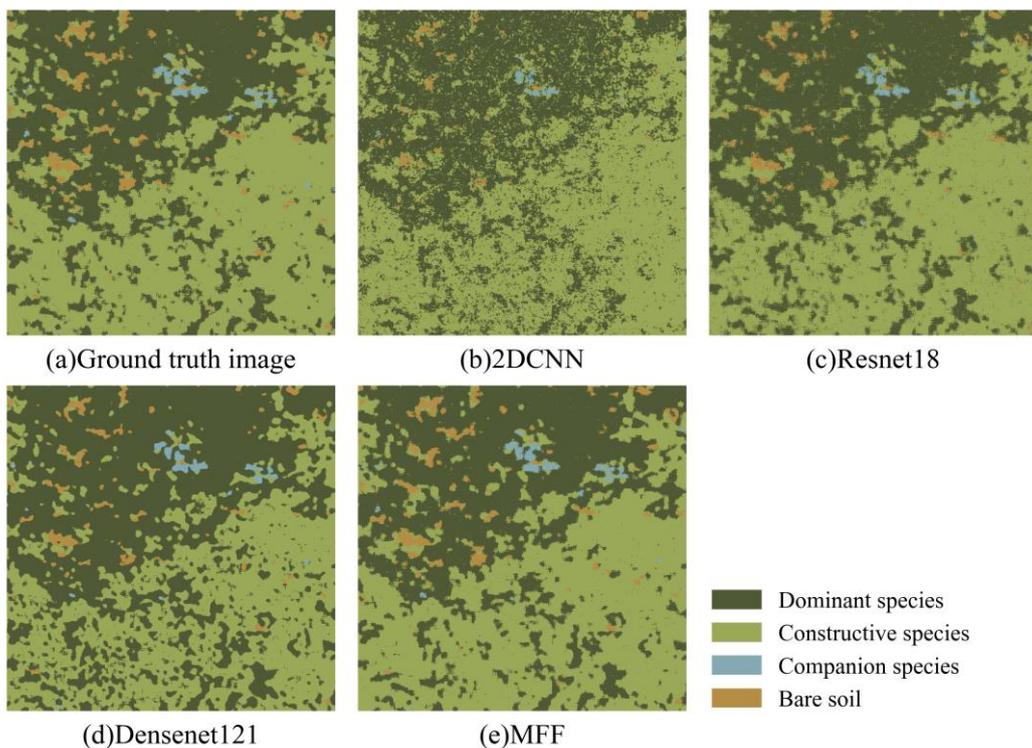

Fig. 6 - Classification effect images of different models

Table 2

Comparison of the classification performance of different models (%)

Category	2DCNN	Resnet18	Densenet121	MFF
Constructive species	82.86	86.70	81.21	**92.28**
Dominant species	77.11	88.07	**97.49**	92.67
Companion species	30.75	48.59	76.59	**78.56**
Bare soil	24.94	40.83	63.48	**75.97**
OA	78.01	85.49	87.76	**91.81**
Kappa	0.5793	0.7269	0.7729	**0.8473**

Efficiency comparison experiment

To further evaluate the computational efficiency of the proposed MFF model, the model parameters, floating point operations (FLOPs), and prediction time are now selected for efficiency evaluation. The comparison model selects 2DCNN, Resnet18, and Densenet121, and the experimental results are shown in Table 3. As can be seen from the table, 2DCNN has the least prediction time and FLOPs, but performs worse in classification performance. The parameter quantity and prediction time of MFF are better than the two classification models of Resnet18 and Densenet121, and the FLOPs are moderate, indicating that MFF has faster model calculation efficiency while ensuring classification accuracy.

Table 3

Comparison of the computational efficiency of different models

Models	Number of participants / M	FLOPs / M	Prediction time / s
2DCNN	5.3	7.3	15
Resnet18	11.2	17.3	67
Densenet121	6.9	134	81
MFF	2.4	18.2	25

CONCLUSIONS

In this paper, we collected data of three types of degradation indicator species in a desert grassland: constructive species, dominant species, and companion species by a UAV hyperspectral remote sensing platform, and proposed a multi-feature fusion (MFF) classification model to achieve high precision and high-efficiency classification performance of fine features in desert grassland. The main findings of this study are as follows:

1) A data set of desert grassland degradation indicator species was established, and a multi-feature fusion (MFF) classification model was proposed, whose overall classification accuracy could reach 91.81% and the kappa coefficient could reach 0.8473.

2) Introduce the idea of spectral correlation and use of LSTM to extract correlation features between long-distance pixels to be classified and neighboring pixels by effective analysis.

3) The analysis of ablation experiments concludes that the classification of spectral features, spatial features, and correlation features of hyperspectral images extracted alone is poor, while the classification performance of joint multi-feature extraction is optimal.

4) Introduce vertical convolution, horizontal convolution, and group convolution mechanisms to further reduce the number of model parameters and effectively improve the computational efficiency of the model. The dual-branch horizontal and vertical convolution (DHVC) module is proposed to compensate for the impact of different back and forward sequences of vertical and horizontal convolution on the classification performance of the model.

5) The overall accuracy of the proposed model is improved by 13.8%, 6.32% and 4.05% compared to three classification models, namely 2DCNN, Resnet18 and Densenet121, respectively.

The effective combination of the MFF model and the hyperspectral remote sensing of the UAV proposed in this paper can effectively achieve the identification and classification study of ground objects of grassland degradation indicator, providing a new method for monitoring grassland degradation. In the future, different feature extraction methods will be considered for joint feature extraction, and the model parameters will be optimized to further improve the classification performance and computational efficiency of the model, to provide a reference for the management of desert grassland.

REFERENCES

[1] Briske D.D., Zhao M., Han G., et al., (2015), Strategies to alleviate poverty and grassland degradation in Inner Mongolia: Intensification vs production efficiency of livestock systems. *Journal of Environmental Management*, 152, 177-182.

[2]   He Y., Li X., Yang X., et al., (2021), The Estimation of Actual and Potential Carbon Sequestration in Typical Steppe in Xilingol County, Inner Mongolia. *Acta Agrestia Sinica*, 29(10), 2274-2285.

[3]   Hsieh T., Kiang J., (2020), Comparison of CNN Algorithms on Hyperspectral Image Classification in Agricultural Lands. *Sensors*, 20(6), 1734.

[4]   Jung D., Kim J.D., Kim H., et al., (2022), A Hyperspectral Data 3D Convolutional Neural Network Classification Model for Diagnosis of Gray Mold Disease in Strawberry Leaves. *Frontiers in Plant Science*, 13, 837020.

[5]   Liu K., Yang M., Huang S., et al., (2022), Plant Species Classification Based on Hyperspectral Imaging via a Lightweight Convolutional Neural Network Model. *Frontiers in Plant Science*, 13, 855660.

[6]   Liu Y., Lu C., (2021), Quantifying Grass Coverage Trends to Identify the Hot Plots of Grassland Degradation in the Tibetan Plateau during 2000-2019. *International Journal of Environmental Research and Public Health*, 18(2), 416.

[7]   Pi W., Du J., Bi Y., et al., (2021), 3D-CNN based UAV hyperspectral imagery for grassland degradation indicator ground object classification research. *Ecological Informatics*, 62, 101278.

[8]   Pan Z., Wang Z., Han G., et al., (2016), Responses of methane fluxes on warming and nitrogen addition in Stipa breviflora desert steppe. *Ecology and Environmental Sciences*, 25(2), 209-216.

[9]   Shenming Q., Xiang L., Zhihua G., (2022), A new hyperspectral image classification method based on spatial-spectral features. *Scientific Reports*, 12(1), 1541.

[10]   Szegedy C., Vanhoucke V., Ioffe S., et al., (2016), Rethinking the Inception Architecture for Computer Vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2818-2826.

[11]   Tang Z., An H., Deng L., et al., (2016), Effect of desertification on productivity in a desert steppe. *Scientific Reports*, 6(1), 27839.

[12]   Wang Y., Ren Z., Ma P., et al., (2020), Effects of grassland degradation on ecological stoichiometry of soil ecosystems on the Qinghai-Tibet Plateau. *Science of The Total Environment*, 722, 137910.

[13]   Xu H., Yao W., Cheng L., et al., (2021), Multiple Spectral Resolution 3D Convolutional Neural Network for Hyperspectral Image Classification. *Remote Sensing*, 13(7), 1248.

[14]   Yang H., Du J., (2021), Classification of desert steppe species based on unmanned aerial vehicle hyperspectral remote sensing and continuum removal vegetation indices. *Optik*, 247, 167877.

[15]   Zhang T., Du J., Zhang H., et al., (2022), Research on recognition method of desert steppe rat hole based on unmanned aerial vehicle hyperspectral. *Journal of Optoelectronics·Laser*, 33(02), 120-126.