mathematics

MDPI

Article

A Novel Method of Chinese Herbal Medicine Classification Based on Mutual Learning

Meng Han [1,†], Jilin Zhang [1,*], Yan Zeng [1], Fei Hao [2] and Yongjian Ren [1]

[1] Computer & Software School, Hangzhou Dianzi University, Hangzhou 310018, China; hanm@hdu.edu.cn (M.H.); yz@hdu.edu.cn (Y.Z.); yongjian.ren@infocore.cn (Y.R.)
[2] School of Computer Science, Shaanxi Normal University, Xi'an 710119, China; fhao@snnu.edu.cn
* Correspondence: jilin.zhang@hdu.edu.cn; Tel.: +86-152-5880-1227
† Current address: Hangzhou Economic Development Zone, No. 1158, No. 2 Street, Baiyang Street, Hangzhou 310018, China.

Abstract: Chinese herbal medicine classification is an important research task in intelligent medicine, which has been applied widely in the fields of smart medicinal material sorting and medicinal material recommendation. However, most current mainstream methods are semi-automatic, with low efficiency and poor performance. To tackle this problem, a novel Chinese herbal medicine classification method based on mutual learning has been proposed. Specifically, two small student networks are designed for collaborative learning, and each of them collects knowledge learned from the other one respectively. Consequently, student networks obtain rich and reliable features, which will further improve the performance of Chinese herbal medicinal classification. In order to validate the performance of the proposed model, a dataset with 100 Chinese herbal classes (about 10,000 samples) was utilized and extensive experiments were performed. Experimental results verify that the proposed method is superior to those of the latest models with equivalent or even fewer parameters, specifically, obtaining 3∼5.4% higher accuracy rate and 13∼37% lower loss. Moreover, the mutual learning model achieves 80.8% Chinese herbal medicine classification accuracy.

Keywords: Chinese herbal medicine; classification; mutual learning; deep neural network

MSC: 68T07

Citation: Han, M.; Zhang, J.; Zeng, Y.; Hao, F.; Ren, Y. A Novel Method of Chinese Herbal Medicine Classification Based on Mutual Learning. *Mathematics* **2022**, *10*, 1557. https://doi.org/10.3390/math10091557

Academic Editor: Teng Li

Received: 28 March 2022
Accepted: 27 April 2022
Published: 5 May 2022

1. Introduction

Chinese herbal medicine (CHM) classification plays a vital role in the field of intelligent medicine. Furthermore, it has been widely applied in the fields of smart medicinal material sorting and medicinal material recommendation. The classification of CHM still faces many challenges, such as small datasets and large scale of model parameters.

To achieve promising CHM classification performance, many researchers have carried out a lot of research work. Wang et al. proposed a new Local Linear Embedding Algorithm (LLE) and Linear Discriminant Analysis (LDA) processing techniques to handle high-dimensional nonlinear data in the classification of CHM [1]. However, the dataset employed in their method only contains six classes, which is very small and cannot meet the needs of real-world applications. Zhang et al. utilized a supervised local projection strategy to identify plant leaves and achieved excellent classification performance [2]. Unger et al. applied support vector machine (SVM) with a Fourier feature and a morphological measuring approach to classify the medicinal materials in the two test sets, which consisted of 26 and 17 categories, respectively, with around 10 samples in each category [3]. In the two test sets, their approaches had an accuracy of 73.21% and 84%, respectively. Principal Component Analysis (PCA) and SVM were adopted by Luo et al. to classify Chinese medicinal materials [4]. Experimental results demonstrated that the SVM approach produces superior experimental results than that of the PCA method. The Self-organizing Map (SOM) algorithm was introduced

to classify CHM images in 2016 [5]. Although these methods have achieved promising CHM classification performance, the datasets utilized by them are limited, with only a few examples per category. On the other hand, these methods rely on less robust hand-crafted features and are not feasible for classification on large-scale datasets. In order to further enhance the classification performance, several deep convolutional neural networks, such as AlexNet [6], VGG [7], GoogleNet [8] and ResNet [9,10], have been employed and shown good performance in classification in recent years [11–13]. Liu et al. [14] utilized GoogleNet to improve the performance of CHM classification. Cai et al. [15] used the convolutional neural network with broad learning system method to identify Chinese herbal medicines, and their method achieved good performance. However, the dataset they used was still small, with just 17 categories and 1700 images. Unfortunately, due to the large scale of parameters, these models are not suitable for classification tasks in platforms or applications with small memory, such as mobile phones.

Mutual learning, a realistic method which can produce promising classification results via small yet powerful deep neural networks. Mutual learning starts with a group of students learning collaboratively [16]. Among them, each student is constrained by two loss terms, including the standard supervised learning loss and the mimicry loss respectively, which will encourage the class probabilities of each student to match those of other students, leading to more robust and richer discriminative features.

In fact, traditional Chinese herbal medicine classification methods have lower accuracy. On the other hand, although deep learning methods have higher accuracy, their training speed are usually slow. To address the aforementioned issues and improve the effectiveness and efficiency of CHM classification, this paper proposes to leverage the mutual learning to handle CHM classification approach, inspired by its superiority. Specifically, in the mutual learning framework, two student networks are employed to learn collaboratively, and each student network can receive information from the other one through the whole training process. Consequently, all student networks will learn more robust and richer features in this strategy, resulting in superior performance than that of the corresponding single network with the same or even much more layers.

In summary, our contributions can be concluded as follows:

1. A novel CHM classification approach based on mutual learning has been explored, which has achieved promising CHM classification performance in terms of both effectiveness and efficiency. Specifically, the mutual learning framework in this paper is based on two basic student networks. They perform collaborative learning during the whole training phase, aiming at extracting more robust and rich features without increasing the scale of parameters, and further enhance the performance of CHM classification.
2. Numerous experiments have been done to assess the effectiveness of our method, including the evaluation of the superiority of the mutual learning, the evaluation of the performance of mutual learning based on two identical student networks, the evaluation of the performance of mutual learning based on two different student networks, respectively. Experimental results illustrate that the mutual learning model can obtain significantly superior performance than those of the corresponding single networks without mutual learning with the same or even deep layers, in terms of both accuracy and speed.

The rest of this paper is organized as follows. Detailed information of the materials and methods are illustrated in Section 2. Section 3 depicts the experimental results and analyses. Section 4 concludes this paper.

2. Materials and Methods

2.1. Dataset

In order to validate the effectiveness of the mutual learning model, a CHM classification dataset CHMC with medium-scale [17] has been applied. Specifically, CHMC has 100 categories of Chinese herbal medicines, and each category contains about 100 images.

Therefore, the dataset owns a total of 10,000 images. Among them, 4/5 samples are utilized for training, and the rest are applied for testing. Some examples from the CHMC dataset are presented in Figure 1. Furthermore, most of the samples in the CHMC dataset have natural environment backgrounds, which will motivate their application in real-world scenarios.

Figure 1. Samples of CHMC dataset [17].

2.2. Problem Definition

Assume the dataset X contains $\mathcal{N}$ samples of class $\mathcal{C}$, $X = \{x_i\}_{i=1}^{\mathcal{N}}$, and the corresponding label can be denoted as $Y = \{y_i\}_{i=1}^{\mathcal{N}}$, where $y_i \in \{1, 2, \cdots, \mathcal{C}\}$.

Among them, $\mathcal{C}$ indicates the number of CHM categories, $\mathcal{N}$ denotes the number of Chinese herbal medicine samples. X/Y presents the CHM samples/label set, x_i and y_i indicate the i-th sample and its corresponding label.

The probability $q_1^c(x_i)$ that the ith sample x_i is predicted to be class c in the student network1 ($ConvNet1$) can be calculated by the following definition:

$$q_1^c(x_i) = \frac{\exp(s_1^c(x_i))}{\sum_{c=1}^{\mathcal{C}} \exp(s_1^c(x_i))} \tag{1}$$

where $s_1^c(x_i)$ indicates the logit value from the softmax layer of $ConvNet1$.

The probability $q_2^c(x_i)$ that the ith sample x_i is predicted to be class c in the student network2 ($ConvNet2$) can be obtained via the following equation:

$$q_2^c(x_i) = \frac{\exp(s_2^c(x_i))}{\sum_{c=1}^{\mathcal{C}} \exp(s_2^c(x_i))} \tag{2}$$

where $s_2^c(x_i)$ indicates the logit value from the softmax layer of $ConvNet2$.

The loss functions $\mathcal{L}_{ConvNet_1}$ and $\mathcal{L}_{ConvNet_2}$ of two student networks can be defined as follows:

$$\mathcal{L}_{ConvNet_1} = (1 - \alpha) * \mathcal{L}_{\mathcal{C}_1} + \alpha * D_{KL}(q_2 \| q_1) \tag{3}$$

$$\mathcal{L}_{ConvNet_2} = (1 - \alpha) * \mathcal{L}_{\mathcal{C}_2} + \alpha * D_{KL}(q_1 \| q_2) \tag{4}$$

where $\mathcal{L}_{ConvNet_1}/\mathcal{L}_{ConvNet_2}$ denotes the cross entropy loss of the *ConvNet*1/*ConvNet*2 network. q_1/q_2 indicates the predicted probability of the *Net1*/*Net*2 network. $D_{KL}(q_2\|q_1)$ presents the Kullback Leibler (*KL*) divergence loss, indicates the KL distance from q_1 to q_2. $D_{KL}(q_1\|q_2)$ has the similar meaning with that of $D_{KL}(q_2\|q_1)$. Furthermore, α denotes a hyperparameter which is employed to balance the two loss items.

The cross entropy losses $\mathcal{L}_{C_1}$ and $\mathcal{L}_{C_2}$ can be defined as:

$$\mathcal{L}_{C_1} = -\sum_{i=1}^{\mathcal{N}}\sum_{c=1}^{\mathcal{C}} y_i^c * \log(q_1^c(x_i)) \tag{5}$$

$$\mathcal{L}_{C_2} = -\sum_{i=1}^{\mathcal{N}}\sum_{c=1}^{\mathcal{C}} y_i^c * \log(q_2^c(x_i)) \tag{6}$$

where y_i denotes the ground truth label of sample x_i, y_i^c is the indicator function, when $y_i = c$, $y_i^c = 1$, when $y_i \neq c$, $y_i^c = 0$.

$\mathcal{L}_{C_1}$ and $\mathcal{L}_{C_2}$ represent the cross entropy losses between the ground truth label of the sample and the corresponding network prediction value, which will force the model to predict results that close to the true label for the corresponding sample.

To enhance the generalization ability of *ConvNet*1/*ConvNet*2 on the test samples, the posterior probability q_2/q_1 of another student network *ConvNet*1/*ConvNet*2 is integrated into the model. Further, the matching degree of the posterior probability of the two networks can be obtained by the following equations:

$$D_{KL}(q_2\|q_1) = \sum_{i=1}^{\mathcal{N}}\sum_{c=1}^{\mathcal{C}} q_2^c(x_i) * \log(\frac{q_2^c(x_i)}{q_1^c(x_i)}) \tag{7}$$

$$D_{KL}(q_1\|q_2) = \sum_{i=1}^{\mathcal{N}}\sum_{c=1}^{\mathcal{C}} q_1^c(x_i) * \log(\frac{q_1^c(x_i)}{q_2^c(x_i)}) \tag{8}$$

where $q_1^c(x_i)$ denotes the probability that sample x_i is predicted to be class c in *ConvNet*1 and $q_2^c(x_i)$ indicates the probability that x_i is predicted to be class c in *ConvNet*2.

2.3. Network Structure

The mutual learning model for CHM classification proposed in this paper contains two student networks (shown in Figure 2). Specifically, student network mainly involves four networks: ResNet18, ResNet34, ResNet50, and ResNet101 [18], respectively. During the training process, the two student networks can learn from each other and obtain better features, which improves their accuracy in CHM classification. Concerning the testing phase, each student network can be utilized to perform the CHM classification task. Meanwhile, they can obtain uniformly better performance than that of the single network without mutual learning.

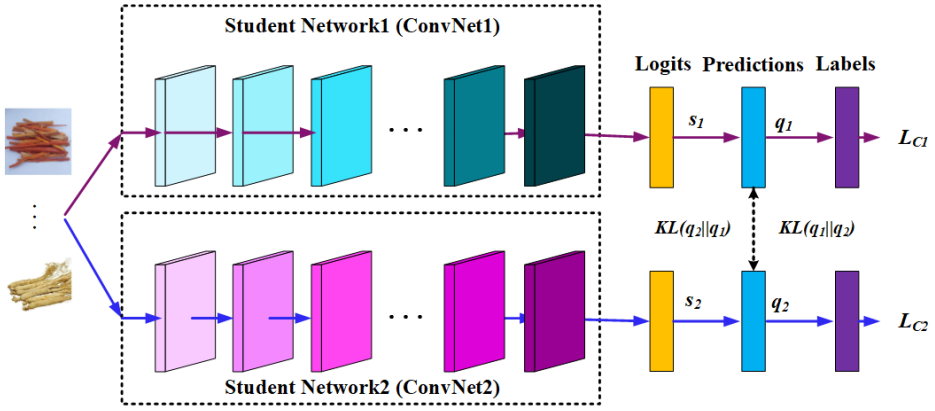


Figure 2. Mutual learning framework for CHM classification.

2.4. Model Training

For fair comparisons, all models mentioned in the experiments utilized stochastic gradient descent (SGD) [19,20] to learn the parameters, with a batch size of 32 (the number of samples input to the model at one time). The initial learning rate is 0.01, which decreased to 0.1 of its previous value after every 80 epochs, and the training phase stopped at 200 epochs. The momentum is set as 0.9, and the hyperparameter α is 0.8.

2.5. Evaluation Criteria

In order to evaluate the effectiveness of different models, several commonly utilized popular evaluation criteria, including Accuracy, Parameters, and Loss are considered. The accuracy represents the classification performance of the model, the higher the better. Parameters indicate the number of parameters of the model and can be utilized to measure the efficiency of the model, the smaller the better. The loss value demonstrates the difference between the model predicted value and the ground truth label of the sample, the smaller the better. Note that all experiments are performed on Nvidia's Titan X GPUs (12 GB), Intel's TMi7-7800X CPUs, 96 GB of RAM, and Ubuntu 18.04.

3. Experimental Results and Analysis

3.1. Evaluation of the Performance of the Mutual Learning Models

In order to verify the performance of the mutual learning model for CHM classification in this paper, we have compared it with several current popular deep learning models, including MobileNetV2 [21], MobileNetV3 [22], ResNet18 [18] and ResNet50 [18] respectively. The results are shown in Table 1. In Table 1, Mul18(50) and Mul50(18) indicate the mutual learning models based on two student networks (ResNet18 and ResNet50). Specifically, Mul18(50) refers to the model that shares the same structure with that of ResNet18 and learns knowledge from ResNet50. Furthermore, Mul50(18) has the similar meaning with that of Mul18(50), which denotes the model that has the same structure with that of ResNet50 and achieves knowledge from ResNet18.

Table 1 illustrates that the two variants of the residual networks, ResNet18 and ResNet50, have obtained superior accuracies than those of MobileNetV2 and MobileNetV3. Furthermore, our model Mul18(50) and Mul50(18) obtain the better CHM classification results than those of ResNet18 and ResNet50, which validate the effectiveness of the mutual learning models.

Table 1. Comparison results of different models.

Model	Accuracy (%)
MobileNetV2	70.75
MobileNetV3	73.05
ResNet18	74.05
ResNet50	76.90
Mul18(50) (Mutual learning model)	**77.95**
Mul50(18) (Mutual learning model)	**80.10**

Moreover, the superiority of our model can be illustrated from the following two aspects. On the one hand, the mutual learning model can achieve surpassing performance without increasing the number of parameters. Specifically, Mul18(50) has 5.3% higher accuracy than that of ResNet18, and the accuracy Mul50(18) is 4.2% better than that of ResNet50, respectively. On the other hand, the shallow mutual learning model can achieve comparable or even better performance than that of the deep single model. For example, Mul18(50) obtained 1.3% better classification accuracy than that of ResNet50 (the parameters of ResNet50 model is nearly twice times of Mul18(50)). The reason is that, during training, Mul18(50) not only learned the knowledge from the ResNet50 model, but also from its own. This learning strategy is comparable to ensemble learning, which integrates several

low-performing weak classifiers (ResNet18 + ResNet50) into a high-performing strong classifier Mul18(50), resulting in a superior outcome.

3.2. Ablation Studies

In order to further assess the effectiveness of the mutual learning model in CHM classification, several ablation studies have been designed in the following. Specifically, one design is to verify the mutual learning model's performance under two identical student networks setting, and the other one is to evaluate the mutual learning model's performance under two distinct student networks setting, respectively.

3.2.1. Evaluation of the Performance of the Mutual Learning Model under Two Identical Student Networks

In order to verify the performance of the mutual learning model under two identical student networks setting, we compare two group models that with or without mutual learning, including Sig*i* and Mul*i*(*i*) respectively. Among them, *i* indicates the model with *i*th layers, $i \in \{18, 34, 50, 101\}$. Sig*i* indicates the single ResNet*i* model without mutual learning and Mul*i*(*i*) denotes the mutual learning models with two identical student networks ResNet*i*. Mul*i*(*i*)-1 indicates the first student network and Mul*i*(*i*)-2 represents the second student network. Table 2 shows the compared results.

Table 2. Comparison results of the mutual learning models based on two identical student networks.

Model	Accuracy (%)	Parameters (M)	Loss Value
Sig18	74.05	11.69	0.0542
Mul(18)-1	**76.50**	11.69	**0.0440**
Mul(18)-2	**76.25**	11.69	**0.0440**
Sig34	75.00	21.80	0.0548
Mul(34)-1	**78.15**	21.80	**0.0432**
Mul(34)-2	**78.15**	21.80	**0.0432**
Sig50	76.90	25.56	0.0542
Mul(50)-1	**79.25**	25.56	**0.0382**
Mul(50)-2	**79.20**	25.56	**0.0382**
Sig101	77.45	44.64	0.0554
Mul(101)-1	**79.85**	44.64	**0.0373**
Mul(101)-2	**79.80**	44.64	**0.0373**

Table 2 shows that when the two student networks are identical, the mutual learning models can achieve significant better results than those of single network model in terms of accuracy and loss value, without increasing the parameters. Specifically, the accuracy of Mul18(18)-1/Mul18(18)-2 is 3.0% higher than that of Sig18, and the loss value of Mul18(18)-1/Mul18(18)-2 is 18.8% lower than that of Sig18; Mul34(34)-1/Mul34(34)-2 has a 4.2% higher accuracy than that of Sig34, and the loss value of Mul34(34)-1/Mul34(34)-2 is 30.3% lower than that of Sig34; Mul50(50)-1/Mul50(50)-2 achieves 3.1% higher than that of Sig50 in terms of accuracy, and Mul50(50)-1/Mul50(50)-2 is 29.5% lower than that of Sig50 in terms of loss value; Mul101(101)-1/Mul101(101)-2 obtains 3.1% higher than that of Sig101, and Mul101(101)-1/Mul101(101)-2 has 32.7 % lower loss value than that of Sig101, respectively.

Moreover, the shallow mutual learning models can obtain better results than those of deep single models at both accuracy and loss value, with much less parameters. Concretely, Mul18(18) obtains about 2% higher accuracy and 19.7% lower loss value than those of Sig34; Mul34(34) achieves about 1.62% higher accuracy and 20.3% lower loss value than those of Sig50; Mul34(34) acquires about 0.9% higher accuracy and 22% lower loss value than those of Sig101 and Mul50(50) obtains about 2.32% higher accuracy and 31.04% lower loss value than those of Sig101, separately.

Those results verify the effectiveness of the mutual learning model based on two identical student networks in terms of effectiveness and efficiency.

It should be noted that the two student networks of the mutual learning model (such as Mul18(18)-1/Mul18(18)-2) have the same parameter values as those of their corresponding single counterpart without mutual learning (such as Sig18). In this case, the mutual learning models can achieve higher accuracy and lower loss value, indicating that the mutual learning model can allow two identical student networks to learn additional knowledge about the other network during training, thereby obtaining better performance. Although the two student networks have the same structure, their initial parameter values are different, resulting in diverse knowledge throughout the training process, which may be employed to guide the learning of another student network. Furthermore, due to the similar structure, these two student networks are all converge to the same tiny range of values after training, resulting in approximately same accuracy and loss value.

To intuitively compare the performance of different models, Figures 3 and 4 present the accuracies and loss values of different models under different training epochs respectively. These two figures illustrate that during the whole training process, the accuracies/losses of the two student networks of the mutual learning model are higher/lower than those of the corresponding single counterpart models without mutual learning, which further verifies the effectiveness of mutual learning models under two identical student networks setting.

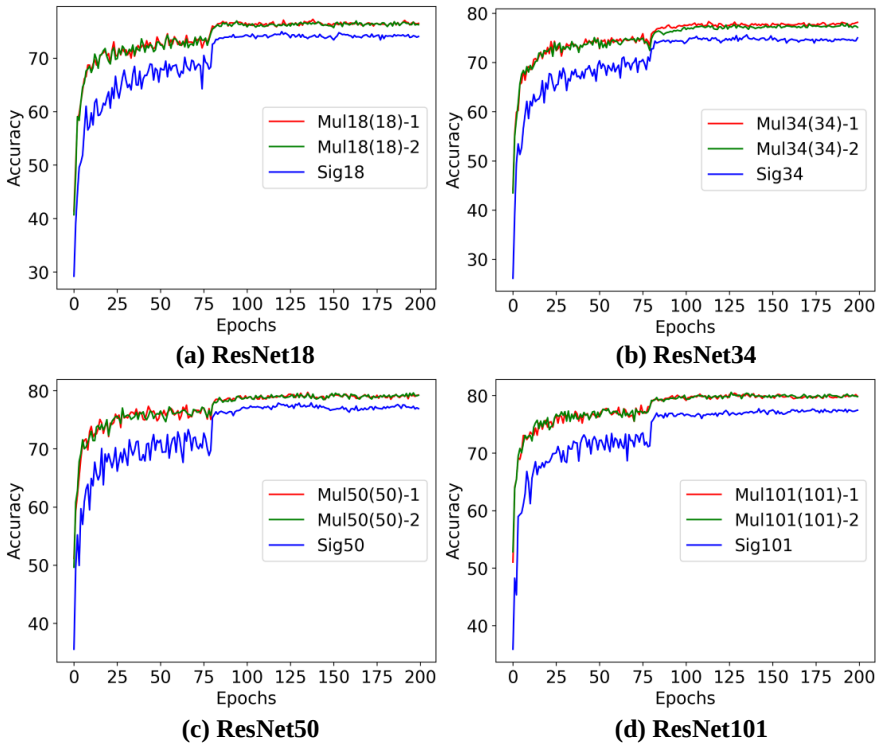
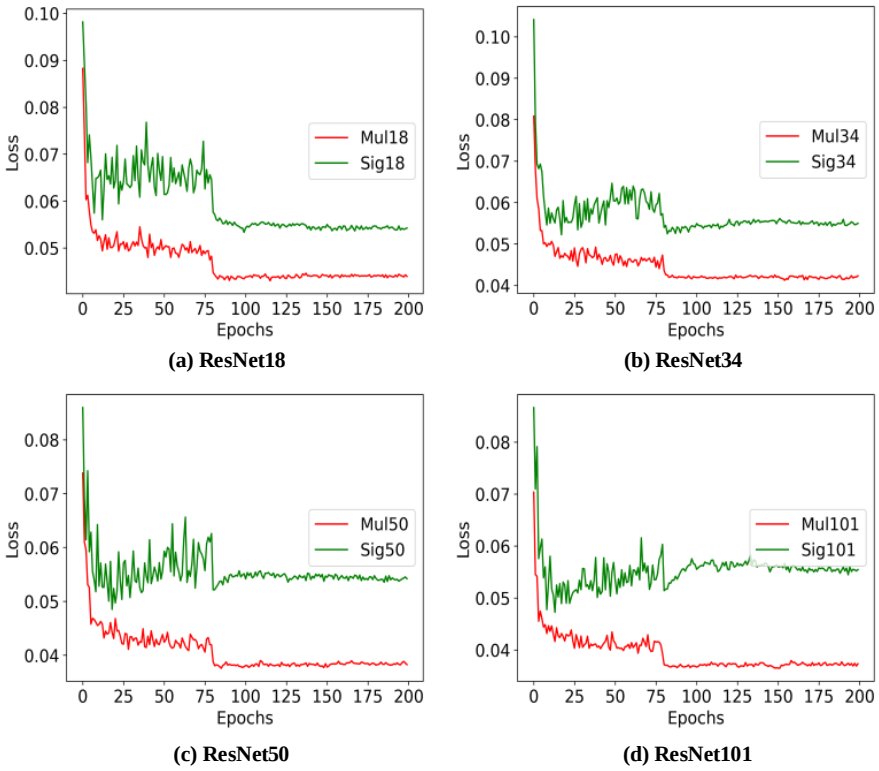


Figure 3. Comparison of the accuracy of two same student networks with or without mutual learning.

3.2.2. Evaluation of the Performance of Mutual Learning under Two Different Student Networks

Considering the performance of the mutual learning model under two different student networks setting, we also compare two set of models that with or without mutual learning, including [Sig$i$, Sig$j$] and [Mul$i$(j), Mul$j$(i)] respectively. Among them, i/j indicates the model with ith/jth layers, $i, j \in \{18, 34, 50, 101\}$, $i \neq j$. Sig$i$ indicates the single ResNet$i$ model without mutual learning, Sig$j$ has the similar meaning with that of Sig$i$. Mul$i$(j) and Mul$j$(i) represent the mutual learning models with two different student networks ResNet$i$ and ResNet$j$. Mul$i$(j) represents the one student network in mutual learning

that has the same structure with Sig$i$ and receives the knowledge from Sig$j$, Mul$j$(i) shares the similar meaning with that of Mul$i$(j). The comparison results are shown in Table 3.



(a) ResNet18 **(b) ResNet34**

(c) ResNet50 **(d) ResNet101**

Figure 4. Comparison of the loss of two same student networks with or without mutual learning.

Table 3 demonstrates that when the two student networks are different, the mutual learning models have achieved better performance than those of their corresponding single networks with the same number of layers. Specifically, Mul18(34)/Mul34(18) receives 4.69%/4.37% higher accuracy and 30.1%/30.8% lower than Sig18/Sig34 in terms of accuracy and loss value; Mul18(50)/Mul50(18) obtains 5.27%/4.16% better accuracy and 32.1%/32.1% lower loss value than those of Sig18/Sig50; Mul18(101)/Mul101(18) achieves 4.86%/3.68% better accuracy and 34.1%/33.4% lower loss value than those of Sig18/Sig101; Mul34(50)/Mul50(34) receives 4.26%/5.07% better accuracy and 34.1%/33.4% lower loss value than those of Sig34/Sig50; Mul34(101)/Mul101(34) obtains 5.40%/3.29% better accuracy and 33.2%/33.9% lower loss value than those of Sig34/Sig101; Mul50(101)/Mul101(50) achieves 3.97%/4.20% higher accuracy and 36.0%/37.3% lower loss value than those of Sig50/Sig101, respectively.

Furthermore, Table 3 also illustrates that the small student networks (such as Mul18(34)) in mutual learning models also receive superior performance than a single network (such as Sig34) corresponding to the large student network in the mutual learning model. Concretely, Mul18(34) obtains 3.33% higher accuracy and 30.1% lower loss value than that of Sig34. Mul18(50) achieves 1.36% better accuracy and 1.36% lower loss value than that of Sig50. Mul18(101) receives a 2.60% better accuracy and 32.3% lower loss value than that of Sig101. Mul34(50) obtains 1.70% superior accuracy and 34.1% lower loss value than that of Sig50; Mul34(101) receives 2.06% superior accuracy and 33.2% lower loss value than that of Sig101; Mul50(101) receives 3.23% superior accuracy and 36.0% lower loss value than that of Sig101, respectively.

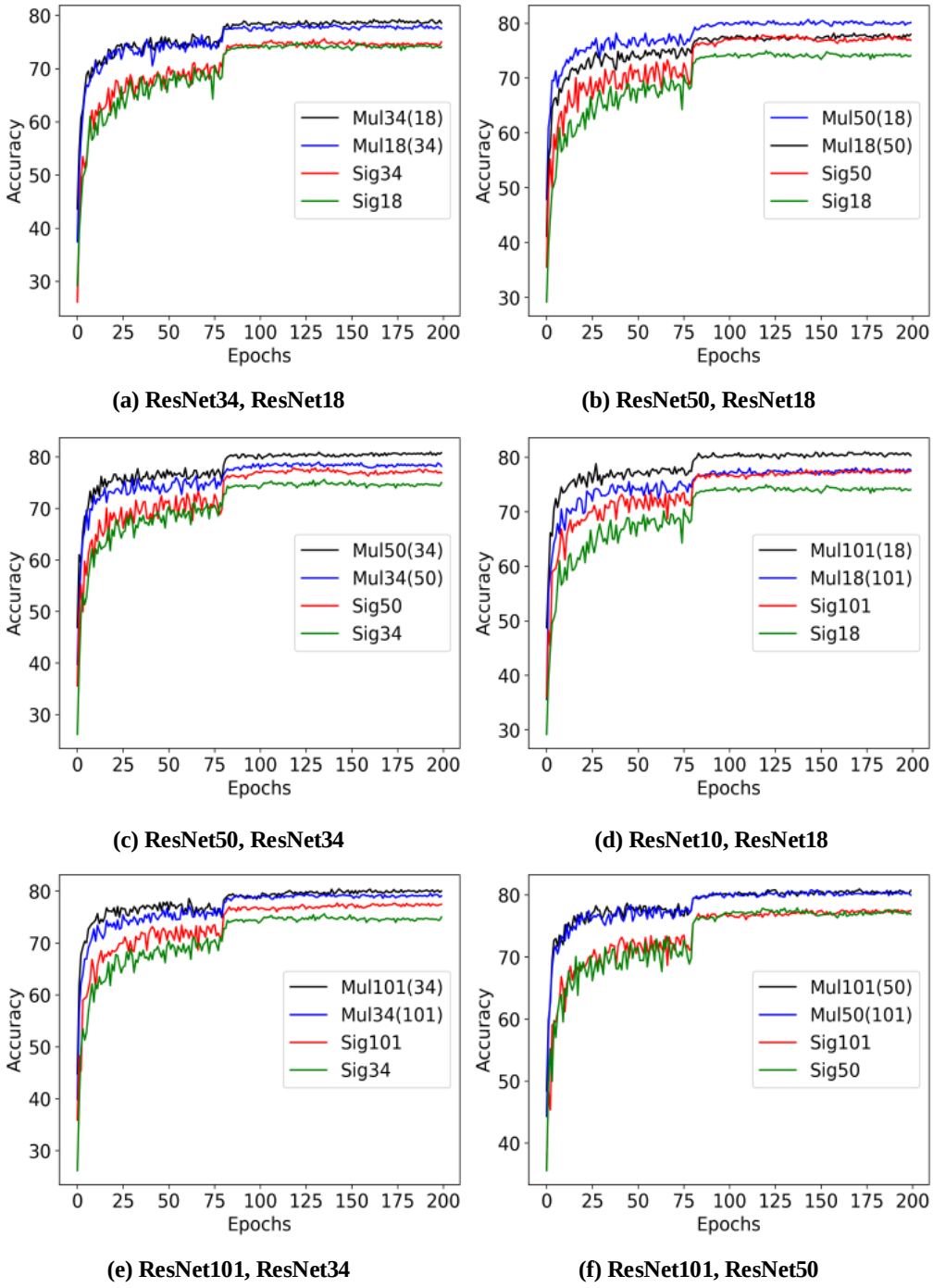
Table 3. Comparison results of the mutual learning models based on two different student networks.

Model	Accuracy (%)	Parameters (M)	Loss Value
Sig18	74.05	11.69	0.0542
Sig34	75.00	21.80	0.0548
Mul18(34)	77.50	11.69	0.0379
Mul34(18)	78.55	21.80	0.0379
Sig18	74.05	11.69	0.0542
Sig50	76.90	25.56	0.0542
Mul18(50)	77.95	11.69	0.0368
Mul50(18)	80.10	25.56	0.0368
Sig34	75.00	21.80	0.0548
Sig50	76.90	25.56	0.0542
Mul34(50)	78.20	21.80	0.0361
Mul50(34)	80.80	25.56	0.0361
Sig18	74.05	11.69	0.0542
Sig101	77.45	44.64	0.0554
Mul18(101)	77.65	11.69	0.0367
Mul101(18)	80.35	44.64	0.0367
Sig34	75.00	21.80	0.0548
Sig101	77.45	44.64	0.0554
Mul34(101)	79.50	21.80	0.0366
Mul101(34)	80.00	44.64	0.0366
Sig50	76.90	25.56	0.0542
Sig101	77.45	44.64	0.0554
Mul50(101)	79.95	25.56	0.0347
Mul101(50)	80.70	44.64	0.0347

Those results verify the effectiveness of the mutual learning models based on two different student networks in terms of effectiveness and efficiency.

It should be noted that although Mul18(34) is the small student network in the mutual learning model, it receives better performance not only than that of small single network Sig18 but also than that of big single network Sig34. Reasons are that, during the whole training process, Mul18(34) collaborative learns knowledge from ResNet18 and ResNet34, which results in richer and much more reliable features, leading to improved performance in CHM classification.

In order to intuitively compare the performance of different models, Figures 5 and 6 show the accuracies and loss values of different models under different training epochs respectively. These two figures demonstrate that during the whole training process, the accuracies/losses of the two student networks of the mutual learning model are significantly higher/lower than those of the corresponding single counterpart models without mutual learning, which further validates the effectiveness of mutual learning models under two different student networks setting. In the future, we will employ a deep no-reference image quality metric [23] to objectively assess the performance of the proposed method in the presence of image artefacts of varying degrees of severity.

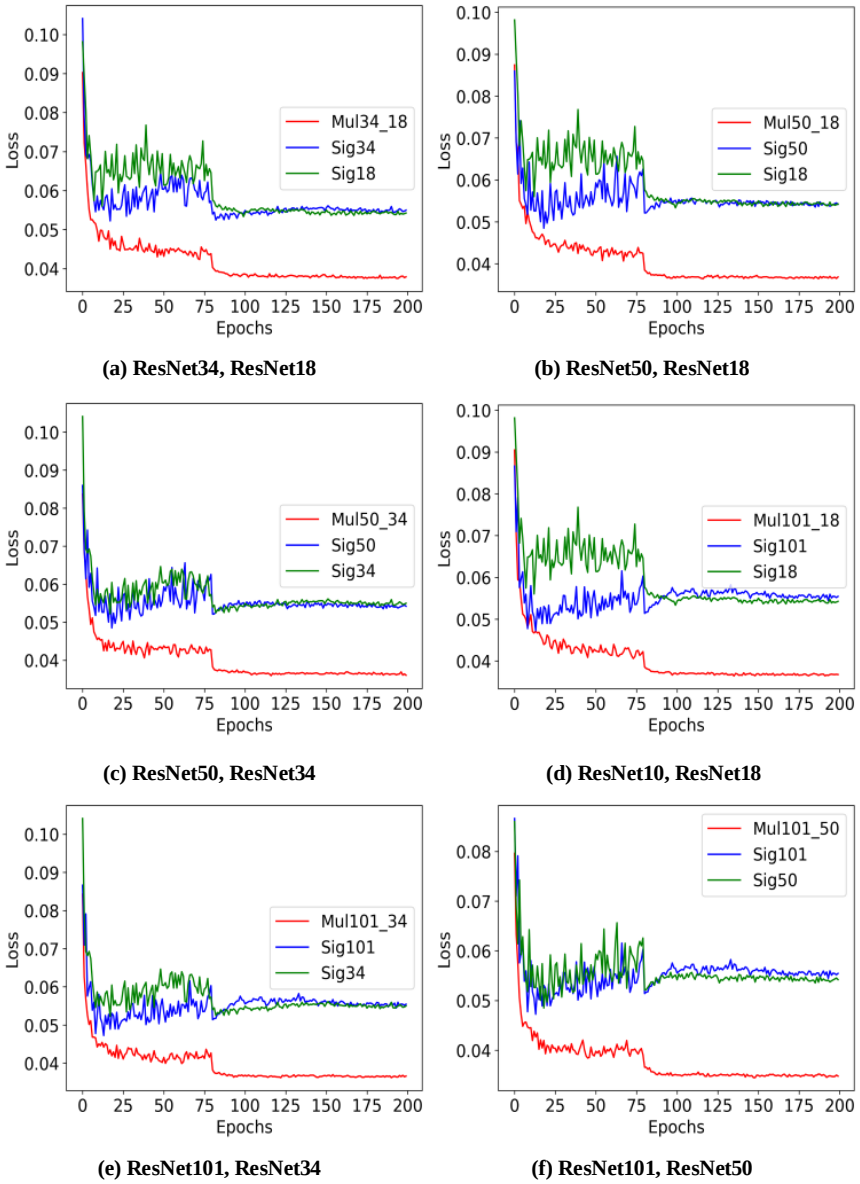


(a) ResNet34, ResNet18 **(b) ResNet50, ResNet18**

(c) ResNet50, ResNet34 **(d) ResNet10, ResNet18**

(e) ResNet101, ResNet34 **(f) ResNet101, ResNet50**

Figure 5. Comparison of the accuracy of two different student networks with or without mutual learning.



(a) ResNet34, ResNet18 **(b) ResNet50, ResNet18**

(c) ResNet50, ResNet34 **(d) ResNet10, ResNet18**

(e) ResNet101, ResNet34 **(f) ResNet101, ResNet50**

Figure 6. Comparison of the loss of two different student networks with or without mutual learning.

4. Conclusions

This paper has proposed to utilize a novel mutual learning model to perform CHM classification. This model can transfer the knowledge from one student network to another one. Furthermore, benefit from the mutual learning, a set of basic student neural networks can cooperatively update parameters and achieve information from each other throughout the training process, resulting in the learning of rich and robust features. Consequently, our model obtains promising CHM classification results with better efficiency and effectiveness. Specifically, experiments achieve 4.16∼13.2% higher accuracy and 8.8∼36.0% lower loss value than those of the non-mutual learning models, which verifies the effectiveness of the mutual learning model in CHM classification.

Author Contributions: Conceptualization, J.Z. and Y.R.; methodology, M.H.; validation, M.H.; formal analysis, F.H.; investigation, M.H.; resources, M.H.; data curation, M.H.; writing—original draft preparation, M.H.; writing—review and editing, M.H. and Y.Z.; supervision, J.Z.; project administration, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (NSFC) under Grant (No. 62072146, No. 61972358); the Key Research and Development Program of China (2019YFB2102101); the Key Research and Development Program of Zhejiang Province (2019C01059, 2019C03135) and Natural Science Basic Research Plan in Shaanxi Province of China (2022JM-371).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: We declare that this paper has no conflicts of interest. Furthermore, we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Abbreviations

The following abbreviations are used in this manuscript:

CHM	Chinese Herbal Medicine
LLE	Local Linear Embedding
LDA	Linear Discriminant Analysis
SVM	Support Vector Machine
PCA	Principal Component Analysis
SOM	Self-organizing Map
CHMC	Chinese Herbal Medicine Classification
KL	Kullback Leibler

References

1. Wang, J.; Wong, R.K.W.; Lee, T.C.M. Locally linear embedding with additive noise. *Pattern Recognit. Lett.* **2019**, *123*, 47–52. [CrossRef]
2. Zhang, S.; Lei, Y.; Dong, T.; Zhang, X.P. Label propagation based supervised locality projection analysis for plant leaf classification. *Pattern Recognit.* **2013**, *46*, 1891–1897. [CrossRef]
3. Unger, J.; Merhof, D.; Renner, S. Computer vision applied to herbarium specimens of German trees: Testing the future utility of the millions of herbarium specimen images for automated identification. *BMC Evol. Biol.* **2016**, *16*, 248. [CrossRef] [PubMed]
4. Luo, D.; Wang, J.; Chen, Y. Classification of Chinese Herbal medicines based on SVM. In Proceedings of the International Conference on Information Science, Electronics and Electrical Engineering (ISEEE), Sapporo, Japan, 26–28 April 2014; pp. 453–456.
5. Wang, M.; Li, L.; Yu, C.; Yan, A.; Zhao, Z.; Zhang, G.; Jiang, M.; Lu, A.; Gasteiger, J. Classification of Mixtures of Chinese Herbal Medicines Based on a Self-organizing Map (SOM). *Mol. Inform.* **2016**, *35*, 109–115. [CrossRef] [PubMed]
6. Mu, L.; Gao, Z.; Cui, Y.; Li, K.; Liu, H.; Fu, L. Kiwifruit Detection of Far-view and Occluded Fruit Based on Improved AlexNet. *Trans. Chin. Soc. Agric. Mach.* **2019**, *50*, 24–34.
7. Chen, Y.Y.; Gong, C.Y.; Liu, Y.Q. Fish Identification Method Based on FTVGG16 Convolutional Neural Network. *Trans. Chin. Soc. Agric. Mach.* **2019**, *50*, 223–231.
8. Xue, Y.; Wang, L.; Zhang, Y.; Shen, Q. Defect Detection Method of Apples Based on GoogLeNet Deep Transfer Learning. *Trans. Chin. Soc. Agric. Mach.* **2020**, *51*, 30–35.
9. Jiao, J.; Wang, W.; Hou, J. Freshness Identification of Iberico Pork Based on Improved Residual Network. *Trans. Chin. Soc. Agric. Mach.* **2019**, *50*, 364–371.
10. Chen, J.; Chen, L.Y.; Wang, S.S.; Zhao, H.Y.; Wen, C.J. Pest Image Recognition of Garden Based on Improved Residual Network. *Trans. Chin. Soc. Agric. Mach.* **2019**, *50*, 187–195.
11. Wang, C.; Zhao, Q.; Ma, Y. Crop Identification of Drone Remote Sensing Based on Convolutional Neural Network. *Trans. Chin. Soc. Agric. Mach.* **2019**, *50*, 161–168.
12. Liu, X.; Fan, C.; Li, J. Identification Method of Strawberry Based on Convolutional Neural Network. *Trans. Chin. Soc. Agric. Mach.* **2020**, *51*, 237–244.
13. Hou, J.; Yao, E.; Zhu, H. Classification of Castor Seed Damage Based on Convolutional Neural Network. *Trans. Chin. Soc. Agric. Mach.* **2020**, *51*, 440–449.
14. Liu, S.; Chen, W.; Dong, X. Automatic Classification of Chinese Herbal Based on Deep Learning Method. In Proceedings of the 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Huangshan, China, 28–30 July 2018.

15. Cai, C.; Liu, S.; Wang, L.; Yang, B.; Zhi, M.; Wang, R.; He, W. Classification of Chinese Herbal Medicine Using Combination of Broad Learning System and Convolutional Neural Network. In Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, 6–9 October 2019.
16. Zhang, Y.; Xiang, T.; Hospedales, T.M.; Lu, H. Deep Mutual Learning. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 4320–4328.
17. Hao, W.; Han, M.; Yang, H.; Hao, F.; Li, F. A novel Chinese herbal medicine classification approach based on EfficientNet. *Syst. Sci. Control Eng.* **2021**, *9*, 304–313. [CrossRef]
18. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26–30 June 2016; pp. 770–778.
19. Bottou, L. Stochastic Gradient Descent Tricks. In *Neural Networks: Tricks of the Trade*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 421–436.
20. Chen, W.; Mirdehghan, P.; Fidler, S.; Kutulakos, K.N. Auto-Tuning Structured Light by Optical Stochastic Gradient Descent. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
21. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
22. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for MobileNetV3. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision(ICCV), Seoul, Korea, 27 October–3 November 2019.
23. Mukherjee, S.; Valenzise, G.; Cheng, I. Potential of deep features for opinion-unaware, distortion-unaware, no-reference image quality assessment. In Proceedings of the International Conference on Smart Multimedia, San Diego, CA, USA, 16–18 December 2019.