



BACNN: Multi-scale feature fusion-based bilinear attention convolutional neural network for wood NIR classification

Zihao Wan¹ · Hong Yang¹ · Jipan Xu¹ · Hongbo Mu¹ · Dawei Qi¹

Received: 26 November 2022 / Accepted: 23 May 2023
© The Author(s) 2023

Abstract Effective development and utilization of wood resources is critical. Wood modification research has become an integral dimension of wood science research, however, the similarities between modified wood and original wood render it challenging for accurate identification and classification using conventional image classification techniques. So, the development of efficient and accurate wood classification techniques is inevitable. This paper presents a one-dimensional, convolutional neural network (i.e., BACNN) that combines near-infrared spectroscopy and deep learning techniques to classify poplar, tung, and balsa woods, and PVA, nano-silica-sol and PVA-nano silica sol modified woods of poplar. The results show that BACNN achieves an accuracy of 99.3% on the test set, higher than the 52.9% of the BP neural network and 98.7% of Support Vector Machine compared with traditional machine learning methods and deep learning based methods; it is also higher than the 97.6% of LeNet, 98.7% of AlexNet and 99.1% of VGG-Net-11. Therefore, the classification method proposed offers

potential applications in wood classification, especially with homogeneous modified wood, and it also provides a basis for subsequent wood properties studies.

Keywords Wood classification · Near infrared spectroscopy · Bilinear network · SE module · Anti-noise algorithm

Introduction

As a biopolymer material, wood has the advantage of being renewable and is widely used in construction, furniture, and aerospace applications (Huang et al. 2020). With the decline in forest resources and shortage of quality wood resources, attention is turning to fast planted forestry wood and wood modification. The functional modification of wood by physical and chemical methods, and therefore the creation of new types of wood characterized by high added value and versatility is important for economic and environmental development (Macior et al. 2022). However, there are considerable variations in the physical and chemical properties of wood among different species, and it is particularly challenging to distinguish between modified wood and traditional wood, and a more effective classification technique is necessary. NIR spectroscopy has been widely implemented in forestry research to can provide information on the internal functional groups of wood, and thus confirm the class of wood and its physical and chemical properties without relying extensively on surface characteristics such as colour and grain. For example, Wang et al. (2015a) used a cluster analysis model, a Bayesian discriminant model and a support vector machine model to classify the NIR spectra of ten woods with an accuracy of 83.3%, 86.7% and 85.0%, respectively. In a following study, Wang et al. (2015b) classified

Corresponding editor: Tao Xu.

Project funding This study was supported by the Fundamental Research Funds for the Central Universities (No. 2572023DJ02). The online version is available at <http://www.springerlink.com>

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11676-023-01652-z>.

✉ Hongbo Mu
mh506@nefu.edu.cn

✉ Dawei Qi
qidw9806@nefu.edu.cn

¹ College of Science, Northeast Forestry University, Harbin 150040, People's Republic of China

the NIR spectra of a total of 296 wood samples of five tree species using a BP neural network model. The classification accuracy reached 100% for species of different genera and more than 85.0% for species of the same genus. However, the traditional method for classification is dependent on pre-processing of the spectra and requires manual extraction of the required features before classification, which is subject to significant interference from human error (Nisgoski et al. 2017).

Convolutional neural networks, as a data-driven modeling approach, have been widely used in NIR spectral feature extraction due to their powerful extraction capabilities. Especially after the emergence of AlexNet (Krizhevsky et al. 2012), various high-precision deep convolutional neural network models have emerged such as LeNet (Lecun et al. 1998), VGGNet (Simonyan and Zisserman 2014), GoogleNet (Szegedy et al. 2015), ResNet (He et al. 2016), and DenseNet (Huang et al. 2017). Convolutional neural networks combined with NIR spectroscopy reflect the internal characteristics of the sample and avoid the drawbacks of redundant information in NIR spectroscopy. They are widely used and their specific applications are in two forms: First, feature extraction is performed directly from high-dimensional raw spectra (Lecun et al. 2015), using the powerful extraction capabilities of convolutional neural network models in one-dimensional spectra to extract their intrinsic features, after which the features are exploited using the traditional exploitation approach in chemometrics. Secondly, a high-level original spectrum is reconstructed by downscaling using preprocessing methods such as PCA, and the reconstructed spectrum is feature extracted using convolutional neural networks which can transform low-dimensional data into high-dimensional abstract features by multi-level non-linear modules. Through layer-by-layer feature extraction, the model can eventually learn complex feature representations. Consequently, convolutional neural networks have an increasingly important role in spectral analysis (Chen and Wang 2018). For example, Jia et al. (2020) built an 8-layer convolutional neural network model to analyze and predict the near-infrared spectra of water quality with a prediction accuracy of over 99.0%. Tang and Chen (2021) used convolutional neural networks to analyse the near-infrared spectra of soils for Ph prediction with an accuracy of 90.0%. Xia et al. (2021) classified plastics using one-dimensional convolutional neural networks with 100% accuracy. Yang et al. (2020) used convolutional neural networks to classify softwood with over 99.0% accuracy. The original NIR spectra of softwood using NIR spectroscopy was classified by Pan et al. (2022) to successfully distinguish 21 wood samples.

This study combined NIR spectra with convolutional neural networks and presents a bilinear attentional convolutional neural network model (BACNN) based on multi-scale feature fusion to classify the NIR spectra of poplar wood

(PW), tung wood (TW), balsa wood (BW), PVA modified poplar wood (PVAW), nano-silica-sol modified poplar wood (SW) and PVA-nano silica sol modified poplar wood (PSW). BACNN mitigates the effect of noise by adding convolutional kernels as 1×7 convolutional blocks (Zhang et al. 2017), after which the features of the spectrum are extracted from different scales using two branches, while adding SE modules (Hu et al. 2018b) to the two branches to obtain better quality features and finally using fully connected layers for classification. The main innovations of this paper are:

- (1) Modification experiments were conducted on poplar wood to obtain the experimental samples, the near-infrared spectra of the samples were collected, and were data enhanced to obtain the data set. The data set contains NIR spectra of different species and NIR spectra of modified woods of the same species obtained by different methods.
- (2) A novel neural network model BACNN is proposed for the data set of this study. BACNN adds a 1×7 block to suppress noise in the spectrum, uses a multi-scale fusion mechanism to extract features from two scales, and adds an SE model in each branch to obtain accurate features. Finally, the superiority of the BACNN model was demonstrated by comparison tests and ablation experiments.

Materials and methods

Specimen preparation

In this experiment, PW, TW, BW, PVAW, SW and PSW were experimental samples. The materials used were free of cracks, knots, discoloration, or other defects to minimize the disturbance of environmental variables. The logs were placed at an ambient temperature of 20 °C and a relative humidity of 65% to reach a state of moisture absorption equilibrium before making the modified woods. Anhydrous ethanol (> 99%), deionized water, sodium sulphite (> 96%), sodium hydroxide (> 96%), polyvinyl alcohol (PVA) and silica nanosol solutions used in the experiments were purchased from the University laboratory. All raw materials were used without further treatment.

The modified poplar wood was first prepared by configuring a 3% aqueous PVA solution and immersing the original wood in the solution for 24 h at room temperature and pressure. After impregnation, the wood samples were removed and washed with deionized water to finally obtain PVAW. Similarly, wood samples were placed into the nano-silica-sol solution and treated to obtain SW. The PVA: nano-silica-sol was mixed at 3:1 (mass ratio), sonicated for 30 min and stirred for eight hours at room temperature with a magnetic

stirrer to obtain the impregnated modified solution. The original poplar wood was impregnated and treated to obtain PSW. The experimental samples are shown in Fig. 1.

NIR spectra measurements

A Nicolet iS10 Fourier transform infrared spectrometer (FT-IR) was used to collect wood spectra with a resolution $> 0.4 \text{ cm}^{-1}$ controlled by the OMSNIC software, allowing the surface of the sample to be scanned at $400\text{--}4000 \text{ cm}^{-1}$. The number of scans per point was set at 32 and the absorption spectrum gathered. The background spectrum was first measured, saved, the wood sample then placed on the spectrometer detector and the absorption spectrum measured. To increase the richness of the spectral samples and have the spectrum fully reflect their characteristics, different parts were collected of each wood, with five pieces of wood were selected for each type of sample. Thirty points were collected for each piece of wood for a total of 906 near-infrared spectral data.

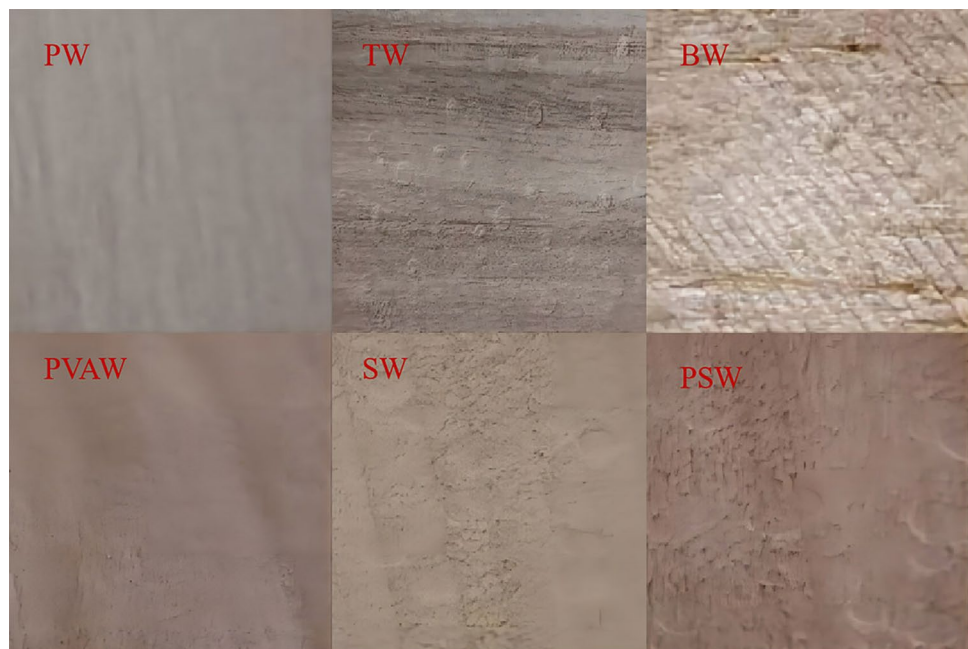
Enhancement and pre-processing of NIR spectra

Deep learning networks require features to be extracted from large volumes of raw data and good performance can only be achieved with a relatively large amount (Gao et al. 2021a). Therefore, augmentation techniques are necessary to enrich the data to improve classification accuracy and prevent overfitting. In this study, tiny random Gaussian noise was added to data augmentation, setting the mean to 0 and variance to 0.02, doubling the data to 1812, and spectrally adding

random Gaussian noise. The enhanced spectrum shows essentially the same spectral trend compared to the initial spectrum, with several important peaks and troughs remaining unchanged. The addition of noise does not change the chemistry of the spectra, but only the fluctuations become larger, so the difficulty of identifying peaks and valleys increases subsequently.

The raw spectrum contained unimportant information such as noise, and also had overlapping peaks (Kauppinen 1983), it was necessary to attenuate the noise and separate the overlapping peaks of the spectrum while downscaling it for reconstruction to reduce the amount of data. This study used S-G convolutional smoothing (Soares et al. 2016) combined with principal component analysis (PCA) (Wang et al. 2017) to preprocess the original spectrum. S-G convolutional smoothing uses polynomials for data smoothing and is based on the least squares method, which can retain useful information in the analyzed signal and eliminate random noise. The PCA method is capable of downscaling the spectra and also of enhancing overlapping peaks, which can be used to solve the problem of spectral overlapping peaks (Kuesel et al. 1996). In the experiments, the best results were obtained when the window length was set to eleven and fitted with second order (Fig. 2). Eight hundred principal components were then extracted from 6950 features using PCA to improve the spectral signal-to-noise ratio, enhance the overlapping peaks and solve the spectral overlap problem (Pachuta 2004). The spectra were reconstructed by dimensionality reduction to further reduce the data volume (Qin et al. 2013). Finally, the data were randomly divided into training and test sets at a ratio of 7:3 (Table 1).

Fig. 1 Experimental samples



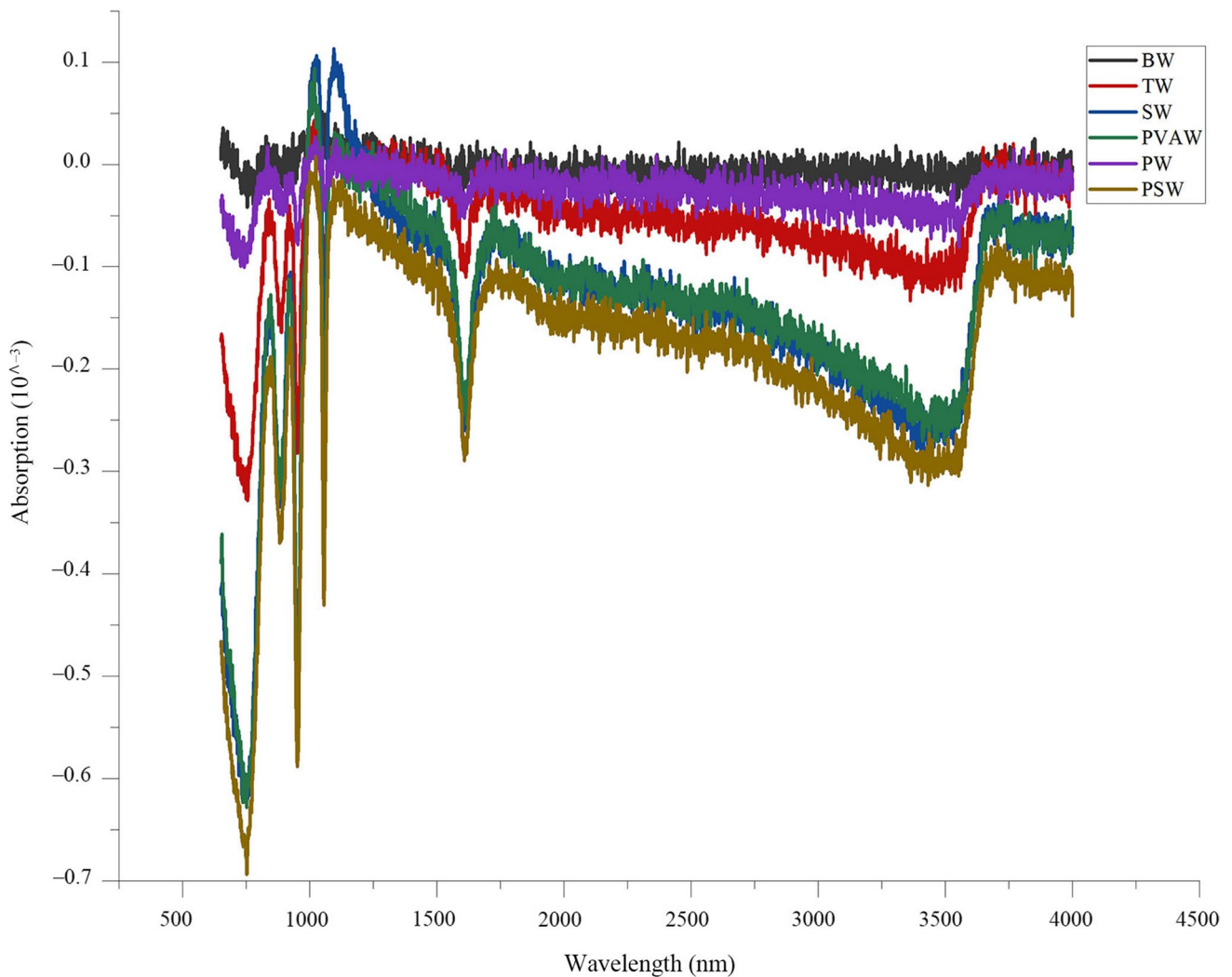


Fig. 2 Spectrograms

Table 1 Number of training and test sets

Class	Training set	Test set	Total
PW	219	83	302
TW	198	104	302
BW	222	80	302
PVAW	207	95	302
SW	221	81	302
PSW	201	101	302

BACNN

Convolutional neural network

The purpose of convolution is to perform feature extraction. It is the filter which is trained to suppress distracting information such as noise and extract the main features for

classification. When a convolutional neural network is used for feature extraction, its formula is Eq. 1.

$$x_i^k = f \left(\sum_{c=0}^{len} W_i^l(c) * x^{k-1}(c) + b_b^l \right) \quad (1)$$

where, x_i^k is the output of channel i of the k^{th} layer of the convolutional kernel; f is the activation function; len is the length of the convolutional kernel; x^{k-1} is the output of the previous convolutional layer; b_b^l is the bias; W_i^l is the weight matrix.

During the training process, the network model continuously adjusted the weight matrix and bias until the loss function was reduced to the ideal value, at which point the input data was given more weight by the convolution layer for the main features and less weight for the noise to achieve good classification. When applying convolutional neural networks to spectral analysis, the features shared by the network parameters help reduce the number of parameters

and prevent overfitting (Jiao et al. 2019). The Max-pooling operation compresses the input feature map and extracts the main features of the spectrum (Graham et al. 2014). The local perception ability of convolution facilitates the extraction of wave peak and trough features in the spectrum by convolutional neural networks, suitable for one-dimensional spectral analysis.

Acquarelli et al. (2017) noted that the convolutional neural network uses the learned convolutional kernel for smoothing and derivative filtering of the input information to solve the problem of noise in the spectrum as well as overlapping peaks. So the dependence of convolutional neural networks on spectral pre-processing operations is greatly reduced. In this study, it was found that the classification accuracy of the model decreased when the first-order derivative spectrum and the second-order derivative spectrum were solved after training, which led to the conclusion that the spectral pre-processing method of S-G convolutional smoothing, combined with PCA, enables the convolutional neural network model to adequately eliminate noise and cope with interference from problems such as overlapping peaks, without the need for additional operations to further pre-process the spectrum.

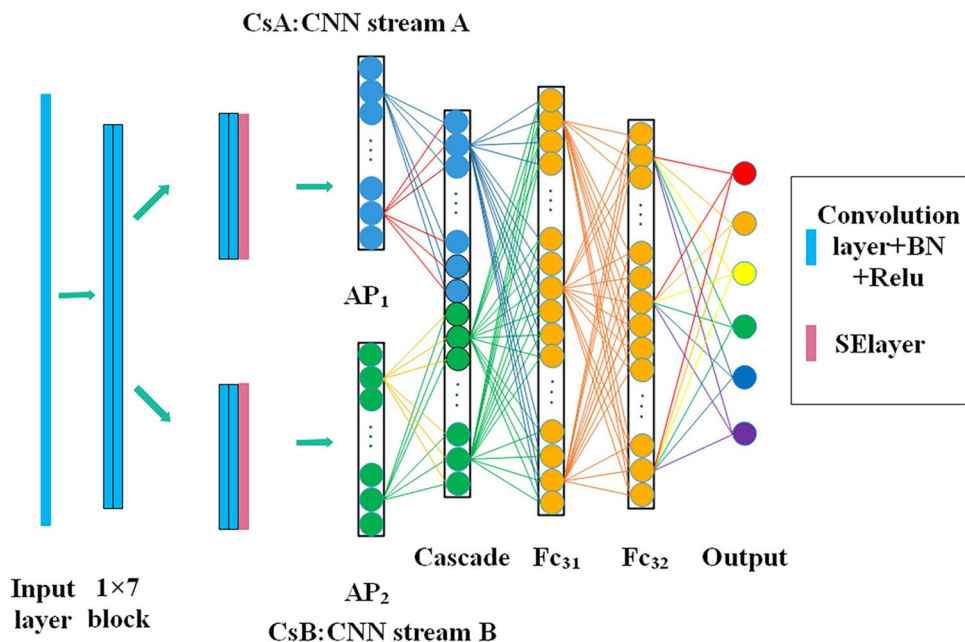
Network structure

Common methods to improve the classification accuracy of neural networks include increasing the depth of the network model and adding an attention mechanism. Due to the large amount of input information in this paper, the required network model is deeper and therefore required

a large amount of computational resources, so a bilinear branching network with multi-scale feature fusion was used instead of increasing the depth of the network model. An attention mechanism was added to obtain more detailed features for classification. The BACNN used in this study is an 11-layer, bilinear convolutional neural network model where two branches with different convolutional kernel sizes are used to extract multi-scale spectral features. The SE module was added to reduce the interference of other information. When the convolutional kernel is set to 3, i.e., relatively small, the accuracy of the network is higher and can accurately identify weak and overlapping peaks when the running speed is slower. When the setting is relatively large, it can filter noise and increase the training speed, but for weak peaks and overlapping summits, it results in misjudgment and omission. Therefore, in this study there were two different scales of the branch to obtain different features and then fusion, which not only reduced the pre-processing of the sample, it further filters noise, obtains more accurate features, and reduces the training time.

The structure of BACNN is shown in Fig. 3. The input spectral signal first passes through two convolutional layers with a 1×7 convolutional kernel to attenuate the interference of noise. The spectral features were then extracted from different scales by two neural network branches, CsA and CsB (Wang et al. 2019). In addition, higher quality features were obtained by adding the SE module, the two features then fused using a bilinear pooling operation, and finally, the one-dimensional vector from the fusion was fed to the fully connected layer for classification. Table S1 shows the detailed parameters of BACNN.

Fig. 3 Network Structure



Multi-scale feature fusion

BACNN extracts features from different angles of the spectrum using two convolutional neural network branches with different convolutional kernels. A 1×3 kernel is assigned to *CsA* to extract coarse features while reducing the parameters of the network, and a 1×5 kernel is attached to *CsB*. Given that a large convolutional kernel increases the perceptual field, *CsB* is able to extract accurate features. The BACNN is represented by Eq. 2:

$$B = (F, CsA, CsB, Fc_{31}, Fc_{32}) \quad (2)$$

where *F* stands for 1×7 convolutional block, *CsA* and *CsB* represent two linear branches, Fc_{31} and Fc_{32} refer to fully connected layers.

The *CsA* branch and the *CsB* branch can be expressed by Eqs. 3 and 4, respectively.

$$CsA = [(C, B, R) \times 2, SE, AP_1] \quad (3)$$

$$CsB = [(C, B, R) \times 2, SE, AP_2] \quad (4)$$

where, *C*, *B*, *R*, AP_1 and AP_2 represent the convolution layer (Pradhan et al. 2021), normalization (Gao et al. 2021b), Relu activation function (Laakmann and Peterson 2021) and adaptive maximum pooling (Hu et al. 2018a) layer of *CsA* and *CsB*.

As fusion requires the same dimensionality of the feature vectors output from both branches, this study adjusts the feature vectors of the SE module to 1×512 using an adaptive maximum pooling layer. The fully connected layer is replaced with a pooling layer to reduce the amount of data while ensuring that both have the same dimensionality. The two are then cascaded along the vertical axis to obtain a 1024×1 feature vector, as shown by Eq. 5:

$$f = \text{Cas}(x_1, x_2) \quad (5)$$

where, x_1 and x_2 represent the output of the two adaptive maximum pooling layers, and *f* the fused feature vector which contains all the features of the two scales. It represents the features more comprehensively, and then connects the fully connected layer with the softmax (Asadi and Littman 2017) layer for classification.

SE module

The SE module, as a channel attention network, has the core focus of modelling the interdependence between channels, assigning different weights to the feature vectors of different channels, and then summing them proportionally to obtain more accurate features. Fig. S1 shows the structure of the SE

module. The SE module contains three parts: squeeze, excitation, and scale. The squeeze module, compresses the feature map into $1 \times 1 \times C$ vector by performing global average pooling on the input vector. Next, the excitation operation, consists of a fully connected layer with a stack of activation functions. Finally, the essence of the scale operation is the multiplication of channel weights. The SE module calculates the weight value of each channel and multiplies it with each channel, thus assigning different proportions to different channels to get better results.

1×7 block

The size of convolutional kernel is important for convolutional neural network models. Networks with small convolutional kernels have the advantages of small computation and fast convergence, but the receptive field is small and easily disturbed by noise. Although convolutional neural networks with large kernels are more computationally intensive, they have a larger field of perception and a suppression effect on noise. Since the input data used in this study were near-infrared spectra of wood, interference from environmental noise was inevitable in the process of obtaining the spectra and at the same time, this study used the addition of random Gaussian noise for data enhancement. There are many noises in the data so this study used two convolutional layers with a convolutional kernel of 1×7 to form a block to suppress the effect of data noise.

Pipeline

The pipeline of wood classification is shown in Fig. S2. First, the original wood sample was prepared and impregnated to obtain the modified sample. Second, the spectrum of the sample was collected using a near-infrared spectrometer, and the spectral data enhanced and pre-processed to obtain the final *I* data set. Finally, the spectrum data was fed into the BACNN network and the classification results outputted by the softmax layer.

Results and discussion

Experimental environment and hyper-parameter

The server used in this experiment was Windows 10, the processor Intel(R) Xeon(R) Bronze 3204, the memory 128 G, and the GPU NVIDIA GeForce RTX 3090 (Table 2). When training the model, the ratio of training to test set was 7:3. The test set contained 544 spectral data and the training set 1268. The hyperparameters include the iterative Epoch, Batch size and Learning rate (L_r), which are set to 200, 800 and $1e-3$, respectively.

Table 2 Experimental environment

Hardware		Software	
Memory	128.00 GB	System	Windows 10
CPU	Inter(R) Xeon(R) Bronze 3204 CPU @ 1.90 GHz/6 core	Python environment	Pytorch-gpu 1.8.0
GPU	NVIDIA GeForce GTX 3090(24G)		+Python 3.8.8+cuda 11.1+ cudnn 8.0.5

Experimental results

The loss function variation curves of the training and test sets during the training are shown in Fig. 4. From Fig. 4, in the training set, the loss function reached a minimum of 0.01 when it was iterated to 100 times. This indicates that the model can converge quickly and can fully learn the features. In the test set, the accuracy reached 99.6% at 61 iterations, and the loss function simultaneously fluctuated around 0.025, indicating that the model had good generalization ability and can make accurate predictions for unknown species of wood in the test set.

Comparison test

To evaluate the superiority of the BACNN, it was compared with BP (Rumelhart et al. 1986), SVM (Ma et al. 2020), AlexNet, LeNet, and VGGNet-11. BP and SVM are traditional machine learning methods with a comprehensive comparison

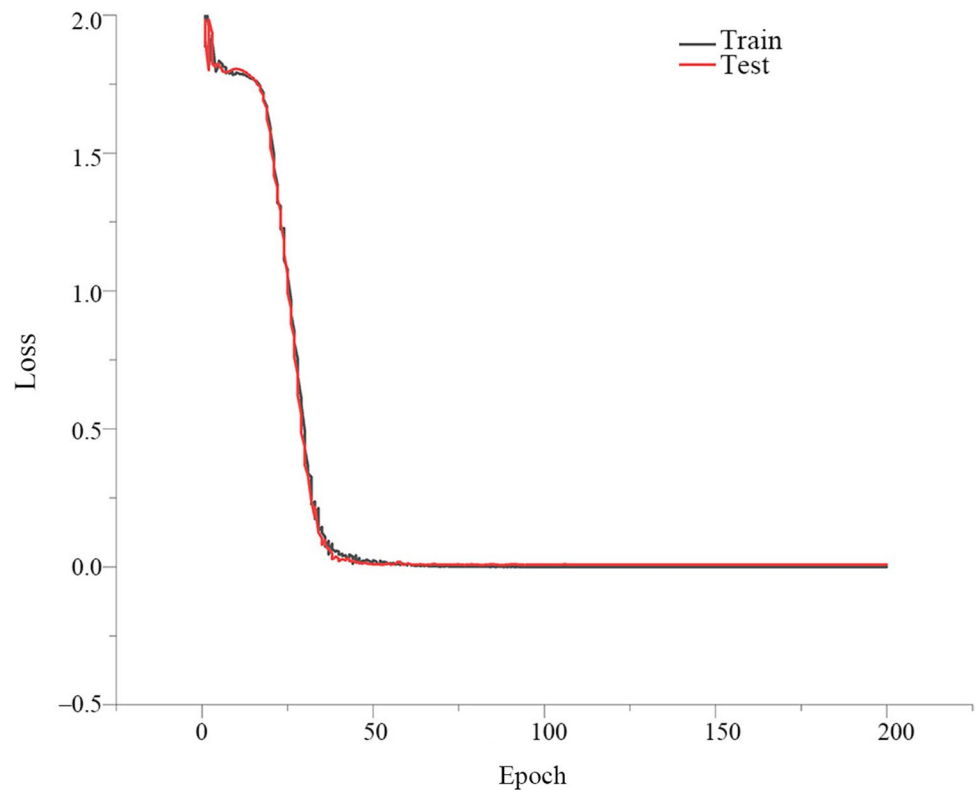
test of P , R , $F1$ and $Accuracy$ for each category; AlexNet, LeNet, and VGGNet-11 are deep learning-based methods that use convolutional neural networks to extract features for classification. This study compared and analyzed the changes of loss function and accuracy curves of these models and their confusion matrices, i.e., compared the combined P , R , $F1$ and $Accuracy$ metrics of these models and drew conclusions.

Evaluation indicators

Using precision (P), recall (R), $F1$ -score ($F1$) and $Accuracy$ as evaluation metrics, all are defined as follows:

$$P = \frac{TP}{TP + FP} \quad (6)$$

$$R = \frac{TP}{TP + FN} \quad (7)$$

Fig. 4 Loss curves

$$F1 = \frac{2 \times P \times R}{P + R} \quad (8)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

where, TP means the prediction is true, the actual is true, FP that the prediction is true, the actual is false, FN that the prediction is false, the actual is true, and TN that the prediction is false, the actual is false.

Comparison with traditional algorithms

Table 3 shows the comparison of BACNN with BP and SVM in four metrics, P , R , $F1$, and $Accuracy$. Bold represents the model proposed in this paper and its associated performance metrics. Since no manual feature extraction was performed, the indicators were not satisfactory using the BP neural network for classification with an accuracy of only 52.9%. Compared with the BP neural network, SVM showed a substantial improvement in all indexes, with an accuracy of 98.7%. Since NIR spectra can reflect the information of functional groups inside the wood, SVM can accurately identify the species of modified popular wood. However, there are errors in classifying different species. When BACNN classified six types of wood material, all indexes perform well and the accuracy rate was as high as 99.6%, and the performance better than BP neural network and SVM.

Table 3 Model prediction analysis

Category	Model	P	R	$F1$	Accuracy
BW	BACNN	100%	97.6%	98.8%	99.6%
	BP	49.4%	59.0%	53.8%	52.9%
	SVM	93.1%	98.7%	95.8%	98.7%
TW	BACNN	98.1%	100%	99.0%	99.6%
	BP	61.7%	55.8%	58.6%	52.9%
	SVM	100%	94.2%	97%	98.7%
SW	BACNN	100%	100%	100%	99.6%
	BP	49.4%	53.7%	51.5%	52.9%
	SVM	100%	100%	100%	98.7%
PVAW	BACNN	100%	100%	100%	99.6%
	BP	51.1%	45.2%	48%	52.9%
	SVM	100%	100%	100%	98.7%
PW	BACNN	100%	100%	100%	99.6%
	BP	47.1%	51.8%	49.4%	52.9%
	SVM	98.8%	100%	99.4%	98.7%
PSW	BACNN	100%	100%	100%	99.6%
	BP	58.2%	52.4%	55.2%	52.9%
	SVM	100%	100%	100%	98.7%

BACNN

Comparison with deep learning methods

LeNet, AlexNet and VGGNet-11 were compared with BACNN to verify the superiority of BACNN. LeNet-5 has five layers, AlexNet has eight, and VGGNet-11 eleven layers. BACNN has eight layers in a single branch and eleven layers in total. The variation of the loss function curves for the four models is shown in Fig. 5. Table 4 shows the accuracy of the prediction set, the time required for training and the size of the resulting network model.

As seen in Fig. 5, when training on the training set, BACNN first converged first and the process was the most stable. LeNet converged the slowest, stabilizing only after 170 iterations; AlexNet and VGGNet-11 converged quickly but were unstable. In the test set, BACNN performed with the highest accuracy, VGGNet-11 s, AlexNet the third and LeNet the lowest (Table 4). The training times and model sizes for the four models are also shown in Table 4, with LeNet, AlexNet, BACNN and VGGNet-11 models increasing in size and training time in this order.

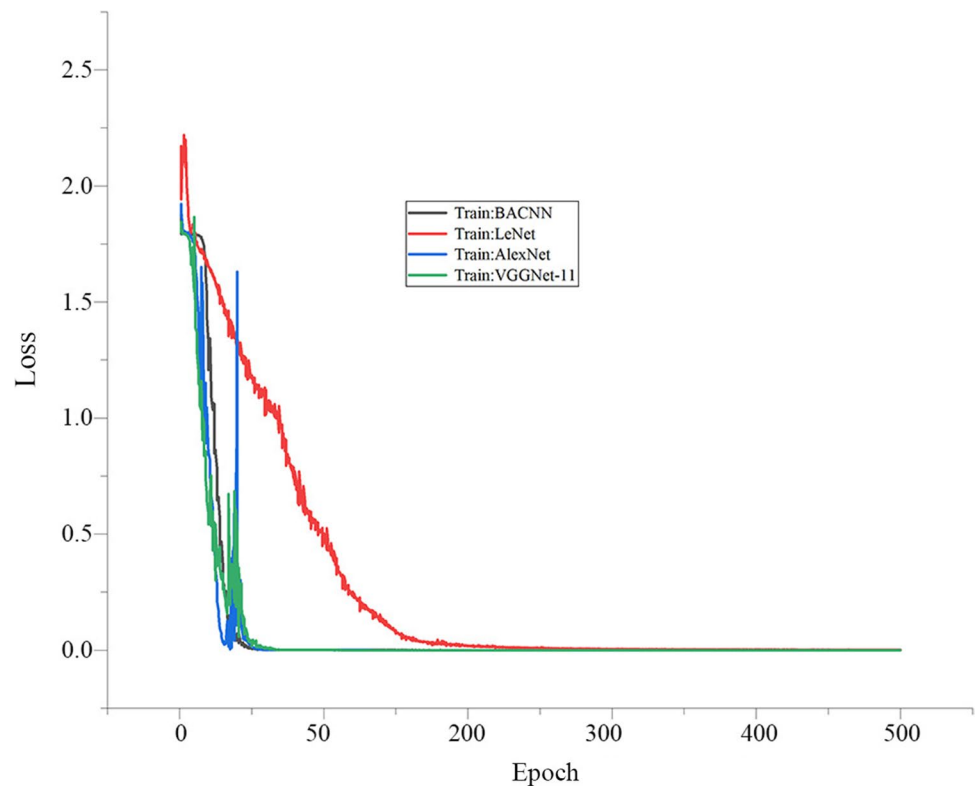
The training time and model size of LeNet, AlexNet and VGGNet-11 increased and the prediction accuracy also increased gradually. The training time and model size of the BACNN were smaller than those of VGGNet-11, but the prediction accuracy was the highest, which shows the advantages of the BACNN.

In the confusion matrix in Table 5, BACNN identified only four samples incorrectly, LeNet identified eleven samples, AlexNet seven samples, and VGGNet-11 five samples incorrectly, indicating that BACNN had the smallest prediction error on the test set.

From these comparison tests (Fig. S3), it can be seen that the classification accuracy was significantly higher than that of BP neural networks and SVMs, compared to traditional machine learning methods because BACNN uses a convolutional neural network model as a feature extractor to automatically perform feature extraction. When comparing BACNN with LeNet, AlexNet and VGGNet-11, LeNet and AlexNet have less training time and smaller models than BACNN, but the training process is unstable, convergence is slower and classification accuracy is lower. In contrast, compared with BACNN, the VGGNet-11 model is larger and takes longer to train, but convergence speed is lower and the classification accuracy is slightly lower. From the above analysis, it can be concluded that BACNN has superior classification results compared to common classification models.

Ablation experiments

To verify the rationality of the structure of the BACNN, ablation experiments were set up with the rationality of the

Fig. 5 Training set Loss**Table 4** Prediction set accuracy

Model	Accuracy	Training time (s)	Number of parameters (KB)
BACNN	99.6%	415	96,709
LeNet	98.0%	50	1241
AlexNet	98.7%	307	69,126
VGGNet-11	99.1%	505	110,025

two-branch network first verified. The BACNN was tested against the CsA the CsB branches to prove the superiority of the two-branch network over the single-branch one. Following this, the rationality of adding the SE module was verified, and by comparing the two cases with and without the SE model, it was demonstrated that adding SE module helped improve the performance of BACNN. It was also shown that adding 1×7 convolution blocks suppressed the noise and enhanced the generalization ability of the BACNN.

Bilinear branching

Bilinear branching extracts spectral features from different scales and then fuses them to obtain more comprehensive and higher quality features. In this section, the effectiveness of the bilinear network was analyzed, and three cases

presented: 1: bilinear model (BACNN); 2: upper branch network (CsA); and 3: lower branch network (CsB). The variation of the loss function for these three cases are shown in Fig. 6.

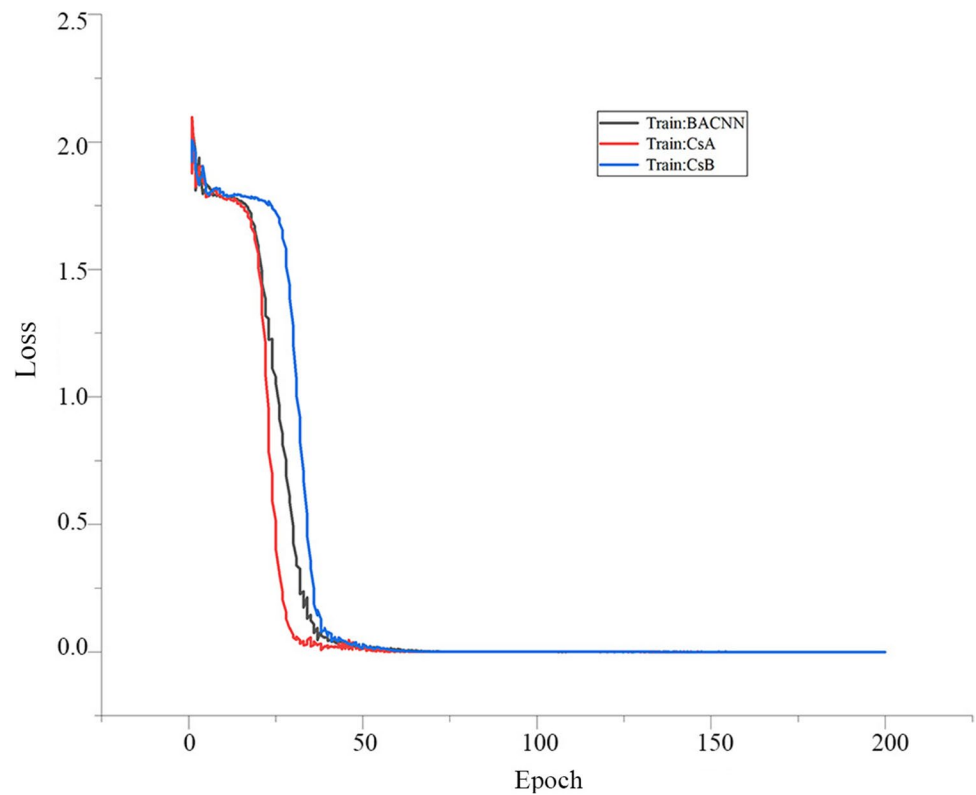
It is noticeable from Fig. 6 that CsA converged the fastest during training for filtering features with smaller convolutional kernels, while CsB was the slowest for extracting features with larger kernels that are more fine-grained. Although the BACNN model is more complex and involves multi-scale feature fusion, it still converged faster than CsB, with an accuracy of 99.6% on the test set, compared to 98.9% for both CsA and CsB. BACNN not only had the highest accuracy and the best generalization capability, but also more stability. The results show that the network structure using two-branch networks to extract features at different scales is superior to that of single-branch networks.

SE model

Because SE modules are 'plug-and-play', they can be easily and effectively used in various networks. In this study, the effects of the presence or absence of SE modules and their location and number were analysed on the performance of the BACNN. As the SE modules require the number of channels to be > 16 , the SE modules were placed in two branches and divided into the following four cases: condition 1: no SE module; condition 2: the SE module placed after the first convolution; condition 3: the SE module placed after the

Table 5 Confusion matrix

Model	Actual category	Prediction category						
		BW	TW	SW	PVAW	PW	PSW	Total
BACNN	BW	82	1	0	0	0	0	83
	TW	1	103	0	0	0	0	104
	SW	0	0	80	0	0	0	80
	PVAW	0	0	2	93	0	0	95
	PW	0	0	0	0	81	0	81
	PSW	0	0	0	0	0	101	101
LeNet	BW	78	5	0	0	0	0	83
	TW	2	101	0	0	1	0	104
	SW	0	0	79	1	0	0	80
	PVAW	0	0	0	94	0	1	95
	PW	0	0	0	0	81	0	81
	PSW	0	0	0	1	0	100	101
AlexNet	BW	80	3	0	0	0	0	83
	TW	1	103	0	0	0	0	104
	SW	0	0	80	0	0	0	80
	PVAW	0	0	0	95	0	0	95
	PW	1	1	0	1	78	0	81
	PSW	0	0	0	0	0	101	101
VGGNet-11	BW	81	2	0	0	0	0	83
	TW	0	104	0	0	0	0	104
	SW	0	0	80	0	0	0	80
	PVAW	0	0	0	93	2	0	95
	PW	0	0	0	0	81	0	81
	PSW	0	0	0	1	0	100	101

Fig. 6 Bilinear branch ablation experiments

second convolution; condition 4: SE module placed after both convolutions. The accuracy of the test set in these four cases is shown in Table 6. Bold represents the position of the SE module when accuracy is highest.

As shown in Table 6, the prediction accurateness of the model was inferior when there were no SE modules. However, the accuracy improved after adding one SE module to each branch, indicating that adding SE modules at this point improves the accuracy of the model. However, the accuracy in condition 4 was the same as in condition 1, indicating that too many SE modules had been added at this point to the detriment of the model's performance. The comparison between conditions 2 and 3 shows that the prediction accuracies of the model were better when the SE module was added to the second convolutional layer. This indicates that, under the experimental conditions of this study, the SE module should be located at a position with a high number of feature channels to be more effective.

Table 6 SE module ablation experiment

Method	Accuracy
condition 1	98.7%
condition 2	98.9%
condition 3	99.6%
condition 4	98.7%

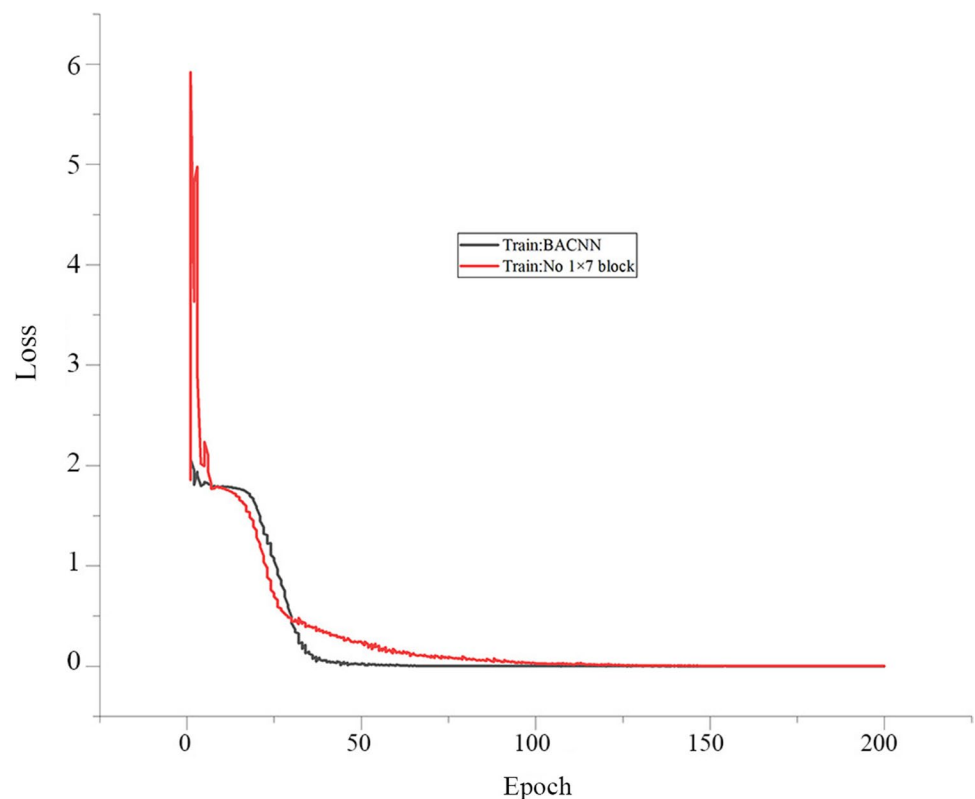
1×7 block

Because the NIR spectral data contained considerable noise, a layer with 1×7 convolutional kernel was used to enhance the noise immunity of the model. To test the effectiveness of the 1×7 block, the loss function of the training set with and without the 1×7 block were compared separately (Fig. 7).

As seen in Fig. 7, the loss function fluctuated considerably at the onset of training by the time the 1×7 blocks were eliminated. Since the network model without 1×7 blocks is more simplified, its initial convergence speed was faster than BACNN, but when the training reached 30 times, the convergence speed became significantly slower with the final speed rather slower than BACNN. With the addition of the 1×7 block, the generalisation of the BACNN model was significantly improved, with the accuracy on the test set 99.6%, higher than 95.8% achieved when the 1×7 block model was removed. In summary, adding the 1×7 block suppresses the noise in the spectral data and improves the stability of the training process. Although the complexity of the network increases, the convergence speed is faster due to less noise interference, which improves the generalization ability of the model and further improves its classification capability.

From the ablation experiment, it may be concluded that setting up a bilinear network for multi-scale feature extraction is reasonable and provides a better performance than a single-branch network. By adding the SE module, it allows

Fig. 7 1×7 block ablation experiments



the BACNN to extract more excellent features and improves the accuracy of classification. Adding 1×7 blocks enhances the stability of the model during training and improves the depth of the model, thus enhancing the network's generalization capability while accelerating the convergence speed during training. In summary, the structure of BACNN classification model is reasonable. It has excellent performance when used for wood NIR spectral classification.

Conclusions

This paper presents a bilinear attention convolutional neural network (BACNN) model based on multi-scale feature fusion to classify the near-infrared spectra of six classes of modified wood with 99.6% accuracy. Comparison tests with BP, SVM, LeNet, AlexNet and VGGNet-11 showed that the BACNN achieved optimal results with both traditional machine learning methods and deep learning methods, verifying its superiority. Ablation experiments showed that the accuracy of the model could be improved by using a two-branch network to extract features from different scales, adding SE modules and 1×7 blocks to the network model, and thus proving the rationality of the BACNN network model structure. Based on the above, the BACNN proposed in this paper can achieve automatic feature extraction and optimal classification results when classifying wood NIR spectra. In addition, the BACNN is also expected to contribute to wood performance prediction.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Acquarelli J, van Laarhoven T, Gerretzen J, Tran TN, Buydens-Marchiori LME (2017) Convolutional neural networks for vibrational spectroscopic data analysis. *Anal Chim Acta* 954:22–31
- Asadi K, Littman M L (2017) In: An alternative softmax operator for reinforcement learning. In: Proc. 28th int'l conf. mach. Learn. Bellevue, WA, 243–252
- Chen YY, Wang ZB (2018) Quantitative analysis modeling of infrared spectroscopy based on ensemble convolutional neural networks. *Chemometr Intell Lab Syst* 181:1–10
- Gao SH, Han Q, Li D, Chen MM, Peng P (2021) Representative batch normalization with feature calibration. *Virtual I*:8669–8679
- Gao MY, Wang F, Song P, Liu JY, Qi DW (2021a) BLNN: multiscale feature fusion-based bilinear fine-grained convolutional neural network for image classification of wood knot defects. *J Sens* 2021:1–18
- Graham B, Engelhardt B, Van den Oord A (2014) Fractional max-pooling. arXiv preprint [arXiv:1412.6071](https://arxiv.org/abs/1412.6071)
- He KM, Zhang XY, Ren SQ, Sun J (2016) Deep residual learning for image recognition. CVPR. Nevada, Las Vegas, pp 770–778
- Hu C, Qu JJ, Xu CP, Zhu AJ (2018a) Garment image recognition based on adaptive pooling neural network. *J Comput Appl* 38(8):2211
- Hu J, Shen L, Sun G (2018b) Squeeze-and-excitation networks. CVPR, Salt Lake City Utah, pp 7132–7141
- Huang PG, Fan Z, Li XP, Guan C, Zhang YF, Wu ZK (2020) Review of computer-based wood feature extraction and identification. *World for Res* 33(01):44–48
- Huang G, Liu Z, Van Der Maaten L, Weinberger Kilian Q (2017) Densely Connected Convolutional Networks. CVPR 2017. Honolulu Hawaii. 4700–4708
- Jia WS, Zhang HZ, Ma J, Liang G, Wang JH, Liu X (2020) Study on the predication modeling of COD for water based on UV-VIS spectroscopy and CNN algorithm of deep learning. *Spectrosc Spectr Anal* 40(9):2981
- Jiao LC, Zhang F, Liu F, Yang SY, Li LL, Feng ZX, Qu R (2019) A survey of deep learning-based object detection. *IEEE Access* 7:128837–128868
- Kauppinen JK (1983) Fourier Self-Deconvolution in Spectroscopy. *Spectrom Tech* 1983:199–232
- Krizhevsky A, Sutskever I, Hinton G (2012) Imagenet classification with deep convolutional neural networks. *Commun ACM* 60(6):84–90
- Kuesel AC, Stoyanova R, Aiken NR, Li CW, Szwergold BS, Shaller C, Brown TR (1996) Quantitation of resonances in biological 31P NMR spectra via principal component analysis: potential and limitations. *NMR Biomed* 9(3):93–104
- Laakmann F, Petersen P (2021) Efficient approximation of solutions of parametric linear transport equations by ReLU DNNs. *Adv Comput Math* Dio. <https://doi.org/10.1007/s10444-020-09834-7>
- Lecun Y, Bottou L, Bengio Y (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
- Lecun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444
- Ma HY, Li XH, Qiang L, Xie XB, Chong J (2020) Zhao CX (2020) Research on identification technology of explosive vibration based on EEMD energy entropy and multiclassification SVM. *Shock Vib* 2:1–10
- Macior A, Zaborniak I, Chmielarz P, Smenda J, Wolski K, Ciszkowicz E, Lecka-Szlachta K (2022) A new protocol for ash wood modification: synthesis of hydrophobic and antibacterial brushes from the wood surface. *Molecules* 27(3):890
- Nisgoski S, DeOliveira A, DeMuñiz G (2017) Artificial neural network and SIMCA classification in some wood discrimination based on near-infrared spectra. *Wood Sci Technol* 51(4):929–942
- Pachuta SJ (2004) Enhancing and automating TOF-SIMS data interpretation using principal component analysis. *Appl Surf Sci* 231(6):217–223
- Pan X, Qiu J, Yang Z (2022) Identification of softwood species using convolutional neural networks and raw near-infrared spectroscopy. *Wood Mater Sci Eng* 1:1–11
- Pradhan T, Kumar P, Pal S (2021) CLAVER: an integrated framework of convolutional layer, bidirectional LSTM with attention mechanism based scholarly venue recommendation. *Inf Sci* 559:212–235
- Qin YH, Ding XQ, Gong HL (2013) High dimensional feature selection in near infrared spectroscopy classification. *Infrared Laser Eng* 42(5):1355–1359

- Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. *Nature* 323(6088):533–536
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv, <https://arxiv.org/abs/1409.1556>.
- Soares SFC, Medeiros EP, Pasquini C (2016) Classification of individual cotton seeds with respect to variety using near-infrared hyperspectral imaging. *Anal Methods* 8(48):8498–8505
- Szegedy C, Liu W, Jia YQ, Sermanet P, Reed S, Anguelov D, Rabinovich A (2015) Going deeper with convolutions. *CVPR*, Boston Massachusetts, pp 1–9
- Tang YS, Chen ZG (2021) Soil pH prediction based on convolution neural network and near infrared spectroscopy. *Spectrosc Spectr Anal* 41(3):892
- Wang XS, Sun YD, Huang AM (2015a) Research on infrared spectrum for timber species identification. *For Eng* 31(6):65–70
- Wang XS, Sun YD, Huang MG, Huang AM (2015b) Back propagation artificial neural network combined with near infrared spectroscopy for timber recognition. *J Northeast for Univ* 43(12):82–85
- Wang QQ, Gao QX, Gao XB, Nie FP (2017) $l(2, p)$ -Norm based PCA for image recognition. *IEEE Trans Image Process* 27(3):1336–1346
- Wang WQ, Zhang J, Wang FL (2019) Attention bilinear pooling for fine-grained classification. *Symmetry* 11(8):1033
- Xia JJ, Huang Y, Li QQ, Xiong YM, Min SG (2021) Convolutional neural network with near-infrared spectroscopy for plastic discrimination. *ECL* 19(5):3547–3555
- Yang SY, Kwon O, Park Y, Chung H, Kim H, Park SY, Choi IG, Yeo H (2020) Application of neural networks for classifying softwood species using near infrared spectroscopy. *J near Infrared Spectrosc* 28(5–6):298–307
- Zhang W, Li CH, Peng GL, Chen YH, Zhang ZJ (2017) A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech Syst Signal Process* 100:439–453

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.