Check for updates

Water chemical oxygen demand prediction model based on the CNN and ultraviolet-visible spectroscopy

Binqiang Ye[1,2], Xuejie Cao[1], Hong Liu[1], Yong Wang[1], Bin Tang*[,1], Changhong Chen[1] and Qing Chen[1]

[1]School of Artificial Intelligence, Chongqing University of Technology, Chongqing, China, [2]School of Microelectronics and Communication Engineering, Chongqing University, Chongqing, China

Excessive levels of organic matter in water threaten ecological safety and endanger human health. As the water resource environment is deteriorating, accurate and rapid determination of water quality parameters has become a current research hotspot. In recent years, the ultraviolet spectrometry method has been more widely used in the detection of chemical oxygen demand (COD), which is convenient and without chemical reagents. However, this method tends to use absorbance at 254 nm to measure COD. It has a good detection effect when the composition of pollutants is single, but in real life, the complex composition of pollutants will seriously affect the accuracy of measurement. Therefore, a COD prediction model based on ultraviolet-visible (UV-Vis) spectrometry and the convolutional neural network (CNN) is proposed. Compared with other traditional COD prediction models, this model makes full use of the absorbance of all ultraviolet and visible wavelengths, avoiding the information loss caused by using specific wavelengths. Meanwhile, this model is constructed based on the shallow CNN, using convolutional layers with different step lengths instead of the traditional pooling layers, which reduces computation and enhances the capture of spectral feature peaks. Additionally, with the powerful feature extraction capability of the CNN, this model reduces the reliance on pre-processing methods and improves the utilization of spectral information. Experiments have shown that our model has better fitting results and accuracy than other traditional COD prediction models such as the principal component analysis (PCA), partial least squares regression (PLSR), and backpropagation (BP) neural network. This study provides a better solution for improving the accuracy of UV-Vis water quality COD detection, which is conducive to real-time monitoring of the water quality, providing data support of water pollution and its development trend for the government's water resource protection policy and promoting biodiversity development.

KEYWORDS

COD, UV-Vis spectroscopy, water quality assessment, CNN, machine learning

Introduction

Water quality is vital to people because it is directly related to the quality of human life and ecological security. Due to the industrial emission of wastewater and the massive use of pesticides and fertilizers, water eutrophication is becoming a serious issue, leading to the decline of dissolved oxygen in the water and the imbalance of ecosystem species distribution, bringing challenges to water resource protection. Achieving real-time monitoring of environmental quality is significant for promoting species conservation and economic growth and development (Bhatti et al., 2022a). Therefore, in order to effectively monitor the water quality and take corresponding measures in time, the optimization of the water quality detection algorithm is imperative.

COD refers to the mass concentration of oxygen in the strong oxidant consumed by dissolved substances and suspended matter in water. As one of the significant parameters to evaluate the pollution level of water bodies, COD is widely used to detect water pollution. The higher COD value indicates a higher concentration of reducing substances in water and the more severe is the water pollution. Currently, the mainstream methods for determining COD have been divided into chemical and physical methods (Li et al., 2018). Chemical methods are mainly titrimetric analysis and electrochemical analysis. The process is complex, slow, and generally also needs to be carried out in the laboratory, which makes it difficult to achieve online detection. The primary physical method is molecular spectroscopy. Spectroscopy is a method that identifies substances and quantifies them by their emission or absorption spectra (Barone et al., 2021). Among them, UV-Vis spectrometry is widely used in the field of water quality monitoring as an environmentally friendly monitoring and analysis method because of its rapid determination and convenience, low cost, high sensitivity, and the possibility of online tracking (Chen et al., 2021). Improving the efficiency of environmental quality monitoring will facilitate the government in obtaining the distribution of pollution, tracking and removing pollutants, and formulating environmental policies timely (Bhatti et al., 2022b). Many scholars have carried out research on spectroscopic methods and proposed various optimization algorithms such as competitive adaptive reweighted sampling (CARS), the successive projection algorithm (SPA), and particle swarm optimization (PSO) to fuse into a new algorithm to select the effective feature wavelength, which improves the accuracy for the following prediction model (Hong-Qiu et al., 2019; Jahandideh-Tehrani et al., 2020; Zhang et al., 2020). Zhao et al. (2016) used the PCA algorithm to improve the efficiency of water quality detection. Bhatti et al. (2021) used Gabor filtering on hyperspectral images to improve the classification accuracy of the images. Mingjin et al. (2019) built a correction model based on PLSR for the main spectral regions of the UV receiver spectrum for the determination of COD in water samples and obtained a good coefficient of determination. Zhu et al. (2022) developed a BP neural network based on elephant herding optimization for COD prediction, which effectively removed the biased data and improved the ability of the algorithm to find the optimal value.

However, the current COD detection algorithms are still inadequate; PLSR has less deviation information on independent variables and larger matching errors, so its prediction accuracy still needs to be improved. The PCA calculation process is complicated and vulnerable to the complexity of organic pollutants, so it is difficult to predict the dynamically changing organic pollutants. BP networks have limitations such as slow convergence speed and easy falling into local extremes. The CNN, a common type of the deep learning model, has better self-learning and self-adaptive capabilities than other traditional prediction models and can better deal with nonlinear problems (Croce et al., 2018). Meanwhile, the CNN can provide better prediction accuracy even without spectral preprocessing due to its powerful feature extraction capability and sample mapping of local features (Zhao et al., 2018).

This article constructs a COD prediction model based on the CNN and UV-Vis spectroscopy to improve the accuracy, inputs the spectral data into the CNN, extracts the high-dimensional information in the spectral information by multi-layer convolution and activation operations, then reduces the dimensionality of the feature map by downsampling, and finally outputs the predicted COD value. The feature extractor used in this algorithm uses a convolutional layer with a step size of 2 instead of the pooling layer of the traditional CNN. It uses convolutional kernels and step size to smooth the one-dimensional spectral data, making it easier for the model to extract more important spectral region information and thus reducing the COD prediction error. Experiments show that this algorithm can effectively predict COD with high accuracy. It has a smaller error than other COD prediction models and provides a better solution to improve the accuracy of UV-Vis water quality COD detection, which is conducive to the real-time online detection of water resources and provides data support for the protection of water resources and the formulation of environmental policies.

Related work

Principle of chemical oxygen demand detection by ultraviolet-visible spectrometry

UV-Vis spectroscopy is a commonly used spectroscopic analysis method based on absorption spectra generated by electron leaps within molecules. Most of its research objects are in the near UV range of 200–380 nm and the visible range of 380–780 nm. The UV-Vis absorption spectrum corresponds to a short electromagnetic wavelength and high energy, reflecting the

FIGURE 1
COD detection principle using the UV-Vis method.
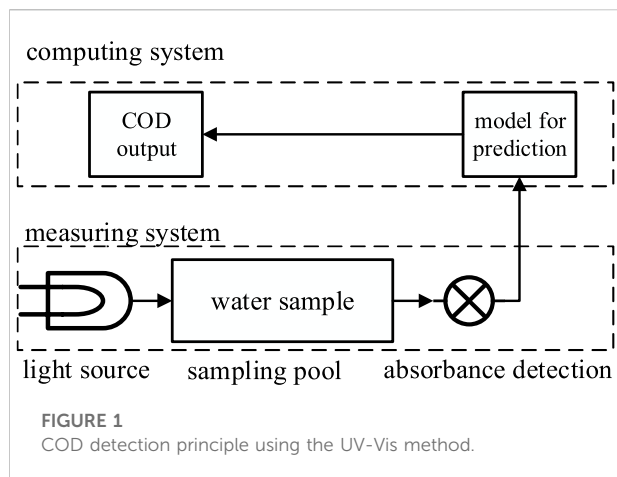
valence electron energy leap in the molecule, and the sensitivity of its determination depends on the molar absorption coefficient of the molecule producing light absorption. The absorption of a molecule in the UV-Vis region is closely related to its electronic structure, and molecules with different structures produce electron jumps of different energies, which are reflected in the UV-Vis absorption spectra and result in specific characteristic peaks, and the structural information on the sample to be measured can be deduced from the position and intensity of the characteristic peaks (Li and Hur, 2017). This principle allows UV-visible spectroscopy to detect COD in water. The detection principle is shown in Figure 1.

The detection system consists of two parts: a spectrophotometer and a computer system. The spectrophotometer inputs the absorbance data on the organic pollutants to be measured in the water sample into the computer by checking the spectral changes of the light source after passing through the sample solution, and a computer system processes the absorbance data through the COD prediction model to finally predict the COD value of the sample.

The COD prediction model is extremely important for the accuracy of the prediction. With further research, more innovative data pre-processing algorithms have been proposed to improve the validity of the data, and a variety of efficient water quality prediction models have been proposed to improve the model's accuracy. These research studies have provided a theoretical foundation for the efficient and accurate detection of the water quality (Guang et al., 2019; Passos and Saraiva, 2019; Sun et al., 2021).

Principle of the convolutional neural network

A typical CNN consists of three parts: convolutional layers, pooling layers, and fully connected layers. Its structure is shown in Figure 2.

Here, the neurons in the convolutional layer are locally connected to their feature surface in the input layer (Baltrušaitis et al., 2018). This locally weighted sum is passed to the activation function to obtain the output value of each neuron in the convolutional layer. The pooling layer reduces the number of connections between the convolutional layers, decreasing the dimensionality of the feature map and the computational complexity of the model. The fully connected layer can integrate the local information with category differentiation in the convolutional and sampling layers (Basha et al., 2020). The output values of the last fully connected layer are passed to the output layer, and finally, classification is achieved by softmax regression.

The CNN can fit multidimensional mapping problems, and the neurons in a multilayer feature extractor can provide enough complexity to simulate the nonlinear nature of the mission. The local connectivity, weight sharing, and pooling operation features of the CNN can reduce the number of training parameters and effectively reduce the complexity of the network while making the model invariant to translation, distortion, and scaling to a certain degree, improving robustness and fault tolerance. Based on these superior properties, it performs better than standard fully connected neural networks in a variety of signal and information processing tasks which have achieved good results in the fields of computer vision, natural language processing, medicine and health, and environmental protection (Bhatti et al., 2019; Olmedilla et al., 2022; Serna et al., 2022).

The proposed method

Data building

Due to the huge variation between water quality samples from different regions and practical water samples that cannot fully cover all valid points in the range of concentrations, the model uses a range of standard COD solutions of 50–500 mg/L, according to the national standard, and then obtains the spectral data on the samples at 200–900 nm through the COD measurement system.

The COD standard solution was prepared according to the national standard of China (GB19914-89) by accurately weighing 1.2754 g of pre-dried, high-purity potassium hydrogen phthalate (HOOCC6H4COOK) in heavy distilled water using a BS243S electronic balance, transferring it to a 1000-ml volumetric flask, and diluting it to the standard line with heavy distilled water to obtain a COD value of 1,500 mg/L of the solution. Then, the COD standard solution was diluted proportionally to obtain a total of 90 samples of 500–50 mg/L of the COD solution decreasing in a 5 mg/L gradient.

The core of the water quality measurement system is the Maya 2000 Pro spectrometer manufactured by Ocean Optics, which is composed of a deuterium–halogen light source,

FIGURE 2
Typical structure of the convolutional neural network.

FIGURE 3
Water quality COD measurement system.
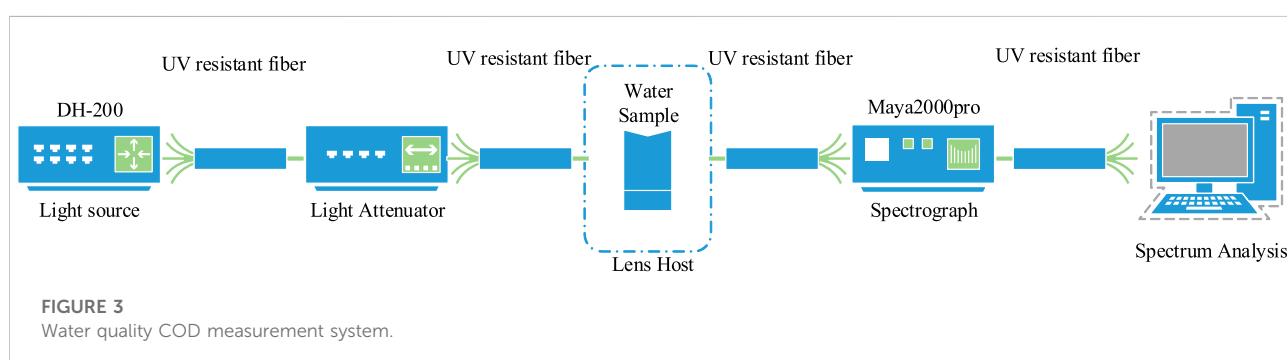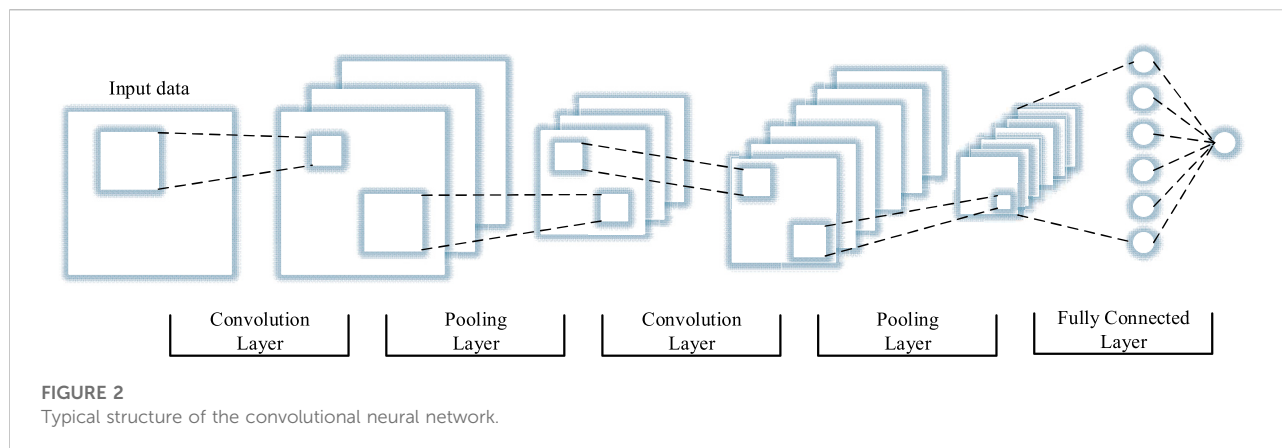
attenuator, sample holder, spectrometer, and PC with acquisition control software being installed; the specific principle is shown in the Figure 3.

To avoid the influence of ambient light on the test results and to ensure the consistency of the acquisition environment, a special sample holder was used to avoid light throughout the measurement. The integration time was set to 80 ms and the average number of scans to 10; each sample collected previously was tested five times, and the spectral data on 200 nm–1000 nm of 450 standard COD solutions were obtained. The spectral data on the samples are shown in Figure 4.

Network model building

Generally speaking, convolutional neural networks can achieve feature extraction of the original data through convolution, activation, and pooling operations, extracting and obtaining important spectral region information, weighing and summing the output of the features from the previous layer through fully connected layers, inputting the results to the activation function, and finally completing the classification of the target (Baltrušaitis et al., 2018). However, considering that the spectral information obtained through

experiments is limited, the training of convolutional neural networks usually requires a large amount of data, while deep neural networks are often prone to overfitting (Roelofs et al., 2019). This article adopts a shallow 1D CNN to construct the COD prediction model, and the specific structure of this model is shown in Figure 5.

This model extracts spectral data features through five feature extractors in series and then implements the prediction of COD through fully connected and activation layers. The feature extractor uses the structure of convolutional, activation, and downsampling layers in series.

In each convolutional layer, the convolutional kernel size is 9×1, 7×1, 5×1, 3×1, and 3×1. The convolutional kernel convolutes the feature map output from the previous layer and constructs the output feature map using the nonlinear activation function. The output of each layer is the result of convolutions of multiple input features, and the convolution process is shown in Eq. 1

$$y^j = f\left(\sum_{i=1}^{n} w^{ij} * x^i + b^i\right) \tag{1}$$

where * denotes the convolution operation, y^j denotes the layer j output feature map, x^i denotes the layer i input feature map, w^{ij}
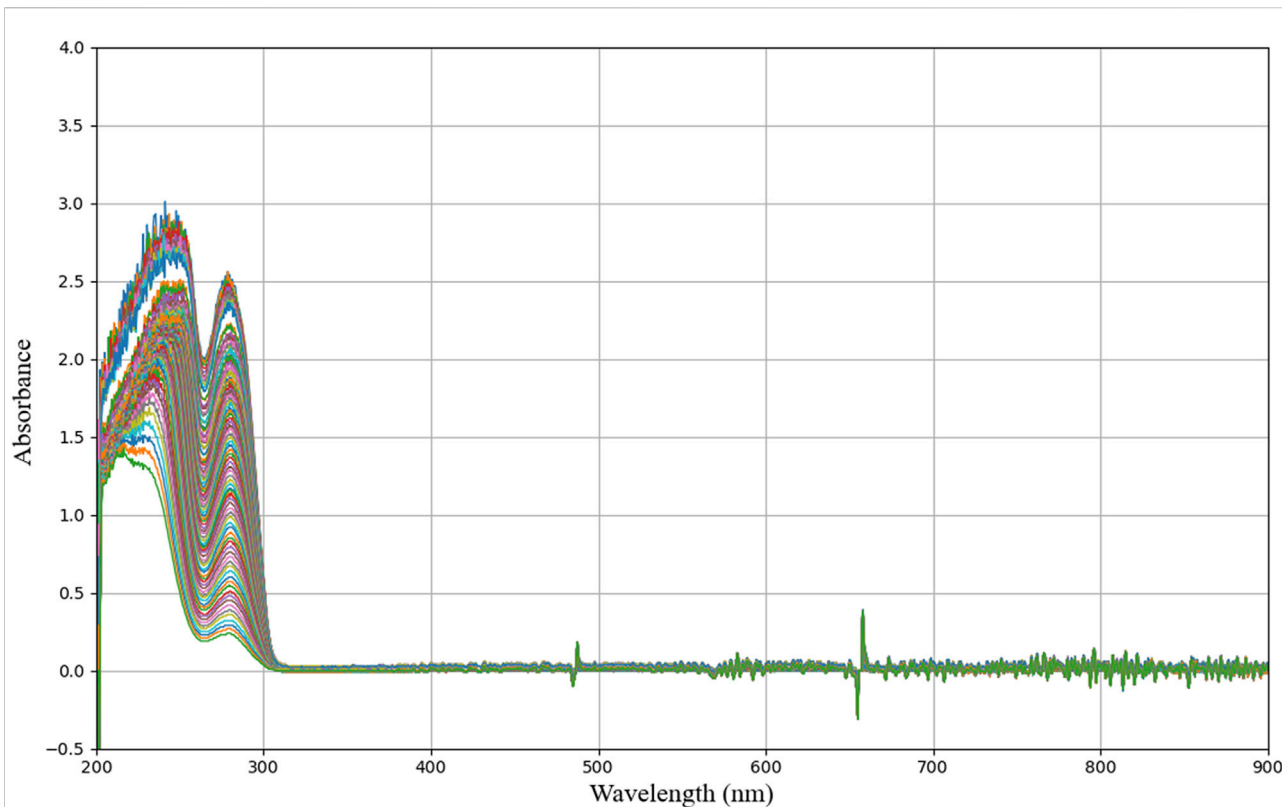
FIGURE 4
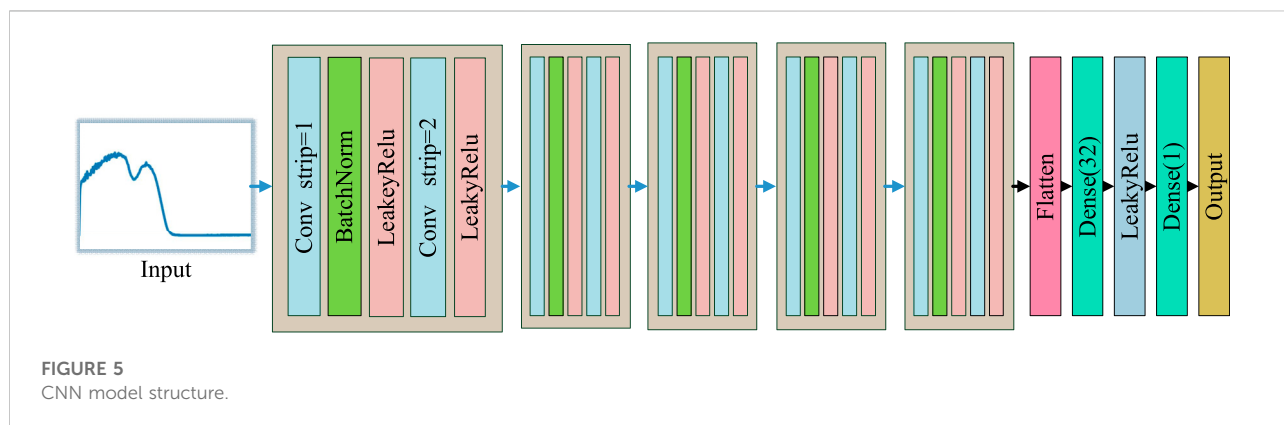UV-visible spectra of the samples.

FIGURE 5
CNN model structure.

denotes the convolution kernel in this layer, and b^i denotes the bias of the layer i feature map.

The activation layer is located after the convolution layer. Leaky ReLU is a variant of the ReLU activation function, the output of which has a small slope for negative inputs and reduces the appearance of silent neurons due to the non-zero derivative, which solves the problem of neurons not learning after ReLU enters the negative regions. Its expression is shown in Eq. 2

$$LeakyReLU(x) = \begin{cases} x & (x > 0) \\ \alpha x & (x \le 0) \end{cases} \quad (2)$$

The downsampling layer is located after the activation layer; since ResNet, many scholars have gradually used a convolutional layer with a step size of 2 instead of a pooling layer with a size of 2. Both can achieve the operation of downsampling the feature map, but for one-dimensional spectral data, instead of using a

FIGURE 6
Flowchart of the study.
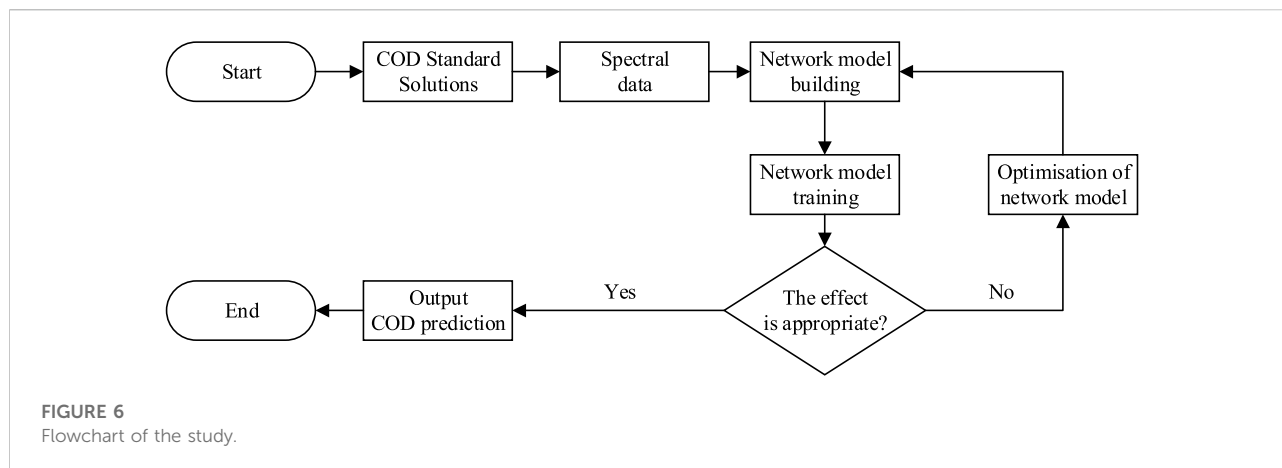
pooling layer, smoothing the data using a convolutional kernel and a step size can make the model easier to use for identifying important spectral regions (Acquarelli et al., 2017). Therefore, in this article, we use a convolutional layer with a step size of 2 and an activation layer instead of the traditional pooling layer to achieve downsampling.

The training and optimization of the convolutional neural network rely on the loss function, which calculates the error between the predicted value and the true value, back-propagates the error from the last layer to each layer of the network, and updates the weights by a back-propagation algorithm. The updated parameters continue to participate in the training, and the cycle repeats until the loss function value is minimized. In this article, our model uses the mean square error as the loss function, and the expression is shown in Eq. 3:

$$L_{mse} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \quad (3)$$

where Y_i denotes the theoretical COD value of sample i and $\hat{Y}_i$ denotes the predicted COD value of sample i.

Network model training

The experimentally obtained spectral sample data are divided into training and test sets by 4:1 while ensuring that the training and test sets are equally distributed. The network learning optimizer used the Adam optimizer, the initial value of the learning rate was set to 0.001, and the learning rate was reduced to the original 0.95 after each training epoch. The batch size was set to 32, and the network model is trained on a TensorFlow 2.3 platform with a GeForce GTX 2080 GPU computer until the loss of the network model on the test set no longer decreased. The flow chart for this study is shown in Figure 6.

Experiments and analysis

Evaluation indicators

To objectively evaluate the COD prediction effect of our model, the coefficient of determination R^2, root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE) are adopted as the performance evaluation indexes of the model prediction accuracy. Among them, R^2 illustrates the proportion of the variance that can be interpreted by the independent variables and is used to evaluate the fit of the model; the closer its value to 1 means the better the model fits. RMSE indicates the error between the predicted and theoretical values, which is more sensitive to outliers. MAPE indicates the relative error between the average predicted and theoretical values on the experimental data set. RMSE, MAE, and MAPE all reflect the error between the theoretical and predicted values. The smaller these values are, the better is the prediction accuracy of models. The calculation formulas of each evaluation index are as follows:

$$R^2 = \left[1 - \frac{\sum_{i=1}^{n}\left(Y_r(i) - Y_p(i)\right)^2}{\sum_{i=1}^{n}\left(Y_r(i) - \overline{Y_p}\right)^2}\right] \times 100\% \quad (4)$$

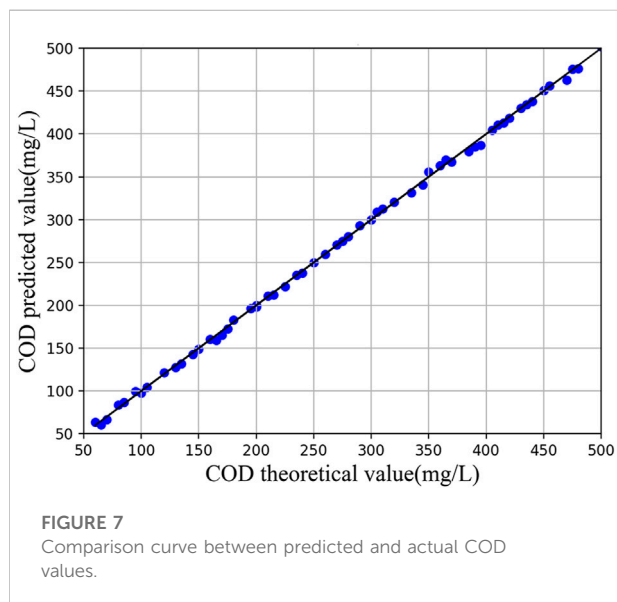$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_r(i) - Y_p(i)\right)^2} \quad (5)$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{Y_r(i) - Y_p(i)}{Y_r(i)}\right| \quad (6)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_r(i) - Y_p(i)\right| \quad (7)$$

where $\overline{Y}_r = \frac{1}{n}\sum_{i=1}^{n}Y_r(i)$, $Y_r(i)$ denotes the theoretical COD value of sample i, and $Y_p(i)$ denotes the predicted COD value of sample i.

TABLE 1 Prediction results of our network model.

Statistics	Concentration of COD (mg/L)									
Theoretical value	175.0	240.0	160.0	410.0	335.0	385.0	60.0	165.0	275.0	405.0
Predicted value	172.7	237.7	160.4	410.8	333.6	385.1	60.8	159.0	274.8	404.7

FIGURE 7
Comparison curve between predicted and actual COD values.

TABLE 2 Evaluation of the model prediction effect.

Method	R^2	RMSE	MAPE	MAE
BP	0.991	4.382	0.063	3.158
PCA	0.989	4.653	0.056	4.227
PLSR	0.985	5.542	0.075	5.650
OURS	0.996	3.899	0.042	2.154

Model performance analysis

After completing the training of the COD prediction network model through the aforementioned experiments, the spectral data on 90 test sets were input to the network to get the COD prediction of the model. After statistics, the average error of the prediction of the COD concentration from the test set data on the model reached 0.9168%. Part of the prediction results is shown in Table 1.

Based on the results of the test set, a comparison curve between the predicted and actual COD values of the model was plotted, as shown in Figure 7.

The comparison curve between the predicted and actual COD values shows that the model has a small error for the prediction of COD. Meanwhile, as the COD concentration of the water sample increases, there is no obvious trend of error growth in this model, which indicates that the model has a uniform error distribution and has high prediction accuracy and precision in the range of 50–500 mg/L. The experiments show that our algorithm effectively utilizes the spectral information on all wavelengths in the UV-visible spectrum.

In order to compare the prediction accuracy and the degree of fit of this algorithm with other COD prediction

models, this experiment compared the COD prediction results of three prediction algorithms, BP (Zhu et al., 2022), PAC (Zhao et al., 2016), and PLSR (Mingjin et al., 2019), on 90 samples from the test set. The coefficient of determination R^2 was used to evaluate the degree of fit of the COD prediction model. RMSE, MAPE, and MAE were used to evaluate the model accuracy. The mean values were obtained after five repeated experimental tests. The results are shown in Table 2.

Table 2 shows that our model has a better fit than PLSR, BP, and PCA using the same dataset, with a goodness of fit R^2 of 0.996, which indicates that the feature extraction ability and local feature mapping of this model are stronger and can better fit the non-linear relationship between the spectral data and COD. The PLSR has a lower fit of 0.985, which indicates that the PLSR cannot fully fit the non-linear relationship between the spectral information and the COD. In terms of model accuracy, the RMSE, MAPE, and MAE of our model are all superior compared to the other three methods, with the RMSE reaching 3.899, the MAPE reaching 0.042, and the MAE reaching 2.154, which demonstrates the high accuracy of the model for COD prediction. The RMSE of the PCA is larger, reaching 4.653, indicating that the PCA has a larger prediction error for individual deviation points in the spectral data, and its robustness needs to be enhanced. The large MAPE of the BP algorithm is due to the large relative error it produces, showing that the tendency of BP to fall into local extremes may affect the overall prediction effect of BP. Overall, the algorithm shows good results in terms of both the degree of fit and prediction accuracy, which demonstrates that our method can make full use of the information in the spectrum and fit the non-linear relationship between the spectral information and the COD values to achieve accurate prediction.

Conclusion

In this article, a method for water COD prediction using the CNN and UV-visible spectroscopy is proposed. The COD prediction value can be acquired by feeding the one-dimensional spectral data on water into the model. To avoid overfitting of the model, it uses a shallow CNN to build the backbone. To make the model extract more representative spectral feature peaks and improve the prediction accuracy, this model adopts the convolutional kernel and step size to smooth the one-dimensional spectral data instead of the pooling layer of the traditional CNN, which reduces the complexity of the model while also enhancing the extraction of information. With the powerful feature extraction capability of the CNN, this model reduces the dependence of traditional COD models on pre-treatment methods. The experiment indicates that our model has a better fitting effect and higher prediction accuracy, which provides a better solution to realize the fast detection of COD. This method can be implemented for the real-time measurement of the water quality, which can help the government to grasp the pollution situation to make environmental policies and take further measures to protect the ecological environment. However, the robustness of the model could be further improved due to the discrepancy between the experimentally obtained water samples and the actual water samples from different regions. In the future, the robustness of the model can be further optimized by collecting data from actual water samples and compressing the size of the model to adapt it to embedded devices to improve the convenience of real-time monitoring devices.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://docs.google.com/spreadsheets/d/1_CE-7XD_Fo5MTnwn8llEt8G9xZbOXH3m/edit#gid=2029270470.

References

Acquarelli, J., van Laarhoven, T., Gerretzen, J., Tran, T. N., Buydens, L. M., and Marchiori, E. (2017). Convolutional neural networks for vibrational spectroscopic data analysis. *Anal. Chim. acta* 954, 22–31. doi:10.1016/j.aca.2016.12.010

Baltrušaitis, T., Ahuja, C., and Morency, L. P. (2018). Multimodal machine learning: A survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* 41 (2), 423–443. doi:10.1109/tpami.2018.2798607

Barone, V., Alessandrini, S., Biczysko, M., Cheeseman, J. R., Clary, D. C., McCoy, A. B., et al. (2021). Computational molecular spectroscopy. *Nat. Rev. Methods Prim.* 1 (1), 38–27. doi:10.1038/s43586-021-00034-1

Basha, S. H. S., Dubey, S. R., Pulabaigari, V., and Mukherjee, S. (2020). Impact of fully connected layers on performance of convolutional neural networks for image classification. *Neurocomputing* 378, 112–119. doi:10.1016/j.neucom.2019.10.008

Author contributions

BY guided the experiments and model construction; XC conducted the experiments of water quality, COD data acquisition, and model construction; HL optimized the model structure and experimental process; YW conducted the training and tuning of the CNN model; BT conducted the comparison between this model and other similar models; CC conducted the processing of experimental data; QC wrote the manuscript.

Funding

Conflict of interest

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

Publisher's note

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

Bhatti, U. A., Huang, M., Wu, D., Zhang, Y., Mehmood, A., and Han, H. (2019). Recommendation system using feature extraction and pattern recognition in clinical care systems. *Enterp. Inf. Syst.* 13 (3), 329–351. doi:10.1080/17517575.2018.1557256

Bhatti, U. A., Wu, G., Bazai, S. U., Ali Nawaz, S., Baryalai, M., Bhatti, M. A., et al. (2022). A pre- to post-COVID-19 change of air QualityPatterns in anhui province using path analysisand regression. *Pol. J. Environ. Stud.* 31 (5), 4029–4042. doi:10.15244/pjoes/148065

Bhatti, U. A., Yu, Z., Chanussot, J., Zeeshan, Z., Yuan, L., Luo, W., et al. (2021). Local similarity-based spatial–spectral fusion hyperspectral image classification with deep CNN and gabor filtering. *IEEE Trans. Geosci. Remote Sens.* 60, 1–15. doi:10.1109/tgrs.2021.3090410

Bhatti, U. A., Yu, Z., Hasnain, A., Nawaz, S. A., Yuan, L., Wen, L., et al. (2022). Evaluating the impact of roads on the diversity pattern and density of trees to

improve the conservation of species. *Environ. Sci. Pollut. Res.* 29 (10), 14780–14790. doi:10.1007/s11356-021-16627-y

Chen, X., Yin, G., Zhao, N., Gan, T., Yang, R., Xia, M., et al. (2021). Simultaneous determination of nitrate, chemical oxygen demand and turbidity in water based on UV–Vis absorption spectrometry combined with interval analysis. *Spectrochimica Acta Part A Mol. Biomol. Spectrosc.* 244, 118827. doi:10.1016/j.saa.2020.118827

Croce, D., Rossini, D., and Basili, R. (2018). "Explaining non-linear classifier decisions within kernel-based deep architectures," in Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, Brussels, Belgium, November 2018 (IEEE), 16–24.

Guang, H., Tong, B., and Li, L (2019). "Chemical oxygen demand soft-measurement method via long short-term memory network," in 2019 Chinese Automation Congress (CAC), Beijing, China, November 2019 (IEEE), 4668–4672. doi:10.1109/CAC48633.2019

Hong-Qiu, Z., Tao, Z., and Yong-Gang, L (2019). An ultraviolet-visible absorption spectrometric method for detection of zinc (II) and cobalt (II) ions concentration based on boosting modeling. *Chin. J. Anal. Chem.* 47 (4), 576–582. doi:10.19756/j.issn.0253-3820.181650

Jahandideh-Tehrani, M., Bozorg-Haddad, O., and Loáiciga, H. A. (2020). Application of particle swarm optimization to water management: An introduction and overview. *Environ. Monit. Assess.* 192 (5), 281–318. doi:10.1007/s10661-020-8228-z

Li, J., Luo, G., He, L. J., Xu, J., and Lyu, J. (2018). Analytical approaches for determining chemical oxygen demand in water bodies: A review. *Crit. Rev. Anal. Chem.* 48 (1), 47–65. doi:10.1080/10408347.2017.1370670

Li, P., and Hur, J. (2017). Utilization of UV-vis spectroscopy and related data analyses for dissolved organic matter (dom) studies: A review. *Crit. Rev. Environ. Sci. Technol.* 47 (3), 131–154. doi:10.1080/10643389.2017.1309186

Mingjin, Z., Li, Y, and Hui, W. (2019). Ultraviolet spectrometry combined with chemometrics used for determination of COD in water samples[J]. *Chin. J. Analysis Laboratory* 38 (12), 1444–1448. doi:10.13595/j.cnki.issn1000-0720.2019.012103

Olmedilla, M., Martínez-Torres, M. R., and Toral, S. (2022). Prediction and modelling online reviews helpfulness using 1D Convolutional Neural Networks. *Expert Syst. Appl.* 198, 116787. doi:10.1016/j.eswa.2022.116787

Passos, M. L. C., and Saraiva, M. L. M. F. S. (2019). Detection in UV-visible spectrophotometry: Detectors, detection systems, and detection strategies. *Measurement* 135, 896–904. doi:10.1016/j.measurement.2018.12.045

Roelofs, R., Shankar, V., and Recht, B. (2019). A meta-analysis of overfitting in machine learning. Proceedings of the 33rd international conference on neural information processing systems. Red Hook, NY: Curran Associates Inc. 32, 11. doi:10.5555/3454287.3455110

Serna, I., Morales, A., Fierrez, J., and Obradovich, N. (2022). Sensitive loss: Improving accuracy and fairness of face representations with discrimination-aware deep learning. *Artif. Intell.* 305, 103682. doi:10.1016/j.artint.2022. 103682

Sun, Y., Brockhauser, S., and Hegedűs, P. (2021). Comparing end-to-end machine learning methods for spectra classification. *Appl. Sci.* 11 (23), 11520. doi:10.3390/app112311520

Zhang, L., Sun, H., Rao, Z., and Ji, H. (2020). Hyperspectral imaging technology combined with deep forest model to identify frost-damaged rice seeds. *Spectrochimica Acta Part A Mol. Biomol. Spectrosc.* 229, 117973. doi:10.1016/j.saa.2019.117973

Zhao, H., Liu, F., Li, L., and Luo, C. (2018). A novel softplus linear unit for deep convolutional neural networks. *Appl. Intell. (Dordr).* 48 (7), 1707–1720. doi:10.1007/s10489-017-1028-7

Zhao, Y. Q., Li, X., Liu, X., Dong, P. F., Wang, L. L., and Wang, X. Q. (2016). [Research on water quality analysis model with PCA method and UV absorption spectra]. *Spectrosc. Spectr. Analysis* 36 (11), 3592–3596.

Zhu, L., Li, M., and Yuan, C. (2022). Prediction model for effluent COD in sewage treatment based on BP neural network optimized by EHO. *J. Chongqing Technol. Bus. Univ. Nat. Sci. Ed.* 39 (3), 26–32. doi:10.16055/j.issn.1672-058X.2022.0003.004