Journal of Computer Science and Electrical Engineering
ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3063

DROUGHT RECOGNITION FOR THE SOYBEAN PLANT BASED ON LIGHTWEIGHT DEEP LEARNING MODEL

WenLiang Hou*, ZhiPeng He, JieWen Gu, JiaLin Zhang, JiaoYing Li
College of Information and Electrical Engineering, Heilongjiang Bayi Agricultural University, Daqing 163319, Heilongjiang, China.
Corresponding Author: WenLiang Hou, Email: houwenliang1024@gmail.com

Abstract: To address the issues of low soybean self-sufficiency and drought - related production constraints, multispectral imaging can non - invasively detect crop characteristics, while deep convolutional networks identify drought from multispectral images. However, large models face mobility limitations due to high computational requirements.This study presents a mobile detection approach integrating ReliefF feature screening and lightweight convolutional neural networks, and develops the "Early Acknowledgment for Soybean Drought" App. It embeds a lightweight model that optimizes 37 - dimensional soybean canopy multispectral features via ReliefF and uses a three - layer 1D convolutional network for drought identification.The model achieves 96.88% classification accuracy on the self - built dataset, with an inference time of 18 ms, a size under 30 MB, and less than 60 MB memory usage on mobile. The APP integrates the multispectral camera SDK and PyTorch inference engine, enabling real - time spectral analysis. Field tests show its one - button operation, low learning curve for farmers, and significant water - saving and yield - increasing effects, offering a lightweight, high - precision mobile solution for soybean drought management and promoting smart agriculture development.
Keywords: Soybean drought; Multispectral image; Feature screening; Recognition model; Mobile app

1 INTRODUCTION

Food security is the strategic cornerstone of national security, although China has realized the absolute security of staple food, but the soybean self-sufficiency rate continues to go down to less than 20%[1], superimposed on the growth of consumer demand and international trade friction intensified, the dependence on imports climbed to the first place in the world. For soybean growth is highly dependent on the characteristics of water, drought stress has become a key adversity factor restricting its production, there is an urgent need to break through the drought-resistant varieties through accurate phenotypic detection technology and water-saving irrigation bottlenecks, in order to enhance the production capacity of soybeans to provide scientific support. As an emerging non-destructive testing method, multispectral imaging technology can simultaneously analyze the external morphology (color, texture) and internal physicochemical properties (water, nutrients) of the crop through the synergistic perception of multiple wavelengths[2], such as visible light and near-infrared light, to realize the multidimensional characterization of the canopy structure and growth status. With the breakthrough of artificial intelligence technology, the high-throughput phenotyping system integrating machine learning and image processing algorithms is becoming a cutting-edge paradigm for precision agriculture research, providing innovative solutions for crop physiological monitoring and stress diagnosis[3].

Gao Shijiao et al innovatively integrated Gaussian filtering and multi-threshold segmentation strategy to construct an efficient extraction model for soybean canopy near-infrared band[4], and the effective segmentation rate was increased to 97.81%; Fu Hongyu et al developed a dynamic monitoring system for physical and chemical traits of Ramie flax by integrating RFE feature selection and multi-temporal remote sensing data[5], and the accuracy of the SVR-LAI model in estimating leaf area index reached R^2=0.737; Fan Xuexing et al proposed a chlorophyll inversion model based on multispectral imaging and PSO-SVR optimization algorithm[6], and realized the prediction of chlorophyll content through the nonlinear mapping of spectral reflectance and SPAD value. The accuracy was improved to R^2=0.91; Han WT et al innovatively adopted the XGBoost-GRA bimodal feature preference strategy combined with RF machine learning model to achieve an inversion of salinity in 0-20 cm soils with an accuracy of R^2=0.820[7], and the accuracy of spatial distribution map was improved to RPI=0.820. The spatial distribution map accuracy is improved to RPIQ=2.273; Huang Linsheng et al fused phase correlation alignment and UNet semantic segmentation network to overcome the problem of multispectral image channel bias[8], so that the lettuce canopy segmentation accuracy reaches 99.19%; Pang Qi integrated hyperspectral imaging and YOLOv3 deep learning framework to break through the bottleneck of hidden defect detection in apples[9], and realized 100% of the F1 scores vs. 68 fps real-time processing performance, which is significantly better than the traditional threshold segmentation method.

In summary, chemometrics methods were crucial in the early development of multispectral image analysis. However, their limited feature extraction and low model complexity severely restrict classification accuracy in complex scenarios. In contrast, deep learning-based methods have emerged as a breakthrough. By constructing deep neural networks with self-learning features, these methods reduce data preprocessing and enable end-to-end analysis. Deep convolutional networks capture multispectral image features at multiple scales, but increased network depth raises model complexity and computational costs, preventing large models from running on mobile devices.This study addresses the inefficiencies of traditional soybean drought phenology detection and the limited field applicability of large devices. By integrating multispectral image feature extraction with deep learning, we developed a portable mobile drought detection system. The system includes canopy segmentation, spectral analysis, and a drought identification model, culminating in the lightweight "Early Acknowedgment for Soybean Drought" App. This enables non-destructive, real-time detection of soybean canopy drought traits, providing a high-throughput tool for breeding and irrigation management. The system's technology roadmap is shown in Figure 1.

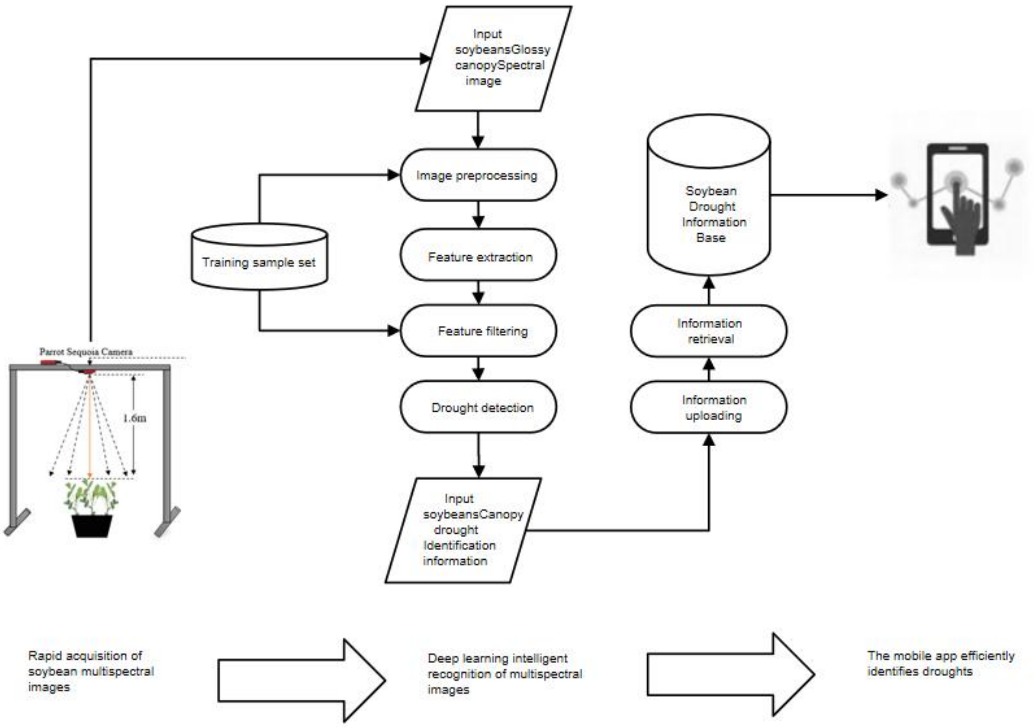


Figure 1 System Technology Roadmap

2 MULTISPECTRAL IMAGE FEATURE SCREENING OF SOYBEAN CANOPY BASED ON RELIEFF ALGORITHM

2.1 Relevant Features Algorithm Principle

As a classical feature evaluation method, Relief algorithm assigns differentiated weight values by analyzing the correlation between each feature and the target category[10]. When the weight evaluation value of a feature is lower than a preset threshold value, the system will automatically reject the feature. The algorithm is concise, effective, and addresses a series of computational and model performance issues derived from the high dimensionality of multispectral features, so the ReliefF algorithm is used as the multispectral feature screening method in this paper. The specific steps are as follows.

Firstly, a sample x_i is randomly selected among the soybean multispectral image samples, and then the distances of the k nearest-neighbor samples in the same class as x_i are computed for $\sum_{j=1}^{k} S(f_i, x_i, H_j)$, and the distances of the k nearest-neighbor samples in a different class from x_i are computed for $\sum_{j=1}^{k} D(f_i, x_i, M_j(A))$, respectively. According to the size of the class spacing and intra-class distance, the size of the weights is adjusted, and the weights are updated by iterating m times, and the feature selection is carried out by the final weights. The formula for calculating the weights is as follows:

$$W^{i+1}(f_l) = W^i(f_l) - \frac{\sum_{j=1}^{k} S(f_i, x_i, H_j)}{mk} + \sum_{C \neq class(x_i)} \frac{\frac{p(A)}{1 - P(class(x_i))} \sum_{j=1}^{k} D(f_i, x_i, M_j(A))}{mk} \tag{1}$$

Where $W^i(f_l)$ is the weight of the l feature f in the first i sample; $H_j(j = 1, 2, \cdots, k)$ is the j th sample among the k nearest neighbor samples in the same class as x_i ; $p(A)$ is the ratio of samples belonging to the class A

in the training samples; $P(class(x_i))$ is the ratio of samples in the same class to the total samples, where $class(x_i)$ is the label of x_i ; $M_j(A)(j=1,2,\cdots,k)$ is the j th sample among the k nearest neighbor samples in the different class from x_i . The distance formula is

$$D_f(x_1,x_2)=\frac{\left|x_{1f}-x_{2f}\right|}{\max(f)-\min(f)} \tag{2}$$

where $D_f(x_1,x_2)$ is the normalized distance between the samples x_1 and x_2 on the f th feature, x_{1f} and x_{2f} are the f th features of the samples x_1 and x_2 , respectively, and $\max(f)$ and $\min(f)$ are the maximum and minimum values corresponding to the feature f in all the samples, respectively.

2.2 Soybean Canopy Drought Recognition Based on 3-Layer Convolution-Relevant Features

In this paper, the multispectral feature dataset of soybean canopy preferred by ReliefF algorithm is used as the input vector to build a multispectral soybean drought convolutional classification model for model simulation[11], and the architecture of the multispectral soybean drought convolutional classification model constructed in this paper is shown in Figure 2.

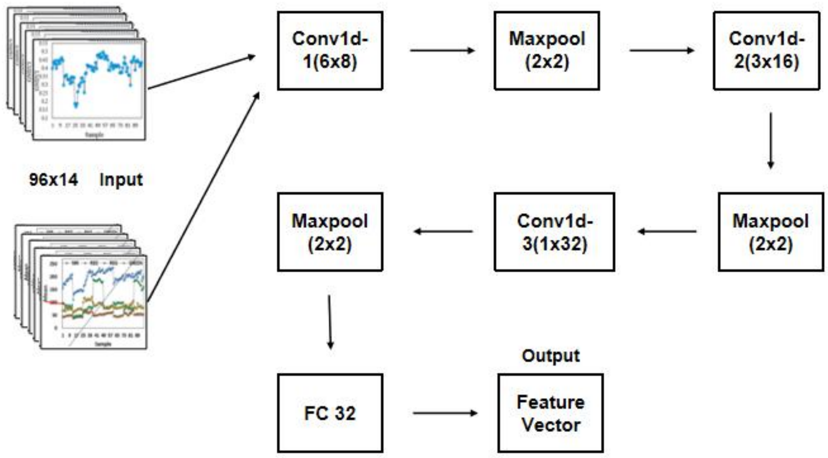


Figure 2 CNN Soybean Drought Classification Model Architecture

Convolutional neural network has powerful feature extraction ability[12], so this paper constructs a soybean canopy drought identification model based on 3-layer convolutional-Relevant Features. This multispectral soybean drought convolutional neural network contains 7 layers, among which there are 3 convolutional layers, 2 pooling layers, 1 fully connected layer and 1 output layer. In this case, since the multispectral soybean canopy in this study is structured data, unlike unstructured image data, the multispectral data features were extracted by a one-dimensional convolutional kernel. The convolutional layers are Conv1d-1, Conv1d-2 and Conv1d-3, with the size of the convolutional kernel of 6, 3 and 1, and the number of convolutional windows of 8, 16 and 32, respectively.The size of the pooling layer is 2, and the number of neurons in the fully connected layer is 32, and the number of the output layer is 1.The principle of the convolutional neural network model constructed in this study and the computational process are as follows:

The whole Convolutional-Relevant Features model consists of three Block modules, where each Block module consists of a convolutional layer, downsampling, and pooling layer. Take the first Block module as an example, the specific process is, first of all, the size of the data set for the 14x1 sample data input to the first convolutional layer, the use of eight 1x6 convolution kernel size, convolution step default 1, feature data by the first layer of the convolution of the size of the shortening of the 9, which is a one-dimensional convolutional operation formula is as follows

$$y_j^l=f\left(\sum_{i=1}^{N}w_{ij}^l * x_i^{l-1}+b_j^l\right) \tag{3}$$

Where y_j^l represents the feature map of the j th convolutional output of the l layer: f represents the nonlinear activation function. $*$represents the convolution operation, N is the number of kernels in the $l-1$ layer, x_i^{l-1} is the feature map of i in the $l-1$ layer, w_{ij}^l represents the weights, and b_j^l is the bias of the j convolution kernel in the l

layer[13]. After completing the convolution operation, the ReLU activation function is used to increase the nonlinear mapping ability of the network. In this paper, Max Pooling with a window size of 2 is used for window sliding calculation for feature compression and information filtering of data. The Max Pooling process used in this paper is shown in the following equation:

$$p_i^{l+1}(j) = \max_{(j-1)w+1 \le l \le jw} \left\{ q_i^l(t) \right\} \tag{4}$$

Where $q_i^l(t)$ is the value of the t neuron corresponding to the i feature quantity in the l layer, w is the pooling layer width, and $p_i^{l+1}(j)$ is the value of the $l+1$ neuron. In order to avoid overfitting, the data are downsampled, the neurons with a random inactivation ratio of 0.2 are randomly inactivated, and the length of the feature map is reduced to 4 and the number of feature channels is changed to 8 after maximum pooling.

3 IMPLEMENTATION OF APP FOR SOYBEAN DROUGHT IDENTIFICATION SYSTEM BASED ON ANDROID TERMINAL

The realization of the overall framework of the APP system of soybean drought identification system based on Android is shown in Figure 3[14]. In the home page, users can log in through the cell phone authentication code and enter the main interface after successful login, which realizes the switching of real-time monitoring, historical data, farmland management and other functions through ViewPage control. In the real-time monitoring module, the SDK of Sequoia multispectral camera is called to obtain multi-band data such as red edge, near infrared, etc. The raw spectral stream is converted to NDVI pseudo-color map and displayed on ImageView control through JNI technology, and at the same time, an asynchronous thread is created to call the PyTorch-trained Conv-ReliefF model to perform the analysis of drought level. The multispectral image data is stored to an Oracle database through a RESTful API and the raw image files are synchronized using MinIO. The server side is built based on Eclipse Java, providing interfaces such as /api/upload to receive images and return model inference results, combined with Redis caching of high-frequency query data to improve response efficiency. The system realizes accurate and safe agricultural drought monitoring through multi-spectral data fusion and HTTPS encrypted communication, and supports the expansion needs of smart farmland management scenarios.

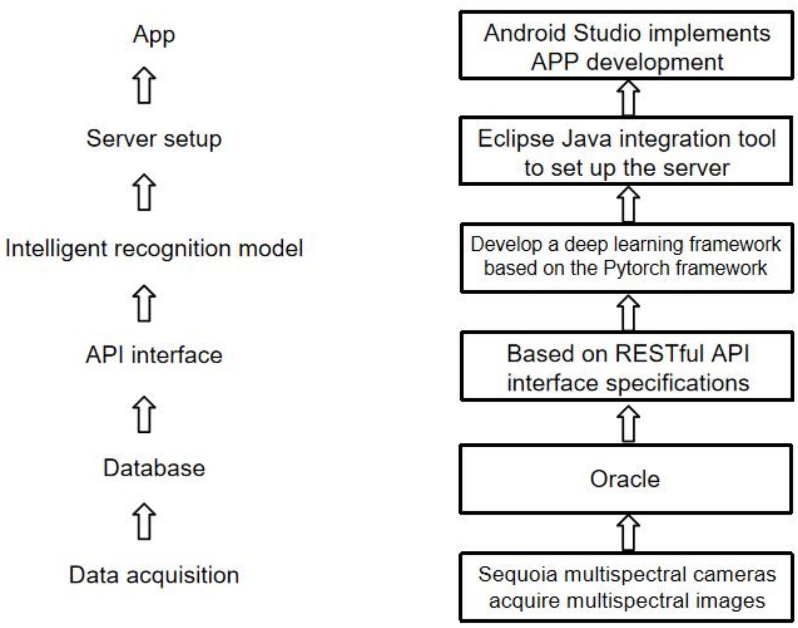
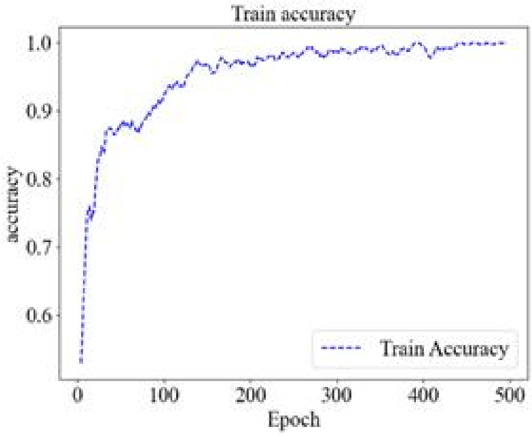


Figure 3 Overall System Framework of the Mobile Soybean Drought Identification System

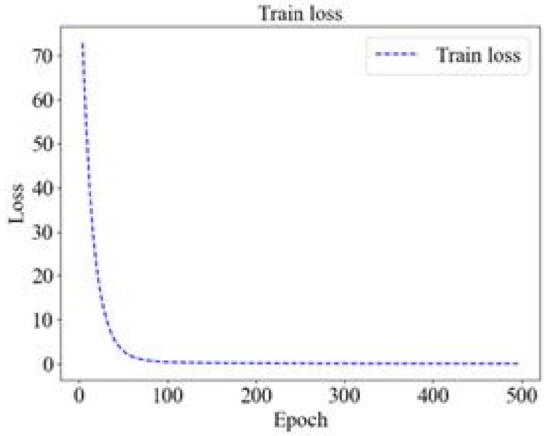
4 RESULTS AND ANALYSIS

4.1 Conv-ReliefF Method Training Results

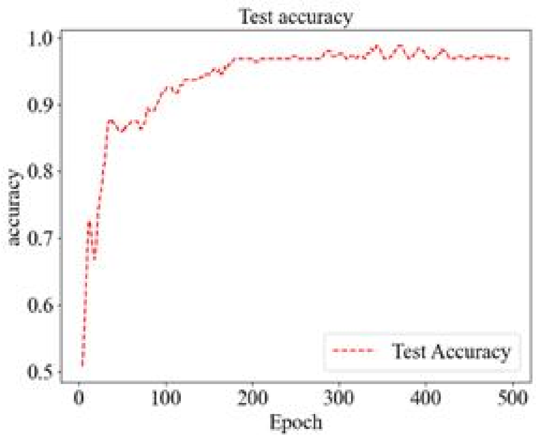
In this paper, the experiments used Parrot Sequoia multispectral camera to collect soybean canopy multispectral data and carry out model training. In the data preprocessing stage, the acquired canopy feature data were divided into a training set and a test set at a ratio of 2:1, in which the training samples totaled 32. In order to eliminate the differences in feature scales, the training set data were normalized using the standardized (Standard-Score) method. For model construction, a one-dimensional convolutional neural network architecture is selected, and the Adam optimizer is used for model optimization (the initial learning rate is set to 0.01), and the L2 regularization strategy is introduced (the coefficient is set to 0.6) to improve the model generalization ability and training stability. During the network training process, binary cross entropy is chosen as the loss function to monitor the model performance, and accuracy is used as the evaluation index to prevent overfitting phenomenon. After 500 rounds of iterative training, the performance evaluation results of the model on the training set and test set are shown in Fig., which specifically demonstrates the trend of the accuracy rate and the loss value.



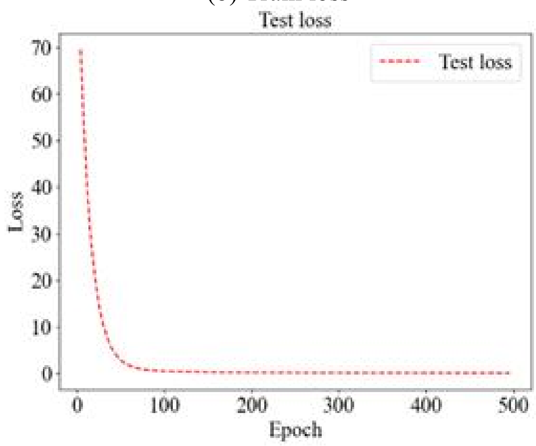
(a) Train accuracy



(b) Train loss



(c) Test accuracy



(d) Test loss

Figure 4 Conv-ReliefF Method Training Results

Observing Figure 4 above, it can be seen that as the number of iterations increases, the overall stabilization is 96.875% at 426 iterations. The model training time is short, only 0.28 s. To evaluate the performance of the trained recognition model, 32 groups of test samples are input into the Conv-ReliefF soybean drought recognition model. To evaluate the performance of the trained identification model, 32 test samples were fed into the Conv-ReliefF soybean drought identification model. The results showed that the model took 0.018 seconds to test, which is more efficient than traditional machine learning models. Among the 32 groups of samples, the model correctly recognizes 31 groups and only 1 group is misclassified, with a prediction rate of 0.96875, which shows good prediction results in the soybean drought classification task.

Focusing on the accurate identification of soybean drought, which is a key issue in agricultural production, this paper constructs soybean drought identification models based on the principles of BP, RBF, RF, SVM and Convolutional neural networks, such as BP, RBF, RF, SVM, OD_Conv and Conv-ReliefF,The accuracy of each model and its equilibrium is shown in Table 1. The study divided the training set and test set in the ratio of 2:1, adjusted the parameters for training in several rounds, and evaluated the training effect with the help of accuracy, running time, and balanced accuracy, and the accuracy of each model reached 100% on the training set. In order to verify the actual performance of the models, 32 sets of test samples were inputted into the trained models and evaluated comprehensively by using the indicators of Accuracy, Precision, F1-score, Recall, etc. The results show that the Conv-ReliefF model has a classification accuracy of 96.88% and a running time of 0.28 seconds, which is higher in recognition efficiency and better in comprehensive performance than the other models, and provides an efficient and effective way to recognize soybean drought. Compared with other models, the Conv-ReliefF model provides an efficient and reliable solution for soybean drought identification. For space reasons, this paper only shows the confusion matrix results of Conv-ReliefF model, as shown in Figure 5.

Table 1 Equilibium Accuracy Table for Each Model

Model	Equilibium accuracy
BP	90.63%
RBF	87.50%
RF	90.63%
SVM	93.75%
OD_Conv	87.50%
ReliefF_Conv	96.88%

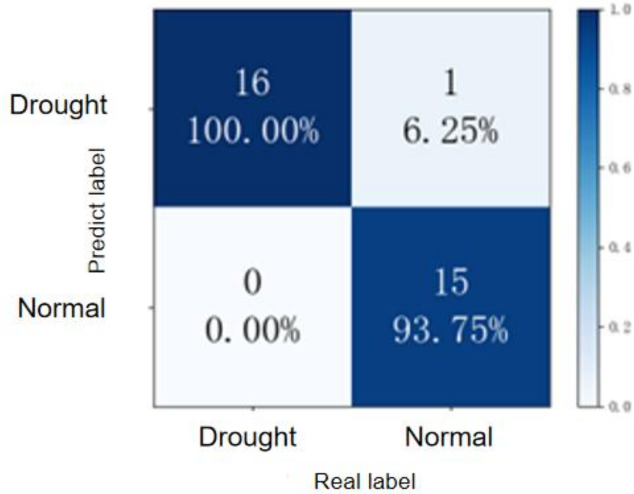


Figure 5 ReliefF-Conv Confusion Matrix Map

4.2Analysis of Application Results Based on Android Conv-ReliefF Soybean Drought Identification System APP

In the operation of the drought recognition module, the original canopy images of the four channels are opened first, and by clicking the Canopy Segmentation button, the system will automatically obtain the target canopy areas of the four channels and calculate the gray value of the canopy in each channel. The segmented image will be displayed in the QLabel control on the right, and the result of the gray value calculation will be presented in the QTextEdit label below. In the Feature Calculation section, click on the Extract All Canopy Features button, the program will calculate the original 37 dimensional feature data and generate an Excel data sheet to be saved to the current path of the program. Users can select the desired features by checking the box, or use shortcut operations such as select all, unselect all, or directly select the model training features to complete the feature selection quickly. The selected feature data is used as input for prediction by the trained model, and the prediction result will be displayed in the model recognition result area below. The specific operation process is as follows: first, select the four channel images to perform automatic crown segmentation, after segmentation, the system automatically calculates the gray value; then click on the training features button to complete the automatic selection of features; and finally click on the trained model, you can get the results of soybean drought recognition in the output box, the recognition effect is shown in Figure 6.

Figure 6 App Specific Implementation Map

Drought identification is realized on a tablet, and using the Conv-ReliefF method, the average computing time for identification is about 280ms, and the accuracy rate is about 96.88%, compared with the traditional machine learning model, the soybean drought classification convolutional neural network established in this paper performs optimally in terms of classification performance, which improves the operational efficiency and the accuracy of model identification. Combined with the application of Android APP, the Conv-ReliefF model shows high efficiency and practicability in actual agricultural production, and the portable device is convenient for farmers to quickly obtain the soybean drought conditions, take timely irrigation measures, effectively reduce the losses caused by drought, and enhance the intelligent level of agricultural production management. It further verifies its superiority in the field of soybean drought identification. Field measurements show that the system supports one-button operation, with a learning cost of less than 10 minutes for farmers, water savings of 15%-20% per hectare, and yield estimation error of <5%, which provides a lightweight, high-precision mobile solution for drought-resistant soybean breeding and precision irrigation, and significantly improves the intelligent monitoring level of smart agriculture.

5 CONCLUSIONS

Aiming at the difficulties of low efficiency of soybean drought phenotype detection and poor applicability in the field, this study innovatively integrates ReliefF feature screening algorithm and lightweight convolutional neural network (Conv-ReliefF) to construct a mobile detection system based on multispectral analysis. By optimizing the screening of 37-dimensional features in the soybean canopy layer through the ReliefF algorithm, combined with a three-layer one-dimensional convolutional network (convolutional kernel 6-3-1), the model achieves a classification accuracy of 96.88% on the self-constructed dataset, with an inference time of only 18 ms for the test set and a training time of 0.28 sec, which significantly outperforms the traditional models such as SVMs, RFs, and so on, in terms of comprehensive performance. The system adopts L2 regularization (coefficient 0.6) to improve the generalization ability, and optimizes the model size through convolution kernel time domain compression (time width=5, step size=2), which is 1/5 of the traditional PC algorithm.The results show that the Conv-ReliefF model size is optimized to be less than 30 MB, and the memory consumption of the mobile terminal is less than 60 MB, which verifies the dual advantages of model lightweighting and high accuracy. The results show that the Conv-ReliefF model is optimized to be less than 30 MB in size and less than 60 MB in memory consumption on the mobile terminal, which verifies the advantages of both lightweight and high accuracy of the model, and is able to meet the demand for real-time detection of soybean drought phenotypes in the field. The "Early Acknowedgement for Soybean Drought" App developed in this paper integrates multispectral data analysis, canopy segmentation (accuracy >97%) and drought diagnosis, and the response time of the whole process is <300 ms, which provides a highly efficient and portable mobile solution for intelligent monitoring of soybean cultivation and precise irrigation, and significantly improves agricultural management. It provides an efficient and portable mobile solution for intelligent monitoring of soybean planting and precise irrigation, and significantly improves the intelligent level of agricultural management.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This research is funded by the National Innovation Training Program for College Students in Heilongjiang Province, Project Name: Design and Implementation of "Early Acknowedgment for Soybean Drought" App, Project No. 202410223078.

REFERENCES

[1] Li Hefeng. Steady increase of domestic soybean production capacity and self-sufficiency. Economic Daily News, 2024-11-22(006). DOI: 10.28425/n.cnki.njjrb.2024.008771.
[2] Shen Panpan. Research on the calculation method of soybean canopy wilt index based on multispectral image processing. Heilongjiang Bayi Nongken University, 2023. DOI:10.27122/d.cnki.ghlnu.2023.000122.
[3] Haoxin Song. Research on citrus information extraction from UAV multispectral images based on deep learning. Guilin University of Technology, 2023. DOI: 10.27050/d.cnki.ghlgc.2023.001051.
[4] Gao Shijiao, Guan Haiou, Ma Xiaodan, et al. A multispectral image extraction method for soybean canopy. Spectroscopy and Spectral Analysis, 2022, 42(11): 3568-3574.
[5] Fu Hongyu, Wang Wei, Lu Jianning, et al. Estimation of physical and chemical traits of ramie based on multi-spectral remote sensing by unmanned aircraft and machine learning. Journal of Agricultural Machinery, 2023, 54(05): 194-200+347.
[6] Fan Xuexing, Zhang Huichun, Zou Yiping, et al. Inversion of plant chlorophyll content based on multispectral imaging and machine learning. Forestry Science, 2023, 59(07): 78-88.
[7] Han Wenting, Cui Jiawei, Cui Xin, et al. Research on salinity estimation of agricultural soil based on feature optimization and machine learning. Journal of Agricultural Machinery, 2023, 54(03): 328-337.
[8] Huang Linsheng, Shao Song, Lu Xianju, et al. Multispectral image segmentation and alignment of lettuce based on convolutional neural network. Journal of Agricultural Machinery, 2021, 52(09): 186-194.
[9] Pang Q. Research on the detection of obvious/hidden defects in apple epidermis based on deep learning and spectral imaging . Shanghai Ocean University, 2022. DOI:10.27314/d.cnki.gsscu.2022.000147.
[10] Bedi S R, Singh A. A Feature Selection Based Relief Algorithm with Fuzzy Logic for Software Effort Estimation.Research Cell: An International Journal of Engineering Sciences, 2018, 30(SP): 1-5.
[11] Jiang C, Sun X, Dai Y, et al. EEG Emotion Recognition Employing RGPCN-BiGRUAM: ReliefF-Based Graph Pooling Convolutional Network and BiGRU Attention Mechanism. Electronics, 2024, 13(13): 2530-2530.
[12] Mahbod A, Saeidi N, Hatamikia S, et al. Evaluating pre-trained convolutional neural networks and foundation models as feature extractors for content-based medical image retrieval. Engineering Applications of Artificial Intelligence, 2025: 150110571-110571.
[13] Yifan D, Chuanbo W, Zheng W. A motor bearing fault diagnosis method based on multi-source data and onedimensional lightweight convolution neural network. Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, 2023, 237(2): 272-283.
[14] Carmona A M ,Sautua J F, Pérez-Hernández O, et al. AgroDecisor EFC: First Android™ app decision support tool for timing fungicide applications for management of late-season soybean diseases. Computers and Electronics in Agriculture, 2018: 144310-313.