sensors

MDPI

Communication

Identification of Multi-Class Drugs Based on Near Infrared Spectroscopy and Bidirectional Generative Adversarial Networks

Anbing Zheng [1], Huihua Yang [1,2,*], Xipeng Pan [2], Lihui Yin [3] and Yanchun Feng [3]

[1] School of Automation, Beijing University of Posts and Telecommunications, 10 Xitucheng Road, Haidian District, Beijing 100086, China; spztf@bupt.edu.cn
[2] School of Computer Science and Information Security, Guilin University of Electronic Technology, No.1 Jinji Road, Qixing District, Guilin 541004, China; pxp201@guet.edu.cn
[3] China Institute for Food and Drug Control, 2 Tiantan Xili, Dongcheng District, Beijing 100086, China; yinlihui@nifdc.org.cn (L.Y.); fyc@nifdc.org.cn (Y.F.)
* Correspondence: yhh@bupt.edu.cn

Abstract: Drug detection and identification technology are of great significance in drug supervision and management. To determine the exact source of drugs, it is often necessary to directly identify multiple varieties of drugs produced by multiple manufacturers. Near-infrared spectroscopy (NIR) combined with chemometrics is generally used in these cases. However, existing NIR classification modeling methods have great limitations in dealing with a large number of categories and spectra, especially under the premise of insufficient samples, unbalanced samples, and sensitive identification error cost. Therefore, this paper proposes a NIR multi-classification modeling method based on a modified Bidirectional Generative Adversarial Networks (Bi-GAN). It makes full utilization of the powerful feature extraction ability and good sample generation quality of Bi-GAN and uses the generated samples with obvious features, an equal number between classes, and a sufficient number within classes to replace the unbalanced and insufficient real samples in the courses of spectral classification. 1721 samples of four kinds of drugs produced by 29 manufacturers were used as experimental materials, and the results demonstrate that this method is superior to other comparative methods in drug NIR classification scenarios, and the optimal accuracy rate is even more than 99% under ideal conditions.

Keywords: near-infrared spectroscopy; drug identification; multi-class classification; deep learning; generative adversarial networks

Citation: Zheng, A.; Yang, H.; Pan, X.; Yin, L.; Feng, Y. Identification of Multi-Class Drugs Based on Near Infrared Spectroscopy and Bidirectional Generative Adversarial Networks. *Sensors* **2021**, *21*, 1088. https://doi.org/10.3390/s21041088

Academic Editor: Simona M. Cristescu
Received: 9 December 2020
Accepted: 28 January 2021
Published: 5 February 2021

1. Introduction

In the drug market, different drugs and different brands will have different pricing. Sellers can use fake packaging on low-cost pharmaceutical products and sell them as high-priced drugs. They may also use inferior brand drugs of the same drug as famous brand products to sell at high prices in the market. Therefore, it is of great significance in drug supervision to identify the true source of drugs by classification and identification of multiple drugs produced by multiple manufacturers.

Near-infrared spectroscopy (NIR) has the advantages of low instrument cost, direct measurement, non-destructive detection, and on-site detection, which is suitable for rapid qualitative and quantitative analysis of drugs [1–3]. It is usually combined with chemometrics methods such as partial least squares discriminant analysis (PLS-DA) [3–5], linear support vector machine (Linear SVM), and other linear classifiers [6–10] and BP-ANN classifier [10,11] in a classification scenario. In recent years, some deep learning methods, such as stack sparse auto-coding (SAE) [12], deep belief network (DBN) [13], deep convolution neural network (CNN) [14], have also been reported in drug identification and classification modeling.

Classification and identification are also very important in the analysis and processing scenarios of sensors' collected data, so the above methods are often used in other sensor data analysis scenarios, not only in the chemometrics domain or NIR domain, and the following problems of NIR classification and identification are also common in the scene of sensor data acquisition, analysis, and processing.

Due to the famous "curse of dimensionality" problem [15] in classification methods and the low quantitative detection limit of near-infrared spectroscopy [2], the combination of traditional chemometrics and near-infrared spectroscopy is not suitable for scenes with a large number of categories and spectra. Meanwhile, the classifiers of deep learning methods are suitable but often need enough samples to participate in training to achieve the desired effect. Thus, both traditional chemometrics methods and deep learning methods require sufficient samples within the class and balanced samples between classes in modeling.

However, in the practice of drug inspection, the number of drug varieties on the market is large, usually involving a large number of categories and a large number of spectra. At the same time, the spectrum samples available are usually unbalanced, and the number of spectra within each category is usually a long-tail trend, that is, the number of intra-class spectra in a few categories is particularly large, while the number of intra-class spectra of more categories is often insufficient.

This will lead to two important adverse consequences: "false high" of accuracy and uneven distribution of model error.

Due to a large number of classes are insufficient in samples, which leads to inadequate extraction of class features, resulting in the low classification accuracy of these classes. However, since the test proportions of them are also low, the urgency of their lower accuracy will never be exposed compared with the dominantly large classes' high accuracy. The accuracy rate of the whole classification may be very high, but for some of the important classes, their accuracies are even too low to be used at all.

Besides, for the finished classifier, due to the imbalance of the number of spectra between classes, the distribution of classification errors is also uneven. Most of the misclassified spectra are classified into the categories with fewer spectra within the class, while the categories with more spectra are very few. Generally, the classes with more intra-class spectra are drugs that are easy to collect samples, such as cheap drugs and common drugs, while those with fewer intra-class spectra are rare drugs that are not easy to obtain samples, such as expensive drugs or newly developed special drugs. If the classifier often classifies common drugs into rare drugs, the consequences are often serious, which means that cheaters cheat successfully, the interests of consumers are greatly damaged, and even the lives of patients are endangered. The classification of rare drugs as common drugs is not so important, because the owners of rare drugs often ask for further inspection and provide other more reliable evidence so that the correct results can always be obtained.

Although the most fundamental way to solve the above problems is to collect more samples, it is usually conditional and expensive, and it is difficult to ensure that a complete, reliable, and sufficient number of samples are collected for specific problems. Under this premise, the second method, which uses high-quality and diversified sample generation methods, generates samples on demand [16] for classification, becomes the most worthy method to be considered.

In recent years, the generative adversarial networks [17] (GAN) method, which is popular in the field of deep learning, is just in line with this idea. Since the original GAN was reported in 2014, it has experienced a long time of rapid development and has evolved many practical and mature sample generation methods [18], such as conditional GAN (C-GAN) [19,20], deep convolutional GAN (DC-GAN) [21], bidirectional GAN (Bi-GAN) [22], cycle GAN [23], etc. In the fields of images [24–27], videos [28–32], voices [33–38] and even natural languages [39–42], excellent sample generation effects have been achieved. "Deep-Fake" programs [43–45] based on a modified GAN can even generate synthetic artificial faces that can't be distinguished from real ones by both humans and machines, which leads to the discussion of artificial intelligence ethics because of its excellent sample generations.

Unfortunately, in the field of near-infrared spectrum detection (including the field of drug near-infrared spectrum detection), and even in the wider field of sensor data processing, there is no relevant report on the application of GAN methods. Therefore, we can only modify an appropriate GAN method that is most suitable for the feature extraction and the generation of category samples on demand. The candidate methods are Info-GAN [46], Bi-GAN, VAE-GAN [47], etc. according to our experience and the implementation difficulties, we finally select Bi-GAN as the modification object to achieve the goal of this paper.

Based on this background, this paper constructs a multi-classification model of drugs based on near-infrared spectroscopy and Bi-GAN sample generation, so that it can correctly classify in the scene of a large number of categories and spectra, and effectively solve the problems of insufficient samples, unbalanced samples, and cost-sensitive classification errors in the classification process.

In the scenario of a large number of categories and a large number of spectra, the problems of insufficient samples, unbalanced samples, and cost sensitivity of classification errors are common in other sensor data classification tasks, so this method can also be applied to data classification tasks obtained by other sensors.

2. Materials and Methods

2.1. Materials

All the materials used in this paper were obtained from the China Institute for Food and Drug Control (Beijing, China). A total of 1721 samples of four drugs (metformin hydrochloride tablets, chlorpromazine hydrochloride tablets, chlorphenamine maleate tablets, cefuroxime axetil tablets) produced by 29 manufacturers were collected. All samples were measured by FTIR spectroscopy (Matrix F spectrometer, Bruker Corporation, Billerica, MA, USA). Before sample collection, the instrument passed a self-diagnosis test and calibration. The wavelength range of data is 4000–11,995 cm^{-1}, and the resolution is 4 cm^{-1}.

The near-infrared spectra of the drugs were recorded using a diffuse reflection optical fiber probe. SMA 905 standard interfaces were used for coupling the optical fiber, light source and spectrometer. The ambient temperature was 18–30 °C, and the air humidity was less than 70%. All samples used the same determination background. The measurement operation followed a unified operation protocol, as shown in Figure 1.

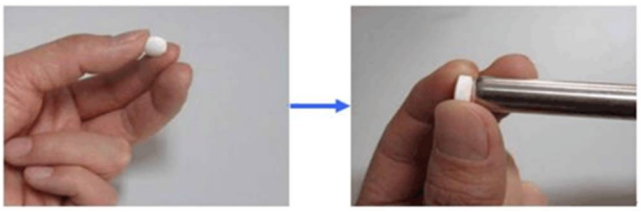

Figure 1. Operation method for spectrometric measurement of ordinary tablets.

Sample information is shown in Table 1. It can be seen from this table that the number of samples is not balanced. Some classes have more samples, reaching 135, while others have fewer samples, only 21. The sample numbers within the class are sorted from high to low to form a column chart, as shown in Figure 2, and the distribution histogram is shown in Figure 3.

Table 1. Drug Names, Manufacturers, Sample Names of the material used in this paper.

No.	Drug Name	Manufacturer	Samples
1	metformin hydrochloride tablets	Xinyi Pharmaceutical Factory Co., Ltd., Shanghai, China	94
2		Hengshan Pharmaceutical Co., Ltd. Shanghai, China	48
3		Zhonghui Pharmaceutical Co., Ltd., Beijing, China	67
4		Yongkang Pharmaceutical Co., Ltd., Beijing, China	21
5		Pacific Pharmaceutical Co., Ltd., Tianjin, China	48
6		Chuanli Pharmaceutical Co., Ltd., Chengdu, China	64
7		Baiyunshan Tangyin Dongtai Pharmaceutical Co., Ltd., Guangzhou, China	27
8		Qilu Pharmaceutical Co., Ltd. Ji'nan, China	35
9		Suzhong pharmaceutical group Co., Ltd., Taizhou, China	48
10		Jingfeng Pharmaceutical group Co., Ltd., Beijing, China	24
11		Zhonglian Pharmaceutical Co., Ltd., Shenzhen, China	68
12		Zhongxin Pharmaceutical Group Co., Ltd., Tianjin, China	97
13		Yabao Pharmaceutical Technology Co., Ltd., Tianjin, China	97
14		Kangrui Pharmaceutical Co., Ltd., Tianjin, China	97
15	chlorpromazine hydrochloride tablets	Xinyi Jiufu Pharmaceutical Co., Ltd., Shanghai, China	59
16		Yungang Pharmaceutical Co., Ltd., Datong, China	94
17		Changzhou Compass Pharmaceutical Co., Ltd., Changzhou, China	58
18		Guangdong Petty Pharmaceutical Co., Ltd., Guangzhou, China	135
19		Jiangsu Tianshili Diyi Pharmaceutical Co., Ltd., Huai'an, China	49
20	chlorphenamine maleate tablets	Taiyang Pharmaceutical Co., Ltd., Beijing, China	39
21		Shanxi Taiyuan Pharmaceutical Co., Ltd., Taiyuan, China	45
22		Shanxi Xinxing Pharmaceutical Co., Ltd. Linfen, China	36
23		Guangdong Nanguo Pharmaceutical Industry, Zhanjiang, China	39
24		Henan Jiushi Pharmaceutical Co., Ltd., Huixian, China	94
25	cefuroxime axetil tablets	Baiyunshan Tianxin Pharmaceutical Co., Ltd., Guangzhou, China	56
26		Beit Pharmaceutical Co., Ltd., Chengdu, China	29
27		Zhijun Pharmaceutical Co., Ltd. Shenzhen, China. (0.125 mg)	27
28		Zhijun Pharmaceutical Co., Ltd. Shenzhen, China. (0.25 mg)	89
29		United Laboratories International Ltd. (Zhongshang branch), Zhongshang, China	37
Total			1721

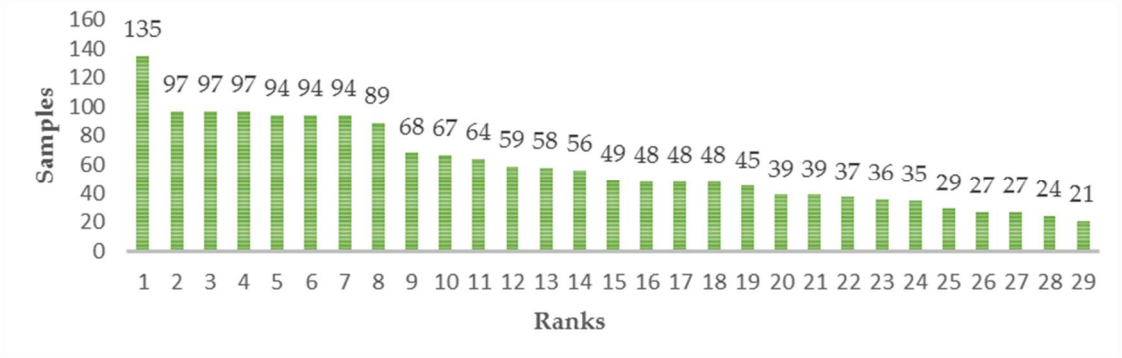

Figure 2. Intra-class samples are sorted from high to low. Every column stands for a class, the height value is the intra-class sample counts.

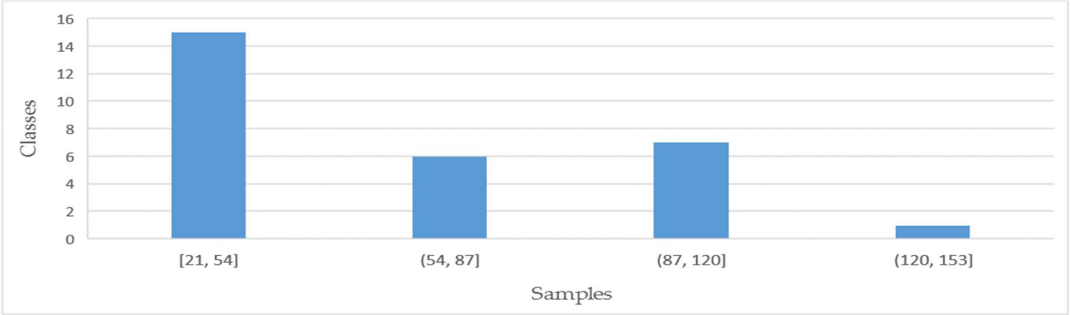

Figure 3. Intra-class samples distribution. Each column represents a range of samples, and the height value is the count of classes belonging to this range.

As can be seen from Figures 2 and 3, most of the samples in the graph are concentrated in the top eight categories, accounting for 46.31% of the weight, and the following 21 categories only represent 53.69% of the weight. For more than half of the classes, the number of samples ranged from 21 to 54, which was less than the average sample size of 59 and far less than the highest sample number of 135. Therefore, it can be asserted that the number of samples in this dataset is extremely unbalanced, and the samples in many categories are insufficient.

The spectra of four kinds of drugs produced by various manufacturers are shown in Figure 4. As can be seen from the figure, the spectra of the same drug produced by different manufacturers according to the Pharmacopoeia of the People's Republic of China (2015 version II) are very similar, and the important bands (peak and valley positions) mostly overlap. Metformin hydrochloride tablets manufactured by two manufacturers (No. 6 and 7) and chlorphenamine maleate tablets manufactured by two manufacturers (No. 18 and 19) were taken from Figure 4 to form Figure 5, and the difference between manufacturers of the spectra could hardly be detected by visual inspection.

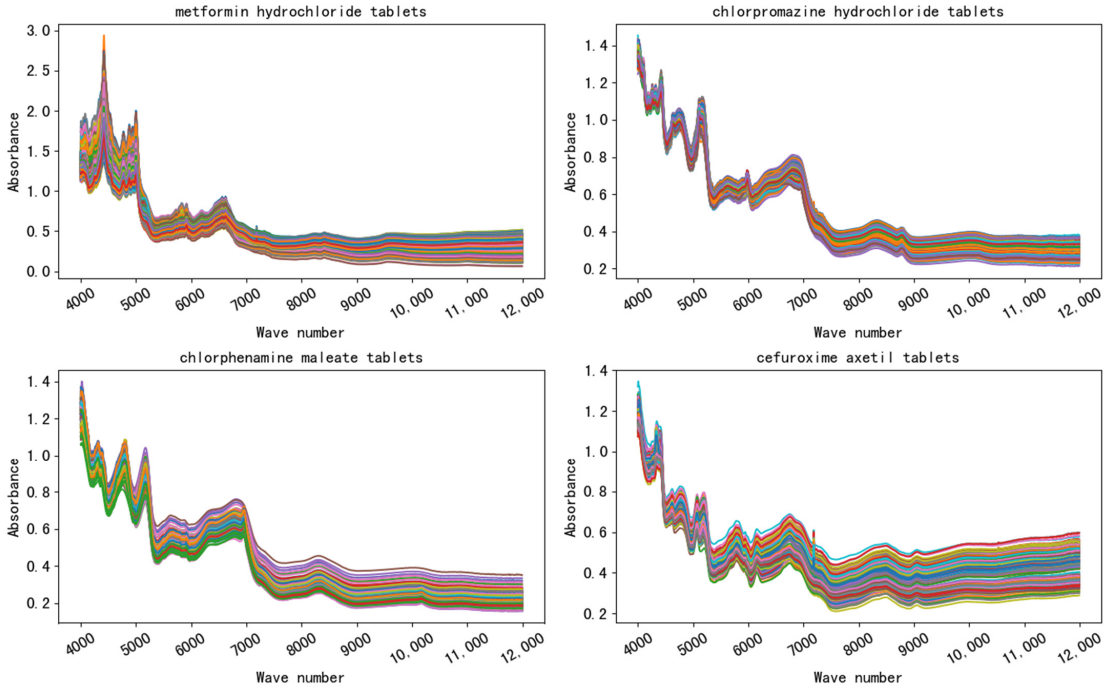

Figure 4. The spectra of four drugs. The same drug produced by different manufacturers according to the Pharmacopoeia of the People's Republic of China (2015 version II) are very similar, and the important bands (peak and valley positions) mostly overlap.

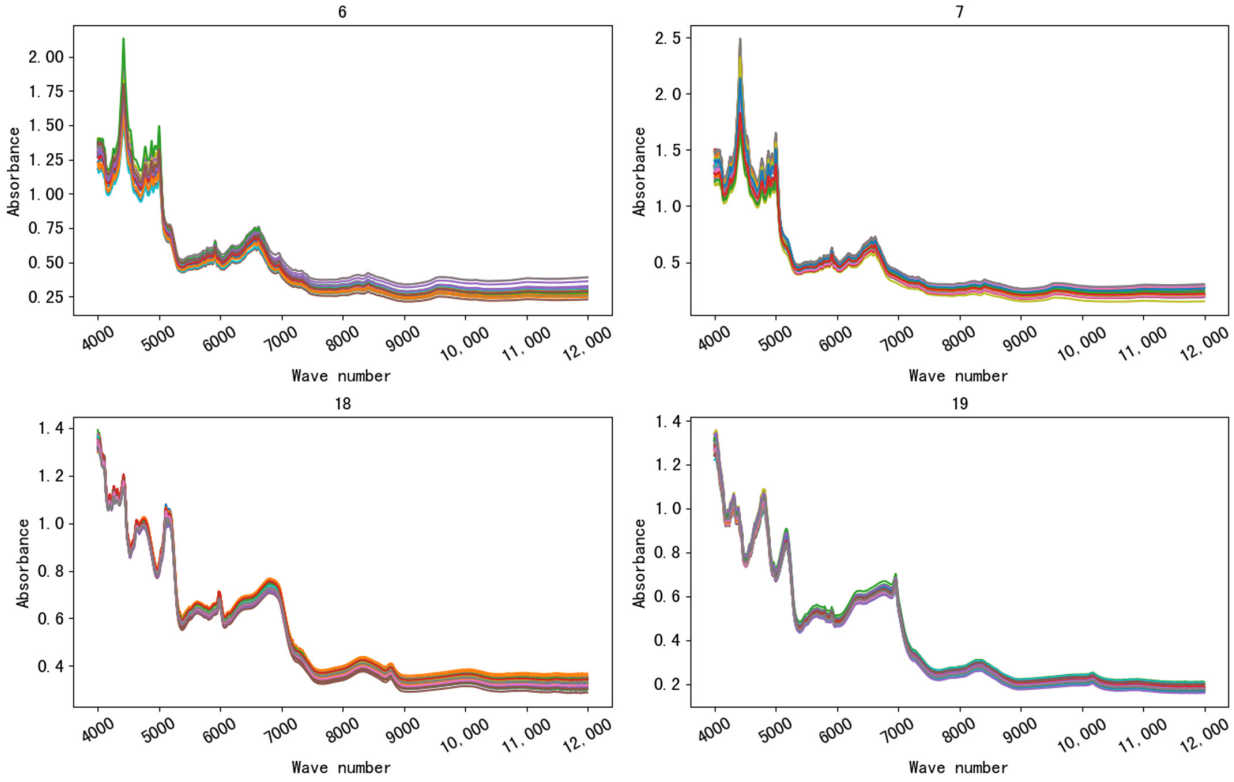

Figure 5. Similar spectra of the same drug produced by different manufacturers. Metformin hydrochloride tablets came from Nos. 6 and 7. Chlorphenamine maleate tablets came from Nos. 18 and 19.

Generally, the time from R&D to final registration of the original drug is about 15 years, and it needs to undergo four-phase clinical trials at a cost of hundreds of millions of dollars. Such drugs cannot be imitated until the patent has expired, and the enjoy the protection of policies such as separate pricing. Generic drugs only replicate the main components of the original drug, and even if a huge investment is invested in the generic process, the price is only about 1/3 even 1/6 that of the original drug. Therefore, it is understandable that generic drugs and the original drug can be as consistent as possible without being distinguished.

This is very challenging for classification algorithm modeling. It requires that the algorithm be able to distinguish subtle differences between the classification features when extracting the class features.

2.2. Methods

Under the above severe classification requirements, we build a classifier based on Bi-GAN generating sample method to achieve fair and accurate classification.

Its main idea is to use artificial a generative adversarial network to generate samples to supplement and improve the sorting of the original samples as shown in Figure 6. Through a fair and reasonable sampling strategy, each category can get enough attention in the model construction, and finally effectively alleviate the shortcomings caused by insufficient intra-class samples and unbalanced inter-class samples in the drug near-infrared spectrum classification method, making the cost of wrong classification problem can be solved effectively.

The key to its realization lies in the modification of the original Bi-GAN. On the one hand, to make original Bi-GAN have the ability to generate specific class samples instead of random "real" samples, and the other is to make classification supervision run through every process of Bi-GAN training, let the training of generator and discriminator be interfered by the classification loss.

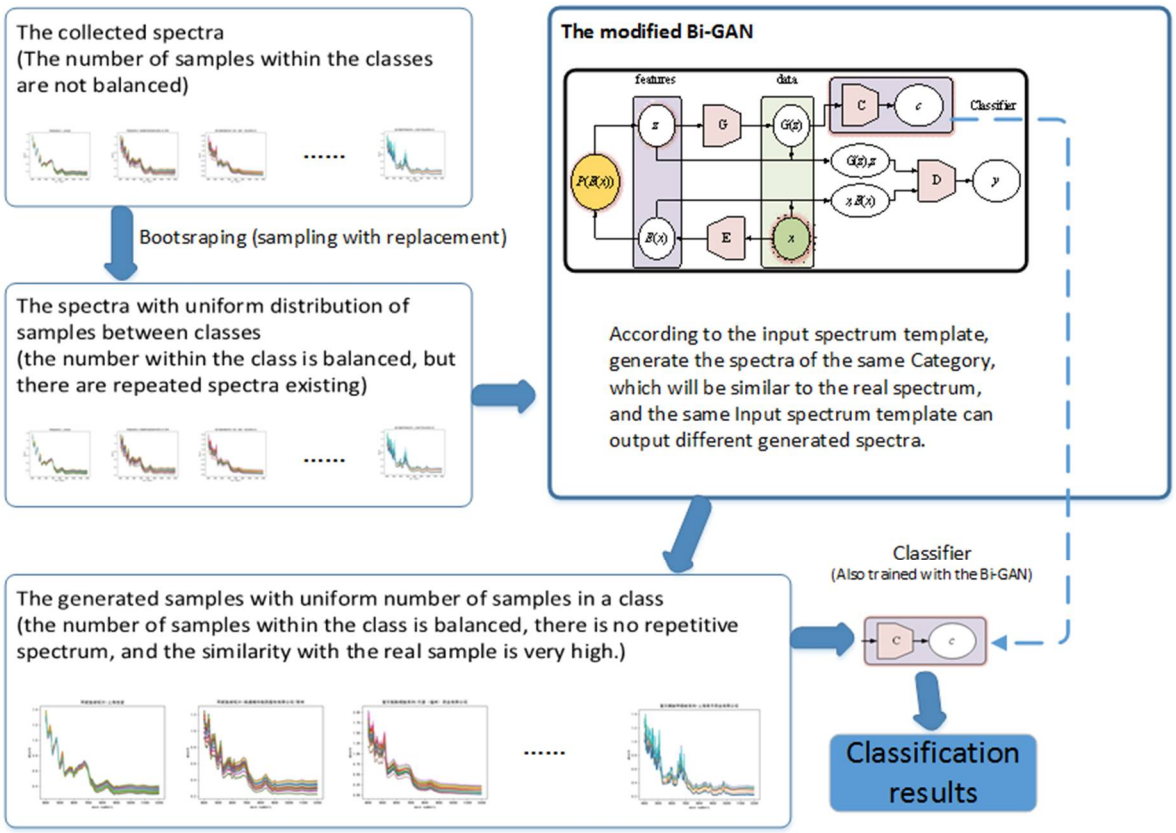

Figure 6. The main process of the proposed method.

2.2.1. Original Bi-GAN

The internal structure of the original Bi-GAN is shown in Figure 7.

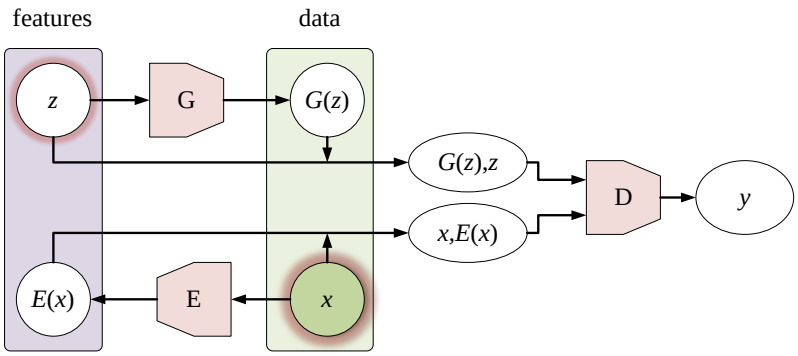

Figure 7. The internal structure of the original Bi-GAN.

Its main objectives are shown in Equation (1):

$$\min_{G,E}\max_{D} V(D,E,G) \tag{1}$$

where:

$$V(D,E,G) := \mathbf{E}_{x\sim p_x}[\underbrace{\mathbf{E}_{z\sim p_{E(\cdot|x)}}[\log D(x,z)]}_{\log D(x,E(x))}] + \mathbf{E}_{z\sim p_z}[\underbrace{\mathbf{E}_{x\sim p_{G(\cdot|z)}}[log(1-D(x,z))]}_{\log(1-D(G(z),z))}] \tag{2}$$

In Equations (1) and (2), G is the generator, which can be regarded as the decoder. D as the discriminator and E as the encoder. x represents the real sample. $E(x)$ represents the representation encoded into the potential space, and it is also the extracted features. z is the random sampling of the prior distribution, and $G(z)$ represents the sample generated by z. y is the data source, if the data to be discriminated comes from the real sample x, then $y = 1$; if it comes from the generated sample $G(z)$, then $y = 0$.

Equation (2) shows that Bi-GAN binds the original spectrum x and its extracted feature $E(x)$, and the generated sample $G(z)$ is bound with its prior distribution sample z, and then the two couples is been labeled with 1 and 0 respectively. The discriminator D is required to distinguish them to the maximum extent, and the generator G is required to prevent discriminator D from distinguishing. After training D and G alternately, generator G and discriminator D reach a Nash equilibrium. At this time, it can be considered that the authenticity of the generated samples has little difference from the "REAL" samples, and G has become a usable sample generator. The effectiveness of the above-mentioned methods in sample generation and feature extraction has been confirmed in reference [22].

However, the prior distribution sampling of the original Bi-GAN is usually the random sampling of the standard normal distribution $N(0,1)$, and the category of the generated sample $G(z)$ is not guaranteed, so it is impossible to determine whether the generated sample is the sample of the specified class. This is not in line with our goal of generating a specific class of samples. Therefore, we need to modify the original Bi-GAN to ensure the generator G can generate demanded random samples of "specified categories".

2.2.2. The Modifications of Original Bi-GAN

The overall modified design based on the original Bi-GAN is shown in Figure 8.

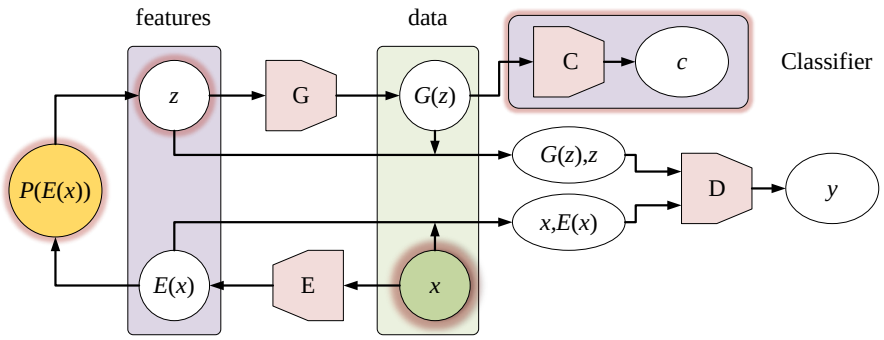

Figure 8. The modification of the original Bi-GAN. This is also the main design of this paper.

In Figure 8, we made the following changes to Bi-GAN:

(1) The sampling of z is limited.

We limit the sampling of z and set the mean and variance of $P(E(x_i))$ as shown in Formula (3). The default values of σ all set to 1 at first, and then they are automatically adjusted according to the previous five history records during the training process:

$$\begin{cases} \mu_i = E(x_i) \\ \sigma_i = \sqrt{\frac{\sum_{j=i-5}^{i-1} (E(x_j) - \mu_i)^2}{4}} \end{cases} \tag{3}$$

when generating the spectrum, the real spectral template x_i must be specified, and then its class label c_i is also determined. x_i is encoded into $E(x_i)$ by encoder E, and then the mean and variance of the prior normal distribution $P(E(x_i))$ is determined according to Formula (3), where the feature vector z could be randomly sampled in the fixed mean and local average variance scope.

(2) Limit the random $G(z)$ in a specified class.

We do this by building a classifier. The classifier C in this paper is composed of MLP and softmax. In the pre-training, the real sample x_i is used as its input, and the generated sample $G(z)$ is used as its input in the formal training. Its output is a predicted class c_i.

The classifier should be pre-trained, and its loss function shown as Equation (4):

$$Loss_{classfication} = -\sum_{i=1}^{k} (c_i \cdot \log(\hat{c}_i)) \tag{4}$$

where c_i is the class label of the real sample x_i in pre-training, and for the generated sample, the class label of its spectral template x_i is taken. k is the total number of drug categories and $\hat{c}_i$ is the predicted category.

G, D, and C are alternately optimized by gradient descent algorithm, and the optimization objective is changed to:

$$\min_{G,E,C} \max_{D} V(D, E, G, C) \tag{5}$$

where:

$$V(D,E,G,C) := \mathbf{E}_{x\sim p_x}[\underbrace{\mathbf{E}_{z\sim p_E(\cdot|x)}[\log D(x,z)]}_{\log D(x,E(x))}] + \mathbf{E}_{z\sim p_z}[\underbrace{\mathbf{E}_{x\sim p_G(\cdot|z)}[(1-\log D(x,z))]}_{\log 1-D(G(z),z)}] + \mathbf{E}_{z\sim p_z}[\underbrace{\mathbf{E}_{y\sim p_C(\cdot|G(z))}[\log C(G(z))]}_{\log C(G(z))}] \tag{6}$$

According to this objective, the loss of discriminator during training is calculated as Equation (7):

$$Loss_{discrimination} = -\sum_{i=1}^{n} (y_i \cdot \log(\hat{y}_i) + (1-y_i) \cdot \log(1-\hat{y}_i)) + Loss_{classification} \tag{7}$$

where y_i represents the data source. If the data to be identified comes from a real sample, then $y_i = 1$; if it comes from the generated sample, then $y_i = 0$. $\hat{y}_i$ is the discriminator's prediction, and the $Loss_{classification}$ is the result of Equation (4).

The loss of generator G during training is calculated using Equation (8):

$$Loss_{generation} = -\sum_{i=1}^{n} \log \hat{y}_i + Loss_{classification} \tag{8}$$

In this way, the classifier involves all G, D training processing. During the iterations, the generator will increasingly tend to generate samples of the same class as the template spectra.

2.2.3. Sampling Strategy in Data Set Processing

In this paper, after the spectra of each category are divided into the training set and test set according to the chosen proportion, they do not directly participate in the training and testing except for the pre-training of classifiers. Instead, the number of spectra of each category participating in the training and testing is determined in an equally fixed number, and they are extracted from data set by random sampling with replacement method.

The advantages of this method are:

Firstly, each class's participating spectra in the training course are equal, so equal attention can be paid to each category, and the categories with fewer samples in the training process will not be ignored.

Secondly, even for the same spectrum template, because there is a random sampling process in the generation phase, the final generated spectrum will be different, so the diversity is guaranteed to a certain extent.

2.2.4. Application of Classifier

After the training of Bi-GAN, three trained networks, *E*, *G*, and *C*, are taken out to form a structure as shown in Figure 9, which is used to predict the categories.

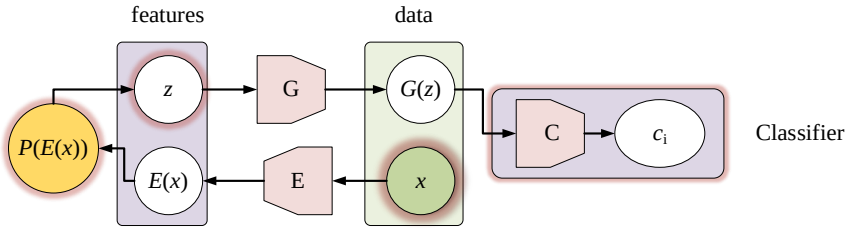

Figure 9. The model structure for practical application. Just input the spectrum to be classified at x, and the predicted category c_i will be output.

For each true spectrum x to be predicted, we can repeat the input a fixed number of times. Due to the existence of $P(E(x))$, the model will produce a different synthetic spectrum each time, which is consistent with the real spectrum in their categories but has good diversity. Most of the prediction results should be consistent with the category of x except for one or two abnormal values. By counting the frequency of output results, we can select the category with the highest frequency, which can be decided as the final prediction result of the x.

In this way, when $P(E(x))$ sampling occasionally appears small probability sampling anomaly, the model will not be disturbed by it, and finally, the correct category is selected.

3. Results

3.1. Experimental Environment

This paper uses the following hardware and software environment for the data modeling experiments:

Hardware environment: CPU Xeon 2678v3 (12 cores, 24 threads), memory 64 GB, SSD 1TB, GPU NVIDIA Tesla V100.

Software environment: operating system Ubuntu 18.04.3 LTS, NVIDIA driver version 440.33.01, CUDA v10.2, cudnn v7.6.5, keras GPU 2.3.1, tensorflow GPU 1.15.0, sci-kit learn 0.19.0.

3.2. Muti-Classification Results

In the experiment, E, G, and C are constructed by multi-layer perception (MLP):

E network MLP uses 2074-120-30 to set up the network, layer 120 and layer 30 are preceded by dropout (0.2), followed by batch normalization layer (BN) decoding. The activation function is RELU.

G network MLP uses 30-360-2074 to set up the network, and the activation function is also RELU.

The C network classifier is designed with 2074-150-30-softmax, and the activation function is sigmoid.

All the networks use RMSprop optimizer, and its parameters are the Keras' default parameters, the batch size is set to 60, trained for 150 epochs.

The experimental spectra were divided into a training set and test set according to 9:1, 8:2, 7:3, 6:4, 5:5, 4:6, 3:7, and 2:8.

The training set is used for modeling, and the test set is used to verify the effectiveness of the model. For example, in the first row of Table 1, for the metaformin hydrochloride tablets produced by Shanghai Xinyi Pharmaceutical Factory Co., Ltd. (Zhengzhou, China), if the training set and test set are divided by 9:1 into 94 samples, 89 samples are randomly selected to be put in the training set for modeling, and the remaining nine samples are put in the test set for verifying the effectiveness of the model. The nine samples in the test set are invisible during the modeling period. They are "non-existent" external samples

for the training process, but they are internal samples for the whole data set because their distribution and internal properties are similar to those of the 89 samples participating in the training.

Each experiment was conducted 10 times and the best results were recorded. The experimental results are shown in Table 2.

Table 2. Experimental results under different training and test set partitions.

Train:Test	Precision	Recall	F1	Accuracy
9:1	0.994	0.993	0.994	0.994
8:2	0.994	0.990	0.992	0.992
7:3	0.996	0.989	0.993	0.993
6:4	0.993	0.988	0.991	0.992
5:5	0.992	0.987	0.990	0.990
4:6	0.908	0.905	0.907	0.907
3:7	0.893	0.861	0.877	0.861
2:8	0.816	0.854	0.834	0.854

As can be seen from Table 2, when the training set's proportion is more than 50%, the classification accuracy rate of the multi-classification model in this paper is more than 99% and when the proportion decreases, the classification accuracy does not decrease accordingly before 4:6, as if it has little relationship with the division of training set and test set, but when the training set is only 40% of the total, the accuracy displays a big drop from 99% to 92%. Since most of the categories in the data set have 30–50 spectra, when the training set accounts for less than 40%, most of the categories obviously begin to reflect missing data. When the training set proportion is only 30%, there are only 6 spectra of the minimum category that can be used for training. While when the training set reaches 20%, there are only four spectra of the minimum category that can participate in the training, and most categories (15 out of 29 categories) have only 10 or fewer data pieces for training.

To investigate the classification errors of each category, we draw the confusion matrix in the case of the most favorable classification (90% of the training set), as shown in Figure 10, and the confusion matrix in the case of the worst classification (20% of the training set), as shown in Figure 11.

From Figure 10, it can be seen that in the case of the most favorable classification (90% of the training set), except for categories 1, 3, and 9 (for tools reasons, the figure count classes from 0, while the information table count categories from 1, we apologize for the inconvenience), all the classes classified perfectly.

The intra-class spectra number of categories 1, 3, and 9 are 94, 67, and 48. Only one category falls into the insufficient data interval 21–54 of Figure 2. However, Category 4, which has the least number of spectra within its class, has a good classification effect. This shows that even if the classification error occurs in this situation, the error is not caused by the lack of intra-class spectra.

As can be seen from Figure 11, those classes with lower classification accuracy are more evenly scattered in the intervals shown in Figure 2, but not in the areas with insufficient data. This shows that the method in this paper has played a due role in eliminating the adverse effect of insufficient spectral numbers in the class.

It can also be seen from Figure 11 that the classification errors are mainly caused by the misclassification among different manufacturers within the same drug. Besides, whether it is in Figure 9 or Figure 10, the classification accuracy of Categories 25–29 (corresponding to cefuroxime axetil tablets) is almost not affected by the decrease of training spectrum proportion. It can be seen that in this method, the most important factor affecting the accuracy of classification is still the inherent characteristics of the spectrum of various pharmaceutical products, which is consistent with our original classification purpose.

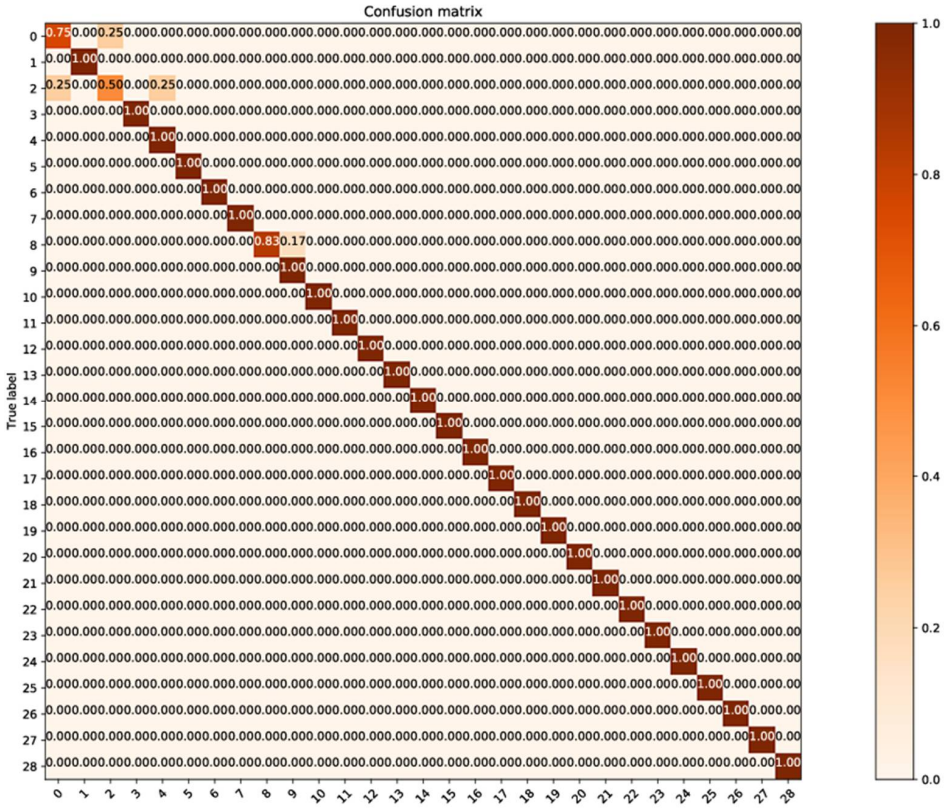

Figure 10. The confusion matrix in the case of the most favorable classification (90% of the training set).

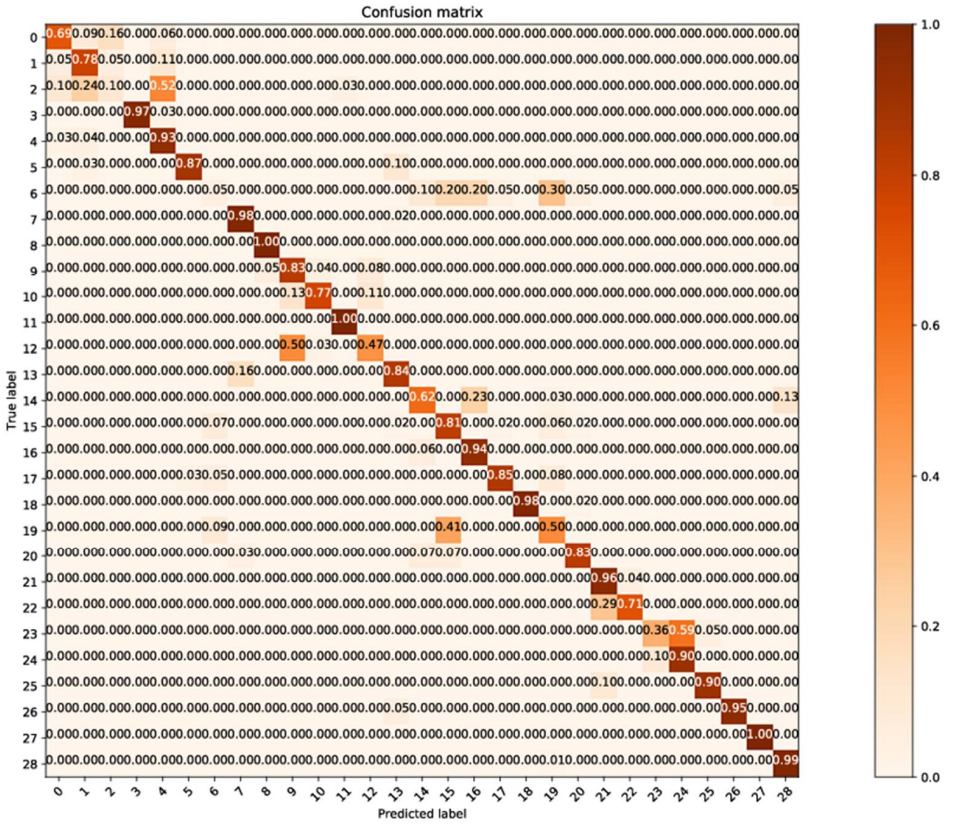

Figure 11. The confusion matrix in the case of the worst classification (10% of the training set).

3.3. Comparative Methods Results

The experimental results are compared with three kinds of algorithms: one is the traditional linear classification algorithm, mainly PLS-DA and linear SVM; the other is the traditional nonlinear classification algorithm, mainly RBF SVM, k-NN, BP-ANN; the third is the deep learning algorithm in recent years, mainly DBN, SAE, and CNN. Among them:

The number of components of PLS-DA is the same as that of z in this paper, which is set to 30. The c value of linear SVM is 1.

The k-NN's k is set to 1. The RBF SVM's Gamma value is set to 0.0001, and the c value is set to 1. The BP-ANN takes two layers, the number of units in each layer is set to 2074-29, and the activation function is sigmoid.

In DBN, only one layer of RBM (1037 units) is set, followed by a full connection layer and a softmax classifier as the output.

The SAE's codec takes two layers, the number of units in each layer is set to 2074-180-30 and 30-180-2074, respectively. The feature layer (30 units) is fully connected to a softmax classifier as the output.

The CNN is constructed according to the optimal method design of the reference [15].

After all the models are well trained, we run the same tests according to our mothed proposed in this paper. The comparison accuracy results are shown in Table 3.

Table 3. Accuracy of various multi-class classification algorithms.

Train:Test	Bi-GAN	PLS-DA	Linear SVM	RBF_SVM	k-NN	BP-ANN	DBN	SAE	CNN
9:1	0.994	0.957	0.943	0.923	0.846	0.910	0.933	0.945	0.991
8:2	0.992	0.950	0.902	0.946	0.853	0.906	0.912	0.923	0.981
7:3	0.993	0.944	0.922	0.925	0.904	0.897	0.900	0.914	0.987
6:4	0.992	0.933	0.929	0.917	0.811	0.883	0.923	0.910	0.979
5:5	0.990	0.926	0.922	0.902	0.798	0.878	0.912	0.804	0.963
4:6	0.907	0.911	0.909	0.850	0.823	0.828	0.863	0.813	0.910
3:7	0.861	0.908	0.889	0.852	0.743	0.795	0.894	0.781	0.872
2:8	0.854	0.809	0.795	0.797	0.741	0.816	0.832	0.778	0.845

It can be seen from Table 3 that:

Overall, the accuracy of the Bi-GAN classifier is better in accuracy than the others. DBN, SAE, CNN, and other deep learning algorithms take second place, PLS-DA and linear SVM still available since the traditional linear classification algorithm also has a certain effect in discriminating the composition of drugs.

Except for PLS-DA, when the partition of the training set and test set is extreme, each algorithm will encounter the inflection point of classification accuracy when encountering the lack of necessary class data. Among them, the sensitivity of the nonlinear algorithms is higher than that of the linear algorithm.

In the drug multi-classification algorithms, the old PLS-DA algorithm still has good performance, and it is still worthy of attention when only focusing on the influence of drug components without considering the nonlinear influence factors.

Although Bi-GAN has high accuracy, it has the highest sensitivity in data missing. Once the data missing is serious, it is easy to deviate. The training time and inferring time of the algorithms are shown in Table 4.

It can be seen from the table that:

Except for Bi-GAN, the training time of all algorithms decreases with the decreasing of training set proportion, while the inferring time shows an upward trend because test sets are expanding.

The k-NN algorithm has the least training time, but its accuracy is the worst, and its inferring time is longer than most of the others.

Although PLS-DA, linear SVM, RBF-SVM, k-NN, and BP-ANN use CPU to calculate, their training can be completed in less than 1 s because of their simple structure. But,

the inferring time of linear SVM and RBF-SVM is longer than that of nonlinear algorithms, including the deep learning algorithms.

The deep learning algorithms' training time is longer than that of linear algorithms, but their inference time is shorter.

Among the deep learning algorithms, the cost of training time and inference time of our method is average. Its cost is fixed and does not vary greatly with the division of training set and test set, and that is a merit.

Table 4. Training and inferencing time of each algorithm (in second, training time/inferring time).

Train: Test	Bi-GAN	PLS-DA	Linear SVM	RBF-SVM	k-NN	BP-ANN	DBN [1]	SAE [2]	CNN [3]
9:1	21.324/0.020	0.585/0.004	0.933/0.271	3.793/0.330	0.080/0.142	18.211/0.002	231.729/0.015	9.843/0.027	23.790/0.005
8:2	21.082/0.029	0.522/0.008	0.787/0.476	3.183/0.595	0.066/0.267	16.779/0.003	204.728/0.029	10.319/0.044	21.368/0.007
7:3	21.433/0.041	0.425/0.011	0.668/0.674	2.700/0.856	0.050/0.369	13.352/0.004	184.345/0.038	9.536/0.061	21.066/0.011
6:4	21.210/0.062	0.340/0.014	0.583/0.791	2.217/0.998	0.041/0.458	8.924/0.006	159.686/0.046	9.967/0.094	21.385/0.012
5:5	20.552/0.067	0.263/0.017	0.433/0.886	1.760/1.052	0.027/0.622	13.784/0.006	140.062/0.066	9.960/0.122	22.253/0.015
4:6	20.496/0.087	0.221/0.022	0.318/0.901	1.186/1.090	0.020/0.604	19.103/0.008	119.054/0.086	8.898/0.136	42.056/0.019
3:7	20.992/0.100	0.180/0.026	0.218/0.821	0.862/0.983	0.014/0.681	12.730/0.010	93.467/0.090	10.028/0.151	36.602/0.023
2:8	20.064/0.117	0.125/0.030	0.123/0.724	0.427/0.827	0.008/0.563	12.947/0.010	73.280/0.184	10.985/0.189	43.297/0.022

[1] DBN's RBM training is using the CPU algorithm, not the GPU algorithm, so its training time is very long. [2] SAE uses early stopping, so epoch training is uneven. [3] In CNN training, from 4:6, epochs are extended from 200 to 500, so the training time is longer.

PLS-DA, Linear SVM, RBF-SVM, k-NN, and BP-ANN use scikit-learn software to train and test, scikit-learn uses the CPU for calculation and only uses a single core and a single thread.

4. Discussion

Construction of the NIR classification model for multi-variety and multi-manufacturer drugs involves much data and complex categories and the application scenarios are challenging. This paper starts with the analysis of the problems of insufficient samples within the class and unbalanced samples between classes in the near-infrared spectrum classification of drugs. Then analyze the cost-sensitive problems of the incorrect classification caused by these problems. Through the modified Bi-GAN, the quantitative generated samples are used instead of the original uneven real samples as the classification training basis, which can effectively solve these problems above to an extent.

By constructing the appropriate network connection, using the appropriate combination of cost functions and a fair sampling strategy, we have achieved excellent classification results in the experiments. The experimental results demonstrate that in this scenario, the proposed method can achieve a classification accuracy of more than 99% in most cases where the training set accounts for more than 50% of the whole data set. Moreover, although the accuracy of this method will be greatly reduced when the proportion of the training set is reduced to 40%, the classification accuracies are relatively stable before that. As for time cost, although the training and inferring time cost of this method are at an average level compared with other deep learning methods, its cost is relatively constant and it also does not fluctuate with the increase or decrease of the number of samples of the dataset.

By comparing the traditional and three new kinds of drug classification algorithms, we can assert that this method has successfully achieved our expectation by solving the pre-set problems, the accuracy and stability of the method for the identification of multi-class drugs by near-infrared spectroscopy are also improved to a certain extent, which can provide a useful reference in the similar scene of near-infrared spectrum analysis and sensor signal data processing.

5. Conclusions

We propose an improved Bi-Gan method to classify the near-infrared spectra of drugs given the problem of insufficient samples within the class and imbalance of samples between classes. By limiting the mean and variance of latent variables, adding the classification loss constraint, and using the fair strategy of sampling with replacement. We achieve

the desired results. The experimental results showed that the best classification accuracy of 1721 NIR spectra of four kinds of drugs produced by 29 manufacturers was significantly 99.4%. Compare to the other eight NIR multi-classification methods in recent years, this method has obvious advantages.

The problems in this paper may also exist in other sensor data classification processing, so the method we propose can be a useful reference for readers in dealing with the multi-classification problems in other scenarios.

Author Contributions: Conceptualization, A.Z. and H.Y.; methodology, A.Z.; software, A.Z.; validation, X.P., L.Y. and Y.F.; formal analysis, H.Y.; investigation, A.Z.; resources, L.Y. and Y.F.; data curation, A.Z.; writing—original draft preparation, A.Z.; writing—review and editing, H.Y.; visualization, A.Z.; supervision, H.Y.; project administration, H.Y.; funding acquisition, H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China, grant number 2018AAA0102600, National Natural Science Foundation of China, grant number 61906050, Guangxi Technology R&D Program, grant number 2018AD11018.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We are grateful for the support of the China Institute for Food and Drug Control.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Prajapati, P.; Solanki, R. A brief review on NIR spectroscopy and its pharmaceutical applications. *IJPCA* **2016**, *3*, 117–123. [CrossRef]
2. Chu, X.-L. *Molecular Spectroscopy Analytical Technology Combined with Chemometrics and Its Applications*; Chemical Industry Press: Beijing, China, 2011; Volume 5, pp. 259–302.
3. Biancolillo, A.; Marini, F. Chemometric Methods for Spectroscopy-Based Pharmaceutical Analysis. *Front. Chem.* **2018**, *6*, 576. [CrossRef]
4. Yong, N.; Ni, W. Near-infrared spectra combined with partial least squares for pH determination of toothpaste of different brands. *Chin. Chem. Lett.* **2011**, *12*, 91–94.
5. Fu, H.Y.; Huang, D.C.; Yang, T.M. Rapid recognition of Chinese herbal pieces of Areca catechu by different concocted processes using Fourier transform mid-infrared and near-infrared spectroscopy combined with partial least-squares discriminant analysis. *Chin. Chem. Lett.* **2013**, *24*, 639–642. [CrossRef]
6. Elizarova, T.E.; Shtyleva, S.V.; Pleteneva, T.V. Using near-infrared spectrophotometry for the identification of pharmaceuticals and drugs. *Pharm. Chem. J.* **2008**, *42*, 432–434. [CrossRef]
7. Weng, X.; Mao, D. Rapid qualitative analysis model for cetirizine hydrochloride Tablets by NIR using chemometric methods. *Comput. Appl. Chem.* **2012**, *29*, 995–998.
8. Gong, L.P.; Wang, W.J.; Yang, N. Development of NIR method for rapid determination of cefalexin tablet. *Chin. J. Pharm. Anal.* **2011**, *31*, 1571.
9. Rodionova, O.Y.; Titova, A.V.; Balyklova, K.S. Detection of counterfeit and substandard tablets using non-invasive NIR and chemometrics—A conceptual framework for a big screening system. *Talanta* **2019**, *205*, 120150. [CrossRef]
10. Byvatov, E.; Fechner, U.; Sadowski, J. Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1882–1889. [CrossRef]
11. Wu, W.; Massart, D.L. Artificial neural networks in classification of NIR spectral data: Selection of the input. *Chemom. Intell. Lab. Syst.* **1996**, *35*, 127–135. [CrossRef]
12. Zhang, W.-D.; Li, L.-Q.; Hu, J.-Q. Drug Discrimination by Near-Infrared Spectroscopy Based on Stacked Sparse Auto-encoders Combined with Kernel Extreme Learning Machine. *Chin. J. Anal. Chem.* **2018**, *46*, 1446–1454.
13. Yang, H.; Hu, B.; Pan, X. Deep Belief Networks Based Drug Identification using Near-Infrared Spectroscopy. *J. Innov. Opt. Health Sci.* **2016**, *10*, 1630011. [CrossRef]
14. Li, L.-Q.; Pan, X.-P.; Feng, Y.-C. Deep Convolution Network Application in Identification of Multi-product and Multi-Manufacturer Pharmaceutical. *Spectrosc. Spectr. Anal.* **2019**, *39*, 3606–3613.
15. Indyk, P. Approximate Nearest Neighbor: Towards Removing the Curse of Dimensionality. Proc Symposium on Theory of Computing. 1998. Available online: https://users.math.msu.edu/users/iwenmark/Teaching/MTH995/Papers/LSH_THM_4_609.pdf (accessed on 20 December 2020).

16. Zhang, X.H.; Xu, Y.; He, Y.L. Novel manifold learning-based virtual sample generation for optimizing soft sensor with small data. *ISA Trans.* **2020**. Available online: https://www.sciencedirect.com/science/article/abs/pii/S0019057820304018 (accessed on 20 December 2020).
17. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *2*, 2672–2680.
18. Creswell, A.; White, T.; Dumoulin, V. Generative adversarial networks: An overview. *IEEE Signal Process. Mag.* **2018**, *35*, 53–65. [CrossRef]
19. Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets. *arXiv* **2017**, arXiv:1701.04722.
20. Zhao, W.; Chen, X.; Bo, Y. Semisupervised Hyperspectral Image Classification with Cluster-Based Conditional Generative Adversarial Net. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 539–543. [CrossRef]
21. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv* **2015**, arXiv:1511.06434.
22. Zhang, Z.; Liu, S.; Li, M. Bidirectional Generative Adversarial Networks for Neural Machine Translation. 2018. Available online: https://www.researchgate.net/publication/334118012_Bidirectional_Generative_Adversarial_Networks_for_Neural_Machine_Translation/citations (accessed on 20 December 2020).
23. Zhu, J.Y.; Park, T.; Isola, P. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *arXiv* **2017**, arXiv:1703.10593v7.
24. Choi, Y.; Choi, M.; Kim, M. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8789–8797.
25. Zhang, H.; Xu, T.; Li, H. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Honolulu, HI, USA, 21–26 July 2017; pp. 5907–5915.
26. Denton, E.L.; Chintala, S.; Fergus, R. Deep generative image models using a laplacian pyramid of adversarial networks. *Adv. Neural Inf. Process. Syst. arXiv* **2015**, arXiv:1506.05751.
27. Nie, D.; Trullo, R.; Lian, J.; Petitjean, C.; Ruan, S.; Wang, Q.; Shen, D. Medical Image Synthesis with Context-Aware Generative Adversarial Networks. In *Proceedings of the Mining Data for Financial Applications*; Springer Nature: Cham, Switzerland, 2017; Volume 10435, pp. 417–425.
28. Xiong, W.; Luo, W.; Ma, L. Learning to generate time-lapse videos using multi-stage dynamic generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2364–2373.
29. Lucas, A.; Lopez-Tapia, S.; Molina, R. Generative adversarial networks and perceptual losses for video super-resolution. *IEEE Trans. Image Process.* **2019**, *28*, 3312–3327. [CrossRef]
30. Liu, M.Y.; Huang, X.; Yu, J. Generative Adversarial Networks for Image and Video Synthesis: Algorithms and Applications. *arXiv* **2020**, arXiv:2008.02793.
31. Ahsan, U.; Sun, C.; Essa, I. Discrimnet: Semi-supervised action recognition from videos using generative adversarial networks. *arXiv* **2018**, arXiv:1801.07230.
32. Bhattacharjee, P.; Das, S. Temporal coherency based criteria for predicting video frames using deep multi-stage generative adversarial networks. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4268–4277.
33. Roheda, S.; Riggan, B.S.; Krim, H. Cross-modality distillation: A case for conditional generative adversarial networks. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 2926–2930.
34. Gao, Y.; Singh, R.; Raj, B. Voice impersonation using generative adversarial networks. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 2506–2510.
35. Hsu, C.C.; Hwang, H.T.; Wu, Y.C. Voice conversion from unaligned corpora using variational autoencoding Wasserstein generative adversarial networks. *arXiv* **2017**, arXiv:1704.00849.
36. Kameoka, H.; Kaneko, T.; Tanaka, K. Stargan-VC: Non-parallel many-to-many voice conversion using star generative adversarial networks. In Proceedings of the 2018 IEEE Spoken Language Technology Workshop (SLT), Athens, Greece, 18–21 December 2018; pp. 266–273.
37. Hono, Y.; Hashimoto, K.; Oura, K.; Nankaku, Y.; Tokuda, K. Singing Voice Synthesis Based on Generative Adversarial Networks. In Proceedings of the ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 6955–6959. [CrossRef]
38. Sisman, B.; Zhang, M.; Dong, M. On the Study of Generative Adversarial Networks for Cross-Lingual Voice Conversion. 2019. Available online: https://www.researchgate.net/publication/335822104_On_the_Study_of_Generative_Adversarial_Networks_for_Cross-Lingual_Voice_Conversion (accessed on 20 December 2020).
39. Zhu, M.; Pan, P.; Chen, W. Dm-Gan: Dynamic Memory Generative Adversarial Networks for Text-to-Image Synthesis. Available online: https://www.researchgate.net/publication/338513061_DM-GAN_Dynamic_Memory_Generative_Adversarial_Networks_for_Text-To-Image_Synthesis/citations (accessed on 20 December 2020).
40. Nam, S.; Kim, Y.; Kim, S.J. Text-adaptive generative adversarial networks: Manipulating images with natural language. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 42–51.

41. Xu, T.; Zhang, P.; Huang, Q. Attngan: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks. Available online: https://openaccess.thecvf.com/content_cvpr_2018/papers/Xu_AttnGAN_Fine-Grained_Text_CVPR_2018_paper.pdf (accessed on 20 December 2020).
42. Glover, J. Modeling documents with generative adversarial networks. *arXiv* **2016**, arXiv:1612.09122.
43. Korshunov, P.; Marcel, S. Deepfakes: A new threat to face recognition? Assessment and detection. *arXiv* **2018**, arXiv:1812.08685.
44. Kwok, A.O.J.; Koh, S.G.M. Deepfake: A social construction of technology perspective. *Curr. Issues Tour.* **2020**, 1–5. [CrossRef]
45. Fletcher, J. Deepfakes, Artificial Intelligence, and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance. *Theatre J.* **2018**, *70*, 455–471. [CrossRef]
46. Chen, X.; Duan, Y.; Houthooft, R. InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 2172–2180.
47. Mescheder, L.; Nowozin, S.; Geiger, A. Adversarial Variational Bayes: Unifying Variational autoencoders and generative adversarial networks. *arXiv* **2017**, arXiv:1701.04722.