

Research Article

Moisture Content Quantization of Masson Pine Seedling Leaf Based on Stacked Autoencoder with Near-Infrared Spectroscopy

Chao Ni ^{1,2}, Yun Zhang,¹ and Dongyi Wang ²

¹School of Mechanical and Electrical Engineering, Nanjing Forestry University, Nanjing 210037, China

²Bio-Imaging and Machine Vision Lab, Fischell Department of Bioengineering, University of Maryland, College Park 20740, USA

Correspondence should be addressed to Chao Ni; chaoni@njfu.edu.cn

Received 29 May 2018; Revised 17 August 2018; Accepted 30 August 2018; Published 19 September 2018

Academic Editor: Hai Guo

Copyright © 2018 Chao Ni et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Masson pine is widely planted in southern China, and moisture content of the pine seedling leaves is an important index for evaluating the vigor of seedlings. For precisely predicting leaf moisture content, near-infrared spectroscopy analysis is applied in the experiment, which is a cost-effective, high-speed, and noninvasive material content prediction tool. To further improve the spectroscopy analysis accuracy, in this study, a new analysis model is proposed which integrates a stacked autoencoder for extracting hierarchical output-related features layer by layer and a support vector regression model to leverage these features for precisely predicting moisture contents. Compared with traditional spectroscopy analysis method like partial least squares regression and basic support vector regression, the proposed model shows great superiority for leaf moisture content prediction, with R^2 value 0.9946 and root-mean squared error (RMSE) value 0.1636 in calibration set and R^2 value 0.9621 and RMSE 0.4249 in prediction set.

1. Introduction

Masson pine is an important species for forestation in southern China due to its broad distribution. It has been widely applied in pulp and building materials, petroleum extraction, and forestry chemicals [1] because of its desirable features including fast growth, high yield, variable light tolerance, and drought resistance [2]. To ensure the quality of masson pine seedlings, the moisture content of seedling leaves is an important evaluation index and should be assessed before plantation. Physiologically, leaf moisture content plays an important role in photosynthesis, material transport, and maintaining leaf morphological and physiological functions [3]. Traditional moisture measurements are almost chemical-based methods, which are complex, time-consuming, and laborious [4, 5]. As a result, it is urgent to establish a fast, nondestructive, and accurate measurement method for moisture content prediction.

The near-infrared (NIR) spectroscopy technique has been widely used for material content prediction, such as moisture content [6], nitrogen content [7], and chlorophyll

content [8] of plants. To date, many researchers have constructed different NIR spectral calibration models to detect the moisture content of different plants such as sunflower [9], populus tree [10], and eucalyptus tree [11]. Although there are many successful spectroscopy applications, there is no research reported for the moisture content prediction of masson pine seedling leaves.

For spectroscopy data analysis, there are many common used calibration models such as partial least squares regression (PLSR) [12], multiple linear regression (MLR) [13], support vector machines for regression (SVR) [14–16], artificial neural networks (ANNs) [17], and so on. However, the prediction results from these models are hopeful to be improved due to the limitations of the models. PLSR and MLR are widely used methods in the spectroscopy area. Both of them are linear models, which means the model can hardly describe the nonlinear relationship between the input spectroscopy data and output material content value. SVR, a nonlinear model, is suitable for small dataset analysis. It usually has strong generalization ability and self-learning capacity. The performance of SVR model usually depends on

the quality input features, and determined representative data features are the premise for training a good SVR model [16]. ANNs offer a simple solution to model complex nonlinear data relationship [18]. The multiple layers in the ANN model can map the data input to the high-level feature representations. With the increasing number of hidden layers, the imbalance between the trainable network parameters and the limited training samples hinders the network training process in practice [19]. Stacked autoencoder (SAE) [20] is a variant of the ANN model, which usually is composed of an encoder and decoder. The encoder can map the input into few high-dimensional data features, and the decoder is expected to reconstruct the raw data. SAE can be trained without supervision, and the high dimensional data features from the encoder have been successfully applied in different tasks [21–24].

In this paper, to combine the advantages of different existing data analysis methods, a new model is proposed for predicting the moisture content in masson pine seedling leaves. In the model, a SAE is pretrained layer by layer without supervision to get the high-level data features. Then, based on the properties of the ANN model, the high-level data can be fine-tuned in a supervised way. Once the high-level data features are determined, the SVR model is established to describe the relationship between the high-level data features and the moisture content. Compared with other traditional spectroscopy analysis methods, the proposed model shows better performance.

2. Materials and Methods

2.1. Materials. A total of 100 annual masson pine seedling samples were obtained from forest farm in Huangping, Guizhou Province, China. Seventy-five of them were used to establish a calibration model, and the remaining 25 were used for prediction. Before measuring the spectra and moisture content, the leaves were cleaned to remove the impurities such as soil and sand, and then the surface of the leaves were naturally dried to prepare for the later experiment.

2.2. Spectra Acquisition and Moisture Quantitative Analysis. All NIR spectra of masson pine seedling leaves were measured using an MPA Fourier-Transform Near-Infrared spectrometer (Bruker Optics, Inc., Germany), equipped with an PbS detector and controlled by OPUS Analyst version 7.0 using a spectral range from 4,000 to 12,493 cm^{-1} in the reflectance mode. Measurements were conducted with the resolution of 4 cm^{-1} , ensuring an adequate signal-to-noise ratio. There are 2203 wavelength variables in the spectrum. The experiment was carried out at approximately 24°C. The spectral reflectance of the top, middle, and bottom areas of the sample were scanned two times each. The final value of the sample spectral reflectance was obtained by calculating the average of the six scans. From each sample, six spectra were collected and averaged for further analysis. Figure 1 shows the raw absorbance spectra of masson pine seedling leaves.

Total moisture content of masson pine seedling leaves was determined using the HB43-S Halogen Moisture

analyzer (Mettler Toledo, Inc., Switzerland). The instrument is working based on the thermogravimetric principle, i.e., the moisture is determined from the weight loss of a sample dried by heating. After putting a sample in the sampling chamber, the temperature was immediately increased and maintained at 125°C. The halogen lamp of the instrument heated the sample until the sample stopped losing mass; then the moisture content of the samples is automatically given by the instrument. Normally, the measurements were completed in a few minutes. The reference moisture content value for the samples was $64.97\% \pm 2.21\%$, with a minimum of 59.13% and a maximum of 70.74%.

2.3. Spectra Pretreatment. Two preprocessing transformations were applied as a standard preparation for the masson pine seedling absorbance curves. Firstly, a Savitzky–Golay (S-G) smoothing with a second-degree polynomial was used for denoising, and the first-order numerical derivative was made to correct the baseline drift. The window width of S-G smoothing was set as 17. Secondly, the spectral matrix was converted into the number between 0 and 1 by vector normalization, which was used to reduce the orders of magnitude difference among the data of different dimensions.

2.4. The Proposed Method

2.4.1. Autoencoder. The basic structure of autoencoder is a kind of unsupervised ANN with one hidden layer, and it consists of an input layer, a hidden layer, and an output layer as Figure 2. The aim of autoencoder is to reduce the data dimension and to map the input data into high dimensional data features.

Define the input as $x = [x_{(1)}, x_{(2)}, \dots, x_{(d)}]'$, where d is the dimension of the inputs. The encoder maps x into the hidden layer $h = [h_1, h_2, \dots, h_{d_h}]'$ by the function f as follows:

$$h = f(x) = s_f(Wx + b), \quad (1)$$

where d_h is the dimension of the hidden layer variable vector h , W is a $d_h * d$ weight matrix, b is the bias vector, and s_f is the nonlinear activation function, which can be chosen as the sigmoid function or other functions such as the tanh function and the rectified linear unit function. Then, the hidden representation vector h is mapped to the output layer \tilde{x} by the function \tilde{f} .

$$\tilde{x} = \tilde{f}(h) = s_{\tilde{f}}(\tilde{W}h + \tilde{b}), \quad (2)$$

where \tilde{W} is a $d * d_h$ weight matrix, \tilde{b} is the bias vector for the output layer, and $s_{\tilde{f}}$ is the nonlinear activation function of decoder. The aim of autoencoder is to search the parameter set $\theta = \{W, \tilde{W}, b, \tilde{b}\}$ to satisfy the equation $\tilde{x} = \tilde{f}(h) = \tilde{f}(f(x)) \approx x$. Define the training input as $X = \{x_1, x_2, \dots, x_N\}$, where N is the number of the training samples and $x_i = [x_{i(1)}, x_{i(2)}, \dots, x_{i(d)}]$ is the vector data of i th training sample. The loss function is defined as follows:

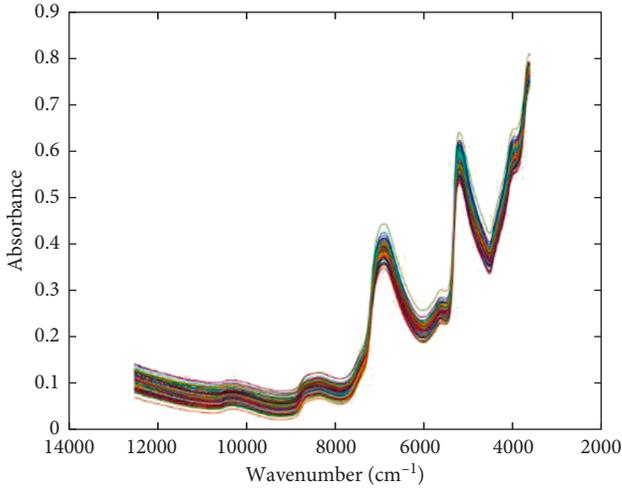


FIGURE 1: Raw absorbance spectra of masson pine seedling leaves.

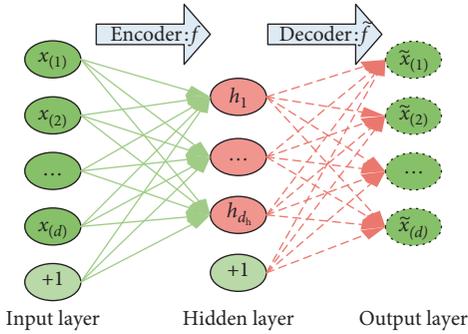


FIGURE 2: The typical structure of an autoencoder.

$$\begin{aligned}
 J(W, \tilde{W}, b, \tilde{b}) &= \frac{1}{2N} \sum_{i=1}^N \|x_i - \tilde{x}_i\|^2 \\
 &= \frac{1}{2N} \sum_{i=1}^N \sum_{j=1}^d (x_{i(j)} - \tilde{x}_{i(j)})^2.
 \end{aligned} \quad (3)$$

Then, the parameter set $\theta = \{W, \tilde{W}, b, \tilde{b}\}$ is updated with the gradient descent algorithm.

In practice, multiple basic autoencoder structures are usually stacked together, and a new network structure is constructed, named SAE, for better dimensional reduction and feature extraction performance. The extracted data feature can be further used for different tasks [25, 26].

2.4.2. Support Vector Machine Regression. Support vector machine (SVM) is a powerful and robust method based on the principle of structural risk minimization, which has an advantage of computational efficiency for the data sets that have many more variables than observations. Compared with ANN, SVM can obtain more reliable and better performance under same training conditions. Recently, SVM has been successfully extended to SVR, especially in chemometrics for quantitative analysis due to its excellent ability of dealing with nonlinearity and small sample size.

The purpose of SVR is to find the underlying relationship between input and output, and a regression function $y(x) = f(x) + b = w^T \phi(x) + b$ is used with an ε -insensitive loss function as follows:

$$|\xi|_\varepsilon = \begin{cases} 0 & \text{if } |\xi| \leq \varepsilon, \\ |\xi| - \varepsilon & \text{otherwise.} \end{cases} \quad (4)$$

Then, the regression problem is equivalent to the following formulations:

$$\begin{aligned}
 &\min \left(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \right), \\
 &\text{s.t.} \begin{cases} y_i - w^T \phi(x) - b \leq \varepsilon + \xi_i, \\ -y_i + w^T \phi(x) + b \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0, \end{cases} \quad (5)
 \end{aligned}$$

where ξ_i and ξ_i^* are slack variables, and the regularization parameter $C > 0$ is used to avoid overfitting. Further, the optimization problem can be converted into the new optimization problem as the following:

$$\begin{aligned}
 &\max \left(\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) \right. \\
 &\quad \left. - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) \right), \\
 &\text{s.t.} \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*), \\ \alpha_i, \alpha_i^* \in [0, C], \end{cases} \quad (6)
 \end{aligned}$$

where α_i and α_i^* are Lagrange multipliers and $K(x_i, x_j)$ is a kernel function. The kernel function maps the nonlinear optimization problem to a linear problem in higher dimensional space. There are several commonly used kernel functions such as linear kernel, polynomial kernel, and radical basis function kernel. In this study, radical basis function kernel is selected as follows:

$$K(x_i, x) = \exp \left(\frac{-\|x - x_i\|^2}{\gamma^2} \right), \quad (7)$$

where γ determines width of the kernel function. Once radical basis function kernel is selected, there are only three parameters to be decided. The parameter ε is to control the error. A larger ε can speed up the train process with low accuracy. Whereas, a smaller ε can achieve better accuracy and reduce the training speed. In this paper, the parameter is set fixed as 0.004. Besides the parameter ε , the regularization parameter C and the kernel function parameter γ are both application-based parameters, which can greatly affect accuracy of the regression. In this paper, genetic algorithm is utilized to search optimal parameters for SVR [27].

2.4.3. Regression Model Based on SAE with SVR. To estimate moisture content in masson pine seedling leaves, a novel regression model based on SAE with SVR is proposed as shown in Figure 3. The procedure of the proposed method is described as follows:

Step 1. Several AEs are stacked layer by layer to form deep neural networks. The deep neural networks can convert a complex input data into a series of simple high-level features by reducing the dimensionality of the input data. To train the networks, the first AE is trained in an unsupervised manner. After training is finished, its decoder is abandoned, and the output of the hidden layer is used as the input of the second AE with same fashion. Till all AEs have been trained layer by layer, the weights of these AEs had been assigned to initialize the deep neural networks. The topmost hidden layer outputs the elementary high-level features of the input data.

Step 2. To achieve better feature representation of the input data, the supervision method is used to tune the weights. A two-layer neural network is added after the topmost layer, and the output of this neural network is the target output. Then, the back-propagation algorithm is utilized to update the weights layer by layer. At last, the fine-tuning weights have been obtained, and then the improved high-level features have been achieved.

Step 3. Feed the improved high-level features into SVR as input data, and then build the regression model using GA algorithm to select the optimal parameters.

2.5. Evaluation Criteria. As normal, the root-mean squared error (RMSE) and the coefficient of determination are widely used as the evaluating criteria for the calibration model. RMSE is defined as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N_T} (y_i - \hat{y}_i)^2}{N_T}}, \quad (8)$$

where y_i and \hat{y}_i are the actual and predicted target output values of the i th sample, respectively and N_T is the number of the testing samples. RMSE indicates the accuracy of the model, and a small RMSE value shows better prediction performance than larger RMSE.

The coefficient of determination R^2 represents a squared correlation between the actual and predicted output, and the reliability of model can be reflected as R^2 . R^2 is defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N_T} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N_T} (y_i - \bar{y})^2}, \quad (9)$$

where \bar{y} is the mean of the actual output in the testing samples. When R^2 is closer to 1, it indicates good prediction performance of the calibration model. In a word, a better calibration model should have small RMSE and large R^2 .

3. Results and Discussion

3.1. Result and Settings of the Proposed Model. For the proposed model construction and test, 100 samples were divided into 75 samples for calibration dataset and 25 samples for prediction dataset. The first step is to decide the settings of the model. It is evident that the number of the hidden layers and the number of the neuron in each hidden layer can have major influence on the performance of the model. However, there is still no automatic method that can be utilized to achieve the selection of the parameters. In this paper, trial and error is used to choose these parameters, and RMSE is adopted to access the performance of the model. The weights and bias for each AE is obtained by greedy layerwise pretraining technique, and each AE is trained with gradient descent algorithm. After the SAE is layerwise trained, the two-layer neural network is connected to the output of SAE for weight fine tuning of ANN. After that, SVR is connected to the high-level features to output the predicted value. After experiments, the proposed model structure is composed of three-layer AEs, and the hidden neurons number is 1600, 1100, and 512, respectively. The batch size is set as 40 samples, and each AE is trained 500 iterations iteratively. Figure 4 shows the training loss trends with the iteration number for each AE. As shown in Figure 4, the SAE model can converge quickly in 500 iterations for each AE.

After the weights are pretrained and fine-tuned, the SVR is connected to output the predicted value. The detailed prediction results are shown in Figure 5, and C and γ of SVR are set as 4.79 and 1.97, respectively, using GA algorithm. The RMSE and R^2 of the prediction dataset are 0.4249 and 0.9621, respectively. As can be seen from Figure 5, the predicted output can match well with the real values. The prediction error is mainly caused by the points which are located in the interval with few samples in the calibration dataset.

3.2. Discussion. To verify the effectiveness of the proposed method, there are several methods to be used to build the calibration model. Each method is trained with the same calibration dataset, and the average result of 4-folder cross-validation in each method is adopted in Table 1 and Figure 6. The results of MLR, PLS, and SVR are directly obtained using PLS_Toolbox (Eigenvector Research, Inc., USA) running on MATLAB 2016. The other methods are implemented by the deep learning framework Keras using Python. As can be seen in Table 1, MLR and PLS produce the worst prediction results since they cannot deal with nonlinear correlative data for its linear essence. SVR is inferior to other methods and just outperforms above two methods, because it cannot describe the nonlinear data adequately. The last three methods all adopt the same network structure. The ANN method is composed of five layers of neural networks, and the first three layers have the same network structure as SAE. The last two layers have the same structure with neural networks as being connected to SAE. The same two-layer neural networks with the neuron layer structure

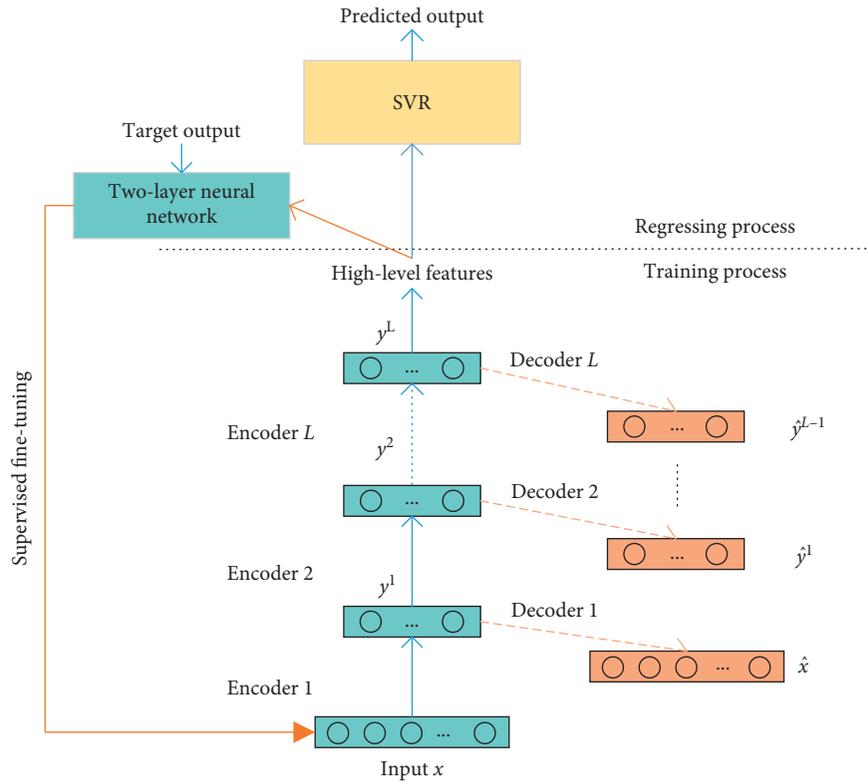


FIGURE 3: The proposed network structure of deep learning with SAE-SVR.

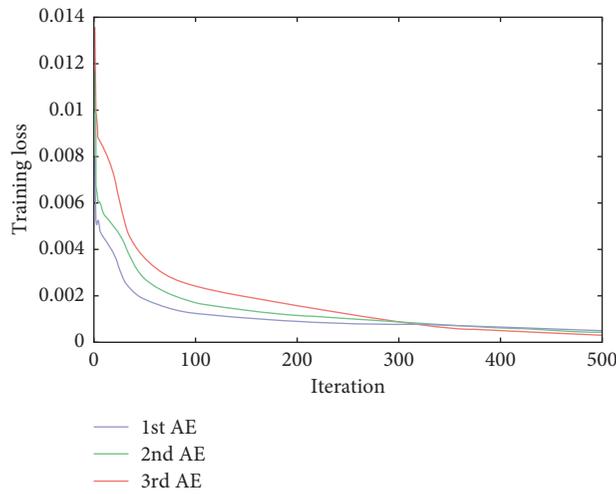


FIGURE 4: The training loss trends of each AE on the calibration dataset.

[512, 128] are connected to SAE for fine tuning the weights. ANN method outperforms above three methods, but it is easily trapped in local optima. By adopting the SAE, high-level abstract features can be extracted layer by layer in SAE. Further, by using the pretrained weights and bias terms in the full connected network, the methods can avoid the inferior local optima and speed up the learning process. Thus, the last two methods can describe the complex data structure more accurately. Besides, a small size of training

samples used in this study may also limit the predictive power of the ANN model. Then, SVR with genetic algorithm is connected to SAE to achieve better performance in small training samples. The regression plot between the predicted moisture content and the reference content is shown in Figure 6. It can visualize the regression performance of different models. If the scatter data points are closer to the perfect correlation line (with slope=1), the prediction model is more accurate (larger R^2 value and smaller RMSE

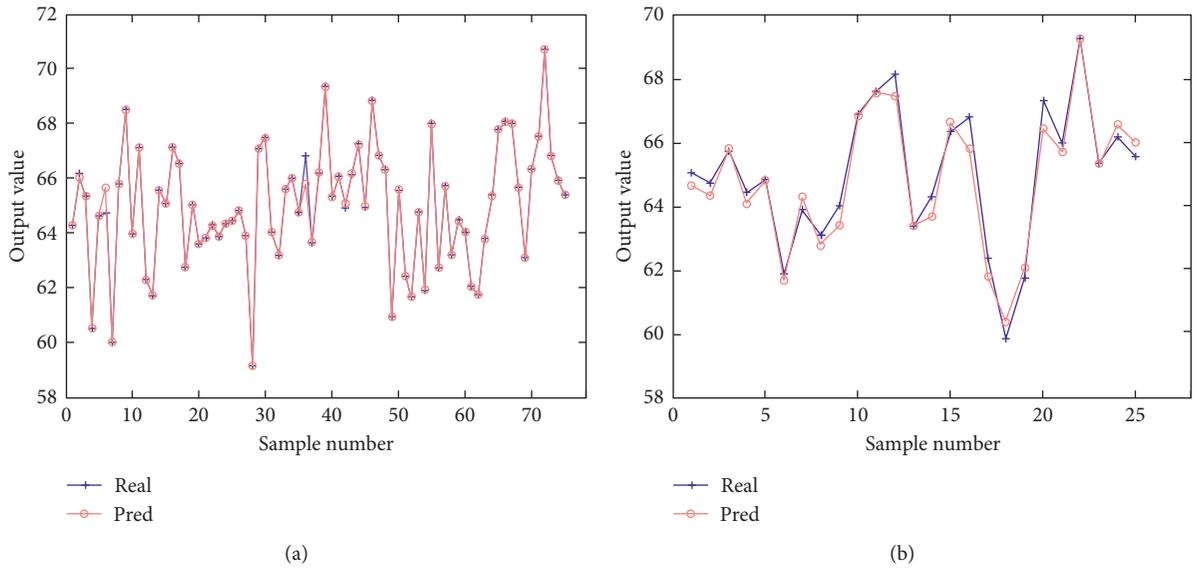


FIGURE 5: The prediction performance on the calibration dataset (a) and the prediction dataset (b).

TABLE 1: Calibration and prediction results of the moisture content in masson pine seedling leaves using different regression models.

Model	Calibration dataset		Prediction dataset	
	R^2_C	RMSEC	R^2_P	RMSEP
MLR	0.9267	0.5619	0.3915	2.879
PLSR	0.9658	0.4079	0.8981	0.6774
SVR	0.9892	0.2285	0.9016	0.7052
ANN	0.9974	0.1123	0.9322	0.5559
SAE-ANN	0.9944	0.1644	0.9421	0.5228
SAE-SVR	0.9946	0.1636	0.9621	0.4249

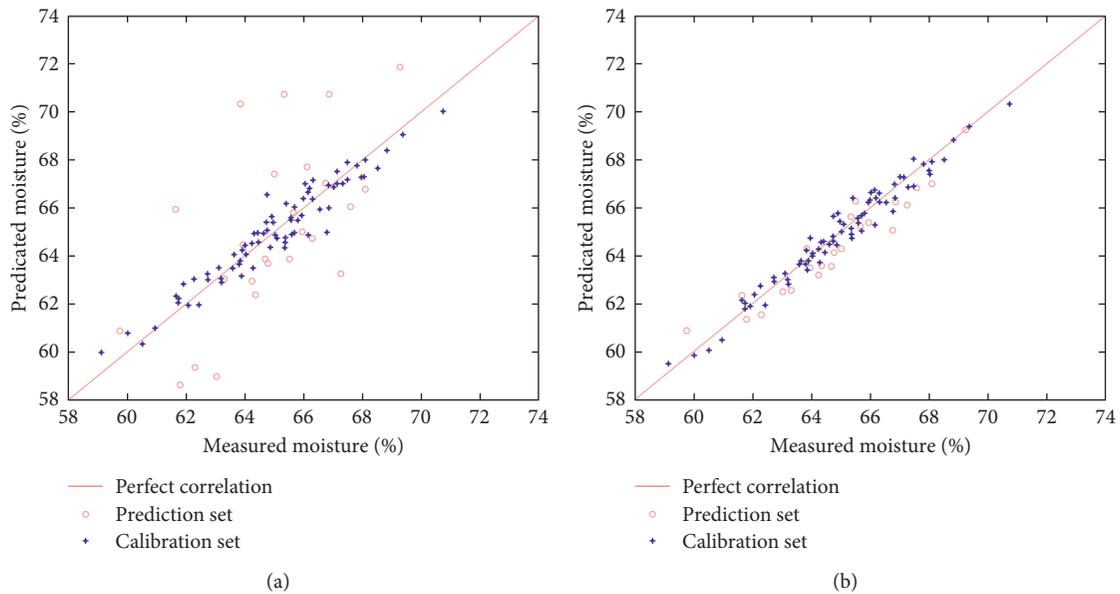


FIGURE 6: Continued.

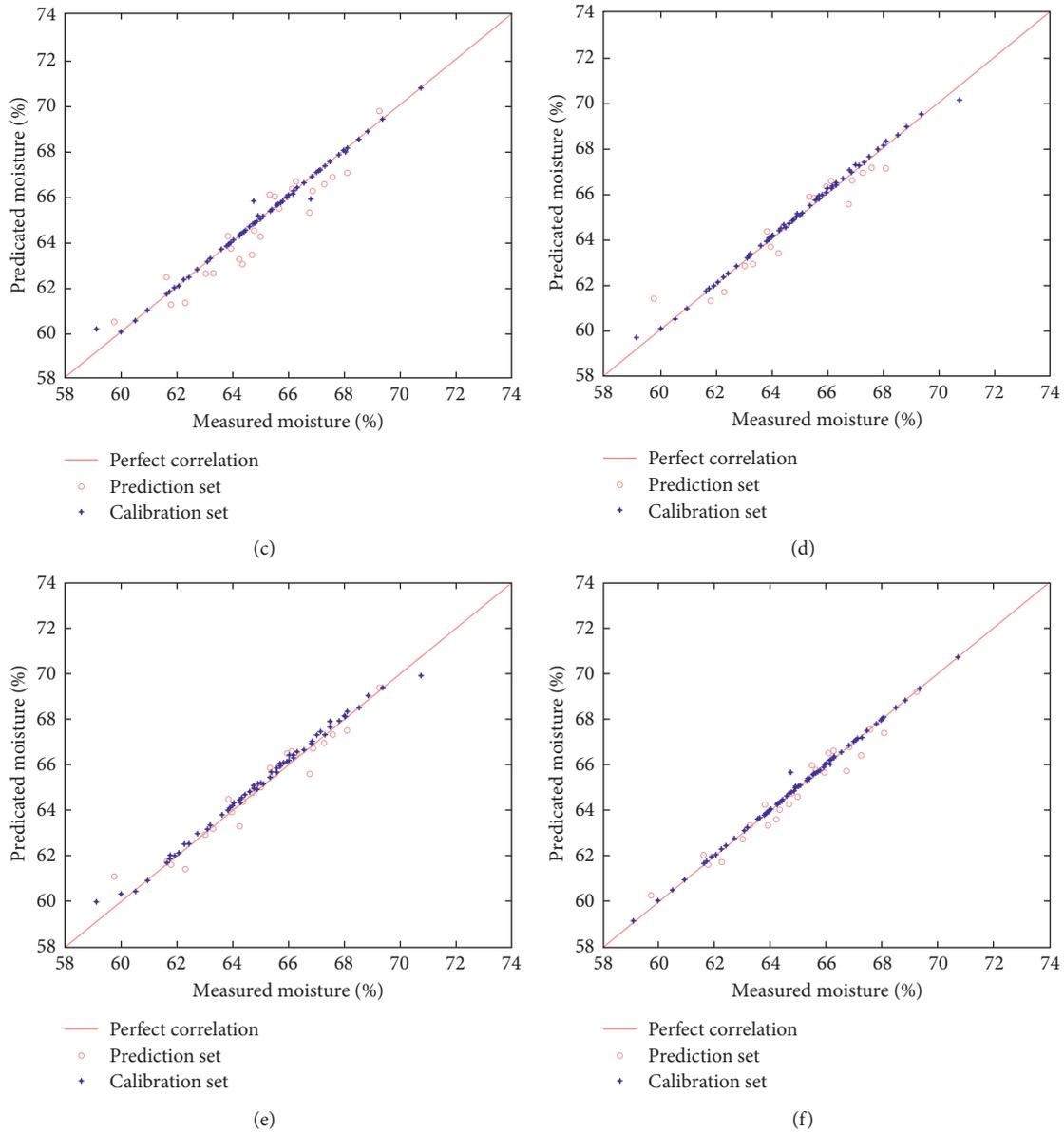


FIGURE 6: Measured vs. predicated moisture content for masson pine seedling leaves by the regression models of MLR (a), PLSR (b), SVR (c), ANN (d), SAE-ANN (e), and SAE-SVR (f).

value). Overall, the proposed method outperforms the other methods for moisture concentration prediction of masson pine seedling leaves.

4. Conclusion

In this study, a new method of stacked autoencoder with SVR is introduced and applied to estimate the moisture content in masson pine seedlings. Compared with MLR, PLSR, SVR, ANN, and SAE-ANN, the SAE-SVR shows superior performances, and gets R^2 value 0.9946 and RMSE 0.1636 in calibration dataset, and R^2 value 0.9621 and RMSE 0.4249 in prediction dataset. The results of this study demonstrate that deep neural networks pretrained by SAE are feasible to be used as the data analysis method for moisture prediction in masson pine seedling leaves.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors gratefully acknowledge the financial support provided by the National Natural Science Foundation of China (NSFC: 31570714) and Jiangsu Overseas Research & Training Program for University Prominent Young & Middle-Aged Teachers and Presidents Program.

References

- [1] Y. Wang, S. Solberg, P. Yu et al., "Assessments of tree crown condition of two Masson pine forests in the acid rain region in south China," *Forest Ecology and Management*, vol. 242, no. 2-3, pp. 530-540, 2007.
- [2] Y. You, X. Huang, H. Zhu et al., "Positive interactions between *Pinus massoniana* and *Castanopsis hystrix* species in the uneven-aged mixed plantations can produce more ecosystem carbon in subtropical China," *Forest Ecology and Management*, vol. 410, pp. 193-210, 2017.
- [3] V. Suguiyama, R. Sanches, S. Meirelles, D. C. Centeno, E. A. Da Silva, and M. R. Braga, "Physiological responses to water deficit and changes in leaf cell wall composition as modulated by seasonality in the Brazilian resurrection plant *Barbacenia purpurea*," *South African Journal of Botany*, vol. 105, pp. 270-278, 2016.
- [4] T. A. Oyehan, I. O. Alade, A. Bagudu, K. O. Sulaiman, S. O. Olatunji, and T. A. Saleh, "Predicting of the refractive index of haemoglobin using the Hybrid GA-SVR approach," *Computers in Biology and Medicine*, vol. 98, pp. 85-92, 2018.
- [5] T. A. Saleh, S. O. Adio, M. Asif, and H. Dafalla, "Statistical analysis of phenols adsorption on diethylenetriamine-modified activated carbon," *Journal of Cleaner Production*, vol. 182, pp. 960-968, 2018.
- [6] J. Posom and P. Sirisomboon, "Evaluation of the moisture content of *Jatropha curcas* kernels and the heating value of the oil-extracted residue using near-infrared spectroscopy," *Biosystems Engineering*, vol. 130, pp. 52-59, 2015.
- [7] R. Tang, X. Chen, and C. Li, "Detection of nitrogen content in rubber leaves using near-infrared (NIR) spectroscopy with correlation based successive projections algorithm (SPA)," *Applied spectroscopy*, vol. 72, no. 5, pp. 740-749, 2018.
- [8] J. Zhang, W. Han, L. Huang, Z. Zhang, Y. Ma, and Y. Hu, "Leaf chlorophyll content estimation of winter wheat based on visible and near-infrared sensors," *Sensors*, vol. 16, no. 4, p. 437, 2016.
- [9] A. J. S. Neto, D. De Carvalho Lopes, T. G. F. Da Silva, S. O. Ferreira, and J. A. S. Grossi, "Estimation of leaf water content in sunflower under drought conditions by means of spectral reflectance," *Engineering in Agriculture, Environment and Food*, vol. 10, no. 2, pp. 104-108, 2017.
- [10] G. Hans, B. Leblon, P. Cooper, A. La Rocque, and J. Nader, "Determination of moisture content and basic specific gravity of *Populus tremuloides* (Michx.) and *Populus balsamifera* (L.) logs using a portable near-infrared spectrometer," *Wood Material Science & Engineering*, vol. 10, no. 1, pp. 3-16, 2015.
- [11] G. Yang, W. Lu, Y. Lin et al., "Monitoring water potential and relative water content in *Eucalyptus camaldulensis* using near infrared spectroscopy," *Journal of Tropical Forest Science*, vol. 29, no. 1, pp. 121-128, 2017.
- [12] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109-130, 2001.
- [13] L. S. Aiken, S. G. West, and S. C. Pitts, "Multiple linear regression," in *Handbook of Psychology*, pp. 481-507, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2003.
- [14] O. Devos, C. Ruckebusch, A. Durand, L. Duponchel, and J.-P. Huvenne, "Support vector machines (SVM) in near infrared (NIR) spectroscopy: focus on parameters optimization and model interpretation," *Chemometrics and Intelligent Laboratory Systems*, vol. 96, no. 1, pp. 27-33, 2009.
- [15] B. Üstün, W. Melssen, and L. Buydens, "Visualisation and interpretation of support vector regression models," *Analytica Chimica Acta*, vol. 595, no. 1-2, pp. 299-309, 2007.
- [16] I. O. Alade, A. Bagudu, T. A. Oyehan, M. A. A. Rahman, T. A. Saleh, and S. O. Olatunji, "Estimating the refractive index of oxygenated and deoxygenated hemoglobin using genetic algorithm-support vector machine approach," *Computer Methods and Programs in Biomedicine*, vol. 163, pp. 135-142, 2018.
- [17] R. M. Balabin, E. I. Lomakina, and R. Z. Safieva, "Neural network (ANN) approach to biodiesel analysis: analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (NIR) spectroscopy," *Fuel*, vol. 90, no. 5, pp. 2007-2015, 2011.
- [18] S. Cui, P. Ling, H. Zhu, and H. Keener, "Plant pest detection using an artificial nose system: a review," *Sensors*, vol. 18, no. 2, p. 378, 2018.
- [19] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527-1554, 2006.
- [20] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, pp. 3371-3408, 2010.
- [21] J. Gehring, Y. Miao, F. Metze et al., "Extracting deep bottleneck features using stacked auto-encoders," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3377-3381, Vancouver, BC, Canada, May 2013.
- [22] C. Tao, H. Pan, Y. Li, and Z. Zou, "Unsupervised spectral-spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 12, pp. 2438-2442, 2015.
- [23] X. Wang and H. Liu, "Soft sensor based on stacked auto-encoder deep neural network for air preheater rotor deformation prediction," *Advanced Engineering Informatics*, vol. 36, pp. 112-119, 2018.
- [24] X. Yuan, B. Huang, Y. Wang, C. Yang, and W. Gui, "Deep learning based feature representation and its application for soft sensor modeling with variable-wise weighted SAE," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3235-3245, 2018.
- [25] P. Vincent, H. Larochelle, Y. Bengio et al., "Extracting and composing robust features with denoising autoencoders," in *Proceedings of 25th international conference on Machine learning*, pp. 1096-1103, Helsinki, Finland, July 2008.
- [26] A. Supratak, L. Li, and Y. Guo, "Feature extraction with stacked autoencoders for epileptic seizure detection," in *Proceedings of 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4184-4187, Chicago, Illinois, USA, August 2014.
- [27] E. Avci, "Selecting of the optimal feature subset and kernel parameters in digital modulation classification by using hybrid genetic algorithm-support vector machines: HGASVM," *Expert Systems with Applications*, vol. 36, no. 2, pp. 1391-1402, 2009.

