Check for updates

Predicting F_v/F_m and evaluating cotton drought tolerance using hyperspectral and 1D-CNN

Congcong Guo[1], Liantao Liu[1]*, Hongchun Sun[1], Nan Wang[1,2], Ke Zhang[1], Yongjiang Zhang[1], Jijie Zhu[3], Anchang Li[1], Zhiying Bai[1], Xiaoqing Liu[1], Hezhong Dong[4]* and Cundong Li[1]*

[1]State Key Laboratory of North China Crop Improvement and Regulation/Key Laboratory of Crop Growth Regulation of Hebei Province/College of Agronomy, Hebei Agricultural University, Baoding, China, [2]College of Mechanical and Electrical Engineering, Hebei Agricultural University, Baoding, Hebei, China, [3]Institute of Cereal and Oil Crops, Hebei Academy of Agriculture and Forestry Sciences, Shijiazhuang, China, [4]Cotton Research Center, Shandong Key Lab for Cotton Culture and Physiology, Shandong Academy of Agricultural Sciences, Jinan, China

The chlorophyll fluorescence parameter F_v/F_m is significant in abiotic plant stress. Current acquisition methods must deal with the dark adaptation of plants, which cannot achieve rapid, real-time, and high-throughput measurements. However, increased inputs on different genotypes based on hyperspectral model recognition verified its capabilities of handling large and variable samples. F_v/F_m is a drought tolerance index reflecting the best drought tolerant cotton genotype. Therefore, F_v/F_m hyperspectral prediction of different cotton varieties, and drought tolerance evaluation, are worth exploring. In this study, 80 cotton varieties were studied. The hyperspectral cotton data were obtained during the flowering, boll setting, and boll opening stages under normal and drought stress conditions. Next, One-dimensional convolutional neural networks (1D-CNN), Categorical Boosting (CatBoost), Light Gradient Boosting Machines (LightBGM), eXtreme Gradient Boosting (XGBoost), Decision Trees (DT), Random Forests (RF), Gradient elevation decision trees (GBDT), Adaptive Boosting (AdaBoost), Extra Trees (ET), and K-Nearest Neighbors (KNN) were modeled with F_v/F_m. The Savitzky-Golay + 1D-CNN model had the best robustness and accuracy (RMSE = 0.016, MAE = 0.009, MAPE = 0.011). In addition, the F_v/F_m prediction drought tolerance coefficient and the manually measured drought tolerance coefficient were similar. Therefore, cotton varieties with different drought tolerance degrees can be monitored using hyperspectral full band technology to establish a 1D-CNN model. This technique is non-destructive, fast and accurate in assessing the drought status of cotton, which promotes smart-scale agriculture.

Introduction

Cotton (*Gossypium hirsutum* L.) is an important cash crop cultivated globally. Drought is major abiotic stress (Cruz de Carvalho, 2008), whose high frequency reduces the average productivity of major crops by up to 50% globally (Lamaoui et al., 2018). According to the World Food and Agriculture Organization, global food output losses caused by drought during the past decade amount to USD 30 billion (Zhang et al., 2021a). Cultivating drought-resistant varieties is not only important for resistance against frequent droughts but also an important current breeding goal. Drought-resistant varieties strongly tolerate drought, with moderate drought stress stabilizing the yields (Wang et al., 2018).

Breeding and screening drought-resistant varieties are usually complex and time-consuming since it depends solely on breeder expertise. Many relevant reports of the classification methods for different genotypes also exist, which mostly focus on fluorescence scanning, protein electrophoresis, deoxyribonucleic acid (DNA) molecular markers (Zhang et al., 2012), the determination of relative water content, net photosynthesis, stomatal conductance electron transfer rate, photochemical quenching, chlorophyll a/b ratio, plant height, and leaf area (Zou et al., 2020). The high-throughput method has gradually become an important technique for selecting drought-resistant varieties from numerous varieties. Drought significantly decreases the leaf water potential, followed by partial leaf stomatal closure, increased leaf temperature, and reduced photosynthetic efficiency (Najafi et al., 2007; Ahmed et al., 2013). Chlorophyll fluorescence kinetic parameters reflect leaf light energy absorption, transformation, transmission, and distribution characteristics (Hikosaka and Tsujimoto, 2021). The maximum photochemical quantum yield (F_v/F_m) in the chlorophyll fluorescence kinetic parameters represents the maximum light energy conversion efficiency in the photosystem II complex (PSII) reaction center. Thus, drought tolerance indicators have subsequently been developed to evaluate the drought adaptability of different plant genotypes. F_v/F_m positively correlates with drought degree (Zou et al., 2020). Therefore, F_v/F_m provides valuable information for evaluating plant physiological changes under drought stress (Lang et al., 2018), hence an efficient drought tolerance index in selecting the best drought tolerant cotton genotype. Measuring the crop drought F_v/F_m is feasible. However, it requires manual measurements and analysis, a 20–30 min plant adaptation period in the dark, which has low efficiency and requires a heavy workload; hence cannot meet plant phenotype analysis needs, such as high flux, automation, and real-time measurement. Therefore, high-throughput evaluation for screening drought-resistant cotton varieties by F_v/F_m warrants further studies. Rapid and efficient methods for screening cotton varieties must be developed by combining high-throughput phenotype methods and drought-resistant variety screening (Shakoor et al., 2017; Feng et al., 2019). This study focused on an accurate and robust prediction of drought-resistant varieties among different cotton genotypes, from small to large spatial scales.

Hyperspectral remote sensing performs fast, non-destructive, and economical data collection. Compared to conventional remote sensing, it produces a large amount of spectral information and has a high resolution and strong spectral continuity. It determines the optimal wave width and effective band from large hyperspectral datasets to obtain the best inversion effect (Yao et al., 2013). In addition, it comprehensively and accurately reflects the inherent spectral characteristics and differences between plants. Compared to the traditional identification method, this technology shortens the analysis time and reduces the material crop consumption, such as wheat (Mahesh et al., 2008; Choudhary et al., 2009), rice (Wang et al., 2015), cotton (Carreiro Soares et al., 2016), and grape (Zhao et al., 2018). Using hyperspectral data to monitor plant growth and development is based on plant spectral characteristics. Based on the spectral reflectance in different wavelength ranges, the spectral index provides a high crop parameter inversion accuracy. The vegetation color, cell structure, and water content determine most plant spectral characteristics. Thus, its successful application depends on a full understanding of the interaction between light and plant matter from the cellular to the canopy scale, the interpretation of reflectance data from different sources and related leaf spectral diversity. However, elucidating the interaction between drought and chlorophyll structural characteristics, cell structures, water, visible light, and near-infrared and short-wave infrared regions is a major challenge due to the inability to separate F_v/F_m from a series of other traits.

Large data volumes and the diversity of analysis methods for hyperspectral data often lead to large data problems (Montesinos-López et al., 2017); hence, advanced algorithms are required for parsing to generate physiological parameters evaluation models. With rapid agricultural artificial intelligence developments (Lu et al., 2017), excellent feature extraction and data inference abilities, and deep learning (DL) algorithms have attracted attention in constructing crop parameter inversion models combined with hyperspectral data (Shah et al., 2019). Machine learning (ML) methods, such as CatBoost, LightGBM, XGBoost, decision trees, Random Forests (RF), Gradient lifting trees (GBDT), adaboost, ExtraTrees, and K-Nearest Neighbor (KNN), are promising for extracting spectral features related to drought resistance by converting original data into new features (Khan et al., 2020). ML usually performs well on a sample-specific basis but loses generalizability when implemented on new data sets with different feature spaces and distributions of different plant species and growth conditions. DL is a new machine learning research field. It was developed to establish and simulate human brain neural networks for analytical learning, and simulates the mechanism of data interpretation in the brain. Thus, it is an unsupervised learning method (Durai and Shamili, 2022; Khan et al., 2022). It derives from artificial neural network research, and its multi-layered perceptron, with multiple hidden layers, which differs from machine learning.

Unlike machine learning, DL has the input, hidden, output layers, and an accepting layer. One-dimensional convolutional neural networks (1D-CNN) are one of the most effective and popular deep learning models. It has the advantage of high recognition accuracy (Ghosal et al., 2018) and provides more general and robust leaf biochemical character retrieval. The network framework includes a convolution, pooling, and full connection layer used for feature extraction, compression, and classification, respectively. Convolutional Neural Networks (CNN) are used in many fields, such as weed and pest identification (Ding and Taylor, 2016), plant disease and stress diagnosis (Ghosal et al., 2018), and agricultural image segmentation (Xiong et al., 2017). Therefore, 1D-CNN has a good developmental history and an advantage in physiological parameter evaluation.

Many studies have used hyperspectral models to analyze and screen crop varieties. For example, Miao et al. (2018) introduced the t-SNE model, pretreated by Procrustes analysis (PA), into the field of hyperspectral imaging (HSI) to classify 800 grains of eight waxy maize varieties. Yu et al. (2021) combined DL and neural networks to classify 18 okra varieties. However, in most studies, the prediction results are based on the spectral information of a single growth stage. Combining the data of each growth stage achieves a higher prediction accuracy. As far as we know, research on screening drought resistant cotton varieties based on hyperspectral reflectance and deep learning at various growth stages has not yet been reported. Therefore, a 1D-CNN regression model with reflectance and F_v/F_m is crucial to screen drought resistant varieties among the different cotton genotypes.

In this study, we aimed to explore the feasibility of F_v/F_m based on 1D-CNN fitting to evaluate drought resistance among cotton genotypes by screening drought-resistant cotton varieties using hyperspectral and deep learning. The F_v/F_m and spectral reflectance of 80 cotton genotypes were measured at the flowering, boll setting, and boll opening stages under drought stress. We hypothesized that deep learning with strong interpretation and stability could be used to interpret the specific spectral responses of drought-resistant cotton genotypes, mainly the leaf reflectance in different genotype diversity and environmental change datasets. The specific objectives were: (1) To compare and analyze the full spectral data and the Successive Projections Algorithm (SPA) dimension reduction data; (2) To compare 1D-CNN with Categorical Boosting (CatBoost), Light Gradient Boosting Machine (LightBGM), XGBoost, DT, RF, Gradient elevation decision trees (GBDT), Adaptive Boosting (AdaBoost), Extra Trees (ET), and K-Nearest Neighbors (KNN); (3) To determine whether F_v/F_m prediction is feasible for screening cotton drought resistant varieties through cluster analysis. Based on Savitzky-Golay (S-G) and 1D-CNN model coupling, an F_v/F_m evaluation model was created, and a model update strategy was proposed to improve accuracy and robustness.

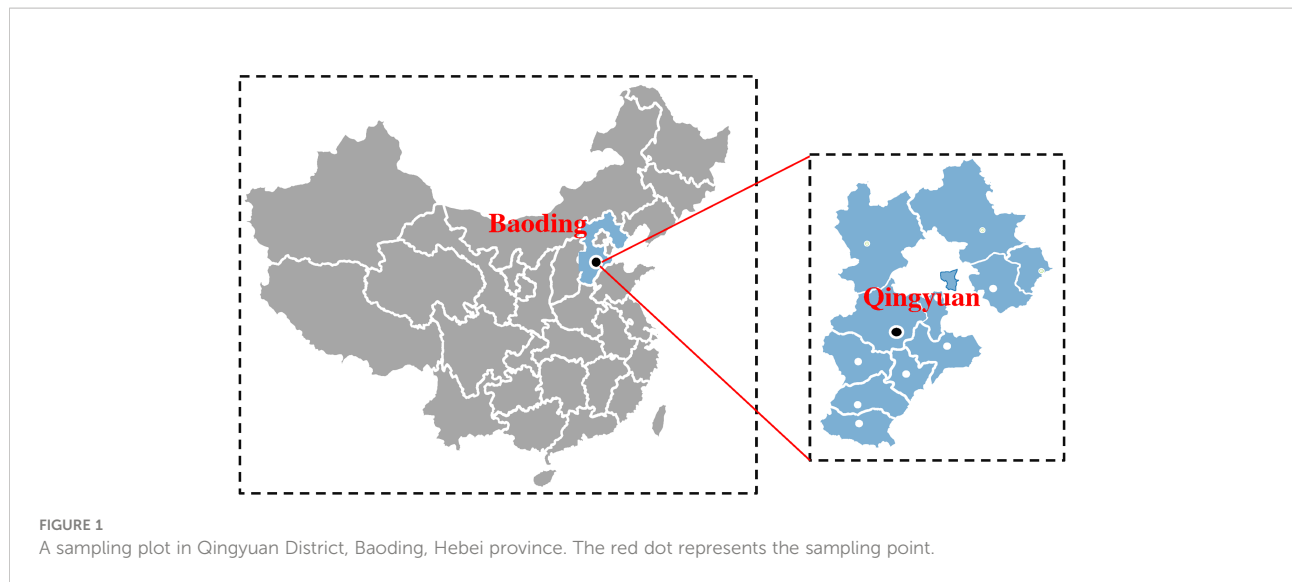# Materials and methods

Plant materials

Eighty cotton cultivars widely cultivated in the Yellow River Basin and the lower reaches of the Yangtze River across different timelines were analyzed in this study, as shown in Supplementary Table 1.

Experimental design and treatments

The experiment was conducted in a cotton field at Qingyuan experimental station of Hebei Agricultural University (38.85° N, 115.30° E, Baoding City, Hebei, China) from April to October 2021. The site information (Qingyuan Experiment Station) is presented in Figure 1. The study location has a temperate continental monsoon climate, with an average annual average temperature of 13°C and 2700 sunshine hours. The annual average precipitation is 532 mm, with about 60% of the precipitation from July to August. The experiment was laid out in a randomized complete block design (Supplementary Figure 1). The experiment had two drought stress levels based on the soil relative water contents (SRWC), including CK (well-watered, 75 ± 5% SRWC serving as the control) and DS (drought stress with 45 ± 5% SRWC) (Gao et al., 2020; Xiao et al., 2020). There were 160 plots per treatment replicated three times totalling 480 plots. The SRWC was monitored by time domain reflectometry (TDR, TRIME TDR series soil moisture meter, IIMKO Company, German) and then watered to maintain the SRWC within the appropriate ranges using micro-sprinkler irrigation.

Selected cotton seeds were sown on 24 April 2021. Four to five seeds were manually sown per hill using the hill-dropping seeding method, with a planting density of 5 plants m^{-2} and a row spacing of 48 cm. Next, mulching was done with a plastic film along the rows. The seedlings were thinned to one vigorous stand per hill upon germination at the two true-leaf stages (Zhang et al., 2021c). Drought stress treatment was induced at the third true-leaf stage. Each plot received 450 kg ha^{-1} of compound fertilizer containing 15% N, 15% P$_2$O$_5$ and 15% K$_2$O as base fertilizer, and 150 kg ha^{-1} urea (46% N) was top-dressed at flowering. In addition, pest control, weed control, chemical control, and plant pruning were performed according to local agronomic practices. The soil texture based on the USDA soil classification standards of the tested soil at different soil layers in the cotton field is shown in Supplementary Table 2.

An electrically powered rain-out shelter was used to protect the plants against receiving precipitation. A rain sensor automatically controlled the rain-out shelter switch. The shelter closed automatically in the event of rain and opened as soon as the rain stopped. Thus, as described previously, any possible interference of natural precipitation with the waterlogging experiment was avoided.

FIGURE 1
A sampling plot in Qingyuan District, Baoding, Hebei province. The red dot represents the sampling point.

Determination of indices and methods

Leaf hyperspectral, F_v/F_m, RWC and LWC were measured on 6 July 2021 (flowering stage), 14 August 2021 (boll setting stage), and 17 September 2021 (boll opening stage). Three representative plants were randomly selected from each plot. The specific determination of indices and methods was as follows:

Hyperspectral data collection

Based on the HR-1024i spectrometer (SVC, USA), the instrument blade clamp light source was used to measure the leaf surface reflection spectrum. The spectrometer had a measurement range of 350–2500 nm and a total of 1024 channels. The spectral resolution was 3 nm, and the sampling interval was 0.6 nm. To ensure a full spectrometer probe view field on leaf samples under the sun, the spectrometer sensor probe was vertically oriented downward, about 0.7 m from the cotton canopy top, and the field angle was set at 25 degrees. White board correction was carried out before each measurement to reduce error. The measurements were carried out in sunny, cloudless, windless, or low wind speed weather, between 10:00 am and 2:00 pm. Three representative, uniform, and pest-free plants were selected from each test plot to measure the reflection spectrum of the top four and fully developed leaves after topping. Before each measurement, the dust on top of the cotton leaves was wiped off to ensure the leaf surfaces were kept clean. Four sample points per leaf were selected, and their average was used as the leaf reflection spectrum. Measurements were taken once per month for three consecutive months. After field spectrum measurements, the top leaf for each plant was marked on its underside and labelled with a serial number for subsequent F_v/F_m measurements to ensure consistency. The detailed determination method and leaf selection are presented in Supplementary Figure 2.

Chlorophyll fluorescence content

A portable FMS-2 fluorometer (Hansatech, King's Lynn, UK) was used to measure the chlorophyll fluorescence characteristic parameter F_v/F_m for newly developed, inverted leaves. Leaf initial (F_o) and maximum fluorescence (F_M) were measured from 0:00 am to 2:00 am. The maximum photochemical quantum yield was then calculated as $F_v/F_m = (F_m-F_o)/F_m$ (Bilger and Björkman, 1990).

Root water contents and leaf water contents

Three plants were selected and uprooted from each plot. Next, their roots and shoots were separated, and the fresh weights were determined. The roots and shoots were then dried 80°C to a constant weight to determine the dry weights. Finally, the water content was calculated as follows:

$$Water\ \ content\,(\%)$$
$$= \frac{Fresh\ \ weight - Dry\ \ weight}{Fresh\ \ weight} \times 100 \qquad (1)$$

Calculation of drought resistance coefficient

The average F_v/F_m was measured to calculate the drought tolerance coefficient as described by Mwadzingeni et al., 2016.

$$Drought\ \ tolerance\ \ coefficient\ \ of\ \ F_v/F_m\,(\%)$$
$$= \frac{Average\ \ value\ \ of\ \ F_v/F_m\ \ DS}{Average\ \ value\ \ of\ \ F_v/F_m\ \ CK} \qquad (2)$$

F_v/F_m is the maximum photochemical quantum yield, *CK* is the normal conditions, *DS* indicates drought stress.

Spectral pretreatment and characteristic wavelength screening

Extraction, reflectance, and spectral pretreatment

The first step was to superimpose and match all spectral curves. In the second step, S-G first-order smoothing was used to eliminate spectral noise and reduce the influence of environmental background interference due to the spectral mutation of the instrument (del Amor et al., 2020). The third step was to remove the file header from the processed data, generate raw data, and save it as a TXT text file. The fourth step was calculating the averages of spectral data and generating spectral data for each ground object type. The fifth step was to interpolate the obtained data because the whiteboard reflectance band did not match the spectral band of each ground object type. The final step was to select the fourth data column (percentage) in the file and multiply the whiteboard reflectance according to the reflectance formula described by Zhao et al., 2022. This test adopted the vertical measurement method using the following formula:

$$Rt = \frac{L}{Lr} Rr \qquad (3)$$

Rt is the reflectivity of the measured object, *Rr* is the reflectivity of the standard version, *L is* the measured value of the measuring object, *Lr* is the standard value of the instrument.

SPA filter characteristic wavelength

A total of 1440 data groups were recorded. The SVC HR (overlay) software was used to extract the wavelength and reflectivity of each sample, and MATLAB was used to perform SPA on all spectra data to extract the characteristic wavelengths. Relevant source code can be found online https://blog.csdn.net/weixin_43637490/article/details/118468559.

Model development

One dimensional convolutional neural network (1D-CNN)

1D-CNN modeling was used to screen the spectral information of cotton drought-resistant genotypes. The main reasons were as follows: (1) the CNN network analyzed one-dimensional data (leaf spectral information) well. (2) It was able to advance the nonlinear mode from the data. (3) It allowed hierarchical spectral data processing to support feature abstraction and extraction. CNN is one of the best algorithms in deep learning, which can be divided into one-, two-, and

three-dimensional. 1D-CNN is a classical deep neural network with high robustness, similar to 2D, with a local connection and weight-sharing characteristics. 1D-CNN was selected to adapt to the nature of spectral data (that is, the spectral reflectance had a one-dimensional data structure) to allow the convolution operation to extract the learning features of patterns. A convolutional neural network is usually used for image recognition, target detection, and classification (Fukushima, 1980). 1D-CNN also performs well in time series prediction and data fitting. In contrast, 2D-CNN is mainly used for image and text recognition, and 3D-CNN is for video recognition and medical applications. Due to its unique structure, CNN processes network structure data characteristics well, effectively solving the data processing difficulties caused by other factors (Liang et al., 2020). The hierarchy proposed in this study includes an input layer, multiple hidden layers (convolution, activation, and pooling layer), and the composition of a full connection (dense) and output layer (Figure 2) (Zhang et al., 2021b):

The convolution layer functions to extract input data features. Different convolution kernels are equivalent to different feature extractors. The main feature is the use of weight sharing and local connections. The operation of one-dimensional convolution is shown in formula (3):

$$y^i = f(\textstyle\sum_i k^{ij} \star x^i + b^j) \qquad (4)$$

where $\star$ represents convolution operation, y^i is the ith output characteristic diagram, x^i is the ith input characteristic diagram, k^{ij} is the convolution kernel used in the layer convolution calculation, and b^j is the offset of the jth characteristic diagram.

For the nonlinear transformation of features extracted from CNN and dense layers, the output of these layers and extracted features were activated using the corrected linear unit (ReLU) function (Cui and Fearn, 2018) (formula (4). The nonlinear activation function ReLU has a low computational cost and fast convergence speed. Its formula is:

$$R(x) = max(0, x) if \begin{cases} x < 0; R(x) = 0 \\ x \geq 0; R(x) = x \end{cases} \qquad (5)$$

where x is the feature of CNN or dense layer calculation.

The pooling layer was abstracted as statistical information extraction to reduce dimensionality and minimize array dimension based on maintaining the original characteristics. The convolution layer significantly reduces the number of network connections. Adding a pooling layer after a convolution layer avoids overfitting to a certain extent. The pooling layer effectively reduces the number of neurons, making the network invariant to small local morphological changes, which creates a larger receptive field. Two types of common pooling functions are recognized: maximum pooling (taking the maximum value of all neurons in a region) and average pooling (taking the average value of all neurons in a region), expressed as
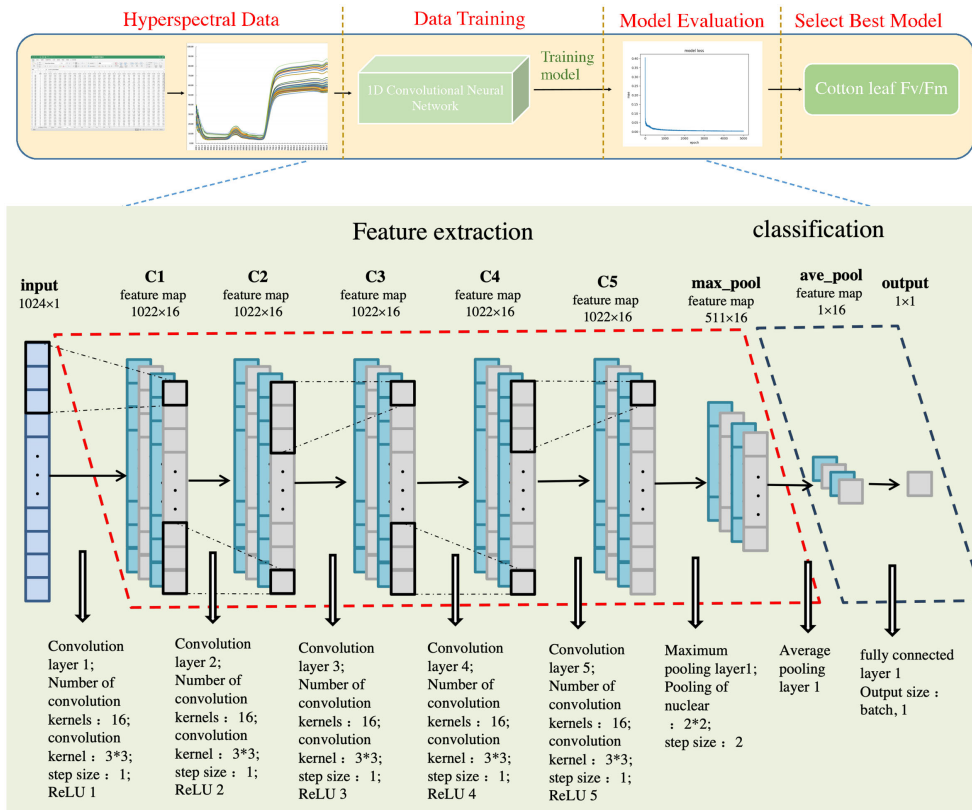
FIGURE 2

Flow chart illustrating 1D-CNN data processing. The training set accounts for 75% and the test set 25%. A matrix was created with 72 rows, 1024 columns, and a dimension added to the training set channel. Another matrix was created with nine rows, 1024 columns, one channel number, zero elements, and a dimension was added to the test set channel. Finally, an empty deep learning model named "model" was defined. Next, five one-dimensional convolution layers were added (corresponding to five ReLU activation functions), including 16 convolution cores measuring 3 x 3 and a step length of one. The first maximum pool layer was added, and the pool core size was 2 x 2, and the step distance was two. The first global average pooling layer was then added, yielding the first full connection layer; the output size was (batch, 1). The RMSProp optimizer was then defined with a learning rate of 0.001, the gradient decay rate of 0.9, the fuzzy factor was zero, and the learning rate decay rate was zero. The MAE loss function and RMSProp optimizer were then used. The framework of the in-depth learning model was integrated, and data was transferred to the defined model by training 5000 epochs; the amount of data in each batch was 5. Finally, the table was given a title, and the image displayed. The test code is the same as the above training code.

in formula (5) and (6), respectively:

$$p^l = max[a^l] \quad (6)$$

$$p^l = \frac{1}{k}\sum a^l \quad (7)$$

where p is the characteristic matrix obtained by pooling, l is the characteristic graph width, and a is the characteristic matrix after convolution layer activation. The maximum and average pooling values calculate the maximum and average values in the adjacent rectangular area, respectively, and location-independent information can be obtained through the maximum pooling value.

The full connection layer is similar to the relationship between one layer and the next layer in the feed-forward network, in which each node of the upper layer and the nodes of the next layer have a weight connection. It is mainly used to complete the final

prediction. Each output neuron of the full connection layer is connected to the neuron in the upper layer, and the input characteristics are combined after the activation function is used to output the prediction results. For the prediction problem, the output layer gives the probability value of the prediction category. Its output is given by formula (7):

$$\delta_i = f(w_i p_i + b_i) \quad (8)$$

where $i = 1, 2,$ and k; δ is the ith output, with a total of K outputs; w_i and b_i are the weights and thresholds of the ith neuron, respectively; and $f(x)$ is the activation function.

In this study, a vector that extracted 1024 spectral features was constructed as the input layer, the F_v/F_m prediction value was used as the output layer (usually, the input vector length is larger than the convolution kernel length), and the hidden layer included 1D-CNN with five convolution layers and two pooling

layers (Figure 2). The spectral data was convoluted, and the convolution filter (also known as the kernel) was used to extract the feature map. The scaler variable was used to accept the entire data normalization process for the following anti-normalization. Data were subsequently normalized in Excel by subtracting the mean and dividing by the variance.

The number of hidden layers, the number of feature maps in each layer, the CNN kernel size, the pool and step size, and the regularization parameters are all adjustable and were optimized by experience to obtain the best value. The optimized architecture specification is presented in Figure 2. Additionally, the proposed architecture was developed as a common architecture for multiple scenarios and case studies (multiple independent data sets), while the existing architecture was evaluated separately on a single data set. The training data set in the CNN model developmental stage was randomly divided into two sub-datasets, calibrated and validated. During feed-forward and backpropagation, these batches were sequentially fed into the network. Once all batches were entered into the model (training era), the validation data set was used to evaluate model efficiency and accuracy on unknown samples. The model was trained on 6 July 2021 (flowering stage), 14 August 2021 (boll setting stage), and 17 September 2021 (boll opening stage) to ensure sample calibration and verification convergence.

Machine learning models

For a more comprehensive model performance and accuracy comparison, nine machine learning algorithms, including CatBoost, LightBGM, XGBoost, DT, RF, GBDT, AdaBoost, ET and KNN, were used for modeling and comparative analysis using 1D-CNN.

CatBoost is a decision tree-based model consisting of an open source software library developed by Hancock (2020) with categorical features in a special way. LightGBM is a distributed gradient boosting framework based on a decision tree algorithm, which supports single-machine multi-threading and multi-machine parallel computing, to quickly process massive data (Meng et al., 2016). XGBoost is an additive model that optimizes only the sub-model in the current step in each iteration (Chen and Guestrin, 2016). DT is a non-parametric supervised learning tool with a tree structure composed of four elements: decision nodes, program branches, state nodes, and probability branches (Sarker et al., 2020). RF is a typical bagging algorithm in ensemble learning (Breiman, 2001), that randomizes the use of variables (columns) and data (rows) to generate many classification trees and then summarizes the results of the classification trees. GBDT was developed by Friedman (2001), and builds on each tree, learning the residual (negative gradient) of the sum of all previous tree conclusions (Kriegler and Berk, 2010). AdaBoost is an algorithm for constructing strong classifiers as a linear combination of simple weak classifiers (Freund and Schapire, 1997; Wang et al., 2022). ET

is directly divided using random features and random thresholds on random features (Geurts et al., 2006; Ahmad et al., 2018). KNN was proposed by Cover and Hart (1967) and is not limited to a fixed number of parameters (Guo et al., 2003).

Model evaluation

To evaluate model performance, leaf samples from each data set were sorted, and 75% of the samples were used as the training data set and the remaining 25% as the test data set. In deep learning, the loss function is used to find errors or deviations in the learning process. However, the loss function uses the same metrics as the training process, which differs in value, to evaluate the performance of the generated model to ensure species fairness in the training and testing data sets (Burnett et al., 2021). Therefore optimization is a key step in comparing prediction and loss functions to optimize input weights. During model training, full-spectrum data is used as input, and model accuracy and loss are recorded simultaneously. The network parameters are fine-tuned based on the results. Therefore, the determination coefficient (R^2), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE) are selected to accurately evaluate test results (Ibrahim et al., 2021).

Set the predicted value to: $\hat{y} = \{\hat{y_1}, \hat{y_2}, \cdots \hat{y_n}\}$ And the true value to $y = \{y_1, y_2, \cdots, y_n\}$. R^2 is the determination coefficient. The higher the model R^2, the higher the accuracy, and the better the fitting effect. The formula is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \in [1, 0] \qquad (9)$$

RMSE is the root mean square error, the difference between the predicted and actual values. The smaller the model RMSE value, the better the model prediction. The calculation formula is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \in [0, +\infty) \qquad (10)$$

MAPE is the mean absolute percentage error; a statistical index used to measure prediction model accuracy. The smaller the model MAPE value, the higher the prediction model accuracy. The calculation formula is as follows:

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}i - yi}{yi}\right| \qquad (11)$$

MAE is the mean absolute error, which is the average of the absolute error between the real and predicted values. It accurately reflects the predicted error value. The larger the model MAE value, the greater the error, indicating a lower prediction model accuracy. The calculation formula is as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|yi - \hat{y}i| \qquad (12)$$

where n is the number of samples, yi is the true values of cotton PH or AGB, $\hat{y}i$ is the predicted values of cotton PH or AGB, and $\bar{y}i$ is the average of the PH or AGB true values.

Results

Effects of drought stress on F_v/F_m in cotton leaves

Generally, when comparing drought stress effects on F_v/F_m (Figure 3), statistical differences were observed among the flowering, boll setting, and boll opening stages ($p \leq 0.05$). The DS and CK were initially increased and then decreased in the three cotton growth stages. DS treatment significantly reduced F_v/F_m ($P < 0.05$), by 2%, 12%, and 3% across the three stages, respectively.

Correlation between F_v/F_m and RWC, and LWC

Correlation analysis between F_v/F_m and RWC, and LWC is illustrated in Figure 4. The results revealed a significant positive correlation between F_v/F_m and RWC, and LWC under DS (Figure 4B) and CK (Figure 4A). Thus, F_v/F_m significantly positively correlated with drought resistance in cotton. F_v/F_m was further used as the input in the model to evaluate the drought resistance of cotton.

Preprocessing of hyperspectral data

The spectrum was pretreated to reduce the influence of the external environment, the dark current of the spectrometer and to eliminate baseline drift, light scattering, and spectrum noise. The Savitzky-Golay technology was applied to preprocess the hyperspectral data, eliminating spectral differences (filtering noise and smoothing waveforms) caused by different scattering levels and enhancing spectral and data correlation. The spectral band peaks and valleys were obvious, overlapping peak interference was avoided, and spectral resolution and sensitivity were improved through Savitzky-Golay pretreatment (Figure 5).

Changes in the cotton canopy reflectance spectrum under different conditions

The cotton canopy spectral reflectance was measured at the flowering (Figures 6A), boll setting (Figures 6B), and boll opening stages (Figures 6C), respectively. The trends for the different varieties at different growth stages were similar, and the differences were obvious under different soil water conditions (Figure 6). In the visible light region (350–750 nm), there were two absorption valleys (370–510 and 600–710 nm) and reflection peaks (520–580 nm). The canopy spectral reflectance increased with drought stress, especially at the "green peak". The higher the soil water content, the better the plant growth, the larger the leaf area index, the higher the chlorophyll content, the stronger the absorption of blue and red light, and the deeper the red valley, leading to an obvious green peak. The opposite
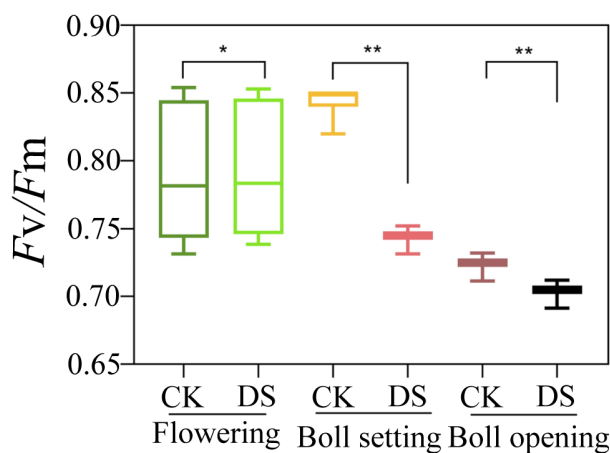

FIGURE 3
Effects of drought stress on F_v/F_m at the flowering, boll setting, and boll opening stages. CK, normal conditions; DS, drought stress. * and ** indicate significance at the 0.05 and 0.01 probability levels, respectively.
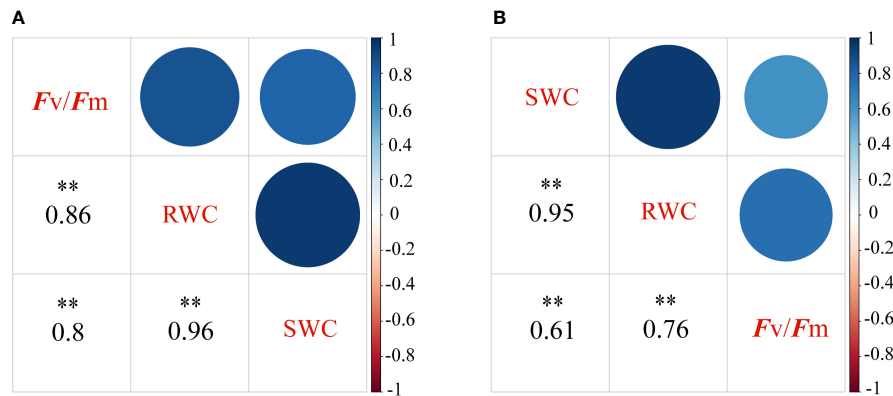
scenario leads to a shallower red valley and, thus, a gentler and less obvious curve at the green peak. However, a reflection platform (760–1250 nm) occurred in the near-infrared region (750–1350 nm), where 1000 nm dropped abruptly. Spectral reflectance decreased with drought stress due to cotton cell structural changes, especially in the "near-infrared platform",

where the difference was significant. A lower spectral reflectance occurs under heavy drought stress. The spectral canopy curves under different drought conditions showed similar trends in the other growth stages. The reflectivity showed a downward trend in the short infrared band (1350–2500 nm), and two water absorption bands occurred at the 1450 and 1950 nm bands.
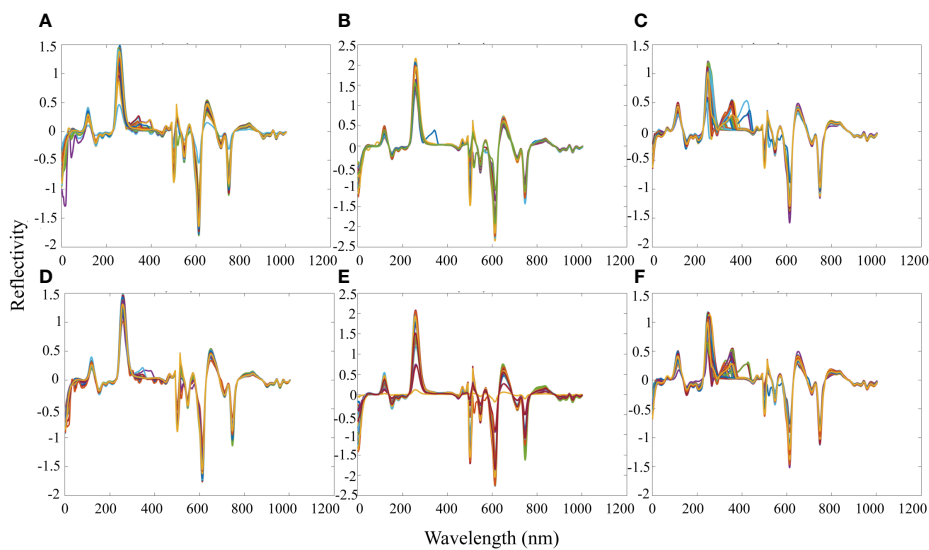
FIGURE 5
Image analysis of spectral data, preprocessed by Savitzky-Golay, under different drought conditions and cotton growth stages. **(A)** The flowering stage of cotton under normal conditions; **(B)** The boll setting stage of cotton under normal conditions; **(C)** Boll opening stage of cotton under normal conditions; **(D)** The Flowering stage of cotton under drought stress; **(E)** The boll setting stage of cotton under drought stress; **(F)** Boll opening stage of cotton under drought stress.

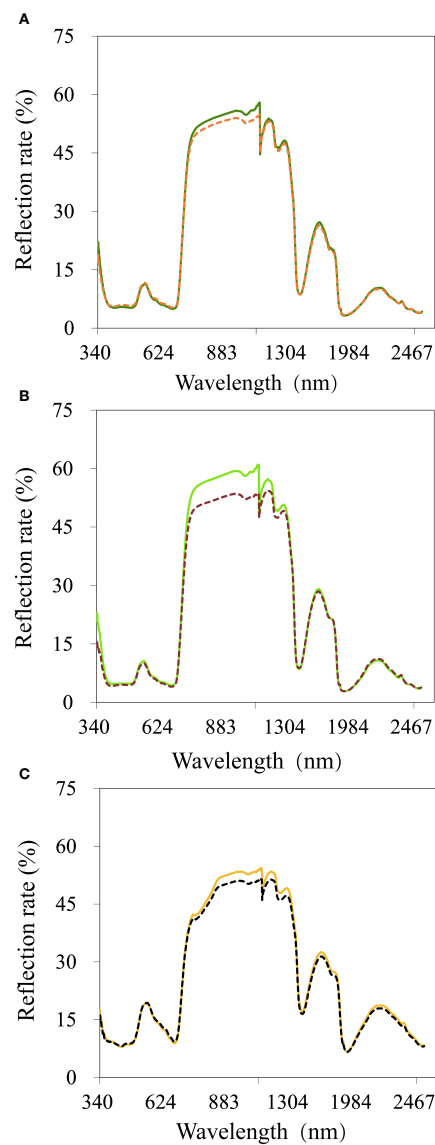**FIGURE 6**
Average reflectance spectra of cotton leaves under different drought conditions and growth stages. **(A)** The flowering stage of cotton under normal conditions and drought stress; **(B)** The boll setting stage of cotton under normal conditions and drought stress; **(C)** The boll opening stage of cotton under normal conditions and drought stress; Each line represents the average value of 240 reflectance spectra of 80 different cotton varieties. Solid lines represent normal conditions; dashed lines represent drought stress.

Full band modeling and analysis and comparative analysis of various modeling methods

To determine the best model algorithm for predicting cotton leaf F_v/F_m, we used the full band and the characteristic wavelengths screened by the SPA algorithm to compare and analyze 1D-CNN, CatBoost, LightBGM, XGBoost, DT, RF, GBDT, AdaBoost, ET, and KNN, respectively. The characteristic wavelengths screened by SPA were inadequate (Supplementary Table 3). Thus, only full band

modeling results are shown here (specifically, training and test sets; Table 1). 1D-CNN, CatBoost, LightBGM, XGBoost, DT, RF, GBDT, AdaBoost, ET, and KNN had a relatively stable model accuracy under the different drought conditions during the flowering stage, but nine of the machine learning algorithms (CatBoost, LightBGM, XGBoost, DT, RF, GBDT, AdaBoost, ET, and KNN) were relatively unstable in estimating F_v/F_m during the boll setting stage under drought stress. 1D-CNN was also relatively unstable in estimating F_v/F_m during the boll setting stage under drought stress. However, the 1D-CNN model had the highest

accuracy and the best effect in the comprehensive evaluation of cotton drought stress. The flowering stage had the highest accuracy when comparing the predictions and analyses of the various stages. The model was more stable under normal conditions.

Furthermore, the loss function of 1D-CNN was observed to decrease rapidly, and the loss rate was low, which improved the accuracy and reduced diagnosis time, leading to a better diagnosis performance (Figure 7).

F_v/F_m as predicted from canopy characteristics

To evaluate the cotton drought tolerance using the spectral features extracted by 1D-CNN, the predicted F_v/F_m value was determined by 1D-CNN and correlated with the actual value (Figure 8). Generally, under sufficient water conditions and drought stress, the correlation between the predicted and measured values was high ($R^2 \geq 0.641$). However, the correlation coefficient was the highest under sufficient water conditions (R^2 of flowering, boll setting, and boll opening stages were 0.908, 0.974, and 0.821, respectively; Predicted and measured of flowering, boll setting, and boll opening stages were 0.7894 and 0.7923, 0.8467 and 0.8439, 0.7246 and 0.7241,

respectively; Figures 8A-C). In addition, the correlation coefficient at the flowering stage was the highest among the treatments (R^2 of CK and DS were 0.908 and 0.959, respectively; Predicted and measured of CK and DS were 0.7894 and 0.7923, 0.7959 and 0.7955; respectively; Figure 8A, D).

Cotton drought tolerance evaluation based on the F_v/F_m drought tolerance coefficient and cluster analysis

Since the above fitting effect was the highest at the flowering stage, the drought tolerance coefficient was used to evaluate cotton drought tolerance. We clustered the F_v/F_m and predicted value drought tolerance coefficients through cluster analysis, thereby highlighting the varieties with strong drought tolerance (Figure 9). We assumed that the higher drought tolerance coefficients for predicted or measured F_v/F_m values indicated enhanced drought resistance. The predicted F_v/F_m classification was similar to the manual measurement classification (Figures 9A, B). The top ten drought tolerant varieties obtained through cluster analysis and evaluation of the measured drought tolerance coefficients were: 38, 24, 6, 56, 25, 58, 8, 43, 71, and 72 (Figure 9A). The top ten drought

TABLE 1 Modeling of drought tolerance at different cotton growth stages with different prediction models.

Prediction model	Conditions	Flowering stage			Boll setting stage			Boll opening stage		
		RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
1D-CNN	CK	0.016	0.009	0.011	0.003	0.005	0.003	0.002	0.005	0.001
	DS	0.010	0.005	0.006	0.006	0.005	0.007	0.002	0.005	0.002
CatBoost	CK	0.021	0.018	2.299	0.001	0.001	0.115	0.003	0.002	0.283
	DS	0.017	0.015	1.849	0.003	0.002	0.315	0.004	0.003	0.454
LightGBM	CK	0.010	0.005	0.629	0.002	0.001	0.167	0.002	0.001	0.162
	DS	0.005	0.003	0.369	0.004	0.003	0.425	0.003	0.002	0.310
XGBoost	CK	0.010	0.003	0.445	0.001	0.001	0.143	0.002	0.001	0.183
	DS	0.007	0.003	0.334	0.005	0.003	0.449	0.003	0.002	0.311
DT	CK	0.010	0.003	0.424	0.002	0.001	0.133	0.002	0.001	0.180
	DS	0.009	0.005	0.636	0.005	0.003	0.405	0.003	0.002	0.275
RF	CK	0.010	0.004	0.528	0.001	0.001	0.089	0.002	0.001	0.177
	DS	0.007	0.004	0.556	0.004	0.003	0.345	0.003	0.002	0.282
GBDT	CK	0.010	0.003	0.390	0.001	0.001	0.086	0.002	0.001	0.174
	DS	0.007	0.004	0.508	0.004	0.003	0.362	0.003	0.002	0.237
AdaBoost	CK	0.010	0.003	0.431	0.001	0.001	0.087	0.002	0.001	0.176
	DS	0.003	0.001	0.171	0.004	0.003	0.343	0.003	0.002	0.313
ET	CK	0.010	0.005	0.608	0.001	0.001	0.086	0.002	0.001	0.157
	DS	0.005	0.003	0.428	0.004	0.002	0.311	0.003	0.002	0.279
KNN	CK	0.038	0.022	2.765	0.001	0.001	0.123	0.003	0.002	0.250
	DS	0.044	0.032	4.011	0.004	0.003	0.424	0.004	0.003	0.365

1D-CNN, One-dimensional convolutional neural network; CatBoost, Categorical Boosting; LightBGM, Light Gradient Boosting Machine; XGBoost, eXtreme Gradient Boosting; DT, Decision Tree; RF, Random Forest; GBDT, Gradient elevation decision tree; AdaBoost, Adaptive Boosting; ET, Extra Trees; KNN, K-Nearest Neighbors; RMSE, Root mean square error; MAPE Mean absolute percentage error; MAE, Mean absolute error.
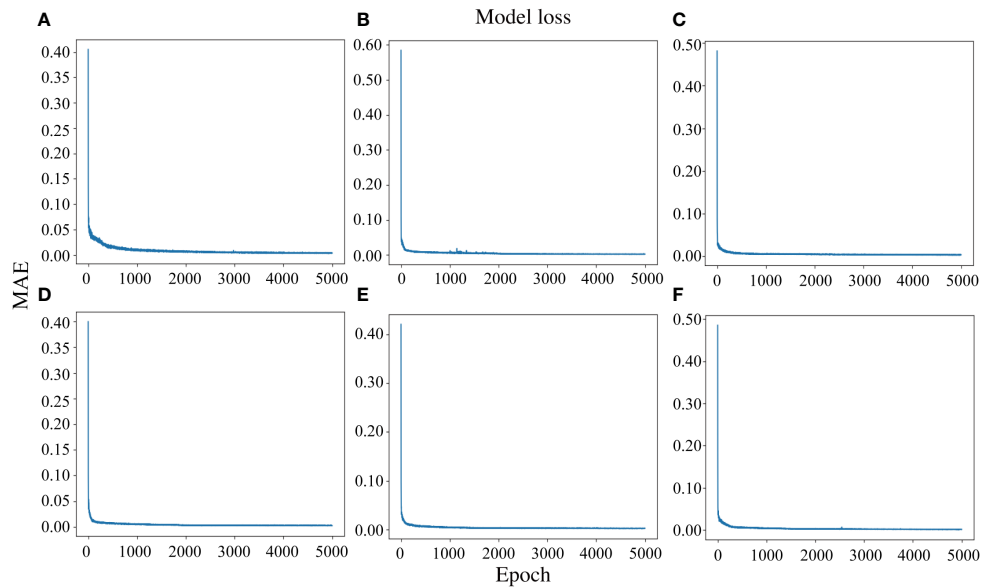
FIGURE 7

The loss rate curve of the 1D-CNN model with different drought conditions and cotton growth stages. **(A–C)**, Flowering stage, boll setting stage, and boll opening stage under normal conditions, respectively; **(D–F)**, Flowering stage, boll setting stage, and boll opening stage under drought stress, respectively.

tolerant varieties predicted were 38, 24, 6, 56, 25, 58, 8, 43 71, and 72 (Figure 9B). The F_v/F_m, drought tolerance coefficient, can be more reliably evaluated from remote sensing data.

Discussion

This study revealed a high correlation between F_v/F_m, RWC and LWC; thus, F_v/F_m can be used as a direct indicator for evaluating the drought resistance of cotton. In addition, F_v/F_m and 1D-CNN models are good at predicting the inversion process of physiological and biochemical cotton indicators and hyperspectral data. The models also achieved the expected effects, and this method can quickly and nondestructively evaluate cotton drought tolerance.

Relationship between measurement parameters under drought stress

Cotton flowering and boll setting stages are extremely sensitive to soil water content and are important for adequate yield, which significantly declines under stress (Bange et al., 2004; Pettigrew, 2004). Therefore, this study evaluated the drought resistance of cotton varieties by investigating the effects of drought stress on cotton plants at the flowering, boll setting, and boll opening stages in the field. Leaf photosynthetic structure is an important index to evaluate plant stress resistance

and plays a key role in plant growth and metabolism, especially for photosystem PSII (El-Hendawy et al., 2019a). PSII maximum photochemical efficiency (F_v/F_m) has widely been used as an indicator for the early detection of different abiotic stresses (Naumann et al., 2008), which directly reflect crop damage under adverse environments. Under normal environmental conditions, F_v/F_m is relatively stable, but under adverse environmental conditions, photosynthetic efficiency is limited, and chloroplasts are protected from light damage, thereby significantly reducing F_v/F_m (Castañeda-Murillo et al., 2022). The findings in this study are consistent with those by Fracheboud (2002), where under drought stress, the F_v/F_m values of cotton varieties decreased during the growth period. Therefore, F_v/F_m values have gained interest as a screening tool to study preliminary and indicative responses to the rapid changes in plant photosynthetic status, and to evaluate the irreversible physiological damage caused by drought tolerance.

Optimizing input variables for the 1D-CNN model is important for hyperspectral inversion of cotton F_v/F_m prediction and drought tolerance evaluation

Numerous studies have mostly used vegetation index as an input to evaluate the degree of stress (Li et al., 2022). However, current vegetation index information is still limited, and the lack
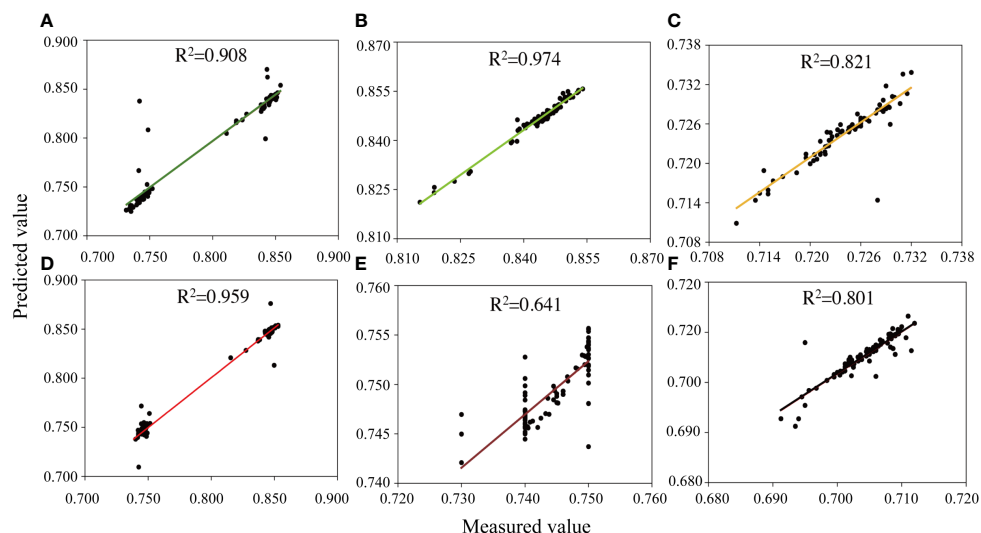
F_v/F_m predicted and measured values from the test data set under **(A−C)** sufficient water conditions and **(D−F)** water stress conditions. The canopy characteristics input by each model is from **(A, D)** 6 July 2021 (flowering stage), **(B, E)** 14 August 2021 (boll setting stage), and **(C, F)** 17 September 2021 (boll opening stage).

of a stable vegetation index closely related to drought stress may eventually reduce model generalizability. However, several specific spectral indices exist that have considerable potential in accurately estimating relevant parameters. SPA is a forward variable selection algorithm that minimizes vector space collinearity (Araújo et al., 2001). Its advantage lies in its extraction of several characteristic wavelengths from the whole band, which eliminates redundant information in the original spectral matrix when screening characteristic spectral wavelengths (Zhang et al., 2019). It is mainly divided into the following steps: firstly, data is imported under different processes; secondly, the Kennard stone algorithm is used to select samples; finally, SPA is used to select variables for multivariable calibration (Zhao et al., 2022). In this study, we used the MATLAB 2019b software to screen the characteristic spectral reflectance wavelengths of each process by coding a continuous projection algorithm, and only 1–2 sensitive wavelengths were screened under normal conditions at the flowering, boll setting, and boll opening stages. This challenged the establishment of a unified spectral index to estimate potential complex factors. Therefore, to improve relevant parameter prediction accuracy, some studies used the full spectrum wavelength (350–2500 nm) (Hansen et al., 2002; El-Hendawy et al., 2019b).

Interestingly, our study revealed that compared to the screening characteristic wavelengths using the continuous projection algorithm, the F_v/F_m predictions in the calibration and validation data sets had additional improvements based on the full band 1D-CNN model analysis. The maximum coefficient

of determination values (R^2) and minimum root mean square error values (RMSE) further revealed that the 1D-CNN model, based on data fusion in all conditions, was the most accurate in predicting F_v/F_m. Rasooli Sharabian et al., (2014) reported similar results. Elsayed et al. (2020) also revealed that compared to a single spectral index, a PLSR model based on spectral index data fusion and canopy temperature improved the GY prediction accuracy of barley and wheat under water stress. This study also revealed that the fusion of full band spectral data further improves the F_v/F_m prediction accuracy of cotton drought tolerance under different conditions. This is because this method can measure potential confounding factors related to environmental conditions. Therefore, it covers all the major physiological plant changes induced by drought stress.

This study supports machine learning and deep learning methods instead of the traditional cotton growth parameter estimation methods. Compared to CatBoost, LightBGM, XGBoost, DT, RF, GBDT, AdaBoost, ET, and KNN, the Fv/Fm remote sensing prediction accuracy inversion model constructed by 1D-CNN was higher and had strong stability. This shows that predicting physiological and biochemical indices and evaluating cotton drought tolerance using hyperspectral technology is feasible. In the field, different varieties have different leaf optical characteristics and canopy structures; thus, spectral interpretation is very complex. Despite these complexities, 1D-CNN achieved high accuracy in independent verification. 1D-CNN has previously been used for image segmentation, weed detection and prediction of other crops (such as rice and soybeans). However, the use of 1D-CNN
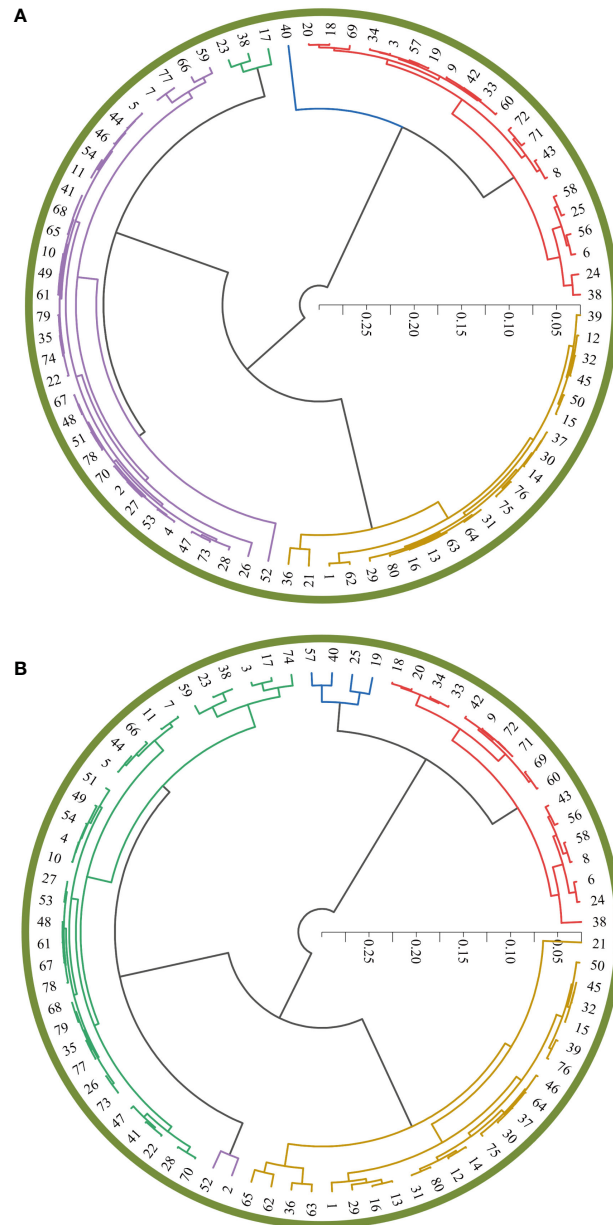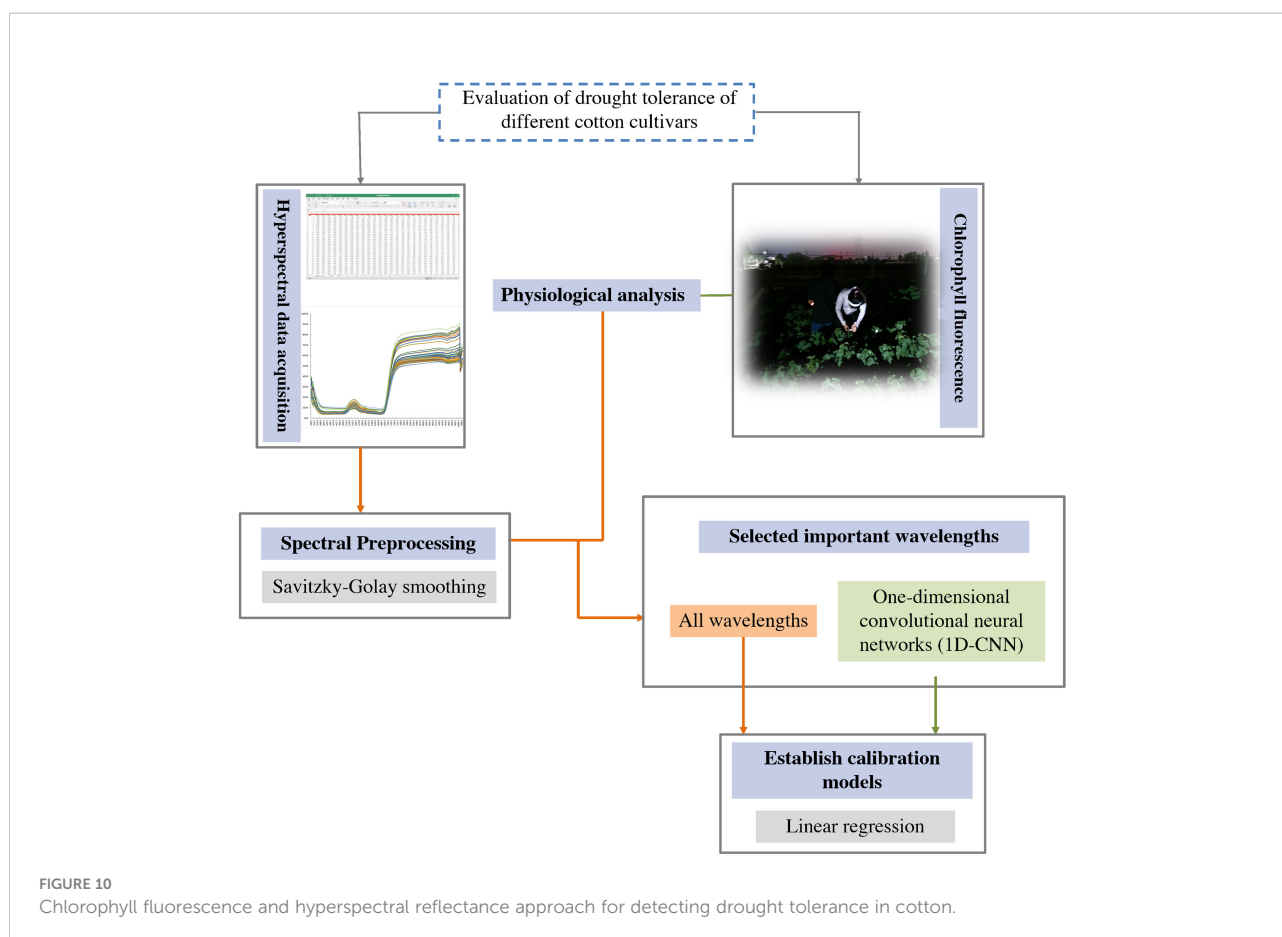
FIGURE 9

Cluster analysis of the drought tolerance coefficients of the F_v/F_m measured values **(A)** and predicted values **(B)** for the 80 cotton varieties. 1, Jifeng 554; 2, Jifeng 103; 3, Jifeng 522; 4, Jifeng 908; 5, Jifeng 914; 6, Jifeng 1982; 7, Jifeng 4; 8, 7886; 9, Cangmian 268; 10, Jimian 315; 11, Han 218; 12, Hannong 12; 13, Han 8266; 14, Han 258; 15, Han 686; 16, YM111; 17, Nongda KZ05; 18, Nongdamian 10; 19, Nongdamian 12; 20, Lumianyan 28; 21, Xuzhou 1818; 22, Zhongmiansuo 41; 23, Shandongxiamian11-42; 24, Zhongmiansuo 12; 25, Yumian 19; 26, Ejing 1; 27, Zhongmiansuo 35; 28, Zhongmiansuo 60; 29, Xinshi 71143; 30, Xinza 15; 31, Xinshi 17; 32, GK39; 33, 0 shi; 34, Zhongmiansuo 94A915; 35, Lumianyan 36; 36, DP33B; 37, Guoxinmian01; 38, Guoxinmian02; 39, Guoxinmian03; 40, Guoxinmian05; 41, Hanwu 216; 42, Zhongmian 100; 43, Zhongmiansuo 79; 44, Cangmian 666; 45, Han 6203; 46, Shikang 126; 47, Cang 198; 48, Ji 228; 49, Guoxinmian 9; 50, K836; 51, Lumian 522; 52, Lumian 5172; 53, K638; 54, Guoxin 4; 55, Jifeng1187; 56, Jifeng 1458; 57, Jifeng 103; 58, Jifeng 914; 59, Jifeng 965; 60, MH335223; 61, Guoxinmian 11; 62, Zhongmiansuo 17; 63, Chunbeibao; 64, Zhongmiansuo 60;65, CG3020-3; 66, Jimian 2016; 67, Ji 1518; 68, Jihang 8; 69, Jimian 262; 70, Ji 178; 71, Ji 172; 72, Yuzaomian 9110; 73, Dexiamian 1; 74, Jicai 6913; 75, Zhongmiansuo 23; 76, Zhongmiansuo 50; 77, Ji668; 78, Zhibao 86-1; 79, Jimian 958; 80, Jifeng 1271.

FIGURE 10
Chlorophyll fluorescence and hyperspectral reflectance approach for detecting drought tolerance in cotton.

for many cotton varieties is rarely reported. Based on our experimental process, the 1D-CNN estimation method used many characteristics and can use cotton spectral values directly as input, automatically learning and selecting features from the training data. Compared to traditional machine learning, 1D-CNN local connection, weight sharing, and hierarchical expression ensure that the network model effectively learns corresponding data features from many samples, avoids the complex feature extraction process and does not require manual feature extraction. Therefore, 1D-CNN improves prediction accuracy and reduces workload.

Possible problems with hyperspectral and 1D-CNN models

Unfortunately, 1D-CNN also has challenges, such as a high square error and high deviation (underfitting), which are mainly caused by inadequate sample number, inconsistent distribution of the training and verification sets, complex network structure (such as 1D-CNN), excessive sample noise interference, poor data quality, and overtraining. From the perspective of variance and deviation, underfitting equates to high training set variance and deviation,

which performs well in the training set. Still, it performs poorly in the test and new data sets. Generally, the main methods required to effectively solve overfitting are to increase the data set size, simplify and regularize the model, increase the drop layer, perform feature selection and sparse learning, delete abnormal noise points, use integrated learning methods, and re-clean the data.

From this study, spectral reflectance alone may not be sufficient to identify the most drought-tolerant cotton lines during screening. Therefore more phenotypic information sources are needed to fully clarify the complexity of drought tolerant genotype responses in cotton.

Influence of time scale differences on model performance

Different time scales and their effects on plant growth must be adopted in agricultural development as a management strategy. The reasons for spectral differences between different time scales are plant growth, phenological development, and environmental changes (Fava et al., 2009; Meerdink et al., 2016; Yang et al., 2016). These differences may be inverted in the relationship between spectra and traits, which is what we

detected in the performance of each independent model for the flowering, boll setting, and boll opening stages. Li et al. (2022) constructed six sorghum genotype models of dry and fresh weight using support vector machines on two separate dates and found that the combined model accuracy was higher than each independent model. Compared to the boll setting and opening stages, the flowering stage model was more robust and accurate (RMSE = 0.016, MAE = 0.009, MAPE = 0.011).

We observed that specific time scales affected accuracy. This study showed that the flowering stage accuracy was higher than the other stages. This may be due to vigorous growth of cotton crops during the early stage, rapid leaf area increases, large pigment accumulation in vegetation tissue, metabolic increase, high photosynthetic activity, strong F_v/F_m absorption, and a gradually enhanced regression equation fitting effect. With the postponement of the cotton growth period and the stress and aging of cotton plants in the later stages, leaves started losing their green coloration, turned yellow, and gradually withered. The F_v/F_m content subsequently decreased significantly until the leaves withered and died, unable to absorb light energy, and dry matter accumulation stopped (Silva Benavides et al., 2013), thus, leading to fitting effect deterioration. This is, therefore, the best period to estimate F_v/F_m.

Conclusion

Full band spectral data was studied here to predict F_v/F_m values and to evaluate cotton drought tolerance, (Figure 10) showed the workflow. The spectral distribution of the 80 cotton varieties at different growth stages and under different water stress conditions had similar trends. However, their near-infrared band reflectance decreased with drought stress and increased then decreased with growth. Compared to CatBoost, LightBGM, XGBoost, DT, RF, GBDT, AdaBoost, ET, and KNN, 1D-CNN models predicted cotton F_v/F_m during the three growth stages, implying that 1D-CNN models have higher accuracy and stability in the large-scale data processing. In evaluating cotton drought tolerance, the predicted F_v/F_m clustering results were similar to manually measured clustering results. Generally, the combined technology of S-G+1D-CNN has been successfully applied to predict cotton variety Fv/Fm values and evaluate drought tolerance. The full spectrum might therefore become an important tool for drought tolerance screening. In this study, it was not necessary to destructively sample all test field indicators, thus greatly reducing cost and time. This accelerated the related processing of phenotypic information for the different varieties and helped to develop a detection system for the high-throughput phenotypic identification algorithm. Therefore, more consideration should be given to spectral data and the computational power of deep learning models to reveal deeper phenotypic information. These models can be used to evaluate and screen out drought-resistant cotton varieties.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

LL, HD, and CL initiated and designed the research. CG, HS, and XL performed the experiments and collected the data. CG, NW, and AL wrote the code and tested the methods. CG, HS, KZ, YZ, JZ, and ZB analyzed the data and wrote the manuscript. All authors revised and approved the submitted version of the manuscript.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1007150/full#supplementary-material

References

Ahmad, M. W., Reynolds, J., and Rezgui, Y. (2018). Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *J. Clean. Prod.* 203, 810–821. doi: 10.1016/j.jclepro.2018.08.207

Ahmed, I. M., Cao, F., Zhang, M., Chen, X., Zhang, G., and Wu, F. (2013). Difference in yield and physiological features in response to drought and salinity combined stress during anthesis in tibetan wild and cultivated barleys. *PloS One* 8, e77869. doi: 10.1371/journal.pone.0077869

Araújo, M. C. U., Saldanha, T. C. B., Galvão, R. K. H., Yoneyama, T., Chame, H. C., and Visani, V. (2001). The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometr. Intell. Lab.* 57, 65–73. doi: 10.1016/S0169-7439(01)00119-8

Bange, M. P., Milroy, S. P., and Thongbai, P. (2004). Growth and yield of cotton in response to waterlogging. *Field Crops Res.* 88, 129–142. doi: 10.1016/j.fcr.2003.12.002

Bilger, W., and Björkman, O. (1990). Role of the xanthophyll cycle in photoprotection elucidated by measurements of light-induced absorbance changes, fluorescence and photosynthesis in leaves of hedera canariensis. *Photosynth. Res.* 25, 173–185. doi: 10.1007/BF00033159

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Burnett, A. C., Anderson, J., Davidson, K. J., Ely, K. S., Lamour, J., Li, Q., et al (2021). A best-practice guide to predicting plant traits from leaf-level hyperspectral data using partial least squares regression. *J Exp Bot.* 72, 6175–6189. doi: 10.1093/jxb/erab295

Carreiro Soares, S. F., Medeiros, E. P., Pasquini, C., de Lelis Morello, C., Harrop Galvão, R. K., and Ugulino Araújo, M. C. (2016). Classification of individual cotton seeds with respect to variety using near-infrared hyperspectral imaging. *Anal. Methods* 8, 8498–8505. doi: 10.1039/C6AY02896A

Castañeda-Murillo, C. C., Rojas-Ortiz, J. G., Sánchez-Reinoso, A. D., Chávez-Arias, C. C., and Restrepo-Díaz, H. (2022). Foliar brassinosteroid analogue (DI-31) sprays increase drought tolerance by improving plant growth and photosynthetic efficiency in lulo plants. *Heliyon* 8, e08977. doi: 10.1016/j.heliyon.2022.e08977

Chen, T., and Guestrin, C. (2016). XGBoost: "A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining.* (San Francisco California USA: Association for Computing Machinery), 785–794. doi: 10.1145/2939672.2939785

Choudhary, R., Mahesh, S., Paliwal, J., and Jayas, D. S. (2009). Identification of wheat classes using wavelet features from near infrared hyperspectral images of bulk samples. *Biosyst. Eng.* 102, 115–127. doi: 10.1016/j.biosystemseng.2008.09.028

Cover, T., and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* 13, 21–27. doi: 10.1109/TIT.1967.1053964

Cruz de Carvalho, M. H. (2008). Drought stress and reactive oxygen species: Production, scavenging and signaling. *Plant Signal. Behav.* 3, 156–165. doi: 10.4161/psb.3.3.5536

Cui, C., and Fearn, T. (2018). Modern practical convolutional neural networks for multivariate regression: Applications to NIR calibration. *Chemom. Intell. Lab. Syst.*, 9–20. doi: 10.1016/j.chemolab.2018.07.008

del Amor, R., Morales, S., Colomer, A., Mogensen, M., Jensen, M., Israelsen, N. M., et al. (2020). Automatic segmentation of epidermis and hair follicles in optical coherence tomography images of normal skin by convolutional neural networks. *Front. Med.* 7. doi: 10.3389/fmed.2020.00220

Ding, W., and Taylor, G. (2016). Automatic moth detection from trap images for pest management. *Comput. Electron. Agric.* 123, 17–28. doi: 10.1016/j.compag.2016.02.003

Durai, S. K. S., and Shamili, M. D. (2022). Smart farming using machine learning and deep learning techniques. *Decis. Analytics. J.* 3, 100041. doi: 10.1016/j.dajour.2022.100041

El-Hendawy, S., Al-Suhaibani, N., Dewir, Y., Elsayed, S., Alotaibi, M., Hassan, W., et al. (2019a). Ability of modified spectral reflectance indices for estimating growth and photosynthetic efficiency of wheat under saline field conditions. *Agronomy* 9, 35. doi: 10.3390/agronomy9010035

El-Hendawy, S., Al-Suhaibani, N., Elsayed, S., Alotaibi, M., Hassan, W., and Schmidhalter, U. (2019b). Performance of optimized hyperspectral reflectance indices and partial least squares regression for estimating the chlorophyll fluorescence and grain yield of wheat grown in simulated saline field conditions. *Plant Physiol. Biochem.* 144, 300–311. doi: 10.1016/j.plaphy.2019.10.006

Elsayed, S., Elhoweity, M., Ibrahim, H. H., Dewir, Y. H., Migdadi, H. M., and Schmidhalter, U. (2020). Thermal imaging and passive reflectance sensing to estimate the water status and grain yield of wheat under different irrigation regimes. *Agr. Water Manage.* 228, 105873. doi: 10.1016/j.agwat.2019.105873

Fava, F., Colombo, R., Bocchi, S., Meroni, M., Sitzia, M., Fois, N., et al. (2009). Identification of hyperspectral vegetation indices for mediterranean pasture characterization. *Int. J. Appl. Earth Obs.* 11, 233–243. doi: 10.1016/j.jag.2009.02.003

Feng, L., Zhu, S., Liu, F., He, Y., Bao, Y., and Zhang, C. (2019). Hyperspectral imaging for seed quality and safety inspection: A review. *Plant Methods* 15, 91. doi: 10.1186/s13007-019-0476-y

Fracheboud, Y. (2002). Identification of quantitative trait loci for cold-tolerance of photosynthesis in maize (*Zea mays* l.). *J. Exp. Bot.* 53, 1967–1977. doi: 10.1093/jxb/erf040

Freund, Y., and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* 55, 119–139. doi: 10.1006/jcss.1997.1504

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.* 29 (5), 1189–1232. doi: 10.1214/aos/1013203451

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* 36, 193–202. doi: 10.1007/BF00344251

Gao, M., Snider, J. L., Bai, H., Hu, W., Wang, R., Meng, Y., et al. (2020). Drought effects on cotton (*Gossypium hirsutum* l.) fibre quality and fibre sucrose metabolism during the flowering and boll-formation period. *J. Agro. Crop Sci.* 206, 309–321. doi: 10.1111/jac.12389

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Mach. Learn.* 63, 3–42. doi: 10.1007/s10994-006-6226-1

Ghosal, S., Blystone, D., Singh, A. K., Ganapathysubramanian, B., Singh, A., and Sarkar, S. (2018). An explainable deep machine vision framework for plant stress phenotyping. *Proc. Natl. Acad. Sci. U.S.A.* 115, 4613–4618. doi: 10.1073/pnas.1716999115

Guo, G., Wang, H., Bell, D., Bi, Y., and Greer, K. (2003). "KNN Model-Based Approach in Classification," In *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE Lecture Notes in Computer Science.* Eds R. Meersman, Z. Tari and D. C. Schmidt (Berlin, Heidelberg: Springer Berlin Heidelberg), 986–96. doi: 10.1007/978-3-540-39964-3_62

Hancock, J. T. (2020). CatBoost for big data: an interdisciplinary review. *J. Big. Data* 7, 94. doi: 10.1186/s40537-020-00369-8

Hansen, P. M., Jørgensen, J. R., and Thomsen, A. (2002). Predicting grain yield and protein content in winter wheat and spring barley using repeated canopy reflectance measurements and partial least squares regression. *J. Agr. Sci.* 139, 307–318. doi: 10.1017/S0021859602002320

Hikosaka, K., and Tsujimoto, K. (2021). Linking remote sensing parameters to CO_2 assimilation rates at a leaf scale. *J. Plant Res.* 134, 695–711. doi: 10.1007/s10265-021-01313-4

Ibrahim, A., Alghannam, A., Eissa, A., Firtha, F., Kaszab, T., Kovacs, Z., et al. (2021). Preliminary study for inspecting moisture content, dry matter content, and firmness parameters of two date cultivars using an NIR hyperspectral imaging system. *Front. Bioeng. Biotech.* 9. doi: 10.3389/fbioe.2021.720630

Khan, M. M., Hossain, S., Mozumdar, P., Akter, S., and Ashique, R. H. (2022). A review on machine learning and deep learning for various antenna design applications. *Heliyon* 8, e09317. doi: 10.1016/j.heliyon.2022.e09317

Khan, N., Sachindra, D. A., Shahid, S., Ahmed, K., Shiru, M. S., and Nawaz, N. (2020). Prediction of droughts over pakistan using machine learning algorithms. *Adv. Water Resour.* 139, 103562. doi: 10.1016/j.advwatres.2020.103562

Kriegler, B., and Berk, R. (2010). Small area estimation of the homeless in los angeles: an application of cost-sensitive stochastic gradient boosting. *Ann. Appl. Stat.* 4 (3), 1234–1255. doi: 10.1214/10-AOAS328

Lamaoui, M., Jemo, M., Datla, R., and Bekkaoui, F. (2018). Heat and drought stresses in crops and approaches for their mitigation. *Front. Chem.* 6. doi: 10.3389/fchem.2018.00026

Lang, Y., Wang, M., Xia, J., and Zhao, Q. (2018). Effects of soil drought stress on photosynthetic gas exchange traits and chlorophyll fluorescence in forsythia suspensa. *J. For. Res.* 29, 45–53. doi: 10.1007/s11676-017-0420-9

Liang, J., Jing, T., Niu, H., and Wang, J. (2020). Two-terminal fault location method of distribution network based on adaptive convolution neural network. *IEEE Access* 8, 54035–54043. doi: 10.1109/ACCESS.2020.2980573

Li, J., Schachtman, D. P., Creech, C. F., Wang, L., Ge, Y., and Shi, Y. (2022). Evaluation of UAV-derived multimodal remote sensing data for biomass prediction and drought tolerance assessment in bioenergy sorghum. *Crop J.* S2214514122000897. doi: 10.1016/j.cj.2022.04.005

Lu, Y., Yi, S., Zeng, N., Liu, Y., and Zhang, Y. (2017). Identification of rice diseases using deep convolutional neural networks. *Neurocomputing* 267, 378–384. doi: 10.1016/j.neucom.2017.06.023

Mahesh, S., Manickavasagan, A., Jayas, D. S., Paliwal, J., and White, N. D. G. (2008). Feasibility of near-infrared hyperspectral imaging to differentiate canadian wheat classes. *Biosyst. Eng.* 101, 50–57. doi: 10.1016/j.biosystemseng.2008.05.017

Meerdink, S. K., Roberts, D. A., King, J. Y., Roth, K. L., Dennison, P. E., Amaral, C. H., et al. (2016). Linking seasonal foliar traits to VSWIR-TIR spectroscopy across california ecosystems. *Remote Sens. Environ.* 186, 322–338. doi: 10.1016/j.rse.2016.08.003

Meng, Q., Ke, G., Wang, T., Chen, W., Ye, Q., Ma, Z.-M., et al. (2016). A communication-efficient parallel algorithm for decision tree. *arXiv* 9. doi: 10.48550/arXiv.1611.01276

Miao, A., Zhuang, J., Tang, Y., He, Y., Chu, X., and Luo, S. (2018). Hyperspectral image-based variety classification of waxy maize seeds by the t-sne model and procrustes analysis. *Sensors* 18, 4391. doi: 10.3390/s18124391

Montesinos-López, O. A., Montesinos-López, A., Crossa, J., de los Campos, G., Alvarado, G., Suchismita, M., et al. (2017). Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. *Plant Methods* 13, 4. doi: 10.1186/s13007-016-0154-2

Mwadzingeni, L., Shimelis, H., Tesfay, S., and Tsilo, T. J. (2016). Screening of bread wheat genotypes for drought tolerance using phenotypic and proline analyses. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01276

Najafi, F., Khavari-Nejad, R. A., Rastgar-jazii, F., and Sticklen, M. (2007). Growth and some physiological attributes of pea (*Pisum sativum* l.) as affected by salinity. *Pak. J. @ Biol. Sci.* 10, 2752–2755. doi: 10.3923/pjbs.2007.2752.2755

Naumann, J., Anderson, J., and Young, D. (2008). Linking physiological responses, chlorophyll fluorescence and hyperspectral imagery to detect salinity stress using the physiological reflectance index in the coastal shrub, myrica cerifera. *Remote Sens. Environ.* 112, 3865–3875. doi: 10.1016/j.rse.2008.06.004

Pettigrew, W. T. (2004). Moisture deficit effects on cotton lint yield, yield components, and boll distribution. *Agron. J.* 96, 377. doi: 10.2134/agronj2004.0377

Rasooli Sharabian, V., Noguchi, N., and Ishi, K. (2014). Significant wavelengths for prediction of winter wheat growth status and grain yield using multivariate analysis. *Eng. Agric. Environ. Food* 7, 14–21. doi: 10.1016/j.eaef.2013.12.003

Sarker, I. H., Colman, A., Han, J., Khan, A. I., Abushark, Y. B., and Salah, K. (2020). BehavDT: a behavioral decision tree learning to build user-ccentric context-aware predictive model. *Mobile. Netw. Appl.* 25, 1151–1161. doi: 10.1007/s11036-019-01443-z

Shah, S. H., Angel, Y., Houborg, R., Ali, S., and McCabe, M. F. (2019). A random forest machine learning approach for the retrieval of leaf chlorophyll content in wheat. *Remote Sens.* 11, 920. doi: 10.3390/rs11080920

Shakoor, N., Lee, S., and Mockler, T. C. (2017). High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field. *Curr. Opin. Plant Biol.* 38, 184–192. doi: 10.1016/j.pbi.2017.05.006

Silva Benavides, A. M., Torzillo, G., Kopecký, J., and Masojídek, J. (2013). Productivity and biochemical composition of phaeodactylum tricornutum (*Bacillariophyceae*) cultures grown outdoors in tubular photobioreactors and open ponds. *Biomass Bioenergy* 54, 115–122. doi: 10.1016/j.biombioe.2013.03.016

Wang, X., Li, J., and Huang, T. (2022). Cnvabnn: an adaBoost algorithm and neural networks-based detection of copy number variations from NGS data. *Comput. Biol. Chem.* 99, 107720. doi: 10.1016/j.compbiolchem.2022.107720

Wang, L., Liu, D., Pu, H., Sun, D.-W., Gao, W., and Xiong, Z. (2015). Use of hyperspectral imaging to discriminate the variety and quality of rice. *Food Anal. Methods* 8, 515–523. doi: 10.1007/s12161-014-9916-5

Wang, Y., Reiter, R. J., and Chan, Z. (2018). Phytomelatonin: A universal abiotic stress regulator. *J. Exp. Bot.* 69, 963–974. doi: 10.1093/jxb/erx473

Xiao, S., Liu, L., Zhang, Y., Sun, H., Zhang, K., Bai, Z., et al. (2020). Fine root and root hair morphology of cotton under drought stress revealed with RhizoPot. *J. Agro. Crop Sci.* 206, 679–693. doi: 10.1111/jac.12429

Xiong, X., Duan, L., Liu, L., Tu, H., Yang, P., Wu, D., et al. (2017). Panicle-seg: a robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization. *Plant Methods* 13, 104. doi: 10.1186/s13007-017-0254-7

Yang, X., Tang, J., Mustard, J. F., Wu, J., Zhao, K., Serbin, S., et al. (2016). Seasonal variability of multiple leaf traits captured by leaf spectroscopy at two temperate deciduous forests. *Remote Sens. Environ.* 179, 1–12. doi: 10.1016/j.rse.2016.03.026

Yao, X., Yao, X., Tian, Y., Ni, J., Liu, X., Cao, W., et al. (2013). A new method to determine central wavelength and optimal bandwidth for predicting plant nitrogen uptake in winter wheat. *J. Integr. Agr.* 12, 788–802. doi: 10.1016/S2095-3119(13)60300-7

Yu, Z., Fang, H., Zhangjin, Q., Mi, C., Feng, X., and He, Y. (2021). Hyperspectral imaging technology combined with deep learning for hybrid okra seed identification. *Biosyst. Eng.* 212, 46–61. doi: 10.1016/j.biosystemseng.2021.09.010

Zhang, X., Liu, F., He, Y., and Li, X. (2012). Application of hyperspectral imaging and chemometric calibrations for variety discrimination of maize seeds. *Sensors* 12, 17234–17246. doi: 10.3390/s121217234

Zhang, Y., Wang, Y., Yi, Y., Wang, J., Liu, J., and Chen, Z. (2021b). Coupling matrix extraction of microwave filters by using one-dimensional convolutional autoencoders. *Front. Phys.* 9. doi: 10.3389/fphy.2021.716881

Zhang, W., Xu, H., Duan, X., Hu, J., Li, J., Zhao, L., et al. (2021a). Characterizing the leaf transcriptome of chrysanthemum rhombifolium (*Ling et c. shih*), a drought resistant, endemic plant from china. *Front. Genet.* 12, 625985. doi: 10.3389/fgene.2021.625985

Zhang, D., Xu, Y., Huang, W., Tian, X., Xia, Y., Xu, L., et al. (2019). Non-destructive measurement of soluble solids content in apple using near infrared hyperspectral imaging coupled with wavelength selection algorithm. *Infrared. Phys. Techn.* 98, 297–304. doi: 10.1016/j.infrared.2019.03.026

Zhang, Y., Zhang, Y., Liu, G., Xu, S., Dai, J., Li, W., et al. (2021c). Nitric oxide increases the biomass and lint yield of field-grown cotton under temporary waterlogging through physiological and molecular regulation. *Field Crops Res.* 261, 107989. doi: 10.1016/j.fcr.2020.107989

Zhao, D., Feng, S., Cao, Y., Yu, F., Guan, Q., Li, J., et al. (2022). Study on the classification method of rice leaf blast levels based on fusion features and adaptive-weight immune particle swarm optimization extreme learning machine algorithm. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.879668

Zhao, Y., Zhang, C., Zhu, S., Gao, P., Feng, L., and He, Y. (2018). Non-destructive and rapid variety discrimination and visualization of single grape seed using near-infrared hyperspectral imaging technique and multivariate analysis. *Molecules* 23, 1352. doi: 10.3390/molecules23061352

Zou, J., Hu, W., Li, Y., He, J., Zhu, H., and Zhou, Z. (2020). Screening of drought resistance indices and evaluation of drought resistance in cotton (*Gossypium hirsutum* l.). *J. Integr. Agr.* 19, 495–508. doi: 10.1016/S2095-3119(19)62696-1