



Qualitative detection of pesticide residues using mass spectral data based on convolutional neural network

Jian Wei¹ · Xuemei Wang² · Zhenyu Wang¹ · Jin Cao²

Received: 18 March 2021 / Accepted: 13 May 2021

Published online: 14 June 2021

© The Author(s) 2021

Abstract

Excessive pesticide residues in crops directly threaten human life and health, so rapid screening and effective measurements of agricultural pesticides residues have important application significance in the field of food safety. It is imperative to detect different pesticide residue types in actual complex crop samples cause mixture analysis can provide more information than individual components. However, the accuracy of mixture analysis can be obviously affected by the impurities and noise disturbances. Purification and denoising will cost a lot of algorithm time. In this work, we used the problem transformation method to convert pesticide residues prediction into multi-label classification problem. In addition, a new convolutional neural network structure Pesticide Residues Neural Network (PRNet) was proposed to solve the problem of multi-label organophosphate pesticide residue prediction. The method of binary correlation and label energy set was used to adapt 35 pesticide residues labels. The Cross Entropy were used as loss functions for PRNet. The comprehensive comparison performances (e.g. 97% optimal accuracy rate) of PRNet is better than the other four models. By comparing the ROC curves of the five models, PRNet performs the best. The PRNet can separate the independent mass spectrometry data by different collision energy applied to phosphorus pesticide compounds through a three-channel structure. No complicated data preprocessing is required, the PRNet can extract the characteristics of different compounds more efficiently and presents high detecting accuracy and good model performance of multi-label mass spectrometry data classification. By inputting MS data of different instruments and adding more offset MS data, the model will be more transplantable and could lay the foundation for the wide application of PRNet model in rapid, on-site, accurate and broad-spectrum screening of pesticide residues in the future.

Keywords Mass spectrometry · Multi-label classification · Convolutional neural network · Pesticides residues

1 Introduction

Pesticides are widely used in agricultural production. They play a significant role in preventing insects or diseases and increasing yields. However, unreasonable use of pesticides occurs occasionally, which could not only cause pesticide residues, but also pollute the environment severely [1]. There are many different types of pesticides based on their

structure, including carbamate, organochlorine, organophosphorus, pyrethroids, heterocycles and amides, etc. Among those, organophosphorus pesticides are the most widely used. There are more than 100 organophosphorus pesticides. Most of them have the irreversible cholinesterase inhibitors for organisms to bring detrimental impacts on human healthy [2]. Therefore, it is a necessary and difficult problem to realize rapid detection and identification

✉ Zhenyu Wang; Jian Wei, zw@ss.pku.edu.cn | ¹School of Software and Microelectronics, Peking University, 24 Jinyuan Road, Daxing District, Beijing 102600, People's Republic of China. ²National institutes for food and drug control, Beijing 100050, People's Republic of China.



of organophosphorus pesticides in vegetables according to the detection limit, sensitivity and accuracy requirements. At present, mass spectrometry (MS), as a strong qualitative function, is one of the most broadly used confirmations in practice. It firstly dissociates the molecules of the substance to ions with different masses, and then uses the different motion behavior of ions in electric field or magnetic field to separate the ions according to the mass charge ratio (m/z) to obtain the mass spectrometry [3]. The qualitative and quantitative information of the samples can be obtained from the MS data. Since it is targeted at the molecular structure of the sample. The information obtained by MS is more accurate and reliable for qualitative analysis. In real life, different compounds are always mixed together. Due to the complexity of MS analysis, the amount of MS data is usually very large. Consequently, most MS peaks are meaningless for the components' identification. Considering that MS data is usually accompanied by some noise to conceal sample ion strength. The introduced confused information can not only make the identification algorithm more computationally intensive, but also increase the possibility of random matching. It results in a decrease in the appraisal results reliability and an increase in the false positive or false negative identifications. Even though, some mathematical methods attempted to detect specific compounds from mixed samples MS data [4].

Over the past several years, many works on Artificial Neural Network (ANN) in the area of MS have been reported. MSnet, created by Curry et al. may be the first generation neural network for MS [5]. It involves a hierarchical system of several neural networks for MS data. The work of Werther et al. uses classical multi-dimensional numerical analysis techniques to compare the performance of ANNs [6]. A. Eghbaldar et al. presented a methodology for optimizing ANN to identify the compounds' structural features from MS data [7]. Ankita thakur uses ANN to mine MS data to detect ovarian cancer [8]. Ion mobility spectra can also be successfully classified by neural networks from the combination of drift time, number, intensity, and peak shape [9].

In general, the dimensionality of MS data inputs is always very large. Although, the performance of the neural network becomes better for the larger data set. Unfortunately, small number of samples and the large dimension of MS data inputs construct the typical dilemma for the realistic datasets. Deep learning can directly extract features from raw large dimension data by neural unit learning data characteristics. Hence, it was explored to predict molecular substructure in the mass spectral data. With the development of deep learning, some models based on recurrent neural network have shown amazing potential when processing sequence data. J. Liu proposed a material

classification system based on short and short time memory (LSTM) [10]. Nevertheless, it is still difficult to detect a class similar molecular inside the mixture sample MS data.

Since the multi-label classification can study each example associated with a set of labels simultaneously. The performance of multi-label classification algorithm in the detection of organophosphorus pesticide residues was investigated in this work. The practical algorithm is based on convolutional neural network (CNN). CNN is the most widely used deep learning method to achieve the desirable classification performance in various classification problems. Compared with other algorithms, CNN can separate the independent MS contributed by different compounds, so as to accomplish better accuracy. In this study, we mimicked the CNN's architecture and use the three-channel architecture as input [11]. Compared with the traditional classification methods, the PRNet (Pesticide Residues Neural Network) can significantly improve the prediction accuracy and performance of mixture samples MS data. This model has a respectable potential in large target MS data analysis.

2 Materials and methods

2.1 Chemical and materials

All high purity pesticide standards with purity greater than 97.0% were purchased from Dr. Ehrenstorfer (Augsburg, Germany) or LGC Standards (Teddington, UK). Stock solutions were prepared in acetonitrile at a concentration of 100 mg/L, stored at -20°C and restored to room temperature and diluted the stock solution to 100 $\mu\text{g/L}$ before detection. Acetonitrile and methanol (Merck, German), formic acid (Fluka) and ammonium acetate (Sigma-Aldrich, Germany) were chromatographically pure. Anhydrous magnesium sulphate (MgSO_4), sodium chloride (NaCl), adsorbents: octadecyl chemically bonded phase silica gel (C18, 43–60 μm) and N-primary secondary amine (PSA, 40–60 μm) were all of analytical grade. A Milli-Q Advantage A10 ultrapure water system from Millipore (Milford, MA) was used to obtain the HPLC-grade water during the analyses. Three fruits or vegetables (grapes, apples and cabbage) were purchased from markets in Beijing.

2.2 Sample preparation and extraction

We precisely weighed 10 g crushed sample of each fruit or vegetable with the accuracy to 0.001 g. Then every sample was put in 50 mL polytetrafluoroethylene (PTFE) centrifuge tube and added 10 mL acetonitrile, extracted in 1 min through high-speed homogenate, added 4 g MgSO_4 and 0.5 g NaCl, vortexed in 1 min, centrifuged in 5 min

under 5000 r/min. Then 6 mL upper acetonitrile phase solution was put into 15 mL centrifuge tube of containing 400 mg C18 + 400 mg PSA + 1200 mg MgSO₄, vortexed in 1 min and centrifuged in 5 min under 5000 r/min; then the supernatant was filtrated with 0.22 μm organic filter membrane for the following determination.

2.3 Preparation of spiked samples

We precisely weighed 10 g targeted sample of grape, apple and cabbage with the accuracy to 0.001 g respectively. Then we added 10 μL monocrotophos and phoxim stock solutions to grape, 10 μL isazofos and methamidophos stock solutions to apples, and 10 μL dichlorvos and chlorpyrifos stock solutions to cabbage. After mixing each sample, we could prepare 0.1 mg/kg spiked samples according to the extraction procedure in above mentioned.

2.4 UPLC-Q-TOF/MS analysis

Ultra-high-performance liquid chromatography-quadrupole-time-of-flight mass spectrometry (UPLC-Q-TOF/MS, Agilent 1290-G6545, USA) equipped with Zorbax Eclipse Plus-C18 column (150 mm (L) × 3 mm with 1.8 μm of particle size, Agilent, USA) was used for chromatographic separation. 2 mM of ammonium acetate and 0.05 % formic acid solution were used as mobile phase A, and methanol aqueous solution which contained 0.05 % formic acid was used as mobile phase B. The column was equilibrated with 90% of mobile phase A and 10% of mobile phase B for 30 min before injection. The amount of mobile phase A maintained unchanged at 90% from 0 to 0.5 min. From 0.5 to 3 min, the amount of mobile phase A decreased from 90 to 50%, from 3 to 20 min to 0%, then maintained for 4 min and at last go back to 90% at 24.1 min. The temperatures of column oven and auto sampler were set at 40 and 4 °C, respectively. The flow rate was 0.4 mL/min and the injection volume were 2 μL.

MS parameters: UPLC-Q-TOF/MS equipped with Dual AJS ESI source was operated in full-scan TOF mode and MS/MS spectra were acquired for further compound identification using auto MS/MS acquisition. MS detection was carried out in positive electro spray ionization mode (ESI⁺). The following operating conditions were used: scan range, 50–1200 m/z; capillary voltage, 3500 V; fragmentor voltage, 120 V; skimmer voltage, 65 V. The temperature of the drying gas and sheath gas were 250 and 325 °C, and the flow rates of the drying gas and sheath gas were set to 7 L/min and 12 L/min. The nebulizer pressure was 35 psi, and the nozzle voltage was 300 V. Auto MS/MS acquisition conditions were set to: mass range, 50–1200 m/z; collision energy (CE) was set at 10 eV, 20 eV and 40 eV. The source

and gas parameters used were the same as those used in full-scan TOF mode.

2.5 Preparation for dataset

The experiments were repeated with different pesticide solutions. MassHunter Acquisition B.03 and MassHunter Qualitative Analysis B.03 were used for the acquisition and treatment of data. In addition, Python 3.7 was used to create pkl file for the dataset of pesticides.

In this work, 35 kinds of common organophosphorus pesticides were mixed randomly, as listed in Table 1. The negative data came from the sample reagents that did not contain the 35 organophosphorus pesticide compounds. The positive data came from the sample reagents that added the mixture of organophosphorus pesticide compounds. There were 1–5 kinds of organophosphorus pesticides in each positive data. With the permission of experimental conditions, over 30,000 mixtures were prepared with more than 97 % purity. A typical mixture sample experimental MS data was shown in Fig. 1.

Additionally, in different conditions, the use of different instruments and experimental operations tend to produce different noises. In order to make experimental MS data closer to the real environment and make the training model more robust, three level of collision energy (10 eV, 20 eV, 40 eV) were chosen to generate the three-channel data besides the experimental MS data according to the assumption of the linear mixed model. The random noise was added in experimental MS data for training model, as demonstrated in Fig. 2. Based on linear composite hybrid approach, 10,000 noise simulation data were also generated to explore the noise impact on the performance of the experimental model eventually.

After finishing the dataset preparation, the total dataset was divided into 32,000 training sets, 4,000 validation sets and 4,000 test sets. For the data label, the Encoding label with one-hot Encoding for the specified compound was set as “1”, and “0” for none specified compound. For the data label, we encode the output label, and set “1” as the existence of the specified compound, and “0” as the absence of the specified compound.

2.6 Algorithm

2.6.1 Support vector machine

Support Vector Machine (SVM) is a common classifier. The core idea of SVM is to find a linear classifier to separate hyperplane with the maximum interval in the feature space. Then SVM can classify the unknown sample set through the hyperplane. SVM uses inner product kernel function instead of nonlinear mapping to

Table 1 Mass spectral data source: 35 organophosphorus pesticide reagents

Compound name	CAS	Molecular weight	Molecular formula	Classification
Glyphosate	1071-83-6	169.0140	C ₃ H ₈ NO ₅ P	Class1
Trichlorphon	52-68-6	255.9226	C ₄ H ₈ Cl ₃ O ₄ P	Class2
Dichlorophos	62-73-7	219.9459	C ₄ H ₇ Cl ₂ O ₄ P	Class3
Parathion	56-38-2	291.0330	C ₁₀ H ₁₄ NO ₅ PS	Class4
Fenitrothion	122-14-5	277.0174	C ₉ H ₁₂ NO ₅ PS	Class5
Parathion methyl	298-00-0	263.0017	C ₈ H ₁₀ NO ₅ PS	Class6
Chlorpyrifos-methyl	5598-13-0	320.8950	C ₇ H ₇ Cl ₃ NO ₃ PS	Class7
Sulfotep	3689-24-5	322.0266	C ₈ H ₂₀ O ₅ P ₂ S ₂	Class8
Fenthion	55-38-9	278.0200	C ₁₀ H ₁₅ O ₃ PS ₂	Class9
Dimethoate	60-51-5	228.9996	C ₅ H ₁₂ NO ₃ PS ₂	Class10
Phorate	298-02-2	260.0128	C ₇ H ₁₇ O ₂ PS ₃	Class11
Isazofos	42509-80-8	313.0417	C ₉ H ₁₇ ClN ₃ O ₃ PS	Class12
Demeton	126-75-0	258.3384	C ₈ H ₁₉ O ₃ PS ₂	Class13
Methidathion	950-37-8	301.9619	C ₆ H ₁₁ N ₂ O ₄ PS ₃	Class14
Isocarbophos	24353-61-5	289.0538	C ₁₁ H ₁₆ NO ₄ PS	Class15
Phoxim	14816-18-3	298.0541	C ₁₂ H ₁₅ N ₂ O ₃ PS	Class16
Cadusafos	95465-99-9	270.0877	C ₁₀ H ₂₃ O ₂ PS ₂	Class17
Methamidophos	10265-92-6	141.0013	C ₂ H ₈ NO ₂ PS	Class18
Isofenphos-methyl	99675-03-3	331.1007	C ₁₄ H ₂₂ NO ₄ PS	Class19
Profenofos	41198-08-7	371.9351	C ₁₁ H ₁₅ BrClO ₃ PS	Class20
Acephate	30560-19-1	183.0119	C ₄ H ₁₀ NO ₃ PS	Class21
Diazinon	333-41-5	304.1010	C ₁₂ H ₂₁ N ₂ O ₃ PS	Class22
Monocrotophos	6923-22-4	223.0610	C ₇ H ₁₄ NO ₅ P	Class23
Phosfolan	947-02-4	255.0153	C ₇ H ₁₄ NO ₃ PS ₂	Class24
Phosalone	2310-17-0	366.9869	C ₁₂ H ₁₅ ClNO ₄ PS ₂	Class25
Phosmet	732-11-6	316.9945	C ₁₁ H ₁₂ NO ₄ PS ₂	Class26
Posfolan-methyl	5120-23-0	226.9840	C ₅ H ₁₀ NO ₃ PS ₂	Class27
Malathion	121-75-5	330.0361	C ₁₀ H ₁₉ O ₆ PS ₂	Class28
Chlorpyrifos	2921-88-2	348.9263	C ₉ H ₁₁ Cl ₃ NO ₃ PS	Class29
Triazophos	24017-47-8	313.0650	C ₁₂ H ₁₆ N ₃ O ₃ PS	Class30
Mecarbam	2595-54-2	329.3700	C ₁₀ H ₂₀ NO ₅ PS ₂	Class31
Ethoprophos	13194-48-4	242.0564	C ₂₀ H ₁₅ NO ₃	Class32
Pirimiphos-methyl	29232-93-7	305.3300	C ₁₁ H ₂₀ N ₃ O ₃ PS	Class33
Omethoate	1113-02-6	213.0225	C ₅ H ₁₂ NO ₄ PS	Class34
Isocarbophos	24353-61-5	289.2900	C ₁₁ H ₁₆ NO ₄ PS	Class35

high-dimensional space [12]. Due to the influence of noise, the classification effect is usually better on small sample set. Meanwhile, SVM uses quadratic programming to solve support vector (the calculation of m-order matrix, M is the number of samples). When the number of M is large, the storage and calculation of the matrix will consume a lot of machine memory and operation time [13]. Therefore, it is difficult to solve the multi classification problem with SVM.

2.6.2 Artificial neural network

Artificial Neural Network (ANN) is an abstract model for describing how the human brain organizes and operates. All neurons contained in a neural network record the

weight corresponding to their inputs [14]. Corresponding to the three processing units, the neural network contains three levels: input layer, hidden layer and output layer [15]. By repeated learning and training of the input information data, the neural network can constantly improve the parameter values of the weights in the neural network to reach the closest conclusion, as represented in Fig. 3.

2.6.3 Extreme gradient boosting

extreme Gradient Boosting (XGBoost) is one of the Boosting algorithms. The idea of the Boosting algorithm is to integrate many weak classifiers together to form a strong classifier [16]. It corrects the residuals of all previous weak

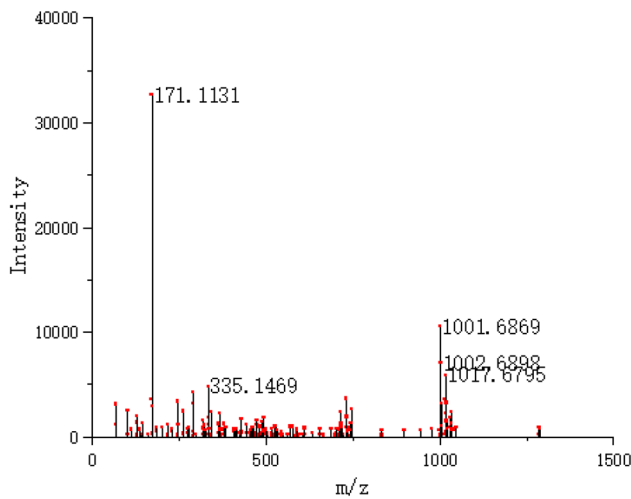


Fig. 1 The MS data (UPLC-Q-TOF/MS) from experiment (the selected MS data were marked by red dots)

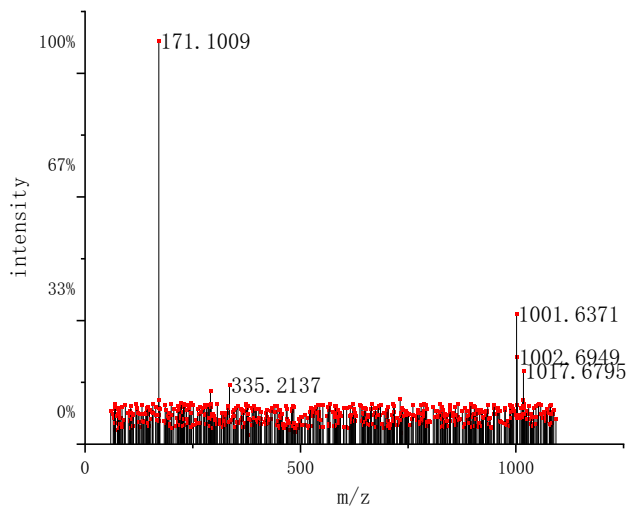
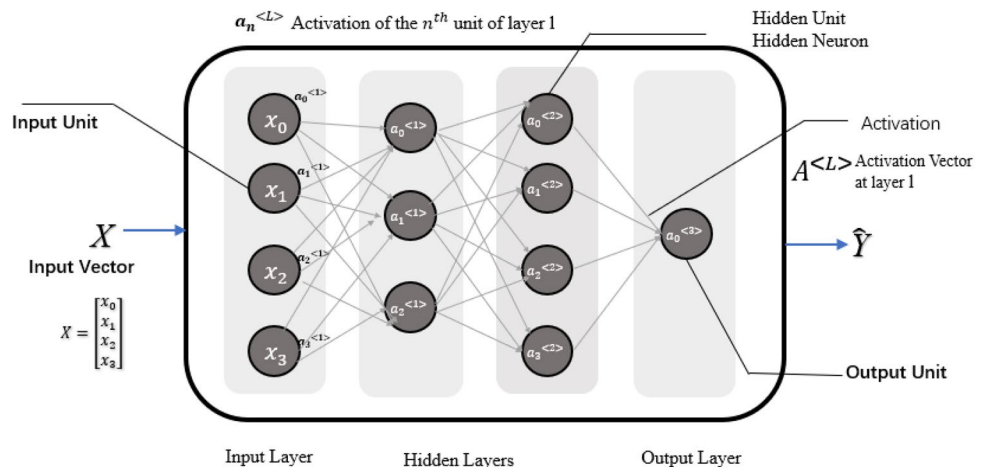


Fig. 2 The MS data (UPLC-Q-TOF/MS) with added noise (the selected MS data were marked by red dots)

Fig. 3 Architecture of multi-layer artificial neural network with error



classifiers by continuously adding new weak classifiers. The final prediction is finally made by adding multiple classifiers. Thus, the accuracy rate will be higher than a single weak classifier. When adding a new model, the gradient boosting algorithm is used to minimize the loss. XGBoost has fast calculation speed and is widely used in small-scale classification tasks [17].

2.6.4 Long short-term memory

Long Short-Term Memory (LSTM) is a kind of time recurrent neural network. This model takes sequence data as input, recurses in the evolution direction of the sequence and connects all nodes in a chain way. The input of the hidden layer is not only related to the output of the input layer, but also related to the output of the hidden layer at the previous moment [10]. LSTM has two transmission states, the memory cell state and the hidden state. This strategy can effectively avoid the phenomenon of gradient disappearance and gradient explosion in the training process. It is very suitable for dealing with problems highly related to time series [18].

2.6.5 Convolution neural network

Convolution neural network (CNN) is a kind of deep neural network with convolution structure to reduce the memory occupation. Its three key operations are local receptive field, weight sharing, and pooling layer [19]. CNN can effectively reduce the number of network parameters and alleviate the over fitting problem of the model, as illustrated in Fig. 4.

The differences between the CNN and the ANN were the usage of convolution layers and the dimension of input [20]. The convolutional layer reduces the number of parameters to be trained by receptive field and weight sharing. The convolutional layer is sometimes followed

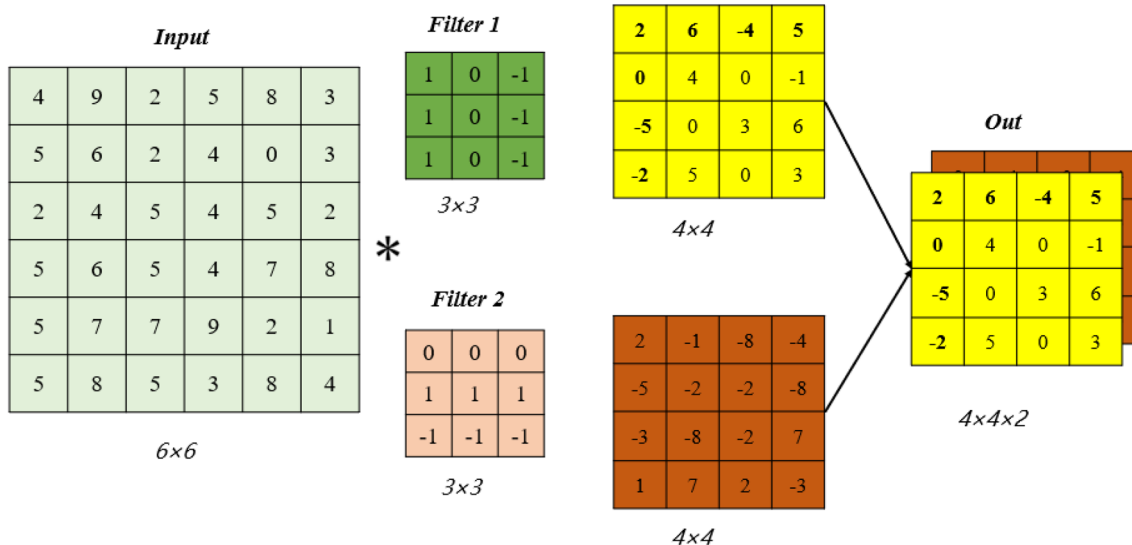


Fig. 4 Convolution operation with multiple filters: By sliding the convolution kernel at the input and calculating the dot product, a matrix called convolution feature is obtained

by a pooling layer, which can reduce the dimension of features, compress the number of data and parameters, reduce overfitting, and improve the fault tolerance of the model. Convolutional neural network has stronger feature learning ability and feature expression ability than traditional neural network [21, 22].

2.6.6 Mass spectrum matrix

Assuming the MS data of a mixture are equal to the weighted sum of the MS of the individual compounds [23]. The MS data of mixtures can be written from the normalized intensities:

$$I_{i,j} = \sum_{k=1}^n a_{ik} s_{kj} \tag{1}$$

where $I_{i,j}$ is the intensity of the i th mass in the j th mixture, n is the number of components in the mixture, a_{ik} is the intensity of mass i in pure compound k , and s_{kj} is the concentration of compound k in the source for mixture j . Equation (1) can be simply expressed as:

$$I_m = AS \tag{2}$$

where I_m is an $i \times j$ matrix, A is an $i \times k$ matrix and S is an $k \times j$ matrix; i, j, k represent the number of different mixtures, masses and pure compounds respectively and I, A, S represent overlapping MS data matrix, pure compound matrix and concentration matrix respectively. The I_m here is the mixture MS data we used for training and testing, and

the information described by the m/z and intensity is the characteristics to be learned by the experimental model.

2.7 PRNet model training

Based on iteration algorithm, PRNet is an optimized multi-target detection CNN model for pesticide residue identification. The architecture of PRNet model is listed in Table 2. The mixture MS data of three energy (10 eV, 20 eV and 40 eV) were flattened as one-dimensional input. Based on the

Table 2 The architecture of PRNet model (N: the number of m/z windows)

Layer (Type)	Output shape	Parameters
conv1d_1 (Conv)	N, 16, 3	160
conv1d_2 (Conv)	N, 32, 3	1568
max_pooling1d_1 (Pooling)	N//2, 32, 3	0
conv1d_3 (Conv)	N//2, 32, 3	3104
max_pooling1d_2 (Pooling)	N//2, 32, 3	0
conv1d_4 (Conv)	N//2, 64, 3	6208
conv1d_5 (Conv)	N//2, 64, 3	12,352
max_pooling1d_3 (Pooling)	N//4, 64, 3	0
conv1d_6 (Conv)	N//4, 128, 3	24,704
max_pooling1d_4 (Pooling)	N//4, 128, 3	0
conv1d_7 (Conv)	N//4, 32, 3	12,320
max_pooling1d_5 (Pooling)	N//4, 32, 3	0
flatten_1 (Flatten)	24*N	0
dense_1 (Full-connected)	N//5	6,912,240
dense_2 (Full-connected)	N//10	28,920
dense_3 (Full-connected)	35	4235

CNN, the established PRNet model had same number of layers and same full-connected layers with Rectified Linear Units (ReLU) activation [24].

Utilizing the twelve hundred m/z windows with 10 ppm interval, in range from 50 to 1200 m/z, the MS data of each energy could be flattened into one-dimensional matrix. With the character of input, one-dimensional convolution layers could further simplify the model. The information between different energy could be recognized by the convolution layers. Max pooling layers were applied to abstract the characteristics of the region and reduce the coupling degree of the model. Convolution layers and pooling layers were used for the feature extraction. The fully-connected layers were applied for the classification. In the last fully-connected layer, sigmoid activation was applied to output the probability of each compound presence. Usually the model threshold was set to 0.5. The threshold value also could be modified according to actual condition (If the goal is to screen for as many pesticide residues as possible, the threshold can be lower down). The operation flow is shown in Fig. 5.

3 Results and discussion

In the classification problem of machine learning, Sigmoid function is the commonly used activation function of the output layer of neural network. Since multiple

classes can overlap each other. Compared with the Softmax, Sigmoid is more suitable for the multi-label classification task. In this work, Sigmoid function was applied as the activation function of the neural network. The binary cross entropy is the corresponding loss function of Sigmoid [25]. The smaller loss function represents the better model robustness. The loss function of this multi-label classification was utilized to estimate the difference between the predicted value and the true value from the model. This loss function can also make the training model as fast as possible and require less memory:

$$BCE = -\frac{1}{m} \sum_{i=1}^m y_i \log f_i(x) + (1 - y_i) \log(1 - f_i(x)) \quad (3)$$

x is the input sample, m is the total number of training data, y_i is the output value of the i th data, and $f_i(x)$ is the predicted value of the i th model.

Because the output of each tag is assumed to be independent. The general configuration for multi-tag binary classification is the BCE and Sigmoid activation functions. Each category outputs a probability between 0 and 1. Each corresponds to a sigmoid function. The Adam algorithm was deployed as the optimizer to update the weights of the neural network iteratively according to the training data [26]. The Model PRNet refers a neural network model for pesticide residue prediction based on the convolutional neural network structure. All models run on an Ubuntu system that based Linux with 24

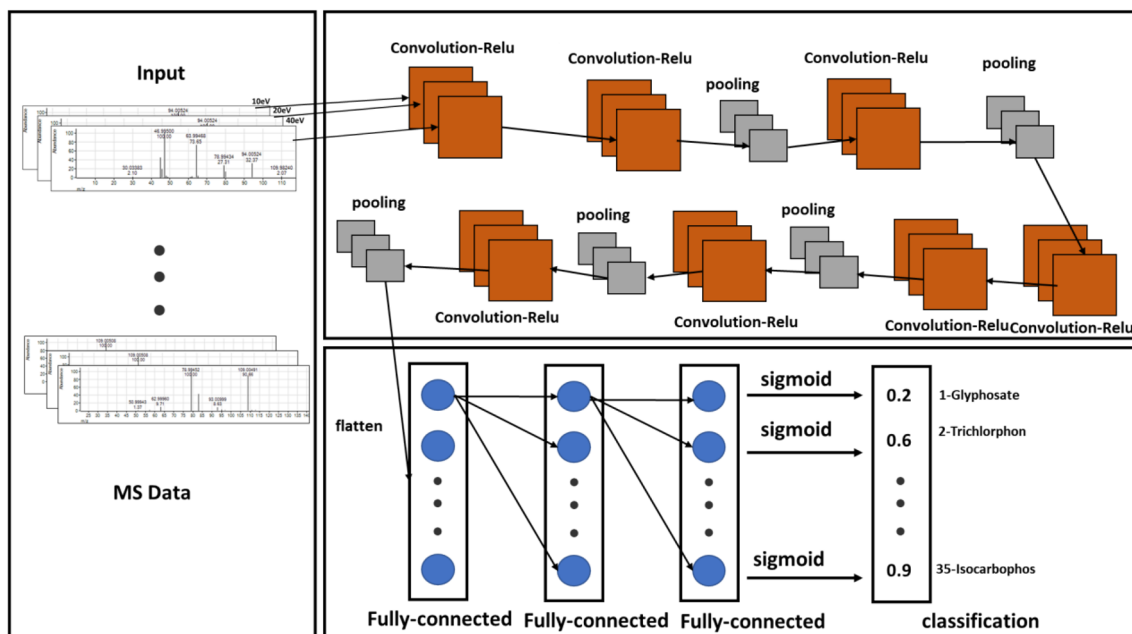


Fig. 5 The architecture of PRNet: Convolution neural network adopts MS data as input, which can effectively learn the corresponding features directly from a large number of samples, reduce

the complex process of data preprocessing, and avoid the complex process of feature extraction

cores, 128 GB RAM and NVIDIA 2080 GPU cards. After 100 epochs of training, the accuracy, recall rate and precision of the test set for each model of the target compound detection are shown in Table 3.

The comparison results show that these five machine learning models can detect multiple target compounds in overlapping samples well. They can effectively learn features from raw MS data. The results obtained by the neural network are usually classification results such as 0.5 and 0.8 instead of labels such as 0 or 1. Therefore, a threshold can be selected to diverge them. When the result is greater than the threshold, the predicted value can be judged as 1. On the contrast, the predicted value is judged to be 0. By increasing the threshold, more confidence can be obtained in the predicted results, thereby improving accuracy. But this will reduce recalls. Otherwise, the number of true cases missed by the model will decrease and the recall will increase. For those models, 0.5 was chosen as the threshold under comprehensive consideration. The ANN model only includes the full connection layers and batch normalization layers, it does not contain other additional structures. The performance effect of ANN is moderate. Due to the good adaptation and outlier processing, PRNet has the respectable feature extraction performance. For the larger mixture MS data predictions, PRNet always obtains higher accuracy than the other four models. Furthermore, for the target compounds detection, recall is an equally important indicator to reflect the proportion of correctly predicted (true) components in all predicted (true and true negative) components. Since the SVM model is difficult to predict a large number of samples with positive labels. The accuracy of SVM is much lower in performance than neural networks for higher number of detection targets. LSTM has a higher accuracy rate with a lower recall rate conversely. It requests a lower threshold. The effect of XGBoost is acceptable. But there is still a certain gap (accuracy and recall) between XGBoost and PRNet. For further observation, the receiver operating characteristic (ROC) [27]

curves were arranged to evaluate the predictive ability of the five investigated models:

The comparison of the five models' ROC curves is illustrated in Fig. 6; Table 4. The AUC (average area under curve) value is the area covered by the ROC curve. The better classification effect classifier has the larger AUC. The ROC curve is farther from the pure opportunity line (the dashed black line), the model has the stronger discrimination ability. Because of the mapping principle of ROC curve, the model based on neural network and other models have different curve shapes. According to Fig. 6, the AUC of PRNet is the largest to indicate the most robust classification performance among the five models.

The average precision (AP) score can also be summarized as a weighted average precision achieved at each threshold to evaluate the five models. In general, the average precision is the mean of the accuracy corresponding to all the values from 0 to 1 across the recall rate:

$$AP = \int_0^1 p(r)dr \tag{4}$$

This integral value is approximately the sum of the precision at each possible threshold multiplied by the change in recall rate:

$$AP = \sum_{k=1}^N p(k)\Delta r(k) \tag{5}$$

In multi-label classification, Mean Average Precision (MAP) is a commonly used evaluation method as well [28]. It measures the quality of the learned model in all categories by taking the average of all AP:

$$mAP = \frac{1}{num} \sum AP \tag{6}$$

From Table 5, the classifier performance of SVM and XGBOOST is also poor. Compared with the ROC curves in Fig. 6, the target detection performance of PRNet model for the mixture MS data is more stable than that of other models.

Table 3 Accuracy(A)/recall(R)/precision(P) for the five models

	SVM			ANN			XGBoost			LSTM			PRNet		
	A	R	P	A	R	P	A	R	P	A	R	P	A	R	P
1 ^a	0.92	0.55	0.73	0.96	0.87	0.52	0.92	0.71	0.62	0.96	0.65	0.61	0.99	0.89	0.82
2 ^b	0.92	0.56	0.74	0.95	0.86	0.51	0.92	0.67	0.62	0.94	0.65	0.61	0.98	0.87	0.83
3 ^c	0.90	0.54	0.74	0.94	0.88	0.48	0.90	0.65	0.60	0.93	0.62	0.62	0.97	0.87	0.82

1^a: model based on MS data of 10 compounds; 2^b: model based on MS data of 20 compounds; 3^c: model based on MS data of 35 compounds

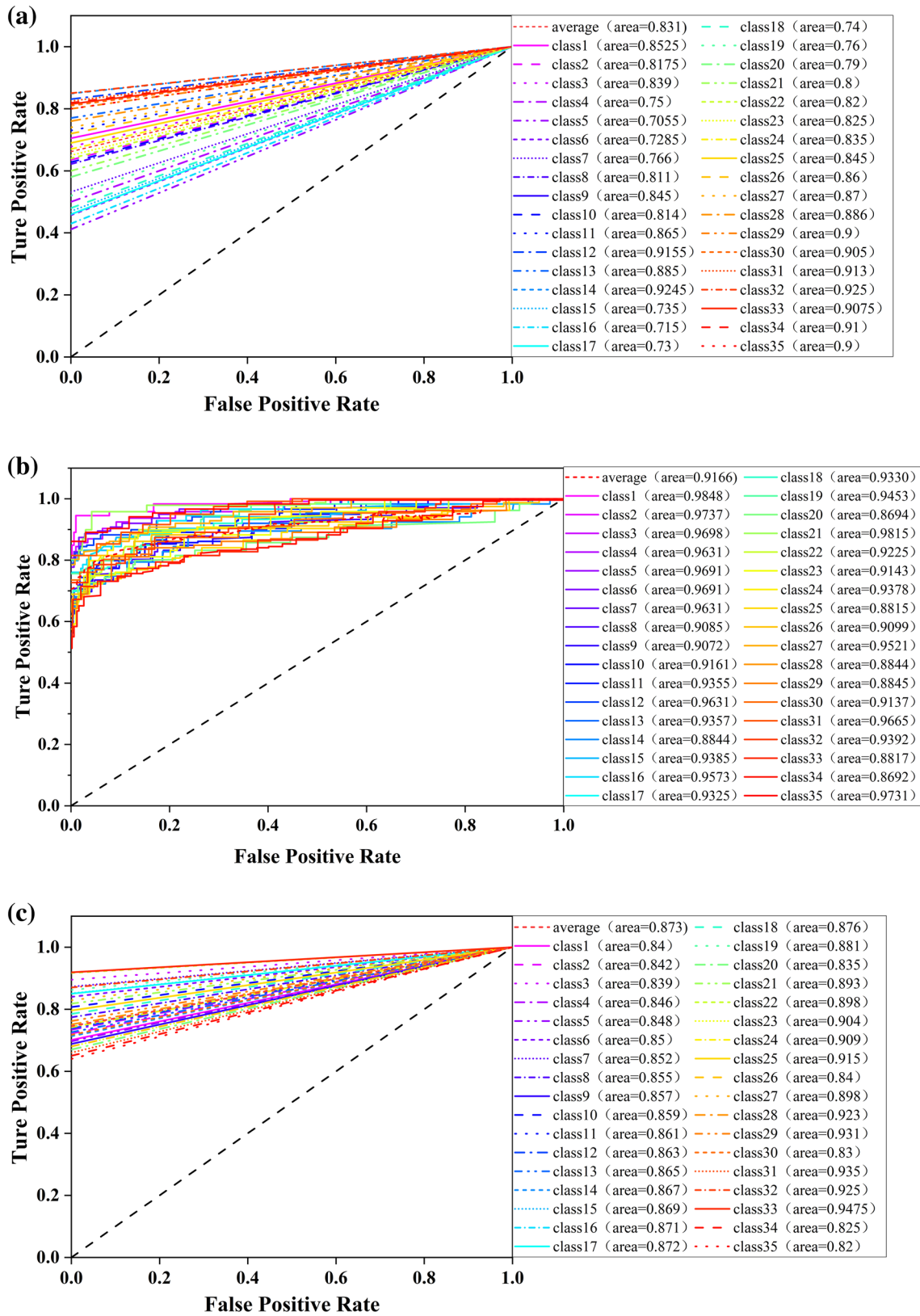


Fig. 6 Receiver operating characteristic (ROC) curves for:(a) SVM; (b) ANN; (c) XGBoost. (d) LSTM; (e) PRNet

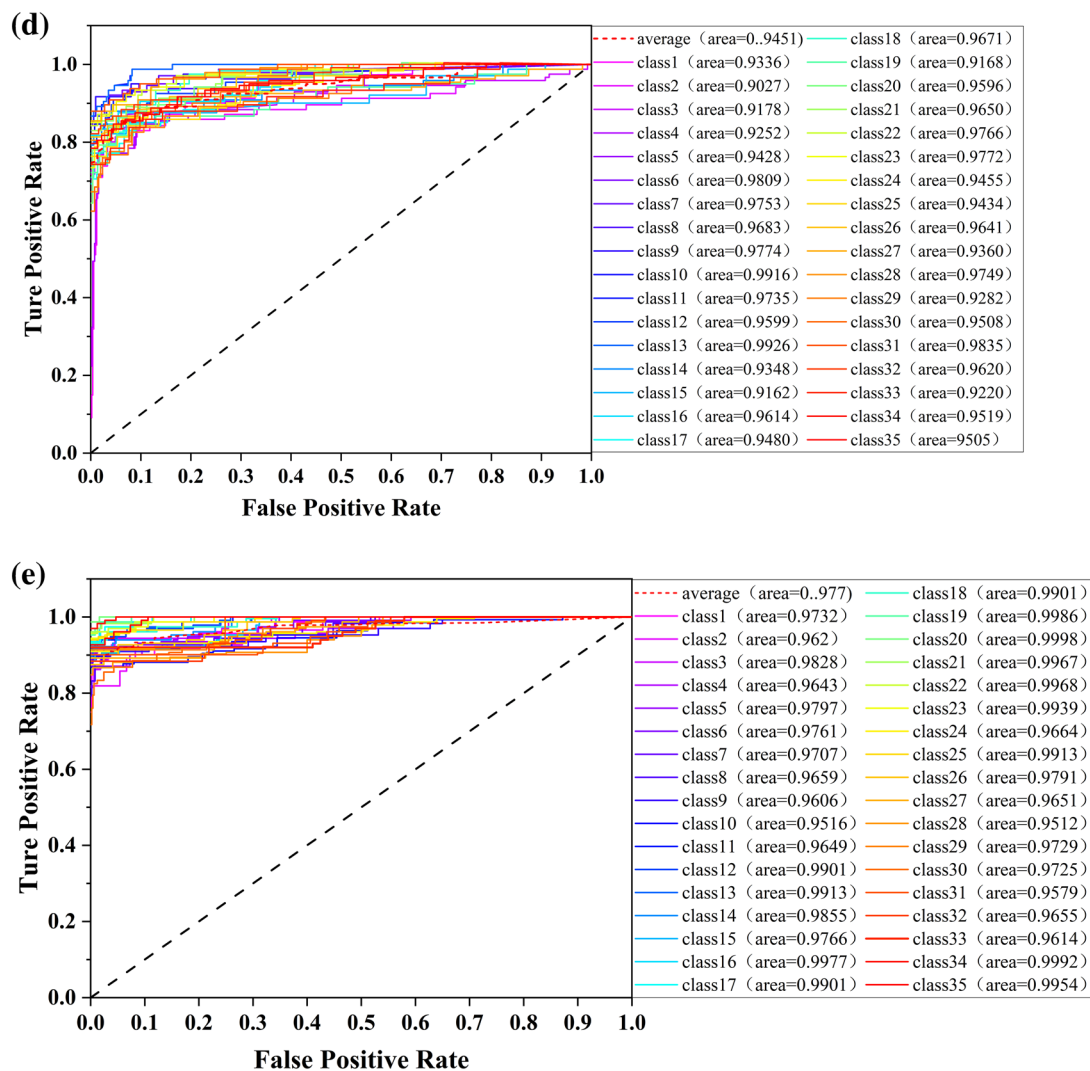


Fig. 6 (continued)

Table 4 The average area under the curve for the five models

Models	SVM	ANN	XGBoost	LSTM	PRNet
mAP	0.83	0.92	0.88	0.94	0.97

Table 5 Mean average precision score for the five models

Models	SVM	ANN	XGBoost	LSTM	PRNet
mAP	0.68	0.78	0.81	0.84	0.95

In general, the detection of pesticide residues involves more types of compounds. Consistent with the Table 3, when more organophosphorus pesticide compounds were added, the accuracy reduction of PRNet was less than that of other models. This indicates that PRNet is more suitable for large-sample detection.

4 Discussion

If the MS data is pre-processed in advance, such as smoothing, baseline correction, and peak picking [29]. The traditional machine learning algorithms, e.g. SVM, can also perform well, especially for single MS data classification. PRNet does not require much preprocessing or denoising, as demonstrated in Table 5.

In the 4000 testing MS data, about 10% testing data are negative reagent data or other type compounds interference data (Multi-classification problems usually do not have negative samples, we treat categories other than 35 organophosphorus pesticides as negative samples). TP means that all compounds were detected. FP means that some compounds have been detected correctly. TN indicates the correct prediction of a negative reagent as the absence of the compound. FN means it falsely predict the presence of a pesticide compound that does not exist.

From Table 6, the performance of PRNet is better than ANN (the ANN is only composed of full connection layers and batch normalization layers), SVM, LSTM or XGBoost in multi-tag target detection for mixture MS data. In fact, SVM and others have a good effect on the double-label classification. As the detection of organophosphorus pesticides increased, SVM cannot work well in the multi-label classification. LSTM is more suitable for time series data analysis [18]. The molecular weight of organophosphorus compounds is usually not very large. So, the MS data of organophosphorus compounds may not reflect strong sequence characteristics, which may be the reason for the general performance of the LSTM model. In the case of more noise impurities, the classification problem with boosting method may appear overfitting. Compared with ANN, PRNet can effectively handle high-dimensional data processing due to the existence of convolution kernel and other structures. Compared with PRNet, the XGBoost model is less effective, which may be related to the characteristics of the MS data. The mixture MS data may have high variability, noise and high dimension. XGBoost is a kind of boosted trees model derived from ensemble learning [30]. High-dimensional sparse features may make the training efficiency of the tree model become extremely low and easy to overfit (Fig. 7).

If there are more MS data of different energies as input, PRNet based on the CNN can handle more relationships between different energies. Comparison with time series, MS energy is an easier controllable variable factor for the input channel of the CNN network structure. Once more MS data of different energies are added as inputs, especially continuous signals on the energy axis, the accuracy of PRNet will decline. Further research on the architecture of PRNet, including the depth of the network, alternating layer and filter size, will be launched to continuously improve the ability of learn and detect complex mixture target compounds.

The size of MS data set is a bottleneck to limit the performance of deep learning. Usually, when the number of detected targets increases, these models need large data sets to learn the characteristics of samples. Therefore, if the

MS data set is very small, deep learning would be limited [30]. Our method based on the CNN can effectively realize the identification of multiple compounds from tandem mass spectrometry data. When the model was modeled with 35 organophosphorus pesticide compounds, the final mAP scores of the five models were 0.68, 0.78, 0.81, 0.84 and 0.95, respectively. This methodology can be applied to classify more pesticide compounds and even other types of compounds (Table 7).

The ★ is applied to compare the average training speeds of the five models in Table 7. In the case of small data samples, the training speed of SVM model is fast. But with the expansion of the training set and the increase of the types of phosphorus pesticides in the classification task, the training speed of SVM becomes very slow. Overall, the comprehensive training speed of SVM model is the slowest. Due to the large parameter space, the training speed of ANN is also slow. As the GPU speeds up and the batch normalization layer is added in the network, the training speed of ANN increases to some extent.

XGBoost has the fastest training speed, but XGBoost's parameter tuning is a little complicated. Thanks to GPU acceleration, the training speed of PRNet model is also moderated. The improvement of training speed obtained by LSTM through GPU acceleration is not obvious. Since LSTM is a kind of time series model, time t depends on the information at time $t-1$ and cannot be executed in parallel, its training speed is slower than the PRNet.

MS data sometimes have high variability, noise and high dimensionality [31], compared with traditional algorithms, deep learning is more universal and less vulnerable to MS data. The local receptive field of CNN can extract the subtle features of mass spectrometry data. CNN has the ability to learn low-level features from complex inputs. The CNN feature detection layer is learned from training data, avoiding explicit feature extraction. At the same time, due to the robustness of the filter, the model based on CNN is less affected by noise. In practice, the MS data may vary considerably depending on the operator, instrument, and laboratory environment. Controlling variables and capturing large amounts of data is crucial. The training model set by data obtained

Table 6 Number of true positive (TP)/true negative (TN) / false positive (FP)/ false negative (FN) /test data for the five models

	Models				
	SVM	ANN	XGBoost	LSTM	PRNet
TP	2017	2677	2433	2835	3521
TN	383	400	393	400	400
FP	1583	923	1167	765	79
FN	17	0	7	0	0

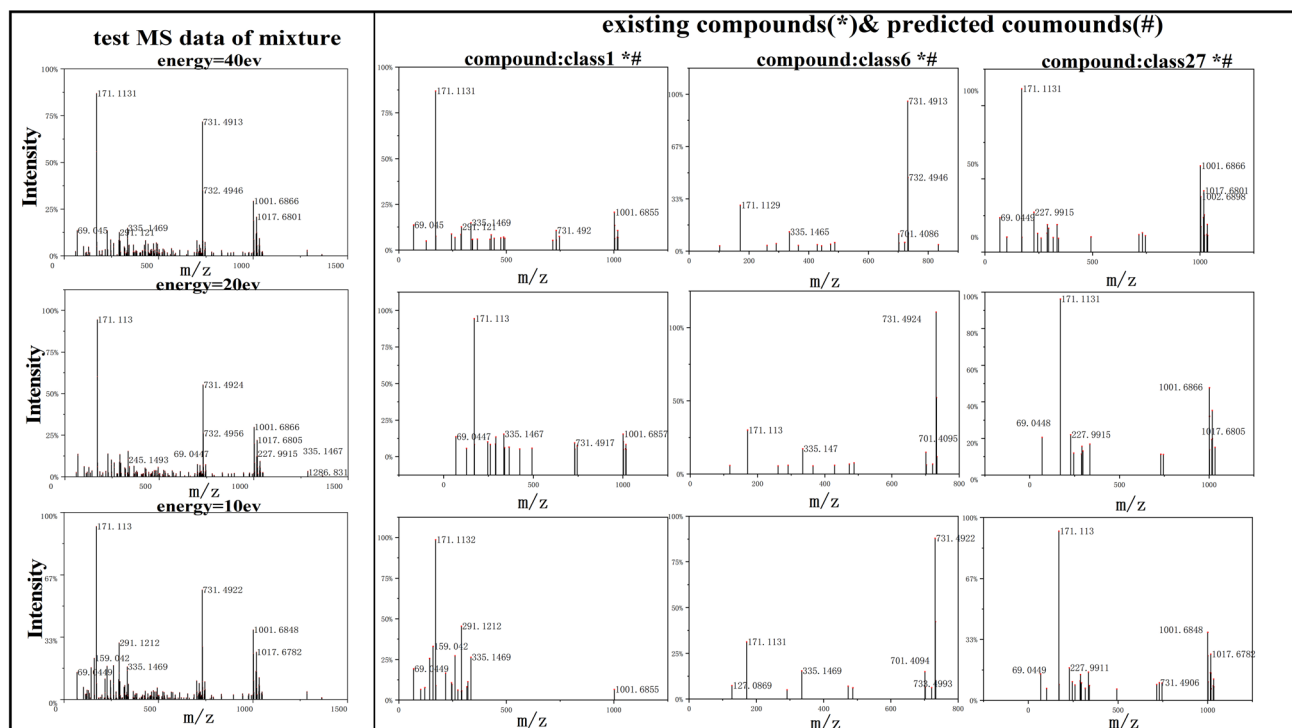


Fig. 7 Positive MS data were truly identified by the model (the selected MS data were marked by red dots)

Table 7 the average training speed of the five models

SVM	ANN	XGBoost	LSTM	PRNet
★	★★	★★★	★★	★★★

under the same conditions and data extension realizes the input of MS data from different instruments. Adding more offset in MS data can brand the model more transplantable.

5 Conclusions

In this paper, we propose a CNN-based method: PRNet. This model can directly detect a variety of organophosphorus pesticide compounds from mixture sample MS data obtained by UPLC-Q-TOF/MS. PRNet can directly extract features from the original high-dimensional MS data without complicated data preprocessing process. By evaluating different neural network structures, the average accuracy of the PRNet can reach 97%. Compared with traditional machine methods such as SVM, ANN, XGBoost and LSTM, PRNet has the best performance in accuracy, recall rate and precision. The multi-component identification and detection of pesticide residues mass

spectrometry data through convolutional neural networks would be an efficient way to help professionals and non-professionals detect pesticide residues quickly and correctly in the future. However, the presented PRNet model can only aimed at the existing pesticide species. If possible, the component concentration prediction would be considered in the future research and more conditions (e.g. spectral tilt) need to be explored to optimize quantitative analysis conditions. Benefit from the excellent transplantation and generalization of neural network, the model can be easily applied to mobile terminals for the pesticide residues detection. The model performance can be significantly improved through further study of input data and selection of structural components.

Acknowledgements The School of Software and Microelectronics at Peking University provided infrastructure support for this work. The research was granted by the Key research and development programs of the ministry of science and technology of China and the project “research and development of intelligent on-site rapid detection technology and related products for chemical pollutants in food” (No. 2018YFC1602800).

Author Contributions Jian Wei carried out the survey and the experimental operations and wrote the manuscript. Xuemei Wang contributed the sort of data set. Zhenyu Wang contributed the research ideas and the article composition. Jin Cao contributed to the article review. All authors discussed results and approved the final manuscript.

Declarations

Conflict of interest The authors have no conflicts of interest to disclose.

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

1. Badawy SM (2020) Optimization of reaction time for detection of organophosphorus pesticides by enzymatic inhibition assay and mathematical modeling of enzyme inhibition. *J Environ Sci Health Part B*, pp. 1–8
2. Li W, Xu K, Wang Y, Lei Z, Zhang Z (2004) Investigation on the detection of pesticide residue in vegetable based on infrared spectroscopy. *Guang pu xue yu Guang pu fen xi= Guang pu* 24(10):1202–1204
3. Smith RM (2004) *Understanding mass spectra: a basic approach*. Wiley
4. Eide I, Neverdal G, Thorvaldsen B, Grung B, Kvalheim OM (2002) Toxicological evaluation of complex mixtures by pattern recognition: correlating chemical fingerprints to mutagenicity. *Environ Health Perspect* 110(suppl 6):985–988
5. Curry B, Rumelhart DE (1990) MSnet: a neural network which classifies mass spectra. *Tetrahedron Comput Methodol* 3(3–4):213–237
6. Werther W, Lohninger H, Stancl F, Varmuza K (1994) Classification of mass spectra: a comparison of yes/no classification methods for the recognition of simple structural properties. *Chemometr Intell Lab Syst* 22(1):63–76
7. Eghbaldar A, Forrest T, Cabrol-Bass D (1998) Development of neural networks for identification of structural features from mass spectral data. *Analytica Chimica Acta* 359(3):283–301
8. Thakur A, Mishra V, Jain SK (2011) Feed forward artificial neural network: tool for early detection of ovarian cancer. *Sci Pharm* 79(3):493–505. <https://doi.org/10.3797/scipharm.1105-11>
9. Bell S, Nazarov E, Wang Y, Eiceman G (1999) Classification of ion mobility spectra by functional groups using neural networks. *Analytica Chimica Acta* 394(2–3):121–133
10. Liu J, Zhang J, Luo Y, Yang S, Wang J, Fu Q (2019) Mass spectral substance detections using long short-term memory networks. *IEEE Access* 7:10734–10744. <https://doi.org/10.1109/access.2019.2891548>
11. Ciampi M, Silvestri S, Gargiulo F (2018) Deep convolution neural network for extreme multi-label text classification. Paper presented at the Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies
12. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297
13. Cui J, Chen Q, Dong X, Shang K, Qi X, Cui H (2019) A matching algorithm with isotope distribution pattern in LC-MS based on support vector machine (SVM) learning model. *RSC Adv* 9(48):27874–27882
14. Yegnanarayana B (2009) *Artificial neural networks*. PHI Learning Pvt. Ltd.
15. Francelin R, Gomide F, Lancas F (1993) Use of artificial neural networks for the classification of vegetable oils after GC analysis. *Chromatographia* 35(3–4):160–166
16. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H (2015) Xgboost: extreme gradient boosting. R package version 04 – 2 1 (4)
17. Chen T, Guestrin C, Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (2016) pp 785–794
18. Sundermeyer M, Schlüter R, Ney H LSTM neural networks for language modeling. In: Thirteenth annual conference of the international speech communication association, 2012
19. Albawi S, Mohammed TA, Al-Zawi S Understanding of a convolutional neural network. In: 2017 International Conference on Engineering and Technology (ICET) (2017) Ieee, pp 1–6
20. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition (2017) pp 4700–4708
21. Teng S, Chen G, Liu G, Lv J, Cui F (2019) Modal strain energy-based structural damage detection using convolutional neural networks. *Appl Sci* 9(16):3376
22. Teng S, Chen G, Gong P, Liu G, Cui F (2020) Structural damage detection using convolutional neural networks combining strain energy and dynamic response. *Meccanica* 55(4):945–959
23. Rasmussen G, Hohne B, Wieboldt R, Isenhour T (1979) Identification of components in mixtures by a mathematical analysis of mass spectral data. *Anal Chim Acta* 112(2):151–164
24. Agarap AF (2018) Deep learning using rectified linear units (relu). *arXiv preprint arXiv:180308375*
25. Ramos D, Franco-Pedroso J, Lozano-Diez A, Gonzalez-Rodriguez J (2018) Deconstructing cross-entropy for probabilistic binary classifiers. *Entropy* 20(3):208
26. Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*
27. Bewick V, Cheek L, Ball J (2004) Statistics review 13: receiver operating characteristic curves. *Crit Care* 8(6):1–5
28. Henderson P, Ferrari V End-to-end training of object class detectors for mean average precision. In: Asian Conference on Computer Vision (2016) Springer, pp 198–213
29. Trygg J, Gabrielsson J, Lundstedt T (2009) *Data Preprocessing: Background Estimation, Denoising, and Preprocessing*
30. Tchistiakov V, Ruckebusch C, Duponchel L, Huvenne J-P, Legrand P (2000) Neural network modelling for very small spectral data sets: reduction of the spectra and hierarchical approach. *Chemometr Intell Lab Syst* 54(2):93–106
31. Skarysz A, Alkhalifah Y, Darnley K, Eddleston M, Hu Y, McLaren DB, Nailon WH, Salman D, Sykora M, Thomas CLP, Soltoggio A (2018) Convolutional neural networks for automated targeted analysis of raw gas chromatography-mass spectrometry data. Paper presented at the 2018 International Joint Conference on Neural Networks (IJCNN)

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