Application of Auto-encoder and Attention Mechanism in Raman Spectroscopy

Yunyi Bai, Mang Xu, and Pengjiang Qian(✉)

School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, Jiangsu, China
qianpengjiang@jiangnan.edu.cn

Abstract. Conventional Raman spectroscopy, which is based on the qualitative or quantitative determination of substances, has been widely utilized in industrial manufacture and academic research. However, in traditional Raman spectroscopy, human experience plays a prominent role. Because of the massive amount of comparable information contained in the spectrograms of varying concentration media, the extraction of feature peaks is especially crucial. Although manual feature peak extraction in spectrograms might reduce signal dimensionality to a certain extent, it could also result in spectral information loss, misclassification, and underclassification of feature peaks. This research solves the problem by extracting a feature dimensionality reduction method based on an auto-encoder-attention mechanism, applying a deep learning approach to spectrogram feature extraction, and feeding the features into a neural network for concentration prediction. After rigorous testing, the model's prediction accuracy may reach a unit concentration of 0.01 with a 13% error, providing a reliable aid to manual and timely culture medium replenishment. And through extensive comparison experiments, it is concluded that the self-encoder-based dimensionality reduction method is more accurate compared with the machine learning method. The research demonstrates that using Raman spectroscopy to deep learning can produce positive outcomes and has great potential.

Keywords: Raman spectroscopy · Deep learning · Auto-encoder · Attentional mechanisms

1 Introduction

Raman spectroscopy is currently widely utilized in industry, food, and biotechnology as an accurate material detection tool with easy data capture, speed, and high accuracy for qualitative or quantitative study of material composition. The processing of spectrograms and the analysis of spectrum data are critical steps in the quantitative measurement of substances, and the regression algorithms used to do so have a direct impact on the spectral data's accuracy. It has been used to analyze substance concentrations in both qualitative and quantitative ways.

The typical Raman spectroscopy processing technique includes numerous steps, and many of them, such as de-baseline correction [1] and smooth-denoising, rely on human

D.-S. Huang et al. (Eds.): ICIC 2022, LNAI 13395, pp. 698–708, 2022.
https://doi.org/10.1007/978-3-031-13832-4_57

expertise to complete the experiments. To discover the feature peaks in the spectrum map, start by downscaling the high-dimensional data to get the key features, then use principal component analysis, random forest, or other approaches to finish particular tasks. The purpose of dimensionality reduction is to extract the features that are useful for the regression task and discard those that are of useless. In addition, there is serious covariance between the wavelength points [2], using all dimensions of the spectrogram not only increases the complexity of the model computation but also introduces unnecessary noise to affect the prediction results. As a result, the selection of feature peaks is crucial. Traditionally, there are three types of feature peak selection, the first one is to explore the feature peak intervals one by one, using the statistical information related to the model, and then decide which feature peak intervals are needed, such as interval partial least squares (IPLS) [3], moving window partial least squares (MWPLS), etc. [4]. The second type of method is to select the peaks according to their covariance index, regression index, and other indicators, such as competitive adaptive reweighted sampling (CARS) [5], partial least squares uninformative elimination (UE-PLS) [6]. The third category is the algorithms for optimization problems. For instance, genetic algorithm (GA), and simulated annealing (SA) to select feature peaks.

In the past few years, With the evolution of hardware techniques, deep learning has continued to develop which has been widely used and achieved rich results in image processing, autonomous driving, speech recognition, etc. A CNN is an important part of deep learning, which is a kind of feed-forward neural network and has shown powerful classification and regression ability in many fields. For example, classical convolutional neural networks: Alexnet [7], VGG [8], Resnet [9], etc. With the advancement of deep learning, auto-encoder [10] Compared with the traditional classical PCA [11] algorithm, auto-encoder is an unsupervised deep learning algorithm for dimensionality reduction, which can learn the nonlinear feature representation that cannot be learned by PCA and can better learn advanced semantic features for high-dimensional data. This work utilizes the self-learning feature of neural networks to obtain the corresponding weight information of each feature peak and achieves the purpose of feature peak selection by reconstructing the input information to retain the large-weighted feature peaks and eliminate the small-weighted feature peaks, meanwhile, further improves the experimental effect when predicting the concentration with the attention mechanism. Our proposed self-encoder-attention mechanism method does not require human manual feature peak selection, and the prediction results are more stable than some traditional machine learning methods, and better results are achieved in the experimental results.

2 Related Work

2.1 Raman Spectra of Single Component Glucose Medium

A total of 17 concentrations of glucose medium concentrations were selected, and their Raman spectra ranged from 0 to 3400 cm^{-1}, as the concentrations were different, the corresponding characteristic peaks were different, which led to different Raman spectrograms for each concentration. With a total of 170 samples, the key to effective identification of the spectrograms of different concentrations lies in the selection of the characteristic peaks. As shown in Fig. 1, which shows the Raman spectra of the single-component medium samples, it is easy to see that in some regions, the curves of different concentrations are nearly overlapping, and the features in these overlapping regions are not useful for the experiment, i.e., they are features to be discarded when performing feature extraction. In other regions, the curves do not overlap, and the features in these regions are the features that are useful for the experiments, which are also called feature peaks, and it is critical to make the most of these usable features for feature extraction.

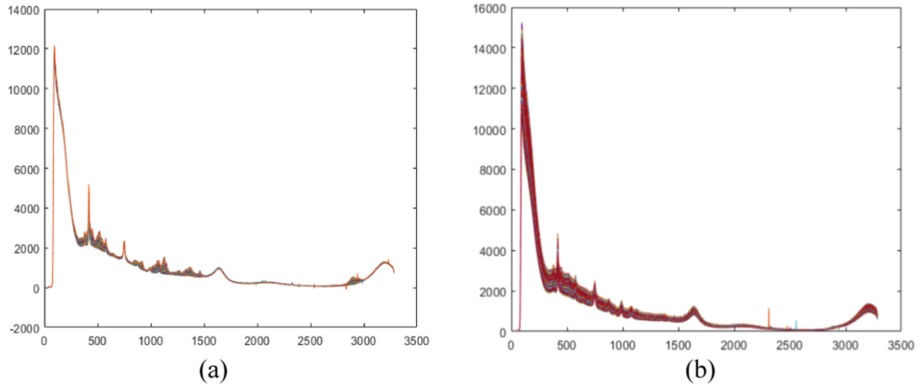


Fig. 1. Raman spectra of single component glucose medium samples.

2.2 Spectrograms of Multi-component Mixtures

In the previous section, we introduced the Raman spectrogram of single-component glucose medium, based on which we extended the original data and used a Raman spectroscopy detector to obtain multi-component Raman spectroscopy data. In this batch of data, besides containing glucose, bacterial substances were also added. As the experiment proceeded, the glucose concentration in the medium was gradually decreasing while the bacterial content was showing an increasing trend. The Raman spectrometer is real-time detection of component concentrations in the medium, generating seven spectral data per minute, and we selected the spectral data every 30 min and recorded the concentration of each component. As shown in Fig. 1(b), the spectra of the mixtures overlap less, indicating that the diversity of the components has a greater impact on the spectra and the multi-component data is more challenging for the model, while at the same time there may be interactions between the components, resulting in some noise characteristic peaks.

2.3 Auto-encoder

The encoder's job is to convert the high-dimensional input x into a low-dimensional implied variable, which allows the network to learn the most valuable characteristics out of the many available. As for the decoder, the role of the decoder is to reduce the implied variable *a to the* initial dimension, i.e., to obtain the reconstruction $\boldsymbol{x^R}$. A good self-encoder is one in which the output of the decoder is almost a complete approximation of the original input. A 3-layer stacked self-encoder is utilized in this study, as illustrated in Fig. 2, with 128, 64, and 128 neuron connections in the hidden layer, respectively.

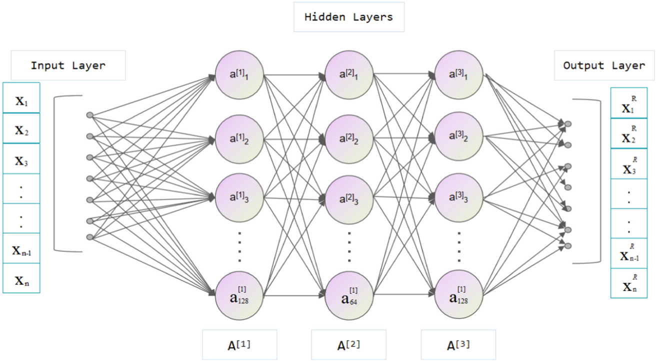


Fig. 2. The hidden layer is a 3-layer Auto-encoder model

The original data x is encoded from the input layer to the hidden layer during the encoding process.

$$a = \sigma(\mathrm{w}_1 x + b_1) \tag{1}$$

Decoding process: from the intermediate layer, i.e. the hidden layer, to the output layer.

$$x^R = \sigma(w_2 a + b_2)x \tag{2}$$

where W1, W2 is the weight parameters, b1, b2 are the bias terms, and **σ** is the activation function, here Relu is chosen as the activation function.

Optimization objective function.

$$MinimizeLoss = \frac{1}{N}\sum_{n=1}^{N}\left\|x - x^R\right\|^2 \tag{3}$$

Adding a nonlinear activation function to the encoded linear combination to reconstruct the input data using the new features obtained after encoding is a very effective and practical means of feature extraction.

2.4 Attention Mechanism

An attention mechanism is commonly known as a resource weighting mechanism, which reallocates resource weights based on the importance of data in different dimensions, and is centered on recalculating to highlight certain important features based on the correlation between the original data. Many kinds of attention mechanisms have emerged, such as the latest attention algorithm [19] and applying self-attention from natural language processing (NLP) [15] Applied to computer vision tasks, the feature map with attention is obtained by weighting and summing the values of the Query, Key vector after similarity calculation with the Value vector. Spatial attention [16] and channel attention [17]. The former retains the key spatial information of the original image while transforming it into another space to focus on the important regions, while the latter assigns weights to the image channel dimensions. Based on this, CBAM emerges [18] to obtain the attentional feature map by tying together the channel attention and spatial attention mechanisms. We provide an attention mechanism in this paper that is comparable to Senet [16], with the variations outlined in Sect. 3.

3 Algorithm Design

Algorithm 1 illustrates the flow of our algorithm design. The original data is the spectral data acquired every minute by the Raman spectroscopy detection probe instrument, and while saving the spectrogram, the concentration of single-component substances and the concentration of individual multi-component substances are recorded. Since our algorithm model cannot read the .spc file format, the .spc spectral file is processed by obtaining and storing the coordinates of each data point of the spectrogram in a two-dimensional array and adding the previously recorded substance concentration values as labels to the two-dimensional array where the spectral data are stored.

```
Algorithm 1: Spectrogram concentration prediction
Require: original Raman spectrogram of all concentrations.spc
        Convert the spectrogram in .spc format to a two-dimensional array $x \in X_i$
    $($i=1,2,3...)
        Original all spectrograms corresponding to concentrations $y \in Y_i$ $($i=1,2,3...)
        Encoder $D$ and decoder $E$
        Attention mechanism $AT$
        The neural network $f(x)$ used for regression prediction and the gradient up-
    date parameters $\theta$
For each epoch over $X_i$ do:
    $x_i \leftarrow X$
    $y_i \leftarrow Y$
    For each mini-batch do:
        $z_i \leftarrow E(x_i)$
        $x'_i \leftarrow D(z_i)$
        $y'_i \leftarrow f(AT(x'_i))$
    End
Loss $\leftarrow \frac{1}{\mathrm{m}}\sum_{i=1}^{m}(y_i - y'_i)^2$
$\theta \leftarrow \theta - \nabla\theta$ L
End
```

In the same way, this paper adds the attention mechanism in deep learning to the model inspired by Senet, which assigns weights to different features based on the channel dimension for the purpose of weight assignment, thus making the neural network pay more attention to the features with higher weights. The original Senet is used for image processing, where the image is usually composed of length, width, and channel. After several layers of convolution, the number of channels increases, and the size becomes smaller, at which point the Senet adaptively pools the length and width to 1 and only weights the channel dimension. Our data does not have the attributes of length and width, so the pooling process is omitted, the weights are directly weighted, and the obtained weights are spliced with the original data to get the weighted data, which will be beneficial for the subsequent prediction or classification tasks.

The next step is the construction of the algorithm model. In selecting the self-encoder, we choose the encoder and decoder based on the Dense layer, and the number of codec layers is all 3 layers, with 128, 64, and 64 neurons per layer, respectively. We also choose the Dense layer-based attention mechanism, with the number of layers set to 2. The model is then back-propagated to update the parameters, using the fully connected layer for prediction, and the predicted value is lost in mean square error with the real value.

4 Experiment

The experimental part of this paper will use multiple algorithms and deep learning methods to compare the results of the algorithmic models by selecting single- and multi-component Raman spectral data of glucose culture media, calculating the root mean square error of each algorithm, by analyzing the fitted curves of the true and predicted values, and by comparing the convergence speed of the neural network model with the fully connected layer and the convolutional layer neural network model.

4.1 Traditional Methods and Neural Networks

To reflect the advantages of the proposed method in this paper, firstly, some classical algorithms from machine learning methods were selected separately for comparison on the Raman spectral data obtained from a single-component glucose medium. As shown in Fig. 3, the results were obtained by CARS, PCA, and NCA [12] by reducing the dimensionality of Raman spectral data to 117, 68, and 49 dimensions, respectively. The overall trend of the basic fitted curves may be noticed is roughly the same despite the different dimensionality of feature extraction, except that there is a lack of inaccuracy, so the fitting effect of the algorithm needs to be further improved. Figure 4(a) shows the same single-component data as Fig. 3, and Fig. 5 shows the fitting results obtained by the self-coding-attention mechanism, and the results are marked enhancement over conventional algorithms.

From the results obtained by the above machine learning method, we added deep learning into it for improvement. Figure 4 shows the fitting curves of multi-component mixtures obtained by machine learning and deep learning algorithms, and it can be

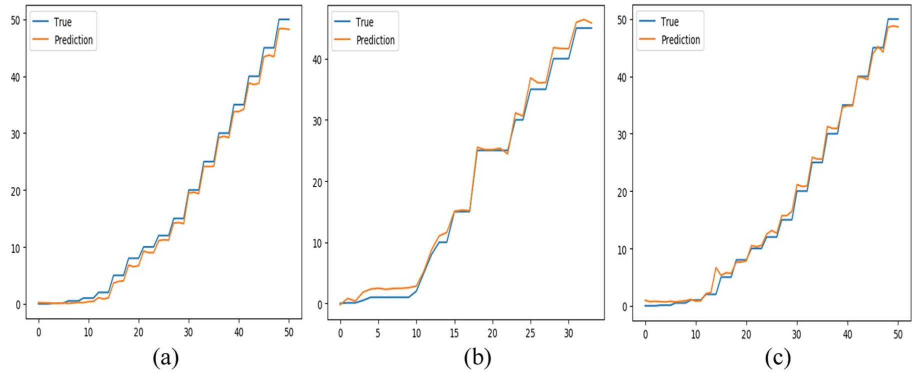


Fig. 3. (a)(b)(c) are the fitted curves of predicted and true values of CARS, PCA, and NCA in order (single component).

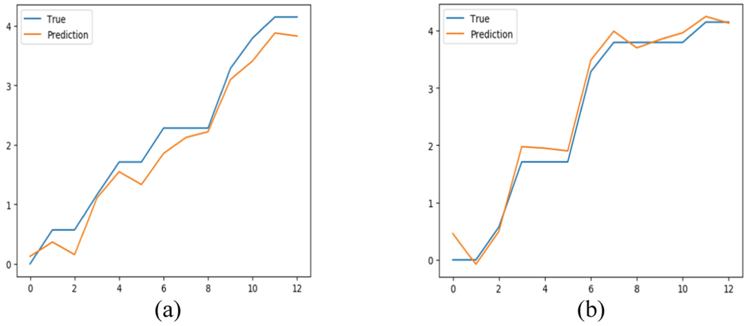
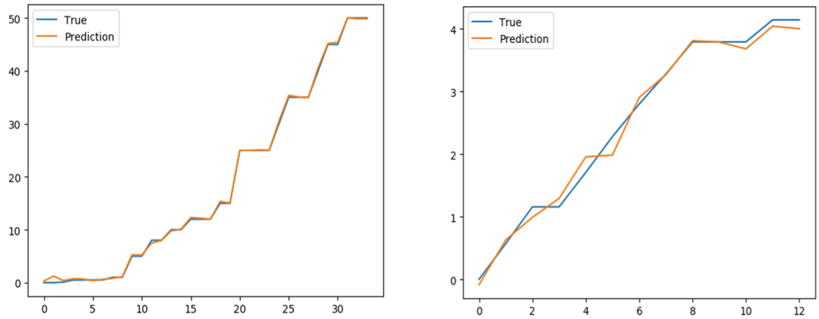


Fig. 4. Fitting curves (multicomponent mixtures) for PCA conventional method (a) and self-encoder (b).

seen that the fitting curves obtained by the deep learning self-encoder method in multi-component glucose medium are better and more accurate. In the experiments, we also found that the conventional PCA method is less stable, the fitting effect is sometimes good and bad, and it is not suitable to handle the real-time Raman spectroscopy detection task, while the deep learning method is not only better but also more stable.

4.2 Self-encoder and Attention Mechanism

In Sect. 4.2, we predicted the concentration of culture-based Raman spectral data, and the results proved that the self-encoder method using deep learning is better than machine learning and traditional neural network methods, but at the same time, it is easy to see that the results obtained by using the self-encoder alone are still lacking in fitting accuracy, so to further improve the fitting effect. In order to further improve the fit, we add the attention mechanism to the self-encoder model, and the selection of the attention mechanism is described in Sect. 3. To demonstrate the reliability and robustness of the model, we also compared the fitting curves of single- and multi-component glucose



(a) (b)

Fig. 5. Self-encoder-attention mechanism fitting curve, single-component (a) multi-component (b).

medium concentrations: Fig. 5 shows that the results of feature extraction based on the self-encoder with the attention mechanism are better than those of the traditional method mentioned above, and after several experiments, we observed that the fitting curves of our method are more stable and there is no model collapse. The predicted and true values are basically on the same curve. Figure 5(b) shows a significant improvement in the fitted curve using the self-encoder-attention mechanism compared to the experimental results in Fig. 4 and Fig. 5. In addition to this, it can be observed that although the fitted curve is intuitively better for the single component, the concentration values for the single component span from 0 to 50 units of concentration, while the multi-component is from 0 to 4 units of concentration.

For further illustration, we compared the results with the real sample concentrations by random sampling: as can be seen from Table 1, among the 10 different concentrations selected for comparison, the results of AE-Attention are the closest to the real values, with errors in the range of 0.3% to 1.7%, and similarly, for the six different concentration species in Table 2, the error can be as small as 0.01 concentration units, which satisfies the error in the practical, Therefore, the experiment has reliable practicality.

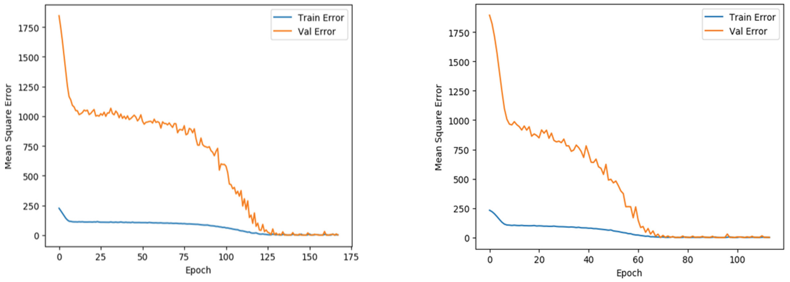
Table 1. Comparison of samples and predicted for different concentrations (single components)

Label	0	0.1	5	15	50
CARS [5]	0.35 ± 0.13	0.33 ± 0.18	4.25 ± 0.64	14.65 ± 1.28	49.24 ± 1.83
PCA [11]	0.7 ± 0.22	0.16 ± 0.09	4.45 ± 1.66	14.55 ± 1.87	48.47 ± 2.21
NCA [12]	1.02 ± 0.37	0.80 ± 0.23	5.48 ± 1.89	15.97 ± 1.65	49.26 ± 0.91
GBR [13]	0.30 ± 0.10	0.37 ± 0.18	5.82 ± 2.37	15.63 ± 2.67	48.93 ± 3.43
RFR [14]	0.18 ± 0.11	0.56 ± 0.35	5.30 ± 2.28	16.50 ± 3.11	49.63 ± 3.03
AE-Attention	**0.11 ± 0.05**	**0.08 ± 0.04**	**5.06 ± 0.53**	**15.34 ± 0.66**	**50.12 ± 0.38**

Table 2. Comparison of samples and predicted for different concentrations (multiple components)

Label	0.27	0.57	1.158	1.708	3.79
CARS [5]	0.22 ± 0.11	0.38 ± 0.14	1.30 ± 0.23	1.55 ± 0.61	3.67 ± 0.27
PCA [11]	0.35 ± 0.20	0.07 ± 0.58	0.87 ± 0.31	1.48 ± 0.27	3.57 ± 0.38
AE-Attention	**0.24 ± 0.08**	**0.62 ± 0.04**	**1.26 ± 0.13**	**1.71 ± 0.04**	**3.78 ± 0.02**

Comparing the convergence speed and the number of parameters and accuracy of fully connected layer neural network (NN) and convolutional neural network (CNN) in the prediction task: In the classification stage after extracting the features, we used CNN and NN for comparison experiments, Fig. 6 illustrates that the accuracy obtained by using N is higher than that obtained by using convolutional neural network (CNN) for classification, and the corresponding convergence speed is As the parameters of NN are more than those of CNN, the accuracy of NN is also a little higher than that of CNN, and in the context of accuracy, we prefer to use NN to complete the task.



(a)

(b)

Fig. 6. Comparison of convergence speed of NN (left panel) and CNN (right panel).

The root means square error (MSE) is a measure that responds to the degree of difference between the predicted and true values, and the actual effect of the model can be visualized by calculating the root mean square error of different algorithms. As shown in Table 3, comparing the five algorithms for extracting features and the use of fully connected layer neural networks (NN) and convolutional neural networks (CNN) for comparison in prediction, it can be seen that the method based on the self-encoder + attention mechanism is the best in both NN and CNN conditions. The same Table 4 for the multicomponent MSE error yields the same experimental results as Table 3. At the same time, also from Table 5, the results obtained without dimensionality reduction feature extraction are poor because the presence of many noisy and useless features has an impact on the results.

Table 3. Root mean square error values of different algorithms for a single component

MSE	CARS [5]	PCA [11]	NCA [12]	Select-Percentile	AE-Attention
NN	0.272 ± 0.12	0.67 ± 0.14	0.539 ± 0.22	4.464 ± 2.43	**0.19 ± 0.08**
CNN	0.4 ± 0.18	0.76 ± 0.26	1.609 ± 0.89	3.55 ± 2.02	**0.18 ± 0.072**

Table 4. Root mean square error values of different algorithms for multi-component

MSE error	PCA [11]	NCA [12]	AE-Attention
NN	0.060 ± 0.021	0.053 ± 0.014	**0.019 ± 0.008**
CNN	0.065 ± 0.014	0.058 ± 0.011	**0.026 ± 0.010**

Table 5. MSE metrics for GBR and RGR

	GBR [13]	RFR [14]
MSE error	0.642 ± 0.38	0.433 ± 0.27

5 Results and Discussion

The deep learning approach in this paper is done in the TensorFlow framework based on Python, using an Intel(R) Core(TM) i5-8500 CPU.

In this work, we propose a Raman spectral processing algorithm based on a self-encoder-attention mechanism, which transforms some methods of total deep learning image processing applied to regression modeling in several different concentrations of culture media, comparing with several commonly used traditional algorithms, and applying deep learning methods in which the consumption of substances in glucose medium can be observed in time in biological fermentation experiments. In order to facilitate timely replenishment and recording of data, yielding practically meaningful results. The method can next be used to predict or classify Raman spectrograms of more complex multi-component mixtures or other substances in the medium, which has good research value.

References

1. Zheng, Y., Zhang, T., Zhang, J., et al.: Study on the effects of smoothing, derivative, and baseline correction on the quantitative analysis of PLS in near-infrared spectroscopy. Spectrosc. Spectral Anal. **2004**(12), 1546–1548 (2004)
2. Martens, H., Naes, T.: Multivariate Calibration. Wiley, Hoboken (1992)
3. Norgaard, L., Saudland, A., Wagner, J., et al.: Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. Appl. Spectrosc. **54**(3), 413–419 (2000)

4. Chen, H., Tao, P., Chen, J., et al.: Waveband selection for NIR spectroscopy analysis of soil organic matter based on SG smoothing and MWPLS methods. Chemometr. Intell. Lab. Syst. **107**(1), 139–146 (2011)
5. Li, H., Liang, Y., Xu, Q., et al.: Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. Anal. Chim. Acta **648**(1), 77–84 (2009)
6. Polanski, J., Gieleciak, R.: The comparative molecular surface analysis (CoMSA) with modified uniformative variable elimination-PLS (UVE-PLS) method: application to the steroids binding the aromatase enzyme. ChemInform **34**(22), 656–666 (2003)
7. Technicolor, T., Related, S., Technicolor, T., et al.: ImageNet classification with deep convolutional neural networks [50] (2012)
8. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. Computer Science (2014)
9. He, K., Zhang, X., Ren, S., Sun, J.: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)
10. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. Science **313** (2006)
11. Wold, S., Esbensen, K., Geladi, P.: Principal component analysis. Chemometr. Intell. Lab. Syst. **2**(1–3), 37–52 (1987)
12. Roweis, S.: Neighborhood component analysis (2006)
13. Wang, J., Peng, L., Ran, R., et al.: A short-term photovoltaic power prediction model based on the gradient boost decision tree. Appl. Sci. **8**(5), 689 (2018)
14. Breiman, L.: Random forest. Mach. Learn. **45**, 5–32 (2001)
15. Vaswani, A., Shazeer, N., Parmar, N., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, pp. 5998–6008 (2017)
16. Jaderberg, M., Simonyan, K., Zisserman, A.: Spatial transformer networks. Adv. Neural Inf. Process. Syst. **28**, 2017–2025 (2015)
17. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
18. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: CBAM: convolutional block attention module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 3–19. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_1
19. Guo, M.H., Xu, T.X., Liu, J.J., et al.: Attention mechanisms in computer vision: a survey. arXiv preprint arXiv:2111.07624 (2021)