



CONSTRUCTION OF AN INTELLIGENT IDENTIFICATION MODEL FOR DRUGS IN NEAR INFRARED SPECTROSCOPY AND RESEARCH ON DRUG CLASSIFICATION BASED ON IMPROVED DEEP ALGORITHM

JIULIN XIA*

Abstract. Near-infrared spectroscopy has important applications in drug and food identification. Combining machine learning with near-infrared spectroscopy to achieve intelligent identification of drugs has become a research hotspot in recent years. To solve the problem of machine learning's inefficiency in classifying small-scale data, a drug identification model based on near-infrared spectroscopy combined with a random fading depth belief network is proposed. Aiming at the problem that the training time of the machine learning algorithm is too long, the extreme learning machine is used to replace the back propagation algorithm to optimize the stack sparse auto-encoder network. Additionally, the stack sparse auto-encoder algorithm based on extreme learning machine algorithm is constructed. The study found that the precision of the Dropout Deep Belief Network model was 99.12%, which was higher than the other three models. Additionally, the area under the curve value of the Dropout Deep Belief Network model was 0.87, which was 0.04 higher than the binary whale optimization algorithm model, 0.26 higher than the factor decomposition machine and depth neural network model, and 0.05 higher than the random forest network model. The sparse auto-encoder algorithm based on the extreme learning machine algorithm achieved a precision of 99.72%. The study proposes two algorithm models that can effectively identify drugs using near-infrared spectroscopy. This has a positive impact on the medical industry and the safety of patients' lives and health.

Key words: Machine learning; Near-infrared spectroscopy; Deep belief network; Extreme learning machine; Dropout-DBN; ELM-SAE; Drug identifications

1. Introduction. Counterfeit and substandard drugs not only cause huge economic losses, but also cause harm to human body and endanger the health and safety of patients. Therefore, the identification of fake and substandard drugs has always been the focus of scholars from all walks of life [1, 2]. Near-infrared Spectroscopy (NIRS) can detect samples efficiently and accurately without damaging the samples to be tested, so it is used in the testing of pharmaceutical, food, chemical and other industries [3, 4]. However, current drug identification technology is mostly limited to identifying genuine and fake drugs in the second classification. Its effectiveness in identifying multiple varieties of drugs is poor. To solve this problem, a drug identification model based on Dropout Deep Belief Network (DBN) and NIRS is proposed. In addition, aiming at the problem that the training time of machine learning algorithm is too long, a stack Sparse Auto-encoder Algorithm based on Extreme Learning Machine (ELM-SAE) is proposed. The paper has realized the high-precision and efficient identification of drugs, which has a certain role in promoting the development of China's medical industry.

2. Review. The application achievements of NIRS in various fields have gradually matured, and it is also well applied in drug identification. This part lists the application of NIRS technology in drug recognition and explores the latest progress of drug classification models based on deep learning algorithms. The relevant study contents are presented in Table 2.1. In summary, near-infrared spectroscopy technology provides strong support for intelligent drug recognition. When combined with deep learning algorithms, drugs can be effectively classified. This paper reviews relevant literature and elaborates on the important application of NIRS technology in drug recognition. It emphasizes the value of drug classification research based on improved deep learning algorithms. Future research can optimize models to improve recognition accuracy, providing powerful tools for drug quality control and regulation.

*Basic Teaching Department, Chongqing College of Science and Creation, Chongqing, 402160, China (jiulin_xia@outlook.com)

Table 2.1: Literature review

Literature	Author	Time	Time method	conclusion
The importance of wavelength selection in on-scene identification of drugs of abuse with portable NIRS [6].	R. F. Kranenburg, Y. Weesepeel, M. Alewijn, S. Sap, W. F. Peter, A. V. Esch	2022	Use of NIRS for the identification of drugs of abuse	Through visual inspection of the near-infrared spectra, the results yielded new insights into the usability of individual spectrometers
Vibrational spectroscopy in analysis of pharmaceuticals : Critical review of innovative portable and handheld NIR and Raman spectrophotometers [7].	R. Deidda, P. Y. Sacre, M. Clavaud, L. Coic, H. Avohou, P. Hubert, E. Ziemons	2019	Study of portable/handheld near infrared and Raman spectrophotometers in drug identification	Infrared spectroscopy and Raman spectroscopy had consistent advantages in drug analysis, demonstrating enormous potential
Multi-manufacturer drug identification based on near infrared spectroscopy and deep transfer learning [8].	L. Li, X. Pan, W. Chen, M. Wei, Y. C. Feng, L. H. Yin, C. Q. Hu, H. H. Yang	2020	Drug recognition using NIRS combined with deep learning algorithms	CNN-based drug recognition method achieved higher classification accuracy and scalability in multi-species and multi-manufacturer near-infrared spectral classification experiments

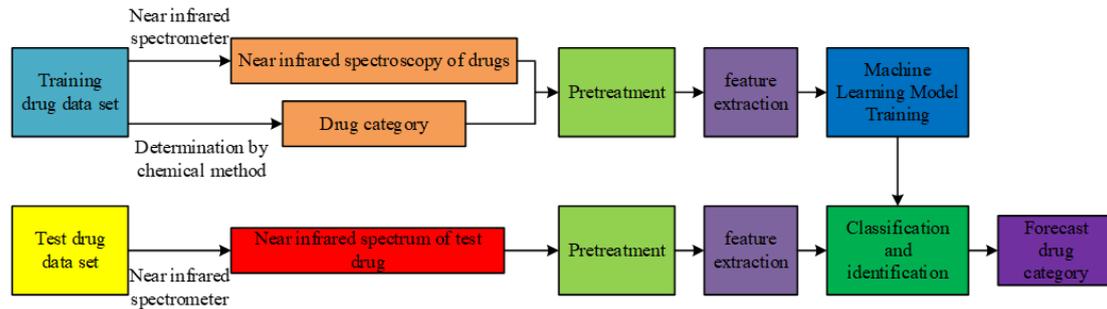


Fig. 3.1: Drug identification process combining NIRS technology and machine learning algorithm

3. Construction of Drug Identification Model Combined with NIRS.

3.1. Construction of Dropout DBN Model. Counterfeit and substandard drugs will not only cause economic losses to patients, but also seriously threaten the life and health safety of patients. Therefore, drug identification has received extensive attention. To avoid the drawbacks of low efficiency, high cost, and potential errors in traditional manual identification methods, many researchers have combined NIRS technology with machine learning algorithms to achieve intelligent drug identification in large quantities with high precision [9, 10]. The drug identification process combining NIRS technology and machine learning algorithm is shown in Figure 3.1.

In Figure 3.1, there are five main steps, namely: NIRS sample data collection, sample data preprocessing, sample data feature extraction, machine learning model construction, and identification, classification and prediction of drugs to be tested. The extraction of sample data features is one of the most important links, which relates to the precision of drug identification models. At present, feature extraction of sample data mainly depends on various algorithms, such as binary whale optimization algorithm (BWOA) [11], random forest network model (RFNM) [12], factor decomposition machine and depth neural network (FM-DNN) [13]. However, the drug identification precision and speed of these algorithms are not ideal, and they need to be further improved [14, 15]. Because of deep learning deep network structure and nonlinear activation function, deep learning model has an excellent performance in high-dimensional, nonlinear big data modeling. Therefore, the application effect of the DBN in drug identification is discussed. However, the DBN model is more suitable for training large-scale data. It is less effective for sample sets with fewer sample numbers and data feature dimensions, and is prone to over-fitting [16]. In drug identification, due to the complexity and difficulty of sample collection and chemical analysis, the number of samples is often small. Therefore, the Dropout technology is introduced into the DBN model. Dropout technology means that during the training process, part of the neural nodes in the hidden layer of the deep learning network and their weights are temporarily removed from the network randomly, as shown in Figure 3.2.

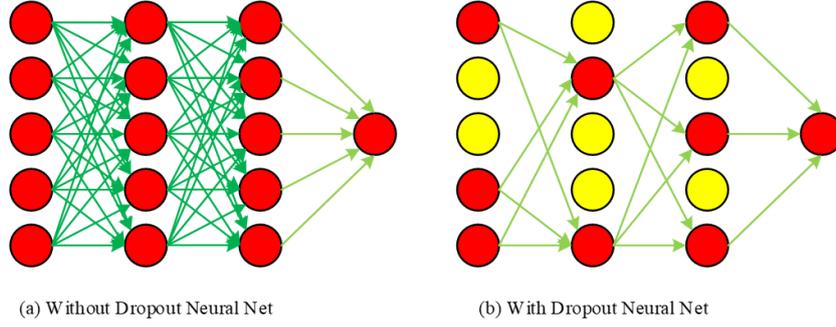


Fig. 3.2: Application effect of Dropout technology

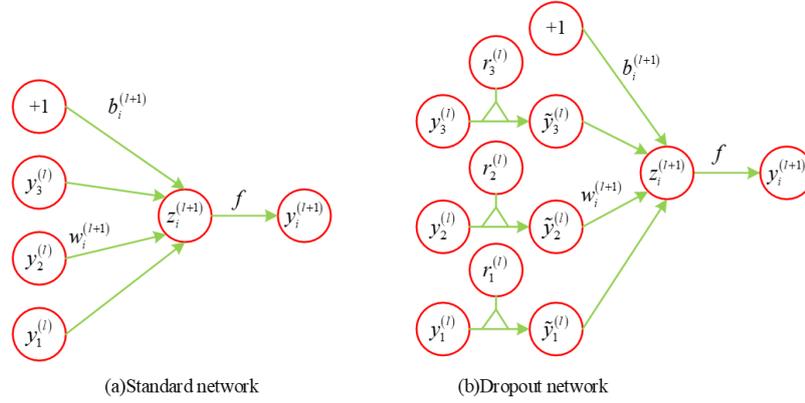


Fig. 3.3: Common network and Dropout network

Through the process shown in Figure 3.2, the Dropout technology can avoid the synergism of some similar features, thus avoiding the decline in model precision caused by over fitting. During training, Dropout technology randomly selects a subnetwork from the learning network containing N nodes to generate $2n$ subnetwork sets. The weight of the sub network set is shared, so its time complexity can still be considered as $O(2n)$. For $2n$ sub network set, only partial training is conducted. Supposing that the hidden layer of a deep learning network has an L layer, we have the index number of the hidden layer, the input vector of layer, the output vector of layer, and the connection weight and offset value of layer, respectively. At this time, the forward network can be described as Formula 3.1.

$$y_i^{l+1} = f(z_i^{(l+1)}) = W_i^{(l+1)} + b_i^{(l+1)} \tag{3.1}$$

In Formula (1.1), $f()$ is the activation function. After the Dropout technology is introduced, the forward propagation neural network and its changes can be shown in Figure 3.3.

In Figure 3.3(b), $r^{(l)}$ is a vector in the hidden layer, and all elements in $r^{(l)}$ obey Bernoulli distribution, which can be expressed as Formula 3.2.

$$r_j^{(l)} = \text{Bernoulli}(p) \tag{3.2}$$

In Formula 3.2, p refers to the sampling probability of Dropout. Using $r^{(l)}$ to sample the output y of the upper layer network, a sub network \hat{y} is obtained as the input of the lower layer network, as shown in Formula 3.3

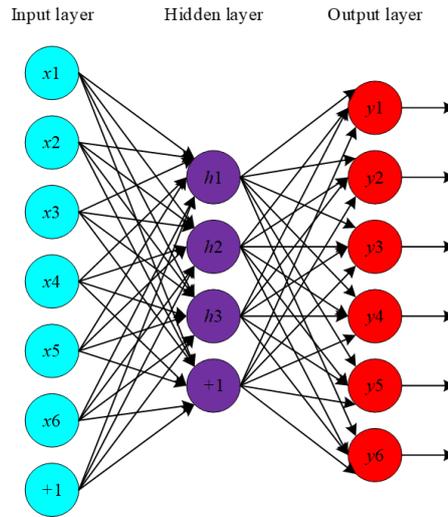


Fig. 3.4: Approximate structure of AEN

and Formula 3.4.

$$\tilde{y}^l = r(l) * y(l) \quad (3.3)$$

$$y_i^{l+1} = f(z_i^{l+1}) = W_i^{l+1} \tilde{y}^l + b_i^{l+1} \quad (3.4)$$

In terms of structure, Dropout DBN is composed of multi-layer restricted Boltzmann machine (RBM) and a layer of BP neural network (BPNN), so it has excellent performance in prediction and classification of high-dimensional feature vectors [17]. The application process of Dropout DBN model in drug identification is as follows: The first is to preprocess the NIRS spectral data of drugs, and input them into the network as training data. Then to set the RBM network structure and number of hidden layers, the learning rate of RBM, and the activation function. Using the training data to train the RBM in the DBN network. After completing the training, to fine-tune the parameters through backpropagation to improve precision. In this process, Logistic classifier is used for the second classification output, and softmax classifier is used for the multi classification output.

3.2. Optimization of Auto-encoder Network. The Dropout DBN model has a good processing effect on small-scale data, but when the drug data is large, its identification accuracy is not ideal. Therefore, for large-scale data, Auto-encoder Network (AEN) network is generally used as the identification model. AEN is the basic model of deep learning algorithms [18]. AEN utilizes an artificial neural network (ANN) to construct a three-layer symmetric network with equal input and output layers. The network is then trained to minimize the error of input and reconstruction data, optimizing the connection weight and offset value of SAE to obtain the internal structural characteristics of the data [19, 20]. The general structure of AEN is shown in Figure 3.4.

The input layer of AEN is set to be x , and there are d input neural units in the input layer. Assuming that the hidden layer of AEN is y , in which there are h nerve units. The input layer of AEN is set as z , which contains the same number of nerve units as the input layer. During training, the data of the model input layer is mapped to the hidden layer, which is called coding, as shown in Formula 3.5.

$$y = f(x) = s(W_y x + b_y) \quad (3.5)$$

In Formula 3.5, $s(\cdot)$ is a nonlinear mapping function, usually represented by a sigmoid function. b_y is the offset value of the hidden layer, and W_y is the weight matrix between the input layer and the hidden layer. After

encoding, the data characteristics of the hidden layer are reconstructed and mapped to the output layer to complete decoding, as shown in Formula 3.6.

$$z = g(x) = s(W_z x + b_z) \quad (3.6)$$

In Formula 3.6, b_z represents the offset value of the output layer, and W_z represents the weight matrix between the output layer and the hidden layer. The eigenvalue z obtained after decoding can be approximately regarded as the eigenvalue of the input data. Therefore, in the process of decoding and reconstruction, weight binding is required to make the weight matrix $W_y = W_z = W$. Therefore, in the training of AEN model, only three groups of parameters, namely W , and b_y , need to be learned. When processing large-scale data with the AEN model, the batch random gradient descent method is typically utilized to obtain the error of small batches of data, which is then used to update the network's weight and bias. Therefore, the AEN model can be expressed as the solution of the optimization problem in Formula 3.7.

$$L(\theta) = \arg \min_{\theta} \frac{1}{n} \sum_{i=1}^n [x_{ik} \log(z_{ik}) + (1 - x_{ik}) \log(1 - z_{ik})] \quad (3.7)$$

In Formula 3.7, $L()$ is the reconstruction error function expressed by cross entropy function, $\theta = \{W, b_y, b_z\}$, that is, three groups of parameters to be learned. x_{ik} is a symbolic function. z_{ik} represents the probability that sample ik is predicted to be positive. The gradient descent method can be used to solve θ , to obtain the required parameters. However, if only the information of input data is saved, AEN model cannot learn and obtain effective feature representation. When the dimensions of the hidden layer and the input layer of the AEN model are the same, it is only necessary to learn an identity mapping to achieve zero error data reconstruction. But this kind of identity mapping does not have enough expression ability for high-level abstract representation. In drug identification, AEN model is required to learn a more complex nonlinear function. To address this issue, constraints must be added to AEN. One approach is to decrease the number of nodes in the hidden layer to reduce the dimensionality of the data. Another option is to include a penalty factor at the input of the network to filter out data noise. To solve these problems, a Sparse Auto-encoder Network (SAEN) is proposed by introducing the sparse idea into DBN. The basic idea of SAEN is that when a hidden layer node has a high probability of being activated, the node represents relatively little information. SAEN can add constraints to the activation function of hidden layer neurons, so that hidden layer neurons are in a state of inhibition for a long time, thus expressing more information. Generally speaking, the activation function of neurons adopts the sigmoid function. At this time, the average activation value of hidden layer node j in all samples can be expressed as Formula 3.8.

$$\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^m f_{\theta}^j(x^{(i)}) \quad (3.8)$$

In Formula 3.8, m is the number of training samples. In AEN, $\hat{\rho} = \rho$ is used to limit the activation probability of hidden layer nodes. ρ is a sparsity parameter, which is generally close to 0. After the AEN is limited by the sparse parameter, the sparse penalty term should be added to the loss function of AEN. The relative entropy function, namely KL divergence algorithm, is used to add Formula 3.9 to the loss function of AEN.

$$KL(\rho || \hat{\rho}) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \left(\frac{1 - \rho}{1 - \hat{\rho}_j} \right) \quad (3.9)$$

After adding restrictions, the self-coding cost function of AEN can be expressed by Formula 3.10.

$$L_{\rho}(\theta) = L(\theta) + \beta \sum_{j=1}^h KL(\rho || \hat{\rho}) \quad (3.10)$$

In Formula 3.10, β is the weight parameter of the sparse penalty. Due to the sparse penalty term, the sigmoid function causes the activation value of most hidden layer neural units to be close to 0, with only a few data

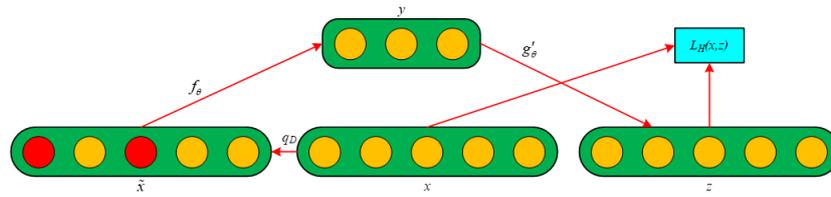


Fig. 3.5: Noise reduction AEN

points producing large activation values. In this case, AEN can learn higher and more complex abstract features. To improve the robustness of AEN, a noise reduction AEN is proposed. The basic idea is to input noisy sample data at the input layer, and then encode and decode the noisy data. In this process, the output sample data after AEN reconstruction should keep the original information of the input data as much as possible. Through the above operations, the robustness of the hidden layer can be improved to better learn high-level abstract features. If the original input vector is, after the noise is added, the input vector is represented as \tilde{x} . At this time, the coding and decoding of the noise reduced AEN can be represented by Formula 3.11.

$$\begin{cases} y = f(\tilde{x}) = s(W_y x + b_y) \\ z = g(y) = s(\widetilde{W}_z x + b_y) \end{cases} \quad (3.11)$$

In Formula 3.11, \widetilde{W} is the weight after adding noise. There are generally two ways to add noise in noise reducing AEN. The first is to add Gaussian noise to the input sample data, as shown in Formula 3.12.

$$\tilde{x} = x + \epsilon \quad (3.12)$$

In Formula 3.12, ϵ is Gaussian noise. The second method is to randomly assign part of the vectors of the input sample data to 0 to add binary masking noise. The specific process of this method is to set a scale at first, then to select the components of this scale in the input sample data, and yo assign these components to 0. The details are shown in Figure 3.5.

3.3. Stack SAE Based on ELM. To enable AEN to learn high-level abstract features, to achieve drug identification, AEN has been improved and a sparse noise reduction AEN has been proposed [21, 22]. SAE is a neural network model. SAE adopts layer by layer greedy training method, and trains each layer of SAE from front to back. The hidden layer representation of the SAE layer is used as the input vector for the next layer of SAE during training. The weight and offset values of SAE are obtained layer by layer. Higher-level abstract features can be extracted from the above operations. At the top level of SAE, Logistic or Softmax classifiers are introduced to achieve binary or multi classification of data. Finally, BPNN algorithm is used to fine-tune the entire neural network. However, the way of BPNN algorithm to optimize weights and offsets is based on gradient descent method, so all parameters need to be modified in all iteration processes, and the training time is extremely expensive. In addition, the neural network trained based on the gradient descent method is easy to obtain local optimal solutions [23]. To solve this problem, ELM is introduced to replace the BPNN algorithm to adjust and optimize SAE, reduce training time, and improve training efficiency and practicality. During training, ELM can obtain the output weight of the hidden layer as long as the number of hidden layer neurons is set in advance [24]. Therefore, as long as the number of neurons in the hidden layer is set in advance, the output weight of the hidden layer can be obtained. The ELM model is shown in Figure 3.6.

In Figure 3.6, N groups of independent input samples are set, with and . When the ELM model has M hidden layer nodes, the ELM model can be expressed as Formula 3.13.

$$f(x) = \sum_{k=1}^M \beta_k G(W_k X_j + b_k) = o_j, 1, 2, \dots, N \quad (3.13)$$

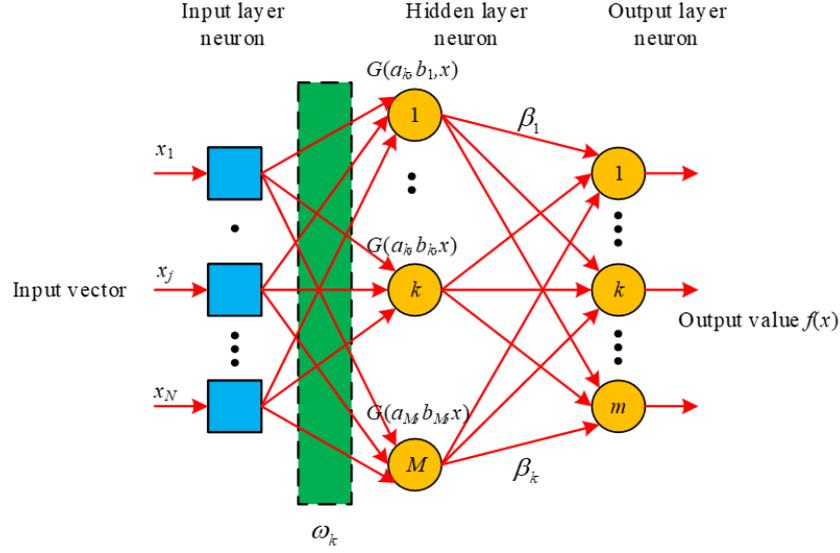


Fig. 3.6: ELM model

In Formula 3.13, $G()$ is the activation function. $W_k = [\omega_{k1}, \omega_{k2}, \dots, \omega_{kn}]^T$ is the input weight matrix. $\beta_k = [\beta_{k1}, \beta_{k2}, \dots, \beta_{kn}]^T$ is the output weight matrix. b_k is the k -th offset value in the hidden layer. o_j is the network output value. Compared with the BPNN algorithm, ELM has excellent generalization ability and training efficiency, and does not require too much manual intervention [25]. Therefore, ELM is used to adjust SAE and construct the ELM-SAE algorithm. The output of the last hidden layer of the SAE model is set to be H_{n-1} , the n th hidden layer contains \hat{N} nodes, and the $n - 1$ th hidden layer contains m nodes. Then the neural network function can be expressed by Formula (1.14).

$$f(x) = \sum_{n=1}^{\hat{N}} \beta_n G(W_n H_{n-1} + b_n) = o_j, 1, 2, \dots, m \quad (3.14)$$

In Formula 3.14, W_N represents the connection weight between the $n - 1$ -th hidden layer and the n -th hidden layer. β_n is the connection weight between the n -th hidden layer and the output layer. b_n represents the offset value between the $n - 1$ -th hidden layer and the n -th hidden layer. For SAE, its training goal is to minimize the error between the actual output and the expected output, as shown in Formula 3.15.

$$\sum_{n=1}^{\hat{N}} = \|o_j - t_j\| \quad (3.15)$$

Formula 3.15 can be converted to Formula (1.16).

$$H_n \beta = T \quad (3.16)$$

In Formula 3.16, H_n is the output matrix of hidden layer nodes. β is the output weight. T is the desired output. To minimize the output error of the ELM-SAE algorithm, the $\widehat{W}_k, \widehat{b}_k, \widehat{\beta}_k$ value is obtained through continuous updating and iteration to make it meet the Formula 3.17.

$$\left\| H_n(\widehat{W}_k, \widehat{b}_k) \widehat{\beta}_k - T \right\| = \min_{w, b, \beta} (W_k, b_k) \beta - T \quad (3.17)$$

In Formula 3.17, $k = 1, 2, \dots, \hat{N}$. Formula 3.17 can be equivalent to the minimum loss function, as shown in Formula 3.18.

$$E = \sum_{j=1}^m \left(\sum_{k=1}^{\hat{N}} \beta_k G(W_k H_{(n-1)j} + b_k) - t_j \right)^2 \quad (3.18)$$

Randomly to initialize the connection weight between the $n - 1$ -th hidden layer and the n -th hidden layer, and the offset value between the $n - 1$ -th hidden layer and the n -th hidden layer to obtain the unique output matrix H_n of the hidden layer. The training process of the ELM-SAE algorithm can be converted into a solution problem of a linear system, as shown in Formula (1.16). At this time, the output weight can be determined, as shown in Formula 3.19.

$$\hat{\beta} = H_n^T T \quad (3.19)$$

In Formula 3.19, $H_n^T T$ is the Moore penrose generalized inverse of the hidden layer output matrix, which can be solved by the generalized inverse theorem and singular value decomposition method. Based on the above contents, the ELM-SAE algorithm is constructed to make up for the long training time of machine learning.

4. Analysis of the Application Effect of Dropout DBN and ELM-SAE in Drug Identification.

4.1. Training Effect Analysis of Dropout DBN Model. In drug identification, the amount of sample data is generally small. To solve the problem of binary classification and multi-classification modeling of NIRS in small sample datasets, a Dropout DBN algorithm based on machine learning is proposed. To verify the identification effect of the Dropout DBN algorithm on infrared spectrum drugs in a small sample data set, the research uses the data collected by the China Institute for Food and Drug Control to test the performance of the algorithm. 500 spectral sample data of erythromycin ethylsuccinate were selected as the sample data, 300 of which were used as the training sample set, and the other 200 were used as the test sample set. The performance of the Dropout DBN algorithm with BWOA, RFNM, FM-DNN and other common drug identification algorithms were compared. During the training process, the loss values of several algorithms [26, 27, 28] change as shown in Figure 4.1.

In Figure 4.1, after 100 iterations, the loss change rate of the four models slows down. Among them, the loss value of the Dropout DBN model is always lower than that of the other three models. After 150 iterations, the loss value of the four models is the lowest, and the loss value of the Dropout DBN model is 0.02. The loss value of the BWOA model is 0.05, which is 0.03 higher than that of the Dropout DBN model. The loss value of RFNM model is 0.02, but the number of iterations when it reaches the lowest value is significantly more than that of Dropout DBN. The loss value of FM-DNN is 0.03, which is 0.01 higher than the Dropout DBN model. This shows that the Dropout DBN model has better convergence and stability, higher prediction precision and a more ideal model.

4.2. Classification Accuracy Analysis of the Dropout DBN Model. After training, the classification precision of the test sample set for the four models was used, and the test results are shown in Figure 4.2.

In Figure 4.2, the precision of the Dropout DBN model is significantly higher than the other three algorithms. When the number of iterations is less than 50, the precision rate of the four models is constantly improving. After the number of iterations reaches 100, the precision rate of the four models slows down. After the precision becomes stable, the precision of the Dropout DBN model reaches 99.12%. The precision of the BWOA model is 98.63%, 0.49% lower than that of the Dropout DBN model. The precision of the RFNM model is 96.82%, and that of the FM-DNN model is 94.12%, 2.30% and 5.00% lower than the Dropout DBN model respectively. The above results show that when the number of samples is small, the Dropout DBN model has high precision in drug identification. The performance of the above four models was tested using ROC, as shown in Figure 4.3.

In Figure 4.3, the Area under the Curve (AUC) value of the Dropout DBN model is obviously superior to the other three models. The AUC value of the Dropout DBN model is 0.87, and that of the BWOA model is 0.83, 0.04 lower than that of the Dropout DBN model. The AUC value of the FM-DNN model is 0.61, 0.26 lower than that of Dropout DBN model. AUC value of RFNM model is 0.82, which is 0.05 lower than that of Dropout DBN model. This shows that the Dropout DBN model proposed in the study can effectively identify the infrared spectrum of drugs.

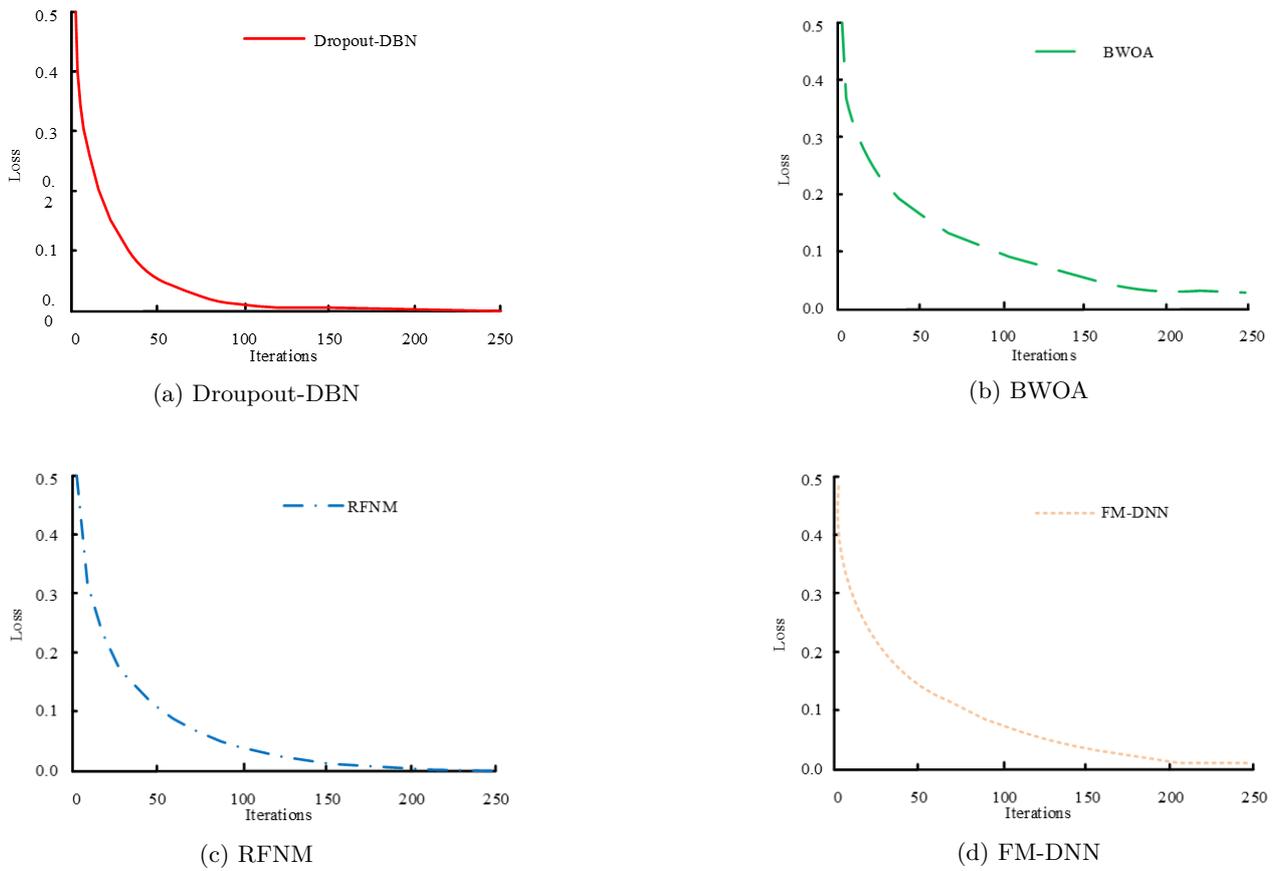


Fig. 4.1: Loss value change of several algorithms

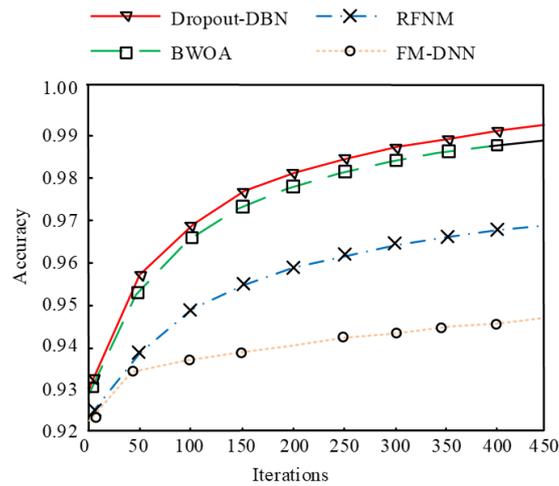


Fig. 4.2: Drug identification precision of four models

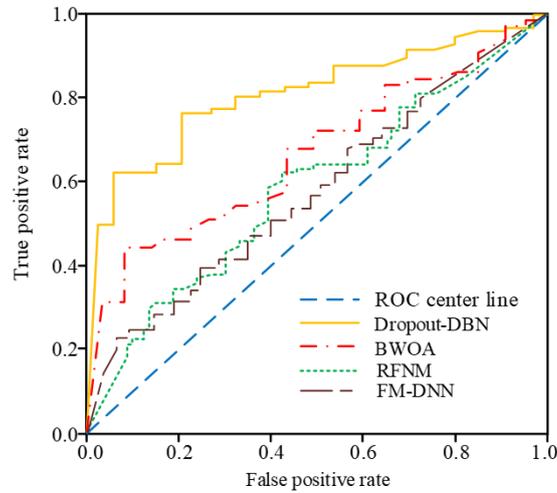


Fig. 4.3: AUC values of four models

Table 4.1: Classification precision of three algorithms for erythromycin ethylsuccinate data sets

Training Sets/Test Sets	Precision/%		
	ELM	SAE-BPNN	ELM-SAE
60/189	91.31	94.48	95.86
80/169	92.78	97.30	97.34
100/149	95.56	98.08	98.26
120/129	95.81	99.05	98.68
140/109	97.32	99.42	98.85
160/89	98.05	99.03	99.34
180/69	98.41	98.96	99.72

4.3. Classification Accuracy Analysis of the ELM-SEA Algorithm. The training time of the machine learning algorithm is slow. To solve this problem, ELM is used to optimize SAE and construct the ELM-SAE algorithm. To verify the performance of the ELM-SEA algorithm, the data of erythromycin ethylsuccinate collected by China Institute for Food and Drug Control was used to test it. ELM algorithm, SAE-BPNN algorithm and ELM-SAE algorithm are respectively used to process and identify erythromycin ethylsuccinate data sets of different orders of magnitude, and the recognition precision and learning efficiency of several algorithms are compared. The classification precision of the three algorithms for erythromycin ethylsuccinate data sets is in Table 4.1.

Table 4.1 shows the classification precision of the SAE-BPNN model is the highest, reaching 99.05%, except when the test set/training set is 120/129. In other data sets, the classification precision of the ELM-SAE model is the highest among the three models. The precision of both ELM-SAE and ELM models increases with an increase in training samples. The highest precision achieved by ELM-SAE is 99.72%, while the highest precision achieved by the ELM model is 98.41%. The classification precision of the SAE-BPNN model starts to decline when the training samples exceed a certain number. This is because the SAE-BPNN model has produced an over-fitting phenomenon and is trapped in local optimization, resulting in a decline in model precision. The training time of the three algorithms on data sets of different orders of magnitude is shown in Table 4.2.

In Table 4.2, among the three models, the training time of ELM model on each dataset is much lower than that of the other two models. However, Table 1 shows that the drug classification precision of ELM model is not ideal. In Table 2, the training time of ELM-SAE model on each data set is significantly lower than that of

Table 4.2: Training time of three algorithms on data sets of different orders of magnitude

Training Sets/Test Sets	Training time/s		
	ELM	SAE-BPNN	ELM-SAE
60/189	0.011	28.78	19.34
80/169	0.013	26.84	17.43
100/149	0.014	27.87	18.64
120/129	0.016	28.68	19.25
140/109	0.016	29.66	20.08
160/89	0.016	42.89	30.13
180/69	0.016	46.56	33.10

SAE-BPNN model. This shows that using ELM algorithm to optimize SAE can greatly improve the training efficiency and reduce the training time of near-infrared spectral data on the premise of ensuring the precision of drug identification. It is an effective method for near-infrared spectral drug identification. To sum up, the two algorithm models proposed in the study can effectively achieve NIRS drug identification, indicating that machine learning has broad application prospects in NIRS drug identification.

5. Conclusion. Machine learning has important applications in various fields. The application ways and effects of machine learning in NIRS drug identification were studied. Aiming at the problem of binary and multi-classification modeling of NIRS in small sample datasets, a Dropout DBN algorithm based on machine learning was proposed. Aiming at the shortcomings of low learning efficiency and long learning time of machine learning algorithms, an ELM-SAE algorithm was proposed. After testing the model with experimental data, the Loss value of the Dropout DBN model was 0.02, 0.03 lower than the BWOA model and 0.01 lower than the FM-DNN model. The precision of the Dropout DBN model reached 99.12%, 0.49%, 2.30% and 5.00% higher than the BWOA model, RFNM model and FM-DNN model, respectively. On the data sets of each order of magnitude, the classification precision of the ELM-SAE model was the highest among the three models. The precision of both ELM-SAE and ELM models increased with the increase of training samples. The highest precision of ELM-SAE and ELM model was 99.72% and 98.41%. However, the classification precision of the SAE-BPNN model started to decline after the training samples exceeding a certain number, which indicates that it has produced an over-fitting phenomenon. The training time of the ELM-SAE model on each data set was significantly lower than that of the SAE-BPNN model. To sum up, the two algorithm models proposed in the study can effectively achieve NIRS drug identification, indicating that machine learning has broad application prospects in NIRS drug identification. The study was conducted only with an erythromycin ethylsuccinate data set without testing the identification effect of the model on other drugs, which may lead to deviation in the experimental results. Therefore, the scope of the study needs to be expanded in the future.

Fundings. The research is supported by Science and Technology Research Program of Chongqing Municipal Education Commission: Preprocessing and Comparison Analysis of Near Infrared Spectrum Data Based on Different Objects (No. KJQN202205402); Natural Science Foundation of Yongchuan District, Chongqing: Research on Food and Drug Identification Model Based on Near Infrared Spectrum and HHT Transform (No. 2022yc-jckx20041).

REFERENCES

- [1] Kademi, H., Ulusoy, B. & Hecer, C. Applications of miniaturized and portable near infrared spectroscopy (NIRS) for inspection and control of meat and meat products. *Food Reviews International*. **35**, 201-220 (2019)
- [2] Zhang, H., Feng, J., Chen, S., Zhao, Z., Li, B., Wang, Y., Jia, J., Li, S., Wang, Y., Yan, M., Lu, K. & Hao, H. Geographical patterns of nirS gene abundance and nirS-type denitrifying bacterial community associated with activated sludge from different wastewater treatment plants. *Microbial Ecology*. **77**, 304-316 (2019)
- [3] Vesoulis, Z., Mintzer, J. & Chock, V. Neonatal NIRS monitoring: recommendations for data capture and review of analytics. *Journal Of Perinatology*. **41**, 675-688 (2021)

- [4] Paul, A., Wander, L., Becker, R., Goedecke, C. & Braun, U. High-throughput NIR spectroscopic (NIRS) detection of microplastics in soil. *Environmental Science And Pollution Research*. **26**, 7364-7374 (2019)
- [5] Han, C., Müller, K. & Hwang, H. Enhanced performance of a brain switch by simultaneous use of EEG and NIRS data for asynchronous brain-computer interface. *IEEE Transactions On Neural Systems And Rehabilitation Engineering*. **28**, 2102-2112 (2020)
- [6] Kranenburg, R., Weesepeel, Y., Alewijn, M., Sap, S., Peter, W. & Esch, A. The importance of wavelength selection in on-scene identification of drugs of abuse with portable near-infrared spectroscopy. *Forensic Chemistry*. **30** (2022)
- [7] Deidda, R., Sacre, P., Clavaud, M., Coic, L., Avohou, H., Hubert, P. & Ziemons, E. Vibrational spectroscopy in analysis of pharmaceuticals: Critical review of innovative portable and handheld NIR and Raman spectrophotometers. *TrAC Trends In Analytical Chemistry*. **114** pp. 251-259 (2019)
- [8] Li, L., Pan, X., Chen, W., Wei, M., Feng, Y., Yin, L., Hu, C. & Yang, H. Multi-manufacturer drug identification based on near infrared spectroscopy and deep transfer learning. *Journal Of Innovative Optical Health Sciences*. **13** pp. 50016-20500 (2020)
- [9] Mehmood, A., Kaushik, A., Wang, Q., Li, C. & Wei, D. Bringing structural implications and deep learning-based drug identification for KRAS mutants. *Journal Of Chemical Information And Modeling*. **61**, 571-586 (2021)
- [10] Ding, Y., Tang, J. & Guo, F. Identification of drug-target interactions via fuzzy bipartite local model. *Neural Computing And Applications*. **32**, 10303-10319 (2020)
- [11] Yusof, N., Muda, A., Pratama, S. & Abraham, A. A novel nonlinear time-varying sigmoid transfer function in binary whale optimization algorithm for descriptors selection in drug classification. *Molecular Diversity*. **2022** pp. 1-10 (2022)
- [12] Zhao, X., Chen, L., Guo, Z. & Liu, T. Predicting drug side effects with compact integration of heterogeneous networks. *Current Bioinformatics*. **14**, 709-720 (2019)
- [13] Wang, J., Wang, H., Wang, X. & Chang, H. Predicting drug-target interactions via FM-DNN learning. *Current Bioinformatics*. **15**, 68-76 (2020)
- [14] Dargan, S., Kumar, M., Ayyagari, M. & Kumar, G. A survey of deep learning and its applications: a new paradigm to machine learning. *Archives Of Computational Methods In Engineering*. **27**, 1071-1092 (2020)
- [15] Ma, W., Liu, Z., Kudyshev, Z., Boltasseva, A., Cai, W. & Liu, Y. Deep learning for the design of photonic structures. *Nature Photonics*. **15**, 77-90 (2021)
- [16] Wang, Y., Pan, Z., Yuan, X., Yang, C. & Gui, W. A novel deep learning based fault diagnosis approach for chemical process with extended deep belief network. *ISA Transactions*. **96** pp. 457-467 (2020)
- [17] Li, L., Pan, X., Yang, H., Zhang, T. & Liu, Z. Supervised dictionary learning with regularization for near-infrared spectroscopy classification. *IEEE Access*. **7** pp. 923-10093 (2019)
- [18] Otto, S. & Rowley, C. Linearly recurrent autoencoder networks for learning dynamics. *SIAM Journal On Applied Dynamical Systems*. **18**, 558-593 (2019)
- [19] Karimpouli, S. & Tahmasebi, P. Segmentation of digital rock images using deep convolutional autoencoder networks. *Computers & Geosciences*. **126** pp. 142-150 (2019)
- [20] Wang, M., Zhao, M., Chen, J. & Rahardja, S. Nonlinear unmixing of hyperspectral data via deep autoencoder networks. *IEEE Geoscience And Remote Sensing Letters*. **16**, 1467-1471 (2019)
- [21] Khamparia, A., Saini, G., Pandey, B., Tiwari, S., Gupta, D. & Khanna, A. KDSAE: Chronic kidney disease classification with multimedia data learning using deep stacked autoencoder network. *Multimedia Tools And Applications*. **79**, 35425-35440 (2020)
- [22] Cui, M., Wang, Y., Lin, X. & Zhong, M. Fault diagnosis of rolling bearings based on an improved stack autoencoder and support vector machine. *IEEE Sensors Journal*. **21**, 4927-4937 (2020)
- [23] Li, D., Fu, Z. & Xu, J. Stacked-autoencoder-based model for COVID-19 diagnosis on CT images. *Applied Intelligence*. **51**, 2805-2817 (2021)
- [24] Yaseen, Z., Sulaiman, S., Deo, R. & Chau, K. An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. *Journal Of Hydrology*. **569** pp. 387-408 (2019)
- [25] Manoharan, J. Study of variants of Extreme Learning Machine (ELM) brands and its performance measure on classification algorithm. *Journal Of Soft Computing Paradigm (JSCP), V Qzol.* **3** pp. 02 (2021)
- [26] Gouda, W., Almurafteh, M., Humayun, M. & Jhanjhi, N. Detection of covid-19 based on chest x-rays using deep learning. *Healthcare*. **10**, 343 (2022)
- [27] Khalil, M., Humayun, M., Jhanjhi, N., Talib, M. & Tabbakh, T. Multi-class segmentation of organ at risk from abdominal ct images: A deep learning approach. *Intelligent Computing And Innovation On Data Science: Proceedings Of ICTIDS 2021*. pp. 425-434 (2021)
- [28] Aldughayfiq, B., Ashfaq, F., Jhanjhi, N. & Humayun, M. Yolo-based deep learning model for pressure ulcer detection and classification. *Healthcare*. **11**, 1222 (2023)

Edited by: Chiranji Lal Chowdhary

Special issue on: Scalable Machine Learning for Health Care: Innovations and Applications

Received: Jul 4, 2023

Accepted: Feb 19, 2024