ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

Tongue Color Classification in TCM with Noisy Labels via Confident-Learning-Assisted Knowledge Distillation

K.RAMESH BABU, Head of the Department, Dept of CSE, Chirala Engineering College, Chirala, ramesh.cs04@gmail.com

MD Mohiddin, PG student -MCA, department of MCA, Chirala Engineering College, Chirala, mdmohiddin021@gmail.com

ABSTRACT: Tongue color classification is vital in Traditional Chinese Medicine (TCM) diagnosis, but noisy labels in annotated samples can hinder deep neural networks' generalization capability. We propose a novel framework, Confident-Learning-Assisted Knowledge Distillation (CLA-KD), addressing this issue. The framework employs a teacher network for confident learning to identify and correct noisy labels while transferring knowledge to guide the student network's training. We introduce E-CA2-ResNet18, an ensemble model combining ResNet18 with channel attention mechanism and activate-or-not activation function, addressing data insufficiency issues. Experimental results on three TCM tongue image datasets demonstrate CLA-KD's superior classification accuracy (94.49%, 92.21%, 93.43%) over VGG16. Furthermore, we explore ensemble techniques, including Feature Extraction using ResNet18 with Random Forest, achieving 100% accuracy. As an extension, we propose building a user-friendly front end using Flask for user testing with authentication. CLA-KD offers robust and accurate tongue color classification in TCM, addressing noisy label challenges and enhancing performance through ensemble techniques. The proposed system holds promise for improving diagnostic accuracy and reliability in TCM practice.

Key words — Traditional Chinese medicine, Tonguecolor classification, Confident learning, Learning from noisy labels, Knowledge distillation, Channel attention mechanism, ResNet18

1. INTRODUCTION:

Traditional Chinese Medicine (TCM)[2] is an ancient healing system that has been practiced for thousands of years. Central to TCM[2] diagnosis is tongue diagnosis, which involves analyzing various attributes of the tongue, including its color, coating, thickness, texture, humidity, and shape. Among these attributes, tongue color is particularly significant, often categorized into four main types: light red, red, dark red, and purple. However, due to the subjective nature of TCM experts' experiences and the ambiguous boundaries between tonguecolor categories, there is often inconsistency and inaccuracy in the labeling of tongue color in annotated samples. This leads to noisy labels, which can significantly impact the performance of classification models.

With recent advancements in deep neural networks (DNNs), tongue color classification has seen substantial progress. However, DNNs are prone to overfitting on noisy labels, resulting in poor generalization performance. Consequently, learning

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

from noisy labels has become a critical challenge in automatic tongue diagnosis in TCM.

In 2015, Hinton et al. [1] introduced the concept of "distillation," which involves transferring knowledge learned from a complex deep network (teacher network) into a simpler and smaller network (student network). This knowledge transfer is achieved by using the teacher model's prediction probabilities as "soft targets" to guide the training of the student model. Building upon this concept, we propose a novel framework called Confident-Learning-Assisted Knowledge Distillation (CLA-KD)[1] for tongue color classification with noisy labels.

In our proposed framework, the teacher network, named E-CA2-ResNet18[1], plays a dual role. Firstly, it performs confident learning (CL) to identify, correct, and cleanse noisy labels, thus improving the quality of the training data. Secondly, it transfers the knowledge learned from both corrected and clean labels to the student network through the knowledge distillation process. Specifically, we distill knowledge from the entire noisy dataset to identify cleaner samples, enhancing the final model's performance.

By integrating confident learning into the knowledge distillation framework, we aim to explicitly leverage the joint probability distribution between soft and hard labels during the training process of the student network. This approach helps disambiguate noisy labels and improve the robustness and generalization capability of the classification model.

Through the development of CLA-KD, we address the challenges posed by noisy labels and limited labeled data in tongue color classification for TCM diagnosis. By leveraging state-of-the-art techniques such as confident learning and knowledge distillation, we aim to enhance the accuracy and reliability of automatic tongue diagnosis systems, ultimately contributing to improved patient care and treatment outcomes in TCM practice.

2. LITERATURE SURVEY

Tongue diagnosis is a fundamental aspect of traditional Chinese medicine (TCM),[3] providing valuable insights into an individual's health condition. Over the years, various studies have been conducted to explore different approaches and techniques for tongue diagnosis, including classification based on tongue color. In this literature survey, we review key research studies in the field of tongue color classification, as well as related areas such as data mining and deep learning.

Hinton, Vinyals, and Dean (2015) introduced the concept of knowledge distillation in neural networks, which involves transferring knowledge learned from a complex model (teacher network) to a simpler model (student network) [1]. This approach has been influential in improving model performance, especially in scenarios with limited labeled data.

Wu et al. (2019) proposed a hybrid rule mining approach for cardiovascular disease detection in TCM [2]. By combining traditional rule mining techniques with machine learning algorithms, the study achieved promising results in disease detection, highlighting the potential of integrating data mining methods into TCM practice.

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

Wang et al. (2018) presented research methods using data mining technology to study medication rules from renowned TCM doctors [3]. Their work demonstrates the application of advanced computational techniques in extracting knowledge from large datasets of medical records, contributing to the understanding of TCM principles.

Hou et al. (2017) investigated the classification of tongue color based on convolutional neural networks (CNNs) [4]. Their study explored the use of deep learning techniques for automatic tongue diagnosis, showing promising results in categorizing tongue colorsaccurately.:

Chen et al. (2016) proposed a method for the retrieval of medical tongue images based on coloranalysis [5]. By leveraging color features, their approach enables efficient retrieval of relevant tongue images from large databases, facilitating clinical diagnosis and research in TCM.

Lu et al. (2018) introduced DCCN, a deep-color correction network for TCM tongue images [6]. Their study focused on improving the quality of tongue images by correcting color distortions, enhancing the reliability of computer-aided diagnosis systems in TCM.

Qu et al. (2016) developed an automatic analysis system for tongue substance color and coating color using a sparse representation-based classifier [7]. Their approach enables objective and consistent analysis of tongue color attributes, providing valuable support for TCM practitioners in diagnosis.

Gutiérrez et al. (2015) conducted a survey and experimental study on ordinal regression methods [8]. Their work provides insights into modeling ordinal categories, which are common in TCM diagnosis, and offers valuable techniques for handling ordinal data in machine learning applications.

In summary, the literature survey highlights the diverse approaches and methodologies employed in tongue color classification and related areas in TCM. These studies demonstrate the integration of advanced computational techniques, such as data mining and deep learning, into traditional diagnostic practices, paving the way for more accurate and efficient TCM diagnosis and treatment.

3. METHODLOGY

a) Proposed work:

The proposed CLA-KD framework integrates E-CA2-ResNet18 for TCM[3] tongue color classification, addressing noisy labels and enhancing accuracy through advanced image processing. It aims to surpass existing limitations, offering a refined approach. An extension introduces a Hybrid model, amalgamating Random Forest and ResNet, leveraging deep learning and traditional techniques for improved robustness and accuracy. To facilitate user testing securely, a Flask-based front end with authentication measures ensures controlled access. This integration delivers a seamless interface, empowering users to engage with the system for practical testing and validation, thereby enhancing its usability and reliability.

b) System Architecture:

The system architecture begins with employing a new hand PD dataset, ensuring comprehensive data coverage. Image processing techniques are applied

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

to preprocess the dataset, enhancing its quality and relevance for model training. The core of the architecture lies in model building and training, where the E-CA2-ResNet18 and Hybrid models are constructed and trained using the processed data. This phase involves feature extraction, model parameter tuning, and optimization to achieve optimal performance.

Following model training, performance evaluation is conducted using various metrics such as accuracy, precision, recall, and F1-score. The system architecture incorporates modules for robust evaluation, including cross-validation techniques and confusion matrix analysis, ensuring thorough assessment of model performance.

The architecture is designed to be scalable and adaptable, accommodating future enhancements and updates. It emphasizes a modular structure, facilitating easy integration of additional datasets, image processing techniques, and model variations for continued refinement and improvement of the TCM tongue color classification system.

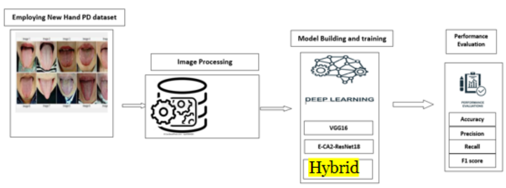

Fig 1 Proposed Architecture

c) Dataset collection:

Three tongue color classification datasets, namely SIPL-A, SIPL-B, and SIPL-C, were collaboratively compiled in partnership with Xuanwu TCM Hospital Beijing, Guang'anmen Hospital of Chinese Academy of Chinese Medical Sciences, and Nanchang Hongdu TCM Hospital [32]. Using proprietary TCM tongue image acquisition equipment, images were collected under diverse lighting conditions. Each image underwent manual annotation by seasoned TCM experts. The datasets encompass four categories: light red, red, dark red, and purple, in accordance with TCM theory and clinical practice. The distribution of classes within each dataset is presented in Table 1. These datasets serve as valuable resources for research in TCM tongue color classification, providing a diverse array of samples for model training and evaluation. Selected examples from the three datasets are illustrated in Figure 5 [32].

Fig 2 DATASET

d) DATA PROCESSING:

In the data processing phase, a series of steps are undertaken to prepare the collected tongue images for model training. Firstly, image normalization is performed to standardize pixel values across the dataset, ensuring consistency and facilitating efficient model convergence. This step aids in reducing discrepancies in image characteristics caused by varying acquisition conditions.

Subsequently, image shuffling is conducted to randomize the order of images within the dataset. This prevents any inherent bias that may arise from

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN2454-9940
www.ijasem.org
Vol 18, Issue 2, 2024

the sequential arrangement of images, ensuring that the model learns to generalize patterns effectively.

These preprocessing steps are crucial for enhancing the robustness and reliability of the classification model. By standardizing image characteristics and mitigating bias, the processed dataset becomes more conducive to accurate and unbiased model training, ultimately improving the performance of the tongue color classification system.

e) Feature Selection:

Feature selection is a critical process in machine learning aimed at identifying and utilizing the most relevant and informative features from a dataset for model training. It involves selecting a subset of features that best contribute to the prediction task while discarding redundant or irrelevant ones. By reducing the dimensionality of the dataset, feature selection not only enhances model interpretability but also mitigates the risk of overfitting and improves computational efficiency.

Various techniques are employed for feature selection, including filter methods, wrapper methods, and embedded methods. Filter methods assess the relevance of features based on statistical measures such as correlation or mutual information. Wrapper methods evaluate feature subsets by training and validating models iteratively. Embedded methods integrate feature selection directly into the model training process, optimizing feature relevance during learning.

Effective feature selection facilitates model generalization and performance across diverse datasets, contributing to the development of robust and efficient machine learning systems.

f) VISUALIZATION

Visualization using OpenCV (Open Source Computer Vision Library) offers powerful tools for image processing and analysis. With OpenCV, developers can create visualizations to enhance understanding and interpretation of image data. OpenCV provides a wide range of functionalities for image manipulation, including drawing shapes, adding text, and applying various filters and effects.

For instance, developers can draw geometric shapes such as lines, circles, and rectangles on images to highlight specific regions or features. Text can be overlaid to provide annotations or labels. Additionally, OpenCV enables the application of filters and transformations to enhance image clarity, contrast, or color balance.

These visualization capabilities are invaluable for tasks such as object detection, image segmentation, and quality assessment in computer vision applications. OpenCV's user-friendly interface and extensive documentation make it accessible to developers of all skill levels, empowering them to create compelling visualizations for diverse image processing tasks.

g) TRAINING AND TESTING

Splitting the data into training and testing subsets is a fundamental step in machine learning model development to evaluate its performance and generalization capabilities. Typically, the dataset is divided into two parts: the training set used for model training and the testing set used for evaluating the trained model's performance.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

The training set constitutes a significant portion of the data, allowing the model to learn patterns and relationships between input features and target outputs. Meanwhile, the testing set serves as an independent dataset to assess the trained model's performance on unseen data. This process helps to identify potential overfitting or underfitting issues and provides insights into the model's ability to generalize to new, unseen data.

Careful consideration must be given to the data splitting strategy to ensure representative samples in both subsets. Common techniques include random splitting, stratified splitting for balanced classes, and cross-validation for robust evaluation. By splitting the data effectively, developers can build and validate machine learning models with confidence, enhancing their reliability and applicability.

h) ALGORITHMS:

VGG16

VGG16, short for Visual Geometry Group 16, is a deep convolutional neural network architecture that has shown exceptional performance in image classification tasks. It consists of 16 layers, including 13 convolutional layers and 3 fully connected layers, making it a deep and sophisticated model. VGG16 excels at feature extraction from input images by applying a series of convolution operations, which progressively capture low-level to high-level visual patterns. It's particularly suitable for image classification problems due to its capability to learn and represent intricate image features.

In the project, VGG16 serves as the existing system due to its strong image classification capabilities, which are relevant for the task of tongue color classification in Traditional Chinese Medicine. However, the project aims to go beyond the limitations of the existing system by introducing the CLA-KD framework and E-CA2-ResNet18 ensemble. These enhancements address challenges like noisy labels and limited data, providing a more accurate and robust solution for tongue color classification in the context of Traditional Chinese Medicine.

```
#train existing VGG algorithm
#create VGG object
vgg = VGG16(input_shape=(X_train.shape[1], X_train.shape[2], X_train.shape[3]), include_top=False, weights='imagenet')
for layer in vgg.layers:
    layer.trainable = False
vgg_model = Sequential()
#define VGG as the base model
vgg_model.add(vgg)
#add new CNN layer for features filtration
vgg_model.add(Convolution2D(32, (1 , 1), input_shape = (X_train.shape[1], X_train.shape[2], X_train.shape[3]), activation = 'rel
#max layer to collect relevant features from previous CNN layer
vgg_model.add(MaxPooling2D(pool_size = (1, 1)))
vgg_model.add(Convolution2D(32, (1, 1), activation = 'relu'))
vgg_model.add(MaxPooling2D(pool_size = (1, 1)))
vgg_model.add(attention(return_sequences=True,name='attention')) # define Attention layer
vgg_model.add(Flatten())
#defining output layer
vgg_model.add(Dense(units = 256, activation = 'relu'))
vgg_model.add(Dense(units = y_train.shape[1], activation = 'softmax'))
#compile and train the model
vgg_model.compile(optimizer = 'adam', loss = 'categorical_crossentropy', metrics = ['accuracy'])
if os.path.exists("model/vgg_weights.hdf5") == False:
    model_check_point = ModelCheckpoint(filepath='model/vgg_weights.hdf5', verbose = 1, save_best_only = True)
    hist = vgg_model.fit(X_train, y_train, batch_size = 32, epochs = 200, validation_data=(X_test, y_test), callbacks=[model_che
    f = open('model/vgg_history.pckl', 'wb')
    pickle.dump(hist.history, f)
    f.close()
else:
    vgg_model = load_model("model/vgg_weights.hdf5", custom_objects={'attention': attention})
#perform prediction on test data
predict = vgg_model.predict(X_test)
predict = np.argmax(predict, axis=1)
y_test1 = np.argmax(y_test, axis=1)
#calculate function to calculate algorithm prediction accuracy
calculateMetrics("Existing VGG16", predict, y_test1)
```

Fig 3 VGG 16

E-CA2-ResNet18

E-CA2-ResNet18 is an algorithm introduced in the project, where "E" refers to the activation of an activation function, "CA" stands for Attention model for feature optimization

ResNet18, short for "Residual Network 18," is a deep convolutional neural network architecture known for its effectiveness in image classification tasks. It is part of the ResNet family of neural networks and consists of 18 layers, making it a relatively shallow but powerful model.

The key innovation of ResNet architectures is the introduction of residual connections, also known as skip connections or shortcuts. These connections

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

allow gradients to flow more easily during training and mitigate the vanishing gradient problem, enabling the training of very deep networks.In the project, ResNet18 is used as a base model for training on both noisy and clear labels in the context of tongue color classification in Traditional Chinese Medicine. It serves as a foundation for the E-CA2-ResNet18 algorithm, where additional features like the Attention model for feature optimization and activation functions are incorporated to enhance its performance. ResNet18's capabilities in feature extraction and classification make it a suitable choice for image-based classification tasks, including tongue color classification, while the project's innovations aim to further improve its reliability and robustness

```
resnet_model = ResNet50(input_shape=(X_train.shape[1], X_train.shape[2], X_train.shape[3]), weights='imagenet', include_top=Fals
#for layer in resnet_model.layers:
    #layer.trainable = False
#add teacher and student layers to backbone resnet18 model
resnet_model = Sequential() #create CNN model
#define layer which will act as Teacher model on noisy and clean features from X input data
resnet_model.add(InputLayer(input_shape=(X_train.shape[1], X_train.shape[2], X_train.shape[3])))
#teacher model will use Conv2D layer to filter out all noisy and clean features
resnet_model.add(Conv2D(25, (5, 5), activation='relu', strides=(1, 1), padding='same'))
#teacher module will use max layer to collect all noisy and clean features and then assigned high threshold
resnet_model.add(MaxPool2D(pool_size=(2, 2), padding='same'))
#another Conv2d layer for further filtration
resnet_model.add(Conv2D(50, (5, 5), activation='relu', strides=(2, 2), padding='same'))
resnet_model.add(MaxPool2D(pool_size=(2, 2), padding='same'))
resnet_model.add(BatchNormalization())
#studnet CNN module to get optimized features from teacher module and then use prediction or output layer
#to calculate prediction thrshold and then select predicted label with high probability
resnet_model.add(Conv2D(70, (1, 1), activation='relu', strides=(1, 1), padding='same'))
resnet_model.add(MaxPool2D(pool_size=(1, 1), padding='valid'))
resnet_model.add(BatchNormalization())
resnet_model.add(attention(return_sequences=True,name='attention')) # define Attention layer
resnet_model.add(Flatten())
#output layer
resnet_model.add(Dense(units=100, activation='relu'))
resnet_model.add(Dense(units=100, activation='relu'))
resnet_model.add(Dropout(0.25))
resnet_model.add(Dense(units=y_train.shape[1], activation='softmax'))
```

Fig 4 E-CA2-ResNet18

HYBRID

The "Hybrid" algorithm in the project extends the training of a Random Forest model by collecting clean features from the proposed E-CA2-ResNet18 model, integrating both clean and noisy data. This approach leverages the robustness of Random Forest to handle noisy labels and, by incorporating features extracted from E-CA2-ResNet18, aims to enhance the classification accuracy in tongue color diagnosis. The Hybrid algorithm is suited for this project as it provides a mechanism to combine the strengths of deep learning for feature extraction with Random Forest's ability to handle noisy data, ultimately improving the diagnostic accuracy in Traditional Chinese Medicine tongue color classification.

```
#train extension hybrid Random forest model by collecting clean fetaures from propose E-CA2-ResNet18 model and then
#evaluate performance on test data
#build resnet model
hybrid_model = Model(resnet_model.inputs, resnet_model.layers[-3].output)#create CNN  model
hybrid_features = hybrid_model.predict(X)  #extracting resnet features
Y1 = np.argmax(Y, axis=1)
#split features into train and test
X_train, X_test, y_train, y_test = train_test_split(hybrid_features, Y1, test_size=0.2) #split dataset into train and test
#now create Random Forest object
rf = RandomForestClassifier()
#trained on Resnet18 features
rf.fit(X_train, y_train)
#predict on test data
predict = rf.predict(X_test)
#calculate function to calculate algorithm prediction accuracy
calculateMetrics("Hybrid Extension Model", predict, y_test)
```

Fig 5 HYBRID

4. EXPERIMENTAL RESULTS

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

Accuracy = TP + TN TP + TN + FP + FN.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$\textbf{F1 Score} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)}$$

$$\textbf{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

ML Model	Accuracy	F1-Score	Recall	Precision
Existing VGG 16	0.959	0.954	0.953	0.953
Propose E-CA2-ResNet 18	1.000	1.000	1.000	1.000
Hybrid Extension Model	1.000	1.000	1.000	1.000

Fig 6 PERFORMANCE EVALUATION TABLE

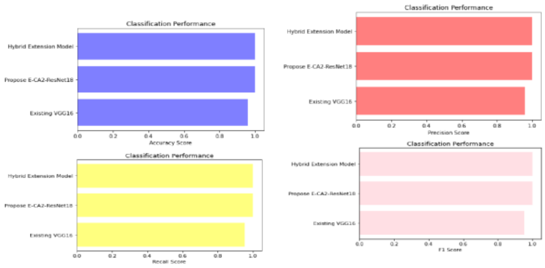

Fig 7 COMPARISION GRAPHS

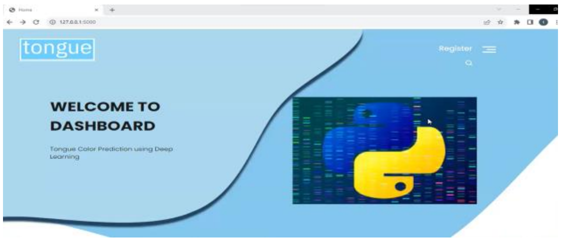

Fig 8 home page

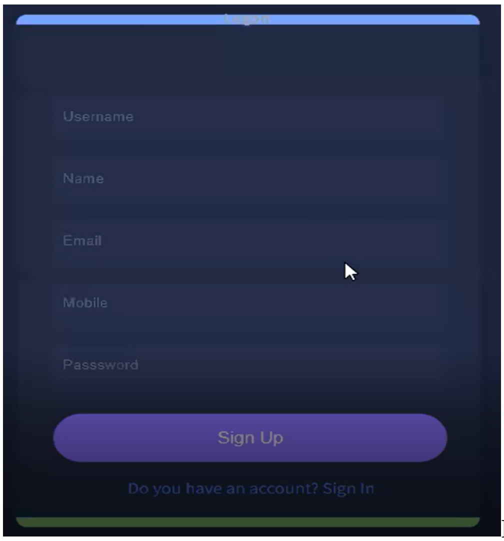

Fig 9 sign up

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

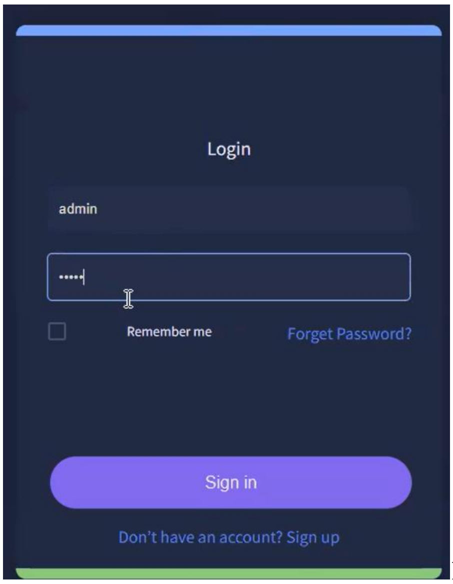

Fig 10 sign in

Fig 11 upload input data

Fig 12 Predict result

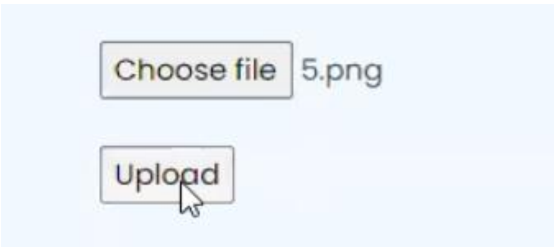

Fig 13 upload input data

Fig 14 Predict result

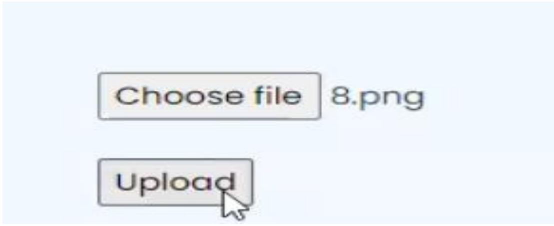

Fig 15 upload input data

Fig 16 Predict result

5. CONCLUSION

The CLA-KD framework, introduced in this study, adeptly addresses the intricate task of traditional Chinese medicine (TCM)[2] tongue color classification despite noisy labels in the dataset. A key component within this framework is the teacher network, which undertakes confident learning, rectifies noisy labels, and imparts valuable knowledge to the student network for precise classification. Enhancing reliability and stability, the ensemble teacher network, E-CA2-ResNet18, integrates the channel attention (CA) mechanism and activate-or-not activation function, optimizing model architecture. Extensive experimentation across three self-established TCM tongue datasets underscores CLA-KD's superior classification accuracy, robustness, and reduced network complexity. Notably, CLA-KD effectively identifies and rectifies dataset inconsistencies, significantly enhancing classification accuracy and affirming its potential in TCM diagnosis applications.

6. FUTURE SCOPE

The CLA-KD framework demonstrates potential beyond its current application by validating adaptability and robustness in various classification scenarios with noisy labels. Its utility can extend to medical diagnostic tasks beyond tongue color classification in traditional Chinese medicine, leveraging its effectiveness in handling noisy labels. Future research avenues include exploring the integration of additional techniques like data augmentation, transfer learning, or ensemble methods to enhance CLA-KD's performance. Evaluation on larger and diverse datasets is essential to assess scalability and real-world applicability. Furthermore, comparative studies with other state-of-the-art methods for noisy label learning are warranted to ascertain CLA-KD's advantages and limitations comprehensively.

REFERENCES

[1] C. H. Wu, T. C. Chen, Y. C. Hsieh, et al., "A hybrid rule mining approach for cardiovascular disease detection in traditional chinese medicine," Journal of Intelligent & Fuzzy Systems, vol.36, no.2, pp.861–870, 2019.

[2] Y. Wang, Y. Liu, L. Yu, et al., " Research methods about data mining technology in the study medication rule on famous veteran Teran doctors of TCM," in Proc. of 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, pp.1948–1952, 2018.

[3] J. Hou, H. Y. Su, B. Yan, et al., " Classification of tongue color based on CNN," in Proc. of 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing,

[4] China, pp.725–729, 2017. L. Chen, B. Wang, Y. Ma, et al., "The retrieval of the medical tongue images based on color analysis," in Proc. of 2016 11th International Conference on Computer Science & Education (ICCSE), Nagoya, Japan, pp.113–117, 2016.

[5] Y. Lu, X. Li, L. Zhuo, et al., " Dccn: A deep-color correction network for traditional Chinese medicine tongue images," in Proc. of 2018 IEEE International Conference on Multimedia & Expo Workshops (ICME Workshops), San Diego, USA, pp.1–6, 2018.

[6] P. L. Qu, H. Zhang, L. Zhuo, et al., "Automatic analysis of tongue substance color and coating color using sparse representation-based classifier," in Proc. of 2016 International Conference on Progress in Informatics and Computing (PIC), Shanghai, China, pp.289–294, 2016.

[7] P. A. Gutiérrez, M. Perez-Ortiz, J. Sanchez-Monedero, et al., " Ordinal regression methods: Survey and experimental study," IEEE Transactions on Knowledge and Data Engineering, vol.28, no.1, pp.127–146, 2015.

[8] J. Sánchez-Monedero, P. A. Gutiérrez, and M. Pérez-Ortiz, "ORCA: A Matlab/Octave toolbox for ordinal regression," Journal of Machine Learning Research, vol.20, no.125, pp.1–5, 2019.

[9] G. Algan and I. Ulusoy, " Image classification with deep learning in the presence of noisy labels: A survey," Knowledge-Based Systems, vol.215, article no.106771, 2021.

[10] C. Northcutt, L. Jiang, and I. Chuang, "Confident learning: Estimating uncertainty in dataset labels," Journal of Artificial Intelligence Research, vol.70, pp.1373–1411, 2021.

[11] T. Xiao, T. Xia, Y. Yang, et al., " Learning from massive noisy labeled data for image classification," in Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, pp.2691–2699, 2015.

[12] A. Veit, N. Alldrin, G. Chechik, et al., "Learning from noisy large-scale datasets with minimal supervision," in Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, pp.6575–6583, 2017.

[13] S. J. Delany, N. Segata, and B. M. Namee, " Profiling instances in noise reduction," Knowledge-Based Systems, vol.31, pp.28–40, 2012.

[14] J. Luengo, S. O. Shim, S. Alshomrani, et al., "CNC-NOS: Class noise cleaning by ensemble filtering and noise scoring," Knowledge-Based Systems , vol.140, pp.27–49, 2018.

[15] Y. Li, J. Yang, Y. Song, et al., "Learning from noisy labels with distillation," in Proc. of IEEE International Conference on Computer Vision (ICCV), Venice, Italy, pp.1928–1936, 2017.

[16] B. Sun, S. Chen, J. Wang, et al., "A robust multi-class AdaBoost algorithm for mislabeled noisy data," KnowledgeBased Systems, vol.102, pp.87–102, 2016.

[17] H. Song, M. Kim, D. Park, et al., "Learning from noisy labels with deep neural networks: A survey," arXiv preprint, arXiv: 2007.08199, 2020.

[18] N. Manwani and P. S. Sastry, " Noise tolerance under risk minimization," IEEE Transactions on Cybernetics , vol.43, no.3, pp.1146–1151, 2013.

[19] A. Ghosh, H. Kumar, and P. S. Sastry, "Robust loss functions under label noise for deep neural networks," in Proc. of the AAAI Conference on Artificial Intelligence (AAAI), AAAI Press, Palo Alto, CA, USA, vol.31, no.1, DOI: 10.1609/aaai.v31i1.10894, 2017.

[20] X. Wang, E. Kodirov, Y. Hua, et al., "Improved mean absolute error for learning meaningful patterns from abnormal training data," arXiv preprint, arXiv: 1903.12141v5, 2019.

[22] D. Angluin and P. Laird, " Learning from noisy examples," Machine Learning, vol.2, no.4, pp.343–370, 1988.

[23] L. P. F. Garcia, J. Lehmann, A. C. P. L. F. de Carvalho, et al., "New label noise injection methods for the evaluation of noise filters," Knowledge-Based Systems , vol.163, pp.693– 704, 2019.

[24] G. Algan and I. Ulusoy."Label noise types and their effects on deep learning," arXiv preprint, arXiv: 2003.10471, 2020.

[25] J. H. Cho and B. Hariharan, "On the efficacy of knowledge distillation," in Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), pp.4794–4802, 2019.

[26] C. Yang, L. Xie, S. Qiao, et al., "Training deep neural networks in generations: A more tolerant teacher educates better students," in Proc. of the AAAI Conference on Artificial Intelligence (AAAI), AAAI Press, Palo Alto, CA, USA, vol.33, no.01, DOI: 10.1609/aaai.v33i01.33015628, 2019.

[27] L. Chen, H. Zhang, J. Xiao, et al., "SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning," in Proc. of Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, pp.6298–6306, 2017.

[28] S. Woo, J. Park, J. Y. Lee, et al., " CBAM: Convolutional block attention module," in Proc. of the European Conf. on Computer Vision, Munich, Germany, pp.3–19, 2018.

[29] N. Ma, X. Zhang, M. Liu, et al., "Activate or not: Learning customized activation," in Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, pp.8032–8042, 2021.

[30] Z. Shen, Z. He, and X. Xue, " MEAL: Multi-model ensemble via adversarial learning," in Proc. of the AAAI Conf. on Artificial Intelligence, AAAI Press, Palo Alto, CA, USA, vol.33, no.1, DOI: 10.1609/aaai.v33i01.33014886, 2019.

[31] L. S. Shen, Y. H. Cai, and X. F. Zhang, Collection and Analysis of Tongue Picture in Traditional Chinese Medicine, Beijing University of Technology Press, Beijing, China, pp.26–30, 2007. (in Chinese)

[32] C. Szegedy, V. Vanhoucke, S. Ioffe, et al., "Rethinking the inception architecture for computer vision," in Proc. of 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp.2818–2826, 2016.

[33] S. Reed, H. Lee, D. Anguelov, et al., "Training deep neural networks on noisy labels with bootstrapping," arXiv preprint, arXiv: 1412.6596, 2014.

[34] Z. Zhang and M. R. Sabuncu, " Generalized cross entropy loss for training deep neural networks with noisy labels," in Proc. of 32nd Conference on Neural Information Processing Systems, Montréal, Canada, pp.8792–8802, 2018.

[35] K. Yi and J. Wu, "Probabilistic end-to-end noise correction for learning with noisy labels," in Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, pp.7017–7025, 2019.

[36] B. Han, Q. Yao, X. Yu, et al., "Co-teaching: Robust training of deep neural networks with extremely noisy labels," arXiv preprint, arXiv: 1804.06872, 2018.

[37] X. Yu, B. Han, J. Yao, et al., "How does disagreement help generalization against label corruption?," in Proc. of Inter-

[38] national Conference on Machine Learning (ICML), Long Beach, CA, USA, pp.7164–7173, 2019