Tongue Colour Diagnosis System Using Convolutional Neural Network

Yi Chen Tan, M. Shahrizal Rusli, Nur Diyana Kamarudin,
Ab Al-Hadi Ab Rahman, Usman Ullah Sheikh,
Michael Tan Loong Peng and Mohd Ibrahim Shapiai

October 20, 2021

Tongue Colour Diagnosis System Using Convolutional Neural Network

Yi Chen Tan[1*], Mohd Shahrizal Rusli[1*], Nur Diyana Kamarudin[2], Ab Al-Hadi Ab Rahman[1], Usman Ullah Sheikh[1], Michael Tan Loong Peng[1], Md. Ibrahim Shapiai[3]

[1]*School of Electrical Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor, email: *tanyichen@gmail.com, *shahrizalr@utm.my*

[2]*Department of System Safety and Communication, National Defence University of Malaysia, Kem Sg Besi, 57000 Kuala Lumpur.*

[3]*Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Jalan Sulatan Yahya Petra, 54100 Kuala Lumpur.*

Abstract. Tongue diagnosis is known as one of the effective and yet noninvasive techniques to evaluate patient's health condition in traditional oriental medicine such as traditional Chinese medicine and traditional Korean medicine. However, due to ambiguity, practitioners may have different interpretation on the tongue colour, body shape and texture. Thus, research of automatic tongue diagnosis system is needed for aiding practitioners in recognizing the features for tongue diagnosis. In this paper, a tongue diagnosis system based on Convolution Neural Network (CNN) for classifying tongue colours is proposed. The system extracts all relevant information (i.e., features) from three-dimensional digital tongue image and classifies the image into one of the colours (i.e. red or pink). Several pre-processing and data augmentation methods have been carried out and evaluated, which include salt-and-pepper noises, rotations and flips. The proposed system achieves accuracy of up to 88.98% from 5-fold cross validation. Compared to the reported baseline Support Vector Machine (SVM) method, the proposed method using CNN results in roughly 30% improvement in recognition accuracy.

1. Introduction

In health care field, there are many ways to evaluate a patient's health condition. Tongue diagnosis is known to be one of the effective and yet noninvasive techniques to evaluate patient's health condition in traditional oriental medicine such as traditional Chinese medicine[1] and traditional Korean medicine[2].

Tongue diagnosis is performed based on the features on the tongue such as tongue body, colour, coating and fur. From the perspective of colour, tongue body colour can be pale, light red, red and deep red. The colour of a healthy person is light-red colour. Pale tongue colour indicates the sign blood vacuity, which is depletion of the blood in Traditional Chinese Medicine (TCM). Red and deep red tongue colour indicates the repletion heat.

The practitioners of oriental medicine inspect the tongue features through visual inspection in order to perform tongue diagnosis. However, the predictions are made based on each practitioner's experience and observation and the prediction may varies. Due to ambiguity, there is a need for automatic digital tongue diagnosis system (TDS) in order to assist the medical practitioner in daily practices.

Like other image classification tasks, TDS requires feature extraction and classification stages as well. In these works [3, 4, 5], both stages are implemented separately. On the other hand, research in TDS is affected by the environmental factors such as brightness [6]. Lastly, there is limited access to tongue image databases for tongue diagnosis system and thus no standard data across researches.

There are many methods implemented for TDS such as active contour based method, threshold tongue segmentation method, and machine learning. Deep learning especially Convolutional Neural Network (CNN) is observed to have a lot of breakthrough recently due to the advancement of computation and its ability of feature learning and representation. Both feature extraction and classification can be done by using CNN. While processing time is not taken accounted on in many works [7, 3, 8], it is found that CNN method takes longer computation time compared with other methods. This technique had been applied in other image processing applications such as facial expression recognition system [9][10] and face recognition [11][12] as well.

The main objective of this paper is to develop a high accuracy TDS based on CNN. This work trains and tests the system using data obtained from these works [5] [13] which consists of 257 labelled tongue images. There are three classes of tongue images in the database (i.e. deep red, red and pink).

2. Related Works

Hsu et al.[7] proposed many colour space related method for many tongue features. The method analyzes the colour contour extracted from the tongue image and classify the feature based on the individual. For example, using HSI colour space, the tongue image is defined to have thick fur if the hue value is below and thin fur otherwise. For tongue colour, this work categorized tongue colour into seven classes based on the R colour spectrum, but the exact value is not defined in the work. The author reported that the method's accuracy is affected by brightness. The reason of brightness invariant in this method is caused by the user whose judgment affected by brightness condition when setting the boundary. Watsuji et al.[14] proposed a tongue inspection supportive system by using fuzzy theory. It consists of members in a set which each member has a degree of grade in membership [15]. It consists of member which each membership has a degree. The system proposed by the author is not fully automated as feature must be inspected by the user manually. The system could classify the disease based on the feature score 10 given by the user. The drawback of these rule-based methods is the requirement of user to define the boundary of differentiating between classes. The performance of such system depends on how well the rule are defined. The work is tedious and time consuming in development to get all required rules.

Gao et al. proposed a seven-class SVM classifier for TDS [3]. The work used 768 tongue images which consists of 665 patients having one of the six common internal diseases and 103 healthy volunteers from Nanjing Traditional Chinese Medicine Hospital and Nanjing University of Traditional Chinese Medicine. The method proposed use "one per class" strategy. This multi-class SVM classifier are consists of multiple classifier which each classifier responsible

for classification of one class. Seven SVMs are trained by using 70 images of the related class and 70 images of other non-related classes by using Gaussian kernel, a kernel for non-linear separation. The colour features are extracted by using chromatic extraction and texture features are extracted by using gray level co-occurrence matrices (GLCM) [16]. The parameters used are penalization coefficient, C=25, and the kernel parameter, σ=0.1. The work compared the proposed method with Bayesian Network. The evaluation result shows that SVM has higher accuracy than Bayesian Network in classification of every disease. This work proven that multi-class SVM can achieve high accuracy in multi-class classification scenario. However, to achieve multi-class classification, multiple models are needed.

On the other hand, Kamarudin et al. proposed a SVM based colour classification TDS [5]. The work uses binary SVM to classify tongue image into deep red or light-red/red. The input images are pre-processed using k-means clustering which cluster a tongue image into four cluster of image: background (black), deep red region, red/light region and transition region. The features are then classified by SVM using different kernel.

Lin et al.[17] proposed a tongue image segmentation using a modified CNN architecture known as residual nets (RESNET)[18]. The connection is to allow each layer to fit a residual mapping explicitly instead of them fitting a desired underlying mapping. This could resolve a degradation problem that arises when the neural networks is getting deeper: the accuracy saturated and then degrades rapidly before over-fitting. The problem causes higher training error when more layers are added to the model. The residual mapping is proven to be able to achieve better generalization performance on recognition tasks as it is easier to optimize the residual mapping [18]. The work shows that increasing complexity of CNN model not necessary increases the accuracy. Hou et al. proposed a CNN based method for TDS [8]. CNN is used for both feature extraction and classification with one single step or model. The work modified the original CaffeNet [19] model architecture, which has multiple layers that progressively compute features from input images [20], by adding Batch Normalization between Convolutional layers. This work also utilizes image pre-processing method, that is normalization on the input image. with pre-processing (normalization), the accuracy is increased and with modification of model, the accuracy is also increased. By combining both normalization and model modification, the accuracy is further increased.

Other recent works that utilize CNN for tongue color classification include the work in [21], where a pre-processed and enhanced image were created and used as datasets, followed by parameter tuning to obtain adequate performance. The work in [22] is also similar, i.e. with the goal of tongue color specification, but focuses on using semantic based CNN and center loss function. These works differ from our work, whereby we have proposed other pre-processing and data augmentation methods to further enhance the robustness of the classification.

SVM and CNN are both suitable methods for TDS as the boundary condition to separate between classes that are learned by the machine. They are both accurate, but CNN is more preferred as it is proven to be robust to brightness, image size and tongue position after it achieved generalization. Data augmentation methods can be adaptively performed to increase the number of samples and improve recognition accuracy [23, 24, 25]. CNN is said to be more sensitive to brightness condition. In addition, CNN is more flexible in modification in term of classes. For example, to modify the CNN model to classify one more class, the modification can be simply done by adding an extra probability calculation neuron. To expand the number of class in SVM method will need another SVM model dedicated for this added class. These properties are important for developing a classification system that is growing in interest like tongue diagnosis.

3. Methodology

This paper proposes a TDS based on CNN. The system recognizes two colour (i.e. pink and red). The objective is to compare with baseline SVM-based method [5] in terms of accuracy and processing time with a low complexity CNN model. The reason for developing low complexity CNN based TDS is to overcome the high processing time and to ease the hardware implementation of the system. The system receives RGB tongue images as input and produces output label.

As shown in the result of related work [20], with pre-processing, the performance of machine learning model could be improved. Thus, pre-processing methods such as image cropping by ratio, resizing and white-coating removal are carried out. The effect of these methods on the developed TDS is investigated.

The database used in this is a small database with 257 tongue images from 3 classes, that are red, deep red and pink with different level of white coating. Thus, data augmentation is expected to be carried out for enlarging the size of the database. This work also investigates the effect data augmentation methods on the developed TDS.

3.1 Image Cropping Ratio

The original image contains information such as tongue parts and background. However, only the tongue information is useful for tongue colour recognition procedures. Information on background is not important and act as noises which reduce the accuracy of the recognition. In order to solve the problem, cropping is needed to eliminate the noises.

Tongue area is not detected automatically. The original image is cropped by a user-defined ratio. The region of interest is obtained by removing parts of the regions based on the width of the original region, w. For simplification of this process, all images are cropped with the same ratio. Through trial and error, as well as manual inspection, the ratio is determined.

The regions within 0.13w from top, 0.02w from bottom, 0.1w from left and 0.1w from right sides are removed. The remaining region contains lesser background and more important related features. Figure 1 shows an example of the image cropped by ratio: (a) original image, (b) cropped by ratio image.

Figure 1. Example of pre-processing methods. (a) Original image, (b) Crop by ratio, and (c) White coating removal.

3.2 Image Resizing

The purpose of resizing the image is important to ensure all input images have the same size. This is an essential step of CNN applications as CNN takes fixed size of input. In this project, the images are resized to various size as size of input is one of the parameters to be tuned.

By minimizing the size of images, the number of computations of convolution process is reduced. Thus, speed of recognition and training is increased. In this work, down-sampling is performed by using cv2.resize function from OpenCV library to reduce the size to various size as size of input image is one of the parameters that to be tuned in optimization flow. The down-sampling uses INTER_AREA interpolation method which re-samples image using pixel area relation.

3.3 Salt-and-pepper Noise

The data was augmented by adding noises. In this project, salt-and-pepper noise which is a form of noise with sparsely occurring white and black pixels is used as a way to augment the database. The noise in white colour (pixel value of 255 at every RGB channel) is known as salt while noise in black (pixel value of 0 at every RGB channel) is known as pepper.

The procedure of adding salt-and-pepper noises is as follows: A threshold value is defined. A random float number from 0 to 1 is generated and if the randomized number is smaller than the threshold value, the pixel is replaced with salt colour; if the randomized number is bigger than 1 - threshold, the pixel is replaced by pepper. The probability if generating salt and pepper are equally the same. The threshold of generating the noises is kept low because too many noises destroy features.

From manual inspection, threshold value is set as 0.005 is chosen. Three additional images are generated from an original image using this configuration. Figure 2 shows an example of image with salt-and-pepper noises added: (a) original image, (b) image with salt-and-pepper noises.

3.4 Flips

Using the concept of flipping, three additional images can be generated. They are the mirrored and synthetic one-sided flips. A mirrored image is generated by horizontally flipping the original tongue image. Synthetic one-sided flips are augmentation techniques that are newly proposed in this work in order to synthetic new images from original image. By using this method, left symmetry and right symmetry can be generated. The procedure starts with cropping the left half and the right half of the tongue image. To generate left symmetry of the tongue, the tongue's left half is duplicated, the duplicate is horizontally flipped, and the flipped duplicate is concatenated to the right of the left half. Similarly, the right symmetry is generated by using the right half of the tongue. An example of images generated by flips are shown in Figure 2 (d) Horizontal flip, (e) One-sided flip (right symmetry), (f) One-sided flip (left symmetry).

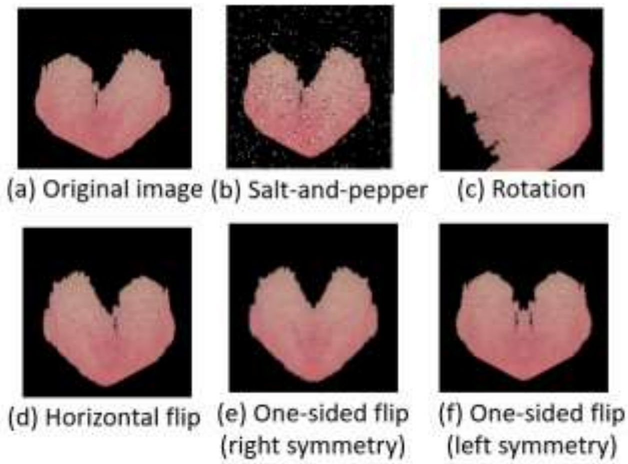

Figure 2. Example of data augmentation methods. (a) Original image, (b) Salt-and-pepper noise, and (c) Rotation, (d) Horizontal flip, (e) One-sided flip right, (f) One-sided flip left.

3.5 Proposed Convolutional Neural Network Model

The architecture consists of six layers, which are two convolutional layers, two sub-sampling layers and two fully connected layers as shown in Figure 3. It is inspired by LeNet-5 architecture [26] and CNN model by Lopes *at el.* [9]. The proposed architecture has one extra fully connected layer compared to the original architecture with the purpose to model global patterns better.

The CNN receives RGB input image (one channel) of size 148×148 pixels. The first layer of the network, which is a convolutional layer perform convolution on the input image with the kernel of 9×9 in size and produces output with 140×140 pixels in size. There are 32 filters and therefore 32 outputs are produced. The second layer is a sub-sampling layer with 2×2 kernel and stride of 2, which performs max-pooling. The layer produces outputs with its size half of the input. The layer is followed by a convolutional layer with 64 filters and kernel size of 9×9. It produces 64 outputs with size of 62×62. The fourth layer is the same with the second layer. The subsequent layer is a fully connected layer with 4096 neurons. The final layer is also a fully connected layer with 1024 neurons.

Drop out is done at the last layer to drop out elements randomly during training. This reduces the overfitting in neural network [27]. According to the author, 20% to 50% of the elements are typically dropped out. Through trial and error, 40% elements are dropped out. The elements are sent to logits layer which produces the raw values of the prediction. The loss is calculated using cross entropy with "softmax_cross_entropy" function. The loss measures how closely the predictions of the model match the target classes. As suggested in [28], stochastic gradient descent [29] is used for training the network because the task is classification, the training data is large (several thousand data) and redundant. The learning rate is 0.001. The activation function of the neurons is ReLU which is defined as $f(z) = max(z,0)$. It is more effectively than the widely used logistic sigmoid [28].

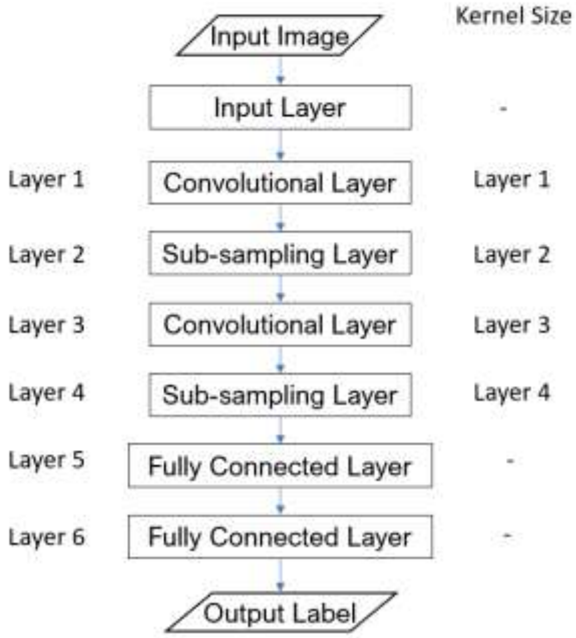

Figure 3. Architecture of the proposed CNN model

4. Experimental Results and Discussion

Another notable contribution from this work is the evaluation method of the model. The work in [30] used k-fold cross validation technique. In our evaluation method, the database is split into different groups for training and testing in a strict subject independent manner. Each subject will only be assigned into one group and will not be assigned to other group. The evaluation without subject overlapping provides fairer and more accurate result as evaluation with subject overlapping provides higher accuracy. Using this technique, the generalization performance to novel subjects can be evaluated. This evaluation method is commonly used in evaluating image processing systems [9, 10].

The experiments compute accuracy using the metric used in approach [9]. The accuracy per expression is calculated as described in Equation (1) where P_E is the number of correctly predicted in the expression and N_E is the total number of samples of that expression. The accuracy, A_{avg} is then computed using the average of accuracy per expression as described in Equation (2) where n_class is the number of facial expressions to be considered. Since this work recognizes two colours only, the n_{class} is equal to 2

$$A_E = \frac{\sum P_E}{N_E} \tag{1}$$

$$A_{avg} = \frac{\sum_{E=1}^{n_{class}} A_E}{n_{class}} \tag{2}$$

The system is implemented using Python 3.7 for general programming, OpenCV 2.0 for image processing and TensorFlow 1.0 library for building the CNN. The experiments were carried out using 2.90GHz Intel i5 processor.

The following parts of the section discuss the experiments carried out to evaluate the TDS in different aspects. Firstly, the impact of white coating removal and data augmentation were investigated. Subsequently, the results of using the proposed architecture and original architecture are presented and discussed.

4.1. Result of the Proposed Convolutional Neural Network

Table 1 shows the confusion matrix of the result achieved by the proposed CNN model. It is the highest accuracy achieved by using the model. The accuracy of recognizing the red colour is 93.75% while the accuracy of recognizing the pink colour is 78.94%. The accuracy of this test is 88.98%. The average accuracy from 5-fold cross validation is 75.41%.

Table 1. Confusion matrix table of the result from the proposed model

	Red	*Pink*
Red	30	2
Pink	4	15

4.2. Comparison with similar work in literature

To benchmark the proposed TDS, comparison against baseline work has been performed. Table 2 shows the comparison of the result. The comparison is made between the result of the reproduced baseline work and this work are compared. As shown from the result, the accuracy achieved in this project is much higher than the baseline work. This work achieved the highest accuracy at 88.98% and average accuracy of 75.72% by using synthetic one-sided flip. The result achieved in this work is much higher than that of the reproduced baseline work. The baseline work achieved an average accuracy of 51.81% and the highest accuracy of 59.62% (without initial pre-processing) using the same data input as the proposed work.

On the other hand, baseline work has much faster processing speed. From this comparison, it can be seen that CNN is much more accurate than SVM. This proves the strength of CNN to extract feature and classify image. However, CNN is a lot slower than SVM. From this work, CNN is 280 times slower than the baseline work. This shows that the technique can enrich the database with new useful feature and reducing the chances of over-fitting. These lead to the greater generalization power of the CNN model after being trained with the augmented database.

Table 2. Result comparison against baseline work

Work	*Average* A_c (%)	*Highest* A_c (%)	*Processing time (ms)*
SVM method [5]	51.81	59.62	0.179
Proposed Work	75.72	88.98	50

5. Conclusion

In conclusion, an accurate CNN-based TDS has been developed by a series of data augmentation flow. The developed TDS can differentiate red and pink tongue colour from image accurately. The system achieved accuracy which is up to 88.98% and average of 75.72%. The developed CNN model is relatively simple by having only 6 layers.

Comparing this work with baseline work in [5], it can be concluded that CNN is able to achieve high accuracy with lesser pre-processing. By using original the image without any initial data pre-processing, the developed CNN-based TDS can achieve the significantly higher accuracy with some trade-off in processing time. Despite this, processing speed still remains very fast in the order of miliseconds for our proposed model.

In future work, more pre-processing and augmentation techniques can be evaluated to find the suitable technique to further increase the accuracy. Apart from that, hardware accelerator of this TDS can be proposed so that it could be implemented in real-world application.

6. References

[1] Lo, L.-C., Chen, C.-Y., Chiang, J. Y., Cheng, T.-L., Lin, H.-J. and Chang, H.-H. Tongue Diagnosis of Traditional Chinese Medicine for Rheumatoid Arthritis. *African journal of traditional, complementary, and alternative medicines :AJTCAM*, vol 10, no. 5, pp.360–369, 2013.

[2] Kim, J., Son, J., Jang, S., Nam, D.-H., Han, G., Yeo, I., Ko, S.-J., Park, J.-W., Ryu, B. and Kim, J. Availability of Tongue Diagnosis System for Assessing Tongue Coating Thickness in Patients with Functional Dyspepsia. *Evidence Based Complementary and Alternative Medicine*, vol 2, pp.1–6, 2013.

[3] Gao, Z., Po, L., Jiang, W., Zhao, X. and Dong, H. A Novel Computerized Method Based on Support Vector Machine for Tongue Diagnosis. *2007 Third International IEEE Conference on Signal-Image Technologies and InternetBased System*. pp.849–854, 2007

[4] Guitao Cao, Jie Ding, Ye Duan, Liping Tu, Jiatuo Xu and Dong Xu. Classification of tongue images based on doublet and color space dictionary. 2016 *IEEE International Conference on Bioinformatics and Biomedicine (BIBM).* pp.1170–1175, 2016.

[5] Kamarudin, N. D., Ooi, C. Y., Kawanabe, T., Odaguchi, H. and Kobayashi, F. A Fast SVM-Based Tongues Colour Classification Aided by k-Means Clustering Identifiers and Colour Attributes as Computer-Assisted Tool for Tongue Diagnosis. *Journal of Healthcare Engineering*. vol. 2017. pp.1–13, 2017.

[6] Ko, M. M., Lee, J. A., Kang, B.-K., Park, T.-Y., Lee, J. and Lee, M. S. Interobserver Reliability of Tongue Diagnosis Using Traditional Korean Medicine for Stroke Patients. *Evidence-Based Complementary and Alternative Medicine*, pp.1–6. 2012.

[7] Hsu, Y., Chen, Y., Lo, L. and Chiang, J. Y. Automatic tongue feature extraction. *2010 International Computer Symposium (ICS2010)*. pp. 936–941, 2010.

[8] Hou, J., Su, H., Yan, B., Zheng, H., Sun, Z. and Cai, X. Classification of tongue color based on CNN. *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, 2017. pp. 94–101, 2010.

[9] Lopes, A. T., de Aguiar, E., Souza, A. F. D. and Oliveira-Santos, T. Convolutional Neural Networks based Method for Improving Facial Expression Recognition. *Pattern Recognition*, vol 6, no. 1, pp.610–628, 2017

[10] Mollahosseini, A., Chan, D. and Mahoor, M. H. Going deeper in facial expression recognition using deep neural networks. *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp.1–10, 2016.

[11] Zhan Wu, Min Peng and Tong Chen. Thermal face recognition using convolutional neural network. *2016 International Conference on Optoelectronics and Image Processing (ICOIP)*, pp.6–9, 2016

[12] Wang, M., Wang, Z. and Li, J. Deep convolutional neural network applies to face recognition in small and medium databases. *2017 4th International Conference on Systems and Informatics (ICSAI)*, pp.1368–1372, 2017.

[13] Kamarudin, N. D., Ooi, C.-Y., Kawanabe, T. and Mi, X. Tongue's substance and coating recognition analysis using HSV color threshold in tongue diagnosis. *First International Workshop on Pattern Recognition*, 2016.

[14] Watsuji, T., Arita, S., Shinohara, S. and Kitade, T. Medical application of fuzzy theory to the diagnostic system of tongue inspection in traditional Chinese medicine. *IEEE International Fuzzy Systems. Conference Proceedings*, vol. 1. pp.145–148, 1999.

[15] Aliev, R. A. Fuzzy Sets and Fuzzy Logic, *Springer Berlin Heidelberg*. 1–64, 2013.

[16] Haralick, R. M., Shanmugam, K. and Dinstein, I. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, pp. 610–621, 1973.

[17] Lin, B., Xle, J., Li, C. and Qu, Y. Deeptongue: Tongue Segmentation Via Resnet. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp 1035-1039, 2018.

[18] He, K., Zhang, X., Ren, S. and Sun, J. Deep Residual Learning for Image Recognition. *CoRR*, 2015.

[19] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R. B., Guadarrama, S. and Darrell, T. Caffe: Convolutional Architecture for Fast Feature Embedding. *CoRR*, 2014.

[20] Carvajal, J., Romero, D. and Sappa, A. Fine-Tuning Based Deep Convolutional Networks for Lepidopterous Genus Recognition., vol. 10125, 467–475, 2017.

[21] Hou, J., Su, H., Yan, B., Zheng, H., Sun, Z., & Cai, X., Classification of tongue color based on CNN. *2017 IEEE 2nd International Conference on Big Data Analysis*, pp 725-729, 2017.

[22] Guangyu N., Caiqun W., Bo Y., Yong P., Tongue Color Classification Based on Convolutional Neural Network. *Advances in Intelligent Systems and Computing*, vol 1364. Pp 649-662, 2021.

[23] N. M. Zabidi, A. A. H. Ab Rahman, VLSI Design of a fast pipelined 8x8 discrete cosine transform, *International Journal of Electrical and Computer Engineering*, vol 7, no. 3, 2017.

[24] A. A. H. Ab Rahman, I. Kamisian, A. Z. Sha'ameri, ASIC Modeling and Simulation of Adaptive FIR Filter, *2008 Interantional Conference on Electronic Design*, pp.1-4, 2008.

[25] S. M. Rizvi, A. A. H. Ab Rahman, M. Khalil-Hani, S. O. Ayat, A Low-complexity Complex-valued Activation Function for Fast and Accurate Spectral Domain Convolutional Neural Network, *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, vol 9, No. 1, pp.173-184, 2021.

[26] L. Bottou, Y. Bengio, and Y. Le Cun, "Global training of document processing systems using graph transformer networks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 1997.

[27] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *J. Mach. Learn. Res*., vol. 15, pp. 1929–1958, 2014.

[28] Y. A. LeCun, L. Bottou, G. B. Orr, and K. R. Miiller, "Efficient backprop," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7700, pp. 9–48, 2012.

[29] L. Bottou, "Stochastic Gradient Descent tricks," *Neural Networks, Tricks of the Trade, Reloaded,* vol. 7700, pp. 430–445, 2012.

[30] Shan, C., Gong, S. and McOwan, P. W. Facial expression recognition based on Local Binary Patterns: A comprehensive study. *Image and Vision Computing*,. vol 27, no. 6, pp.803 – 816, 2009.