Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

MBE, 18(2): 1169–1186.
DOI: 10.3934/mbe.2021063
Received: 24 November 2020
Accepted: 21 December 2020
Published: 12 January 2021

Research article

Deep sparse transfer learning for remote smart tongue diagnosis

Xu Zhang[1], Wei Huang[2], Jing Gao[1], Dapeng Wang[3], Changchuan Bai[4] and Zhikui Chen[1,*]

[1] The School of Software Technology, Dalian University of Technology, Dalian 116620, China
[2] Department of Critical Care Medicine, First Affiliated Hospital of Dalian Medical University, Dalian 116620, China
[3] First Affiliated Hospital of Dalian Medical University, Dalian 116620, China
[4] Dalian Hospital of Traditional Chinese Medicine, Dalian 116620, China

* **Correspondence:** Email:zkchen@dlut.edu.cn; Tel: +8613478461921.

Abstract: People are exploring new ideas based on artificial intelligent infrastructures for immediate processing, in which the main obstacles of widely-deploying deep methods are the huge volume of neural network and the lack of training data. To meet the high computing and low latency requirements in modeling remote smart tongue diagnosis with edge computing, an efficient and compact deep neural network design is necessary, while overcoming the vast challenge on modeling its intrinsic diagnosis patterns with the lack of clinical data. To address this challenge, a deep transfer learning model is proposed for the effective tongue diagnosis, based on the proposed similar-sparse domain adaptation (SSDA) scheme. Concretely, a transfer strategy of similar data is introduced to efficiently transfer necessary knowledge, overcoming the insufficiency of clinical tongue images. Then, to generate simplified structure, the network is pruned with transferability remained in domain adaptation. Finally, a compact model combined with two sparse networks is designed to match limited edge device. Extensive experiments are conducted on the real clinical dataset. The proposed model can use fewer training data samples and parameters to produce competitive results with less power and memory consumptions, making it possible to widely run smart tongue diagnosis on low-performance infrastructures.

Keywords: deep transfer learning; domain adaptation; pruning; sparse network; remote diagnosis; traditional Chinese medicine

1. Introduction

With the advent of 5G, people put forward higher and more extreme requirements for the energy consumption and delay of terminal devices, and the edge computing will receive more attention.

Meanwhile, life has dramatically altered, accompanied by more attention are paying to personal health, and a trend is that people want to stay indoors and even use smartphones for remote health check. People are so busy that they can hardly take time off, epidemic disease such as the COVID-19 also isolates people, making it difficult for those in need to go in quest of doctors' advices to improve quality of life. Thus, to meet the high computing and low latency requirements of deep learning applications, artificial intelligence based remote diagnosis with edge computing is playing a more important role in smart medicine [1]. It is promising in the computerization of Traditional Chinese Medicine (TCM) theories and approaches, because not limited in disease, sub-health is bedeviling humanity much since the social and economic advancement lead to unhealthy lifestyles such as stressed work, unbalanced diet and less exercise, and TCM achieves the significant and curative effectiveness in recuperating certain sub-health conditions [2, 3].

Tremendous data foundation in big data age makes data-driven deep approaches pragmatic in smart medicine. To make better use of data, deep learning method such as convolutional neural network (CNN) based models are often employed to produce data driven decisions [4]. However, in order to model intrinsic diagnostic patterns in smart diagnosis, very deep models with numerous parameters are often needed, who can hardly deploy to low-performance terminal devices due to the huge scale and computing cost. Nowadays, the world's leading mobile and IoT device manufacturers and researchers are constantly seeking for smart terminals with perceptual ability [5, 6], and integrate algorithms into terminals is a complement and enhancement [7, 8]. The edge computing system is resource-sensitive that a efficient deep network model is necessary.

The smart tongue diagnosis, named as tongue inspection in TCM, is able to be modeled as a typical multi-label classification based on image recognition. It plays a vital role as the fundamental part, owning to its significant function in TCM-based healthcare. In TCM tongue inspection, doctors find lesions on tongue manifestation to identify etiology, different lesions may co-exist and be used for judging symptoms together. By observing the tongue, it assesses the relationships between tongue and organs deriving with the theory governing exterior to infer interior, and using visceral manifestation theory to establish the disease location.

There are mainly two key factors which restricts modeling efficient deep TCM tongue inspection network:

- A: The scarcity of labeled data samples. There is a lack of existing TCM tongue image dataset. And it is necessary to observe the color, shape, fur of tongue and humidity in tongue inspection to obtain the essential information. Thus, the high-performance camera is needed in gathering tongue manifestation images at the clinical diagnosis from scratch. In addition, these images need to be labeled, and only experienced TCM practitioners may be able to give a reliable judgment, which lead to a higher cost.

- B: The redundancy in deep learning model. In order to extract precise features effectively, a large number of parameter need to be calculated in CNN-based models. The models are always computation-intensive and have numerous redundant parameters, which leading to puffiness networks of large volume. Computing power and memory resources of devices are wasted, let alone using in edge computing.

But it is hard to build an appropriate model to solve both of above problems at the same time,

because:

- For factor *A*, domain adaptation in deep transfer learning is a common strategy for insufficient training data samples in the target domain. However, the unique morphological features of the lesions in tongue inspection make it distinct from most frequently-used data sets in domain adaptation such as PASCAL VOC [9] and ImageNet [10], which increase the difficulty of transfer commonalities from existing models in traditional way, since it has been proved that the differences between source domain and target domain in sample distribution will significantly affect the final performance of the model after domain adaptation [11, 12].

- As for *B*, the structure of neural networks is usually simplified to cope with parameter redundancy. Small efficient neural networks using fewer parameters are designed to avoid dense computing, and memory efficient architectures such as SqueezeNet [13] and MobileNet [14] are usually chosen for reference. However, because of the distinction between network structures, it is difficult for domain adaptation to reuse structures and parameters.

To sum up, model should be designed from the aspect of parameter efficiency to avoid the waste of computational resources. At the same time, advantages of domain adaptation should able to be organically integrated in to address the lack of precious TCM data samples, so as to reduce costs in training model. Newly proposed Lottery Ticket (LT) pruning method [15] in deep learning can find representative sparse subnet from original network, who inherits model's performance and is retrainable using fewer parameters. Our idea is: If the subnet can be proved transferable, the method can be used for generating simplified networks fit with domain adaptation. Thus, compact model with sparse neural network is designed while knowledge can be inherited from existing dense model.

In this paper, a *Similar-Sparse Domain Adaptation* (SSDA) method is proposed, illustrated by the case of smart tongue diagnosis. A compact model with two combined sparse deep neural network (DNN) is introduced to learn the TCM diagnosis patterns. In better training the model while overcoming the lack of clinical data, the first *Similar-stage* generates a new source domain by calculating the similarity between it and target domain, for stronger ability to extract features of tongue lesions. After that, the second *Sparse-stage* aims to simplify the network to reduce redundancy. Inspired by the LT pruning method, a transferable strategy in SSDA is proved for pruning transferable subnet in domain adaptation. Finally, simulation experiments are conducted on the real clinical data set in terms of classification accuracy and amount of parameter computations. The proposed SSDA strategy is able to generate DNN-based simple efficient model using fewer training data samples and parameters to converge, thus produce competitive results with less power and memory consumptions, suitable for edge computing. Concretely, it can use at least only 3.5M parameters, which is similar to the efficient architectures suitable for mobile devices such as MobileNet, to get corresponding performance with a dense ResNet.

The rest of this paper is organized as follows: Section 2 discusses related work about domain adaptation and TCM tongue inspection. Section 3 gives a brief introduction of used real data source. Proposed method, including problem definition, model architecture, strategy of the similar-stage and sparse-stage of SSDA are given in the Section 4. Section 5 provides experimental results and discussion. Finally, Section 6 concludes the whole work and gives future discussion.

2. Related work

2.1. Domain adaptation

The domain adaptation, also known as deep transfer learning, is commonly used in reusing existing knowledge for rapid adaptation to reduce the need for training data in new tasks [16]. Due to its characteristics, the strategy is active in various practical applications. In domain adaptation, the source domain D_S is defined to represent a domain with rich information, whose sample distribution might different from it in target domain D_T where the objective task belongs, while D_T represents the domain in which the test sample is located with few labels. Information-rich source domain samples are used to improve the performance of the target domain model through the transfer of knowledge [17]. The transfer is able to include samples and their distribution, features with projection, network architecture and parameters of models, etc., according to which it can be divided into instance-based, mapping-based, model-based and adversarial-based [18–20].

To date, there are some existing methods focus on the optimization of the transfer strategy in domain adaptation, such as Deep Domain Confusion Maximizing (DDC) [21], Deep Adaptation Networks (DAN) [22] and Deep Coral [23]. However, they put more emphasis on further optimizing reliability and accuracy using dense networks. Comparing with these methods, in this paper, we think more of model simplification that using fewer parameters as much as possible on the basis of guaranteed performance, thereby compressing network to achieve the same effect as dense network, to serve edge computing applications.

2.2. Deep learning based TCM diagnosis

TCM has a good mass foundation in China. In recent years, DNN-based methods have made a series of progress in assisting TCM doctors exploring the association between disease and lesions in smart diagnoses [24, 25]. Many TCM theories and approaches such as the pulse diagnoses [26] and the meridian system [27, 28] have been modeled with deep methods, and TCM-based diagnosis such as type-2 diabetes [29] and breast cancer [30] disease detection model are also designed with an improved efficiency and interpretability. These methods effectively inhibit the subjectivity of doctors in artificial diagnosis, while improving the reliability and repeatability, which is important to further improve the scientific quality and clinical significance of TCM. However, they use dense models without considering design simplified and efficient neural networks to provide the extension possibility for edge computing. At the same time, the insufficient and preciousness of real TCM data also often increases the difficulty of model training.

2.3. TCM smart tongue diagnosis

As for the TCM tongue inspection modeling, early researches are limited to BP neural networks or even shallow machine learning to extract features and classify colors or other syndromes [31–34], and deep learning is emerging more recently [35, 36], lesions such as toothprints on the tongue is identified using reliable deep architecture [37, 38], and as to domain adaptation, [39] uses transfer learning framework for cross-domain tongue segmentation. In most cases of above methods, the tongue image is segmented while key areas and edges features are extracted through DNN, then recombined higher-order features are used in judgment. In this paper, inspired of ensemble learning [40], proposed

method designs a joint network structure for the model that a Region Proposal Network (RPN) based object detection network is introduced as auxiliary task which aimed to accelerate convergence and improve the accuracy and generalization ability.

3. Data sources

For the validation of proposed SSDA method in TCM auxiliary diagnosis, the clinical data samples collected from the Traditional Chinese Medicine Association of Dalian, China, which composed of 607 original tongue manifestation images is used for generating training and testing data set. There are 300 images of the standardized format shot by professional camera for TCM inspection, which have unified size of 768×576 pixies, and environment variables such as light intensity, position and proportion of the tongue body in the picture are basically the same. The other 307 photographs are captured in a more general way such as smartphone camera, and have different original sizes and environment variables, which can better simulate and fit the actual diagnosis process to expand the range of identification, increasing the robustness in training model. All the images have been labeled by experienced TCM doctors.

For these tongue manifestation images used in the paper, only 20 samples can be collected at each working day of the doctor. And because random samples are chosen in collecting patients' tongue images, the unbalanced distribution will tend to make categories who have more samples achieve higher accuracy than others in the training. In order to increase the samples, 1) we searched for some open-source tongue manifestation images of typical cases from web and monographs; 2) the image augmentation is used to balance the data set. Since the horizontal flip transformation on tongue manifestation images will not change the relative position of lesions on tongue body, categories with less number are flipped horizontally to balance the count of input images per category for training the model. The input of category that has more images would be randomly selected from the original and flipped data set in the training. Finally, a data set which has 2000 samples resized to 416×416 is constituted. The illustration in Figure 1 gives a overview of the data set.

4. Methods

4.1. Problem definition

According to the analysis above, the process of TCM tongue inspection is able to be modeled as the multi-label classification based on image recognition. As shown in Figure 1, tongue manifestation images used in this paper are labeled in the form of $A\_B\_C\_D$, in which $A \in \{Red, Purple\}$ indicates the color of the tongue body, while $B \in \{White, Yellow\}$, $C \in \{Thin, Thick, Oily\}$ and $D \in \{full, spalling\}$ indicates the color and status of the fur respectively.

Cause it still does not cover the entire features that might need in diagnosis, an object O is attached in addition for recording other features or the *representative lesions*. Meanwhile, the label "*Oily*" in C is removed due to the lack of opposite value, whose corresponding symptoms are also able to cover by the representative lesions.

The *representative lesions* are obvious morphological changes on the tongue body, which are often used to identify specific abnormal tongue manifestation. In this paper, 5 kinds of representative lesions are defined in accordance with the TCM diagnostic theories including "*cracks*", "*toothprints*",

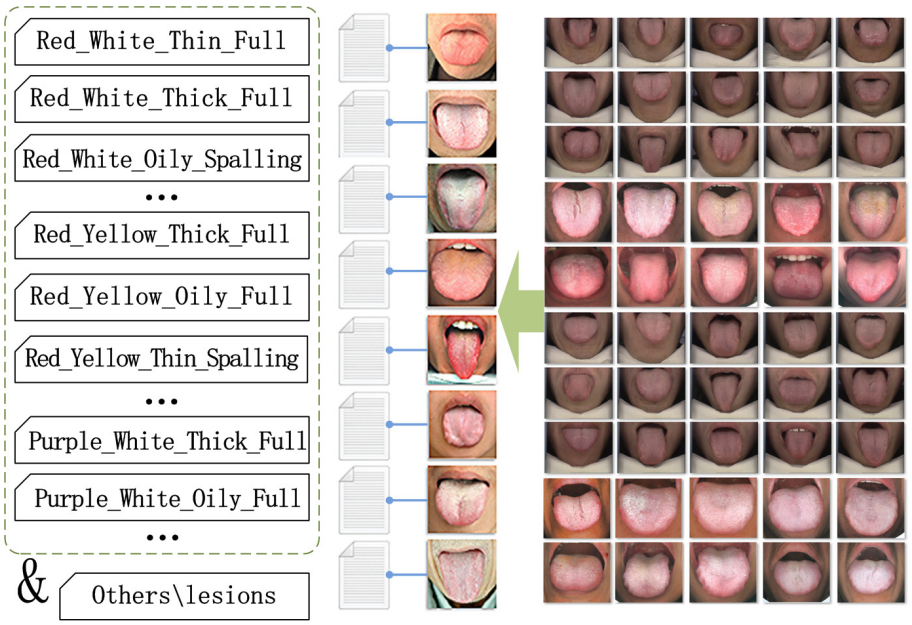

Figure 1. Examples of the labeled data set.

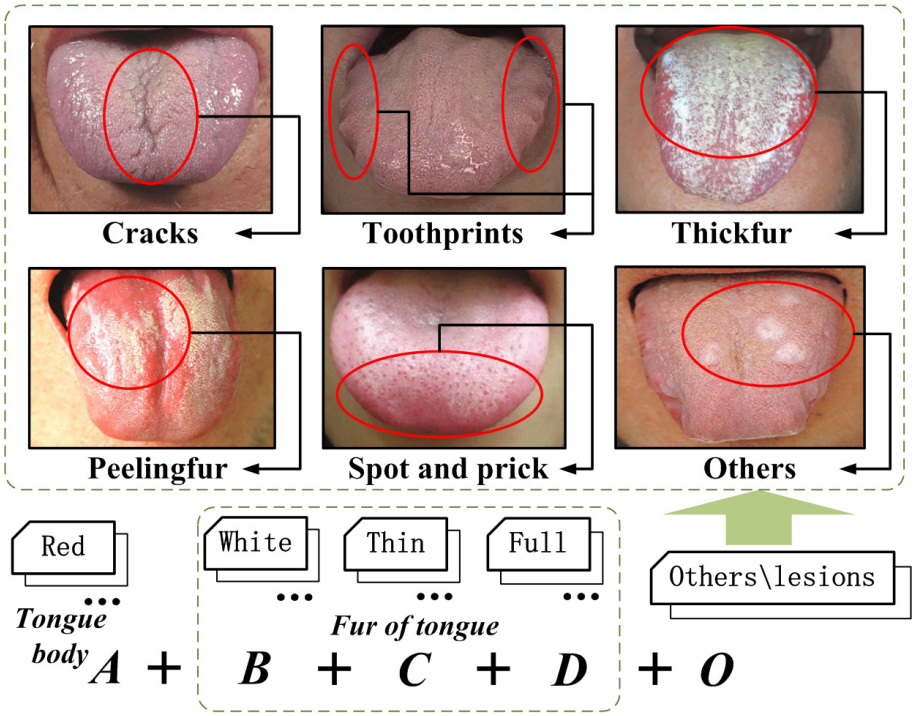

Figure 2. Illustration of labels and representative lesions.

"*thickfur*", "*peelingfur*" and "*spot and prick*". These lesions can be easily found by object detection due to their distinctive feature, and then be used as evidence in judging tongue manifestation, i.e., the final multi-label classification. As illustrated in Figure 2, tongue manifestation images are ultimately labeled in a form of $A\_B\_C\_D + O$. By the combination of labels, the results of final classification can be obtained.

Then, a overall process of the model is shown in Figure 3. In order to transfer required knowledge more effectively, and ensure only the most critical parts will be transferred during domain adaptation: 1) In the similar-stage, multiple similarity measures are used to generate new source domain that approximate the target domain. 2) In the sparse-stage, origin dense network is pruned to generate simplified structure that using in transfer. 3) With the aid of RPN-based target detection, joint architecture is designed to realize the classification of lesions and simulate TCM tongue inspection diagnosis.

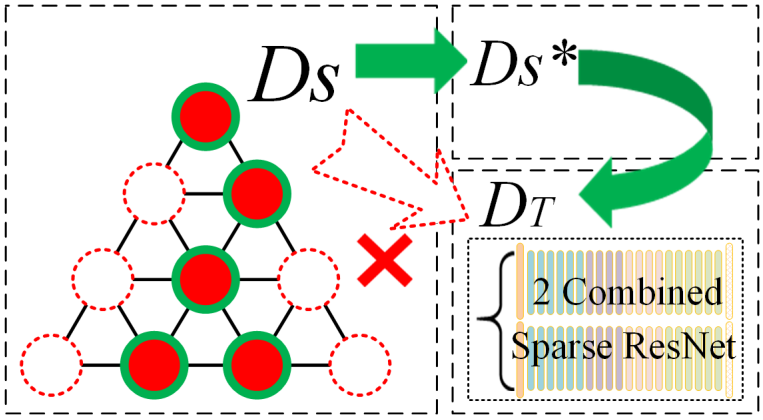

Figure 3. Overall Process of SSDA model.

4.2. Model architecture

The schematic of proposed model's structure is shown in Figure 4. The model jointly uses a combination of two deep CNN to realize the classification based on image recognition, where the ResNet architecture is used as the backbone. The design is based on the idea of ensemble learning [40] that the object detection network is introduced as auxiliary task, which can accelerate convergence, improve the accuracy and generalization ability.

First, collected clinical data set with the defined labels is used to train a ResNet network with 34 convolutional layers (***A*** in Figure 4). Then, for finding extra features and representative lesions in the picture, a Faster R-CNN based Region Proposal Network (RPN) is introduced. As an object detection approach based on CNN, Faster R-CNN extracts feature maps of the input image using a group of basic Conv+Relu+Pooling layers, which are shared for subsequent layers. Therefore, the training of RPN can be built on the trained network with a small computational cost added. In conclusion, the object detection function is added to the first ResNet-34 network to generate the second one (***B*** in Figure 4). Finally, the input images of tongue manifestation will be labeled through the double-networks structure, while the representative lesions or other abnormities are also marked out as the evidences for the final multi-classification, to simulate the judgement of diagnosis.

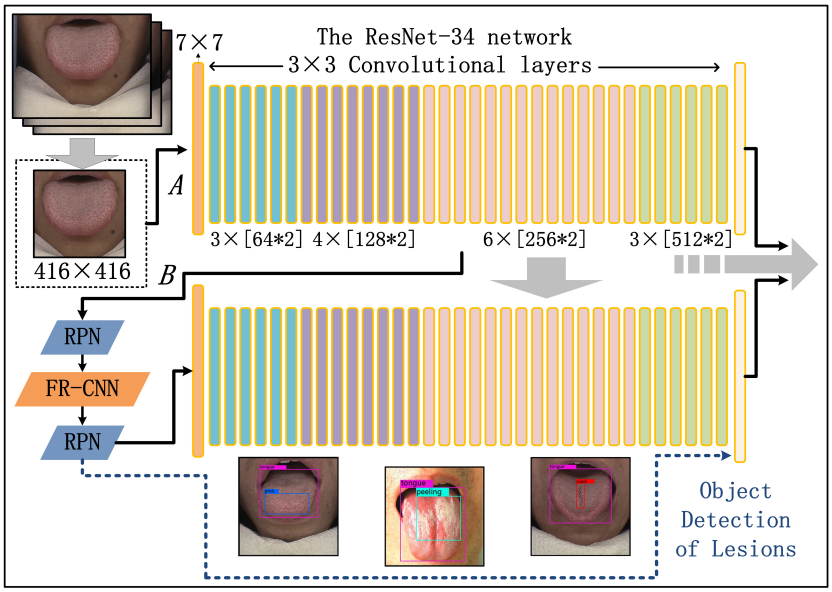

Figure 4. Architecture of TCM tongue inspection model: multi-label classification with object detection.

Thus, the loss function of proposed model is defined as:

$$L = \alpha L_A\{p_n\} + \beta L_B\{p_i, t_i\} \tag{4.1}$$

in which

$$\begin{aligned} L_A &= -\sum_{n=1}^{K} y_n \log(p_n), \\ L_B\{p_i, t_i\} &= \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \end{aligned} \tag{4.2}$$

For K categories and labels y, it uses categorical cross entropy combing with R-CNN loss to balance the ability of classification ($\boldsymbol{A}$) and target recognition ($\boldsymbol{B}$) simultaneously. Concretely, α and β are used to dynamically adjust the weights of two networks. p are predicted probabilities, and t are parameterized coordinates while p^* and t^* are true values. N_{cls} is mini-batch size and N_{reg} is the number of anchor locations that constitute the regression loss together with λ as $\lambda\frac{1}{N_{reg}}$. The value of λ can be used in adjusting proportion of weights between classification loss / regression loss.

4.3. The similar-stage in similar-sparse domain adaptation

It is not enough for the training $\boldsymbol{A}$ from scratch using pure collected data set because of the insufficient clinical data samples. The domain adaptation approach is usually used in this situation, but due to the distinguish of sample distribution between lesions on tongue body and common objects, existing domain adaptation method result in poor performance. Thus, the similar-stage of proposed SSDA method is introduced in this section.

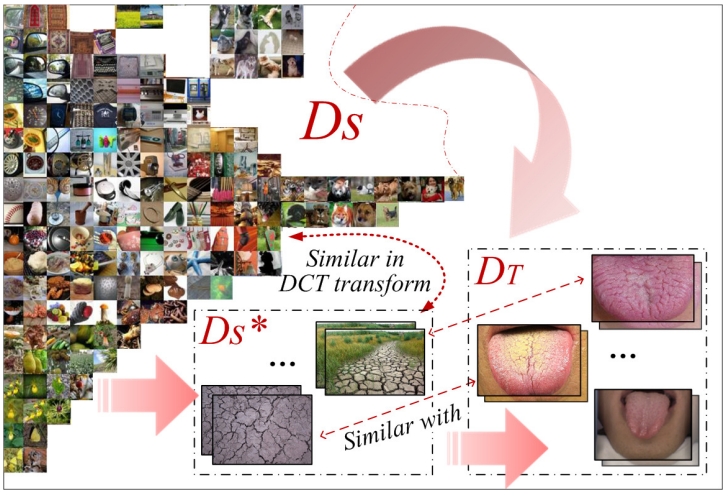

Figure 5. The similar domain adaptation: generate D_S*.

In domain adaptation, a domain can be defined as $D = \{\chi, P(x), x = \{x_1, ..., x_n\} \in \chi\}$, which consists of the feature space χ and the marginal probability distribution $P(x)$. By the transfer of latent knowledge from source domain D_S to adjust the distribution, it can improve the performance of learning task on target domain D_T. The similar-stage of domain adaptation is based on the following idea: some image samples which have look-alike features with the morphological characteristics of lesions can be easier found, and they have a similar effect in training instead of lesions.

As illustrated in Figure 5, for example, the morphological characteristics of the lesion *cracks* are similar to those of the *cracks of the earth*, while the samples of the latter ones are more common. In a similar fashion, a $D_S* \subset D_S(\chi_S* \in \chi_S)$ is generated, whose samples' distribution is tilting toward it in D_T. Then, the model can earn more information for a better performance by replacing D_S with D_S*. It is able to be applied to a variety of tasks that are lack of training samples with choosing D_S* wisely. As for the generation of D_S*, of course the samples can be chosen manually because human beings are good at spotting such similarities without special training. However, for better computerization, a function is designed to measure the similarity of samples between D_S* and D_T, and the top-n samples over the threshold is selected:

$$F(X) = \max\{F_p(X), F_{cos}(X)\} - 0.1F_h(X) \tag{4.3}$$

in which three similarity metrics are considered for evaluation: Pearson correlation coefficient, cosine similarity and hamming distance, that

$$\begin{aligned} F_p(X) &= \frac{(X - \bar{X}) \cdot (X^* - \bar{X^*})}{\|(X - \bar{X}\|_2 \cdot \|(X^* - \bar{X^*}\|_2} \\ F_{cos}(X) &= \frac{X \cdot X^*}{\|X\|_2 \cdot \|X^*\|_2} \\ F_h(X) &= IXI^T - IX^*I^T, I(a,b) = C(a)cos\frac{(b + \frac{1}{2})\pi}{N} \end{aligned} \tag{4.4}$$

The hamming distance calculates the hash value of images and comparing them after the Discrete Cosine Transform (DCT), in which C is the compensation coefficient to make sure that the DCT transformation matrix is orthogonal. The value is the number of different characters in the corresponding positions of two hamming strings, higher value lead to lower similarity, using 0.1 to balance the weight.

4.4. The sparse-stage in similar-sparse domain adaptation

The model should also be designed from the aspect of parameter efficiency, which make it easier to store and run in low-power device systems. So in the sparse-stage, the redundant and low-importance parameters in trained dense networks is pruned to build a simplified model. The stage is between $\boldsymbol{A}$ and $\boldsymbol{B}$, so that when finish training $\boldsymbol{A}$ at similar-stage of SSDA, the resulting network can then transform into a sparse and simplified one. Subsequently, $\boldsymbol{B}$ is also able to train on the sparse network. In this way, the whole model is composed of two sparse ResNet-34, whose total parameters are even far less than a single ResNet-34, but with better performance. The LT pruning method is chosen, because in its original version, trainable representative sparse structure can be found from original network, which provide the structural basis for transfer, whereas other approaches who compress the network by pruning algorithms break the structure and the resulting network cannot be retrained. In this paper, the transferability is proved, so that it can be extended in domain adaptation to discover and handle only the most important parts of the original network.

Reviewing the LT method. In a feed-forward neural network $f(x)$ with initial parameters θi, whose accuracy rate achieves $\alpha\%$ when the network finishes training in j iterations, a sparse subnet defined as $f^*(x, m \odot \theta i)$ can be found by iterative pruning. It satisfied when the subnet is convergence, accuracy rate $\alpha^*\% \geq \alpha\%$, and iteration $j^* \leq j$. The $m \in \{0,1\}^{|\theta|}$ is used as mask function to mark and determine retained weight (set to 1) and pruned weight (set to 0). The subnet can remain performance with less parameters left. In particular, when θ_i from the original dense network is used in fine-tune instead of random initialization, the performance is better.

Define the network on target domain D_T as a task T. When meet conditions, we consider each T with sparse structure as has been pruned from T_S, the dense DNN on D_S* on the basis of transferable LT hypothesis. Vice versa, if only the transferability of T can be proved, transferable LT hypothesis can be extended to generate transferable subnets from corresponding dense network and remains required knowledge from source domain. Since T is a defined generic task, the transferability is easy-proved, and we will verify it in the simulation experiments. The proof of proposed transferable LT hypothesis is shown as follows.

Proof. the transferable LT hypothesis in SSDA

Theorem 4.1. *For a feed-forward neural network $f(x)$,*
with initial parameters θi, max accuracy $\alpha\%$, convergence in j iterations
$\exists$ a sparse subnet defined as $f^*(x, m \odot \theta i)$,
with accuracy rate $\alpha^*\% \geq \alpha\%$, and iteration $j^* \leq j$.

Assumption 4.1. *For the learning task modeled as a DNN, define it as $T = f^*(x, m \odot \theta i)$.*
$\exists$ a large T_S, a corresponding dense network in D_S,it let T be a sparse subnet of T_S.
Considering T_S on D_S*. When it convergence in training, if:

• *Iteration:* $j^* \leq j_S$*;*
• *Accuracy rate:* $\alpha^*\% \geq \alpha_S\%$.
• T *and* T_S *have a same original structure that by pruning* T_S*,* T *can be generated.*
When all of above conditions are met, T *can be regard as the sparse subnet found from* T_S *by LT hypothesis. Vice versa, it can be used in generating transferable subnets from corresponding dense network and remains required ability from source domain.*

Concretely, the steps of the sparse-stage in SSDA are as Algorithm 1 shows. As mentioned above, the sparse structure generated from pruning D_S* based on the Transferable LT Hypothesis is able to preserve cross-domain knowledge in the transfer of domain adaptation.

Algorithm 1 *Steps of the Sparse-stage in SSDA*

Input:
A feed-forward neural network defined as $f(x, \theta i, \alpha, j)$
Whose initial parameters θi, max accuracy $\alpha\%$, convergence in j iterations;
Output:
A sparse neural network $f^*(x, m \odot \theta i, \alpha^*, j^*)$,
With accuracy rate $\alpha^*\% \geq \alpha\%$ in iterations;$j^* \leq j$
And $\mathcal{M}$.
while $\alpha^*\% \leq \alpha\%$ **OR** $j^* \geq j$ **do**
Go through T_S.
An M_i in $\mathcal{M}$ is defined as $\max(0, \frac{\theta_i \theta_n}{|\theta_i|})$, to measure the contribution of each parameter. The θ_n is the final weight of parameters after training.
Sort $\mathcal{M}$ (descending).
Set pruning rate r, the M in top-$r\%$ are set to 1,
while others are 0.
Then use $\mathcal{M}$ to judge whether a parameter be pruned:
for each $M \in \mathcal{M}$ **do**
if $M = 1$ **then**
reset M with θ_i;
else
pruning M.
end if
end for
Train the pruned network for $\alpha^*\%$ and j^*.
end while
Using the sparse net as a new T for the proposed SSDA. In the fine-tune training on D_T, it initialized with θ_i, while the pruned parameters are frozen.

- The D_S can always be a well pre-trained network, or existed mature applications who is modeled as DNN.

- Proposed function $M(\theta_i) = \max(0, \frac{\theta_i \theta_n}{|\theta_i|})$, instead of $M(\theta_i) = |\theta_i|$ in original LT method, is able to avoid when a positive value change into negative one (or vice versa) in the training, the absolute value failed in expressing the trend of change correctly.

- To evaluate an optimal transferable subnet, the relationship between remaining parameters and accuracy needs to be discussed according to actual demand, which will be shown in the simulation experiment.

5. Experiment

In this section, experiments are designed to simulate the method in actual solutions that identify lesions on tongues to help doctors in auxiliary diagnosis. The small-scale clinical data set that we collected in Dalian, China is used in training model and validating results. The ResNet-34 used in building the model is pre-trained on ImageNet dataset in advance to acquire the ability of image recognition in computer vision. At the similar-stage, the model is trained using the SGD algorithm with learning rate 0.1, batch size 128, momentum 0.9 and weight decay 0.0005. In sparse-stage, 10 rounds of iterative pruning are performed, and the pruning rate in each epoch is set to 20% with batch size 128, learning rate 0.01, momentum 0.9 and weight decay 0.0001. When training the model, a server which contains 2 Inter Xeon Silver 4110 8-cores CPUs and 2 NVIDIA Tesla M60 128G GPUs is used. The validation process can also run on a ordinary office computer, since a simplified model with sparse structure is trained by the proposed SSDA method.

Figure 6. The accuracy of defined labels' multi-classification: the confusion matrix.

5.1. Multi-classification: under the help of object detection

As shown in Figure 6(a), after the similar-stage training, the model succeeds in identifying the attributes of color and fur of tongues using defined labels. For the color of tongue body and tongue fur, the accuracy of classification results achieves over 98%, and it in identifying thickness and distribution of fur on the tongue is able to exceed 80%. At the same time, under the help of object

detection approach, the model can find out representative lesions while earning higher accuracy than image-recognition-based method in classification, because these lesions are often small, whose overall-features are not obvious but local-features are obvious. When lesions are detected in the image whose confidence exceeds the threshold, they are added to the labels of the image as a defined O, and then compared with real labels. As Figure 6(b) shows, the overall identification accuracy is acceptable, reaching 88.95% on average and up to 95.8%. According to the confusion matrix, some of the objects are missed in detection, but at a lower rate comparing with the false positives. By combining these detected labels, multi-classification results of input pictures can be obtained.

5.2. Relationship between accuracy / parameters in sparse-stage

In this section, to validate the effectiveness of the sparse-stage and the transferable LT hypothesis in SSDA, experiments are designed. The relationship between the number of used parameters and model performance is mainly investigated.

Based on the experiment results above, the average accuracy in final multi-label classification using the combination of two dense networks achieves 92.34%. Due to the training steps at sparse-stage, after the sparse network is generated, the following RPN training is also able to run on the simplified structure that reducing the amount of computation. Thus, the total parameters of proposed model are approximately twice over the sparse ResNet-34 after pruning. We compare it with training a model using two dense ResNet-34, which will not be pruned when training $\boldsymbol{A}$, and then training $\boldsymbol{B}$ on this basis.

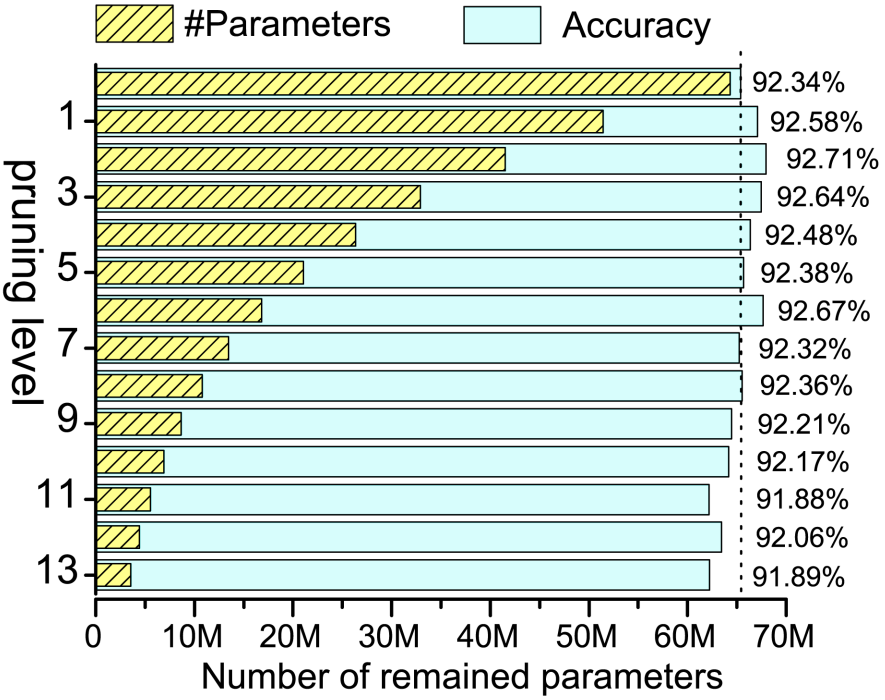

Figure 7. Experimental results: remained parameters with accuracy of different pruning levels.

In each dense ResNet-34, there are over 3.2×10^9 parameters, whose convolution layers can be pruned to reduce the time complexity of computation. In each round of iterative pruning, 20% of the current parameters is pruned, that 10.74% of the original parameters remains after 10 rounds, and 5.5% after 13 rounds. When re-training the sparse network, the weights of parameters are frozen, and only the fully connected layer is fine-tuned. The relationship between remained parameters and model accuracy is as shown in Figure 7.

As the result shows, it obtains a better performance using fewer parameters, because the sparse structure can effectively inhibit redundant and negative-feedback parameters by proper pruning. Comparing with the corresponding dense net, proposed method can achieve similar and sometime better performance, but save lots of parameters costs. The highest accuracy is at 92.71% after the second pruning, when 64% of the parameters in original dense network are remained. And the model can use 26% of the parameters to achieve an accuracy of 92.67%.

And deeper-level pruning can be done depend on actual demands to get smaller models. When only 10.74% of the parameters remains (about 6.9M comparing with original over 64M), the model is able to run on a passable accuracy of 92.17%, and even run on 5.5% of the parameters while achieving 91.89%. That is, in the sparse-stage, 90% of parameters computation cost can be cut down, or 70% for a better performance. It proves that necessary ability can be transferred in a sparse structure. So when the task can accept some performance loss in exchange for generalization ability, the sparse-stage offers the possibility for running deep computing methods on devices with low computational power.

5.3. Other discussion

5.3.1. The effectiveness of transfer in similar-stage

As shown in Table 1, when being trained directly without domain adaptation, model is unavailable due to the lack of clinical data and the difficulty in recognizing unique features of tongue lesions. And when a traditional transfer strategy in domain adaptation is used (by reusing knowledge from ImageNet without generating D_S*), an average accuracy of 80% is achieved, which is significantly lower than proposed method.

Table 1. Performance comparisons of related methods.

Method	Acc	#Parameters
SSDA	92.17(prunelevel-10)	6.9×10^6
	92.67(prunelevel-6)	1.67×10^7
	91.89(prunelevel-13)	3.5×10^6
Dense Model	92.34	6.4×10^7
Without Transfer	<50(unapplicable)	6.4×10^7
Without Similar-stage	80.6%	6.4×10^7
Without RPN	85.71(avg)	3.2×10^7
Tooth-print recognition [37]	91.7(VGG-16)	1.38×10^8
	94.2(ResNet-50	4.62×10^7
CNN-based Tongue identification [38]	Tongue-print only	1.44×10^8

5.3.2. The effectiveness of auxiliary task

Table 1 shows the RPN-net succeed in improving the performance of the model. However, the α and β in loss function can dynamically adjust the contribution weights of the two networks, and we trained the model using the condition that makes the average recognition accuracy best in the experiments. But in fact, in general, when β is someway inhibited, the *toothprints* and *spot and prick* can be better recognized (over 95%). Therefore, in practical application, if the recognition of a certain type of lesions is emphasized, the weight of auxiliary task can be appropriately constrained for better performance.

5.3.3. Comparing with related methods

Although double ResNet-34 is jointly-used in designed model, proposed SSDA can use 20% of the parameters amount to a single dense ResNet-34, but get much better performance, and even 10% with acceptable performance loss. The amount of parameters (6.9M and 3.5M) is similar to MobileNet [14] (4.2M), which is efficient architectures suitable for mobile devices.

It is hard to make a direct transverse comparison because of the different definitions of problems and tongue lesions. The [37] and [38] solved similar problem but only identify toothprints of tongue, while we identified more lesions simultaneously, and used much fewer parameters.

6. Conclusion

In this paper, a Similar-Sparse Domain Adaptation (SSDA) method is proposed for modeling the smart tongue diagnosis of TCM with the limited tongue manifestation images collected from clinical diagnoses. Specifically, a compact deep learning model with two combined 34-layers ResNet and a Faster R-CNN based RPN network is introduced to detect lesions and learn the TCM diagnosis patterns. To train proposed model while overcoming the lack of clinical tongue images, the similar-stage is designed to achieve transfer learning. And the transferable LT hypothesis is proposed in sparse-stage to reduce the parameters computation while keeping the overall accuracy, which generate simplified model structure. So the model consumes less resource to produce competitive results, which is conducive to widely run smart tongue diagnosis on low-performance infrastructures.

Also, the model training with SSDA strategy can be widely used in diagnosis of other diseases which based on image recognition and lack of data, and even further, other recognition tasks whose volume of network need to be compressed such as to accommodate edge computing. Since the representative characteristics of samples in other tasks possibly will be different from those in tongue inspection, it supposed to wisely generate reasonable source domains. For the simplification of model, the suitable network architecture and the balance between accuracy and model volume is also need be sought depend on particular task.

In the future, the SSDA strategy of generating simplified deep models will be studies on more other tasks to overcome the scarcity of data and the redundancy of model parameters, which improving the scope of using deep learning methods.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 62076047), the Fundamental Research Funds for the Central Universities (No. DUT20LAB136 and No.

DUT20TD107), and Dalian Science and Technology Innovation Fund Project (No. 2020JJ26SN049).

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. Z. Ning, P. Dong, X. Wang, X. Hu, L. Guo, B. Hu, et al., Mobile edge computing enabled 5G health monitoring for internet of medical things: A decentralized game theoretic approach, *IEEE J. Sel. Areas Commun.*, (2020), 1–16.
2. D. C. Mainenti, Big data and traditional chinese medicine (TCM): What's state of the art?, in *2019 IEEE International Conference on Big Data (Big Data)*, 2019, 1417–1422.
3. Q. Jiang, X. Yang, X. Sun, An aided diagnosis model of sub-health based on rough set and fuzzy mathematics: A case of TCM, *J. Intell. Fuzzy Syst.*, **32** (2017), 4135–4143.
4. F. Cui, Deployment and integration of smart sensors with iot devices detecting fire disasters in huge forest environment, *Comput. Commun.*, **150** (2020), 818–827.
5. P. H. O. Santos, G. L. Soares, T. M. Machado-Coelho, B. A. G. de Oliveira, P. Y. Ekel, F. M. F. Ferreira, et al., Multi-objective genetic algorithm implemented on a STM32F microcontroller, in *2018 IEEE Congress on Evolutionary Computation (CEC)*, 2018, 1–7.
6. Z. Ning, P. Dong, X. Wang, X. Hu, J. Liu, L. Guo, et al., Partial computation offloading and adaptive task scheduling for 5G-enabled vehicular networks, *IEEE Trans. Mob. Comput.*, (2020).
7. Y. Huang, X. Ma, X. Fan, J. Liu, W. Gong, When deep learning meets edge computing, in *2017 IEEE 25th international conference on network protocols (ICNP)*, 2017, 1–2.
8. Z. Ning, K. Zhang, X. Wang, L. Guo, X. Hu, J. Huang, et al., Intelligent edge computing in internet of vehicles: a joint computation offloading and caching solution, *IEEE Trans. Intell. Transp. Syst.*, (2020), 1–14.
9. S. Vicente, J. Carreira, L. Agapito, J. Batista, Reconstructing pascal voc, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2014), 41–48.
10. Y. Zhang, B. D. Davison, Impact of imagenet model selection on domain adaptation, in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision Workshops*, (2020), 173–182.
11. S. M. Xie, N. Jean, M. Burke, D. B. Lobell, S. Ermon, Transfer learning from deep features for remote sensing and poverty mapping, in *Proceedings of the AAAI Conference on Artificial Intelligence*, **30** (2016).
12. R. Xu, G. Li, J. Yang, L. Lin, Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation, in *Proceedings of the IEEE International Conference on Computer Vision*, (2019), 1426–1435.
13. F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, K. Keutzer, SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size, preprint, `arXiv:1602.07360`.

14. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, et al., Mobilenets: Efficient convolutional neural networks for mobile vision applications, preprint, arXiv:1704.04861.
15. J. Frankle, M. Carbin, The lottery ticket hypothesis: Finding sparse, trainable neural networks, preprint, arXiv:1803.03635.
16. G. Csurka, A comprehensive survey on domain adaptation for visual applications, in *Domain adaptation in computer vision applications*, Springer, Cham, (2017), 1–35.
17. M. Wang, W. Deng, Deep visual domain adaptation: A survey, *Neurocomputing*, **312** (2018), 135–153.
18. X. Huang, Y. Rao, H. Xie, T. Wong, F. L. Wang, Cross-domain sentiment classification via topic-related TrAdaBoost, in *Proceedings of the AAAI Conference on Artificial Intelligence*, **31** (2017).
19. C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, C. Liu, A survey on deep transfer learning, in *International conference on artificial neural networks*, Springer, Cham, (2018), 270–279.
20. H. Chang, J. Han, C. Zhong, A. Snijders, J. Mao, Unsupervised transfer learning via multi-scale convolutional sparse coding for biomedical applications, *IEEE Trans. Pattern Anal. Mech. Intell.*, **40** (2017), 1182–1194.
21. E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, T. Darrell, Deep domain confusion: Maximizing for domain invariance, preprint, arXiv:1412.3474.
22. M. Long, Y. Cao, J. Wang, M. I. Jordan, Learning transferable features with deep adaptation networks, in *International conference on machine learning*, PMLR, (2015), 97–105.
23. B. Sun, K. Saenko, Deep coral: Correlation alignment for deep domain adaptation, in *European conference on computer vision*, Springer, Cham, (2016), 443–450.
24. L. Zhang, X. Li, J. Lai, L. Zhang, Bioinformatics databases for network pharmacology research of traditional chinese medicine: A systematic review, in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, (2017), 1400–1404.
25. Y. Ye, B. Xu, L. Ma, J. Zhu, H. Shi, X. Cai, Research on treatment and medication rule of insomnia treated by TCM based on data mining, in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, (2019), 2503–2508.
26. K. Tago, H. Wang, Q. Jin, Classification of TCM pulse diagnoses based on pulse and periodic features from personal health data, in *2019 IEEE Global Communications Conference (GLOBECOM)*, IEEE, (2019), 1–6.
27. J. Wei, J. Wang, Y. Zhu, J. Sun, H. Xu, M. Li, Traditional chinese medicine pharmacovigilance in signal detection: decision tree-based data classification, *BMC Med. Inf. Decis. Making*, **18** (2018), 19.
28. C. Wu, T. Chen, Y. Hsieh, H. Tsao, A hybrid rule mining approach for cardiovascular disease detection in traditional chinese medicine, *J. Intell. Fuzzy Syst.*, **36** (2019), 861–870.
29. Y. Li, H. Ye, An analysis and research of type-2 diabetes TCM records based on text mining, in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, (2018), 1872–1875.

30. J. Yang, Y. Wen, G. Zhao, J. Duan, Research on association rules of breast cancer and TCM : Syndrome based on data mining, in *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, (2019), 2788–2792.
31. X. Li, Q. Shao, J. Wang, Classification of tongue coating using gabor and tamura features on unbalanced data set, in *2013 IEEE International Conference on Bioinformatics and Biomedicine*, (2013), 108–109.
32. J. Ding, G. Cao, D. Meng, Classification of tongue images based on doublet SVM, in *2016 International Symposium on System and Software Reliability (ISSSR)*, (2016), 77–81.
33. Z. Qi, L. P. Tu, J. B. Chen, X. J. Hu, J. T. Xu, Z. F. Zhang, The classification of tongue colors with standardized acquisition and icc profile correction in traditional Chinese medicine, *BioMed Res. Int.*, **2016** (2016).
34. T. C. Lee, L. C. Lo, F. C. Wu, Traditional chinese medicine for metabolic syndrome via TCM pattern differentiation: tongue diagnosis for predictor, *Evidence-Based Complementary Altern. Med.*, **2016** (2016).
35. W. Liu, C. Zhou, Z. Li, Z. Hu, Patch-driven tongue image segmentation using sparse representation, *IEEE Access*, **8** (2020), 41372–41383.
36. W. Jiao, X. Hu, L. Tu, C. Zhou, Z. Qi, Z. Luo, et al., Tongue color clustering and visual application based on 2D information, *Int. J. Comput. Assist. Radiol. Surg.*, **15** (2020), 203–212.
37. W. Tang, Y. Gao, L. Liu, T. Xia, L. He, S. Zhang, et al., An automatic recognition of tooth-marked tongue based on tongue region detection and tongue landmark detection via deep learning, *IEEE Access*, **8** (2020), 153470–153478.
38. S. Sadasivan, T. T. Sivakumar, A. P. Joseph, G. C. Zacharias, M. S. Nair, Tongue print identification using deep CNN for forensic analysis, *J. Intell. Fuzzy Syst.*, **38** (2020), 6415–6422.
39. L. Li, Z. Luo, M. Zhang, Y. Cai, C. Li, S. Li, An iterative transfer learning framework for cross-domain tongue segmentation, *Concurr. Comput. Pract.*, **32** (2020).
40. H. Yang, J. Zhang, H. Dong, N. Inkawhich, A. Gardner, A. Touchet, et al., DVERGE: diversifying vulnerabilities for enhanced robust generation of ensembles, preprint, `arXiv:2009.14720`.

AIMS Press