DESIGN OF PREDICTIVE MODEL FOR TCM TONGUE DIAGNOSIS IN MALAYSIA USING MACHINE LEARNING

KOE JIA CHI

A project report submitted in partial fulfilment of the requirements for the award of Bachelor of Engineering (Hons.) Electronic and Communications Engineering

Lee Kong Chian Faculty of Engineering and Science
Universiti Tunku Abdul Rahman

April 2020

DECLARATION

I hereby declare that this project report is based on my original work except for citations and quotations which have been duly acknowledged. I also declare that it has not been previously and concurrently submitted for any other degree or award at UTAR or other institutions.

Signature : 

Name : KOE JIA CHI

ID No. : 1503731

Date : 16/05/2020

APPROVAL FOR SUBMISSION

I certify that this project report entitled **"DESIGN OF PREDICTIVE MODEL FOR TCM TONGUE DIAGNOSIS IN MALAYSIA USING MACHINE LEARNING"** was prepared by **KOE JIA CHI** has met the required standard for submission in partial fulfilment of the requirements for the award of Bachelor of Engineering (Hons.) Electronic and Communications Engineering at Universiti Tunku Abdul Rahman.

Approved by,

Signature : 

Supervisor : Dr. Lai An-Chow

Date : 16/05/2020

Signature : 

Co-Supervisor : Dr. Goh Yong Kheng

Date : 16/05/2020

ACKNOWLEDGEMENTS

I would like to thank everyone who had contributed to the successful completion of this project. I would like to express my gratitude to my research supervisor, Dr. Lai An-Chow and Dr. Goh Yong Kheng for their invaluable advice, guidance and their enormous patience throughout the development of the research.

In addition, I would also like to express my gratitude to my loving parents and friends who had helped and given me encouragement.

ABSTRACT

In recent years, Traditional Chinese Medicine (TCM) has gained popularity in Malaysia. There are four diagnostic methods (四诊) in TCM: Inspection (望), Listening and Smelling (闻), Inquiry (问) and Palpation (切). Tongue diagnosis which is part of Inspection is carried out through the observation on patient's tongue body and coating. However, tongue diagnosis is subjective and is lack of objective evaluation criteria as the judgement is made based on the TCM physician's experience, and thus different physicians might have different judgements towards the same patient. The lack of objectivity and standard evaluation criteria in tongue diagnosis have restricted its development. In this project, machine learning algorithm will be applied to design a predictive model for TCM tongue diagnosis. This project is divided into several parts, specifically as follows:

1. The existing tongue image acquisition system has strict requirements on the light source and the camera. However, the portability and popularity of these instruments are still poor, and thus an easy way of taking tongue image by using mobile camera is proposed. Five important rules for taking a tongue image are established to ensure the image quality.
2. Mask R-CNN is trained to segment the tongue from the image. The results show that it is able to segment the tongue under different illumination and even if it is blur or not captured exactly from the front of the tongue.
3. Four tongue features (greasy tongue coating (腻苔), teeth-marks (齿痕), cracks (裂纹), and spots (点刺)) are extracted from each image. YOLO are employed in this project to extract cracks and teeth-marks while Mask R-CNN are used to extract greasy tongue coating and spots. YOLO achieves 100% accuracy in extracting cracks and near 80% accuracy in extracting teeth-marks. Meanwhile, Mask R-CNN achieves 87.5% accuracy in extracting greasy tongue coating,. However, both Mask R-CNN and YOLO do not perform well in extracting spots. Although Mask R-CNN achieves 85% accuracy, its sensitivity and F1-score are just 45% and 47% respectively.

4. Six supervised machine learning algorithms (Linear Regression, Logistic Regression, K Nearest Neighbors (KNN), Decision Trees (DT), Support Vector Machine (SVM) and Random Forest) are used to perform disease prediction. Besides, cross validation and bootstrap are implemented to ensure the robustness and to improve the accuracy of the predictive model. Two predictions are carried out: prediction of healthy/unhealthy and prediction of high blood pressure/no high blood pressure. However, all algorithms perform poorly in predicting healthy/unhealthy as the highest accuracy is just 68% which was obtained using DT. For prediction of high blood pressure/no high blood pressure, all algorithms have really bad performance without bootstrapping, where their accuracies are all around 50% while their sensitivity and F1-score were less than 20%. After bootstrapping, KNN and SVM are able to achieve near 80% in accuracy, sensitivity and F1-score. KNN even achieves 90% sensitivity. In other words, KNN is able to catch most of the positive cases correctly.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS / ABBREVIATIONS

TCM	Traditional Chinese Medicine
CNN	Convolutional Neural Network
FCN	Fully Convolutional Network
KNN	K-Nearest Neighbors
SVM	Support Vector Machine
LOG	Laplacian of Gaussian
FPN	Feature Pyramid Network
RoI	Region of Interest

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 General Introduction

In recent years, Traditional Chinese Medicine (TCM) has gained popularity in Malaysia. Malaysian Chinese have played an important role in the development and wide application of TCM in Malaysia. The ancestors of Malaysian Chinese originate from China. They had moved to Malaya since the 7th century. According to Zheng Jianqiang (2018), in the 15th century, Zheng He, the famous Chinese explorer and fleet commander brought a lot of Chinese herbal medicines such as tea, ginger, rhubarb, etc. when he visited to Malacca. At that time, Malacca Sultanate had maintained a good relationship with the Ming Dynasty.

When Malay Peninsula was colonized by British, a large number of Chinese workers came to the Malay Peninsula to work in tin mines and rubber tree plantations. They relied on Chinese herbal medicine, acupuncture and massage to maintain their health, and also played a role in the spread of TCM in Malaysia.

Malaysia is a multicultural society which embraces the coexistence of different cultures. TCM not only has a place here, but also develops with the support of our government. In 2015, the government had included traditional and complementary medicines in the 11th Malaysia Plan, showing our country's emphasis on traditional and complementary medicine development and its potential to increase national income. Besides, in 2018, the Ministry of Health introduced the “Traditional and Complementary Medicine Blueprint 2018-2027 (Health Care)”, which then accelerates the further development of TCM in Malaysia.

There are four diagnostic methods (四诊) in TCM: Inspection (望), Listening and Smelling (闻), Inquiry (问) and Palpation (切). Inspection (望诊) is carried out through observation on one’s mental state (精神状态), complexion (面部色泽), body shape (形体胖瘦), gesture (动静姿态) and tongue body & coating (舌质舌苔) as TCM physicians claim that these physical

features are closely related to human organ. In order words, one's health condition will be reflected in the changes of these physical features.

Tongue diagnosis which is part of Inspection is carried out through the observation on patient's tongue body and coating. The theory and principle of tongue diagnosis are developed from the experiences of TCM physicians that have been accumulated through thousands of years of clinical practice. Liu Feilong (2014) claims that the discoloration of a particular region of the tongue indicates a lesion in the human organ corresponding to that region. Therefore, tongue diagnosis is based on the theory that each region of tongue is closely related to a particular human organ. By observing on one's tongue, TCM physicians can make judgement if there is a lesion in that particular internal organ. However, tongue diagnosis is subjective and is lack of objective evaluation criteria as the judgement is made based on the TCM physician's experience, and thus different physicians might have different judgements towards the same patient. The lack of objectivity and standard evaluation criteria in tongue diagnosis have restricted its development.

However, with the development of Artificial Intelligent, people are paying more attention to computerized tongue diagnosis. Relying on modern information technology techniques to study the principle of tongue diagnosis, making it more scientific, quantitative and objective, has become an inevitable direction of tongue diagnosis research. Therefore, this project proposes the design of a predictive model for TCM Tongue diagnosis using machine learning.

Machine learning is one of the popular techniques in data mining. As an application-driven domain, data mining incorporates such things as statistics, machine learning, pattern recognition, database and data warehousing, information retrieval, visualization, algorithms, and high-performance computing.

Classification is an important form of data analysis which extracts models that characterize important data classes. This model is called a classifier and it predicts the class label of the classification. For example, we can build a classification model that divides the tongues into healthy and abnormal, and this analysis can help us to interpret the data better.

Data classification mainly consists of two phases: the learning phase and the classification phase. The learning phase can also be seen as learning a

mapping or function $y = f(x)$, which predicts the class label, y of a given tuple, x. Typically, this mapping is provided in the form of classification rules, decision trees, or mathematical formulas. Next, in classification phase, we build model to perform classification. The prediction accuracy of the classifier is first evaluated. It is worth noting that if we use the training set to train and test the classifier at the same time, the accuracy may be very high, but the performance of the classifier is greatly reduced when testing with unknown new data. This scenario is known as overfitting, where the classifier has been trained to fit the training set perfectly, and therefore fail to predict additional or new data.

Therefore, another testing set consisting of testing data with their corresponding labels is needed. The data in testing set should be totally different from that in training set. In other words, the data in testing set aren't used in the training process of a classifier. Today, there is a variety of machine learning algorithms being proposed to perform classification and prediction such as Artificial Neural Network (ANN), Multilayer Perceptron, Random Forest, Support Vector Machines (SVM), etc. These algorithms have also been widely used in fraud detection, performance prediction, target marketing, manufacturing, and medical diagnosis.

In this project, machine learning algorithm will be applied to perform automatic tongue segmentation, tongue regions extraction and segmentation. Besides, a convolutional neural network based automatic tongue feature extraction method is proposed to extract and analyse tongue features. Lastly, based on the features extracted, a predictive model will be built to carry out TCM tongue diagnosis.

1.2 Importance of the Study

During TCM tongue diagnosis, the TCM physician will observe the patient's tongue. The human tongue carries the clues about the health of other organs. Therefore, through the observation on the tongue, sign of any internal disease or lesion in any internal organ can be identified. This diagnosis method is not only fast and effective, but also non-invasive, and causes no harm to the patient, so the tongue diagnosis is widely used in the medical field.

However, tongue diagnosis is subjective and is lack of objective evaluation criteria as the judgement is made based on the TCM physician's

experience, and thus different physicians might have different judgements towards the same patient. Besides, interference from ambient light when the observation is carried out may lead to a wrong diagnosis. In addition, TCM tongue diagnosis has faced difficulty in quantification of the condition of the tongue. For example, the rot (腐腻), moistness (润燥) and thickness of the tongue coating are difficult to quantify. Therefore, the standardization and objectification of TCM tongue diagnosis is an important part of the modernization of TCM.

1.3 Problem Statement

With the development of image processing and pattern recognition technologies, computerized tongue diagnosis system has been improved tremendously. Various algorithms for tongue segmentation, feature extraction and classification have been proposed. However, there are several important problems that have yet to be solved.

First, tongue image acquisition system is the fundamental component of a computerized tongue diagnosis system. Although there has been different design for tongue image acquisition system being proposed, but there is a lack of clear guidelines on taking a qualified tongue image. The existing system use different cameras and lighting sources. For example, Jiang Yiwu et al. (2000) uses a standard color temperature cold light (color temperature value of about 5300 K, brightness of about 3100 Lux) as a light source while Wei Baoguo et al. (2002) uses two Osram full-spectrum L18/72 Biolux D 6500 light with color temperature of 6 500 K. Therefore, the tongue image quality is inconsistent for different system.

Next, the colour of tongue image is device dependent. Due to different equipments used in acquiring tongue image, a colour correction algorithm is needed to make the colour of tongue image consistent. Various colour correction algorithm have been proposed but they are not designed specifically to correct tongue colour. Therefore, current colour correction algorithm has to be improved so that it could be applied in tongue colour correction.

Lastly, tongue feature extraction algorithm has to be improved to make it more objective. There are algorithm designed to extract different tongue features such as tongue coating, tongue colour, teeth-mark, moisture of tongue,

etc. However, these features have been selected based on TCM physicians' experience. Some feature which is highly related to a certain disease could be ignored. Therefore, an automatic tongue feature extraction algorithm which could automatically related features has to be designed.

1.4 Aims and Objectives

The main aim of this project is to design a predictive model for TCM tongue diagnosis that could predict healthy/unhealthy and high blood pressure/no high blood pressure through tongue image. The specific objectives of this research are:

1. To understand the theory and principles of TCM tongue diagnosis
2. To analyse the changes in healthy tongue and tongue of high blood pressure patient
3. To implement a predictive model through machine learning for TCM tongue diagnosis
4. To evaluate the predictive model in terms of accuracy, precision, sensitivity, specificity and F1-score

1.5 Scope and Limitation of the Study

This project will include the design of guideline for taking tongue images using mobile phone, image colour correction algorithm, tongue segmentation algorithm, tongue region segmentation algorithm, automatic feature extraction algorithm and lastly disease prediction. The prediction result will be compared to judgement made by TCM physician, in order to evaluate its accuracy.

There are some limitations in this project which will not be in the project scope. First, other than observing the tongue, TCM physician may listen to the voice of the patients to make judgement. However, the voice analysis will not be covered in this project. Next, the prediction of disease will be based on changes in tongue features, therefore the case where the TCM physicians also could not make judgement by solely carrying out observation on tongue will not be considered. Also, there will be no correction for tongue image rotated in z-axis. As shown in figure below, tongue rotated in z-axis could lose the details

of tongue margin (舌边). Tongue margin is important to observe the existence of teeth-mark (齿痕).

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Tongue is an organ in the mouth which is composed of skeletal muscle fibres. It is long and flat with its surface being covered by mucous membrane. It assists us to stir food in our mouth, swallow, and taste the food. According to Traditional Chinese Medicine (TCM), tongue is divided into four regions: root (舌根), center (舌中), tip (舌尖) and margin (舌边). As shown in Figure 2.1.1, tongue root locates at the back of the tongue, tongue center is at the middle of the tongue while tongue tip is at the front end of the tongue and the tongue margin is at either side of the tongue. The results of TCM clinical trials found that discoloration of a particular region of the tongue indicates a lesion in the human organ corresponding to that region (Liu Feilong, 2014). Tongue tip relates to lungs and heart; tongue center relates to spleen and stomach; tongue margin relates to liver; tongue root relates to kidney.

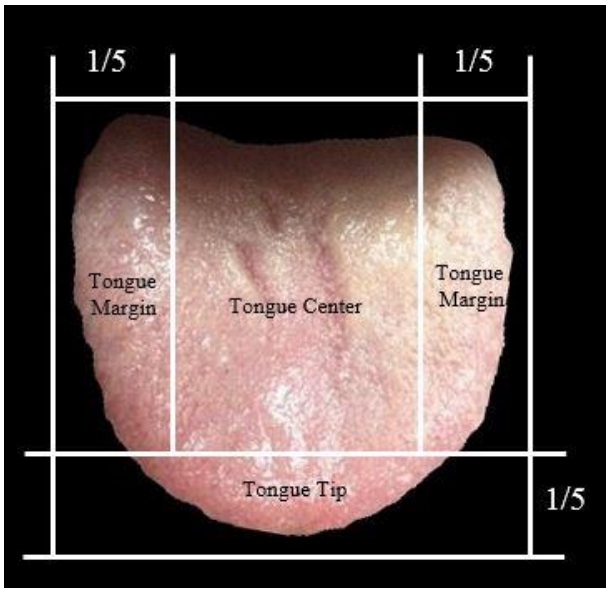


Figure 2.1.1 Tongue regions with their ratios. Adapted from Xu Jiayu (2009).

When the human body is attacked by bacteria and viruses, the immune system will defend our body by producing cells to attack the antigen. When the immune system is triggered, the hypothalamus is also activated, which make us feel cold, causing the metabolic rate to accelerate and blood supply to the cells

to increase. By doing so, our body temperature will rise. When blood supply is increased, the color of the tongue becomes significantly redder than when it is healthy. Also, when we are sick, the activities of the parasympathetic nerve will be affected, leading to a decrease in salivary secretion; tongue coating (舌苔) will become sticky and harder to be observed (Zhang Guangyu, 2018). Therefore, observation on the colour and shape of tongue become an important part of TCM tongue diagnosis.

According to Yu et al. (1994), TCM tongue diagnosis involves observation on color, texture, shape, state and coating of the tongue. The physician will observe the patient's tongue, and make the judgement based on his/her clinical experience. Hu, et al. (2018) state that TCM physician had divided human body into nine constitutions (体质) based on the characteristics of human tongue, as shown in Table 2.1.1.

Table 2.1.1 Characteristics of human tongue with different constitutions

Constitutions	**Characteristics of human tongue**
neutral (平和体质)	Light red tongue with thin white coating
qi deficiency (气虚质)	Light red and swollen tongue with teeth mark at the edges of tongue
yang deficiency (阳虚质)	Red tongue with little moisture and tongue coating
yin deficiency (阴虚质)	Red tongue with little moisture and coating
blood stasis (血瘀质)	The lips are dull or purple with some petechiae on the tongue
phlegm & dampness (痰湿质)	Swollen tongue with white greasy coating
damp-heat (湿热质)	Red tongue with yellow greasy coating
qi stagnation (气郁质)	Light red or dull tongue with thin white or dry and white greasy coating
special constitution (特禀质)	Diverse forms but usually with visible cracks and the condition of coating peeling off

2.2 Tongue Image Acquisition

Traditional tongue diagnosis have been carried out by TCM physician through their observation on patients' tongue. Lighting and observation angle could affect the diagnosis. In addition, different physicians may have different opinions towards the same patient because they have different clinical experience. The lack of objective diagnostic indices in traditional tongue diagnosis has hindered its development. The objectification of tongue diagnosis requires investigation in tongue image acquisition method, image processing, feature extraction and classification. The quality of tongue image is one of the important factors that affects the subsequent processing of tongue image and it forms the foundation of further diagnosis.

The tongue image acquisition system (舌象采集装置) designed by Yu et al. (1994) used a tungsten halogen lamp to ensure the camera's color temperature requirements. The illumination system was designed based on Kohler's illumination principle, where the two illumination systems were at 45 ° angle on both sides of the subject, and the light was evenly projected on the tongue surface. Subsequently, Jiang Yiwu et al. (2000) proposed the establishment of a darkroom to avoid the influence of external light on the tongue image being captured. A head holder (头部固定架) was designed to fix the position of the subject's head and tongue, the light source and the camera. He used a standard color temperature cold light (color temperature value of about 5300 K, brightness of about 3100 Lux) as a light source while taking tongue image.

With the development of computer technology, the quality of digital image had been greatly improved, leading to rapid development in the objectification of tongue diagnosis had been rapidly developed. The tongue image acquisition system invented by Wei Baoguo et al. (2002) used a Kodak DC260 digital camera with image resolution of 1536×1024. Two Osram full-spectrum L18/72 Biolux D 6500 light with color rendering index, Ra = 96 were used. The color temperature was 6 500 K; the illumination geometry was 45/0. Wu Zuchun (2011) proposed that the camera lens used for taking tongue image should have macro function, focal length of 50mm ≤ f ≤ 105mm, and small aperture (f/8~f/11). Meanwhile, he used a digital SLR camera with CCD as the

sensor. Mode II (adobe RGB) was chosen as the colour mode; ISO was set to minimum; At low light conditions, auxiliary light source was used with ISO value not more than 400. Manual preset white balance was used.

Some tongue image acquisition system were designed as a box. Yang (2018)'s tongue image acquisition system (as shown in Figure 2.2.1) had been designed based on average head size of adult. It had a tongue window and a window for camera; the tongue window was tilted for 66 ºwith the horizontal plane in order to ease the capture of complete tongue image including tongue root.

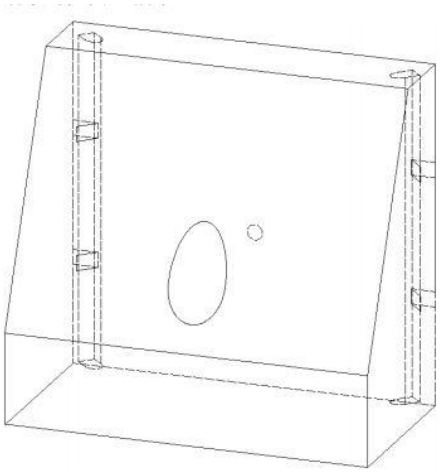


Figure 2.2.1 Tongue Image Acquisition System designed by Yang (2018)

At present, tongue image are usually taken by digital cameras, video cameras and digital SLRs. Undeniably, the image quality is excellent as the tongue features can be clearly captured. However, these tongue image acquisition systems are not portable and are difficult for many people to have exact same device for tongue diagnosis. Therefore, the design has to be improved in order to enhance its popularity and at the same time maintain the good image quality.

2.3 Colour Correction

Images taken using different mobile phone and digital camera will have problem of device-dependent colour space rendering because the image colour information will depend on the imaging specification of the camera. Besides, some noises will be generated together with the images due to variation in environment lighting. Therefore, to order to enhance the accuracy of subsequent

tongue diagnosis, colour correction is one of the most important part during image pre-processing process.

There are many colour correction algorithms being proposed specifically to be used in different area of application. Among them, polynomial-based correction method (Luo, et al., 2001; Cheung, et al., 2004) and neural network mapping (Cheung, et al., 2004) are two popular colour correction algorithms. However, there are only a little number of researches that focus on correcting tongue image colour. Zhang et al. (2005) proposed a novel color correction approach based on the Support Vector Regression (SVR) algorithm, and their experimental results confirmed the effectiveness of the proposed technique. Hu et al. (2016) used the support vector machine (SVM) to predict the lighting condition and the corresponding color correction matrix according to the color difference of images taken with and without flash. Next, Zhuo et al. (2015) proposed a kernel partial least squares regression based method to obtain consistent correction by reducing the average color difference.

However, most of the methods mentioned above require a reference. For example, colour checker or images with and without flash have to be taken as a reference. Besides, there is lack of colour correction method which is able to eliminate the interference of shadow. Chen et al. (2017) proposed a two-stage color correction algorithm to effectively solve two problems. To remove the shadows in the tongue images, Frankle-McCann retinex algorithm was implemented. Then, to restore the whole color distribution of the tongue images as real world, the gray world algorithm was utilized to fine-tune the color values of the tongue images.

2.4 Tongue Segmentation

The acquired tongue image will contain the subject's face and the background. Therefore, further image processing has to be carried out to segment the tongue. The early tongue segmentation algorithms mainly include threshold method, edge detection method and region segmentation method. After that, a variety of new segmentation algorithms had been proposed, such as mathematical morphology, watershed method, fuzzy set theory, clustering algorithm, artificial neural network, etc., which make the segmentation result more and more

accurate and subsequently lay a good foundation for subsequent feature extraction operation and analysis.

Jiang et al. (2017) extracted the information of G, B, and V channels of RGB and HSV colour space from the image, then uses the Otsu threshold method to segment the tongue, and improve the final segmentation result using the morphological opening method. The shortcoming of this algorithm is that the segmentation result after Otsu threshold method contains some non-tongue regions which is wrongly segmented; morphological opening method can remove these non-tongue regions only if the area of tongue regions is larger than that of non-tongue regions.

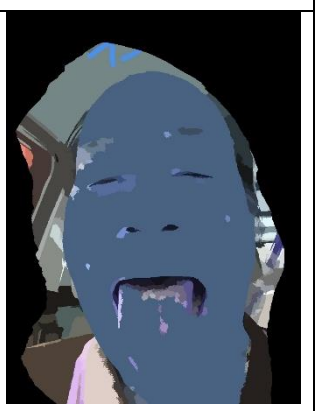
Next, Yu et al. (1994) used fuzzy mathematics to perform cluster analysis to locate the rough tongue region. He first set the threshold for tongue image R, G, B pixel values, then grouped the similar pixels by comparing their R, G, B values. However, this algorithm is highly dependent on colour space information, and the algorithm becomes less accurate when the background is complex or the tongue colour is close to the skin colour.

Wang Sheng (2016)'s tongue segmentation operation had two steps: tongue localization and precise segmentation. Firstly, the skin colour detection algorithm was used to remove the complex background, then translated H channel value in the HSV colour space. After that, the mean shift algorithm was used for filtering and extraction of tongue localization result in the L*a*b* color space. The precise segmentation focus had improved the mark control watershed algorithm. For subsequent precise segmentation operation, the foreground mark obtained through the morphological operation was merged with the tongue positioning result to obtain a new foreground mark; the watershed algorithm was used to obtain the rough segmentation result; the geodesic contour model was used to improve segmentation result. The skin colour detection algorithm used in this document has a strong dependence on the colour space. When the background of the acquired tongue image is complex or the background colour is similar to the skin colour, the skin colour detection algorithm becomes inaccurate.

Xu (2011) had implemented mean shift based clustering to divide the image into a number of clusters based on the color and spatial similarity, then employed Principal Component Analysis (PCA) to fit an ellipse into the cluster.

After that the similarity measurements between the cluster and the fitting ellipse were computed and the cluster was detected as a cluster that contained the tongue if the similarity was greater than a threshold. The tongue was then segmented with Tensor Voting based image segmentation method. The mean shift algorithm used by Xu was proposed by Comaniciu (2002). Comaniciu (2002) defined mean shift procedure which was used as the computational module for robust feature space analysis. The feature space analysis technique was applied to application like discontinuity preserving filtering and image segmentation. Mean Shift is widely used for feature space analysis; it is easy to implement but its performance is closely related to the selection of parameters: spatial radius, h_s, range radius, h_r and minimum density, M. However, trial and error is the only way that we could choose the most suitable parameter values as there are no systematic way of choosing them. Table 2.4.1 shows the mean shift result with $h_s=8$, $h_r=7$, $M=100$ on two different images. Tongue in first image had been successfully grouped into one cluster but not for tongue in second image. This shows that the same set of values may not fit in with different images.

Table 2.4.1 Mean shift result with $h_s=8$, $h_r=7$, $M=100$

	Original Image	Mean-shift result
1		
2		

With the development of machine learning and deep learning, several breakthroughs been made in recent years, and deep learning algorithms are

increasingly used in various fields. Convolutional Neural Network (CNN) has been widely used in image processing and speech recognition. Yan Tingting (2016) proposed a six-layer convolutional neural network based on CNN and mathematical morphology of tongue image segmentation algorithm. The network was trained by a large number of samples to achieve the classification of image pixels. Mathematical morphology was used to improve the results. However, this algorithm has a poor performance on segmentation of tongue from the lips. Chen Feifei (2018) used the gray projection method to locate the segmented tongue image, then constructed a VGG 16-FCN-8s neural network to extract the tongue. Mathematical morphology was then used to optimize the extraction results and realized the segmentation of the tongue image. FCN is commonly used for semantic segmentation where it will group each pixel of same category into a single mask. The segmentation result will be less accurate when the background is having something looks similar with human tongue as it will be segmented together with the tongue.

There are many existing algorithms for tongue image segmentation. However, these algorithms are designed and proposed specifically to deal with tongues images that have been collected using specific tongue image acquisition system. In other words, algorithm proposed by one researcher may not properly segment the tongue from an image acquired using different tongue image acquisition system. The robustness of tongue segmentation algorithm has to be improved. Therefore, we need a tongue segmentation algorithm which is able to perform its job regardless of image brightness and the complexity of the background environment when an image is taken.

2.5 Tongue Feature Extraction

The traditional tongue diagnosis mainly relies on the physician's observation on patient’s tongue to judge and analyse. The lack of objective evaluation criteria restricts the further application and development of the tongue diagnosis. Therefore, modern scientific and technological means has to be implemented to study the principle of tongue diagnosis in order to make it more scientific, objective and quantitative. According to Yang (2018), TCM has investigated several features of human tongue and each feature is divided into several ratings as shown in Table 2.5.1.

Table 2.5.1 Tongue features with ratings

Features	**Ratings**
Colour	light (淡), light red (淡红), red (红), blush (绛红); light (淡), light purple (淡紫), purple (紫), blackish purple (黑紫)
Size	small, normal, large
Thickness	thin, normal, thick
teeth-marks (齿痕), cracks (裂纹), spots (点刺), petechiae (瘀斑)	none, mild, moderate, severe
State (舌态)	soft (萎软), short (短缩), tremor (震颤), skew (偏斜)
Coating colour (苔色)	grayish black, grayish white, white, yellow, black and yellow
Coating texture (舌苔质地)	none, thin, thick
Body fluid (津液)	dry, moist, watery
Greasy tongue coating (腻苔)	none, mild, moderate, severe
The condition of tongue coating peeling off (剥苔)	none, mild, moderate

Considering past work, various feature extraction algorithm had been designed to extract the features as shown in Table 2.5.1. For example, there are several models being proposed to analyse the colour of tongue body and its coating. Zhou Yue (2002) used HSI (hue, saturation, intensity) model to distinguish tongue body (舌质) and tongue coating (舌苔), and used Gaussian model for statistical analysis of confusing areas, leading to effective tongue body and tongue coating separation and their colour identification. Then, according to the energy distribution of 2D Gabor wavelet coefficients and the relationship between orientation and tongue traces (舌纹), the invariant moment method was used to qualitatively describe the number of tongues traces, which then provided important quantitative information for tongue diagnosis.

Next, Liang Jinpeng et al. (2017) proposed a classification algorithm toward six common types of tongue body and tongue coating based on their color characteristics with improved KNN algorithm. The KNN algorithm assigned different weights to the neighbor samples, increased the weight of the nearer samples, reduced the weight of the farther samples, and thus achieved accurate classification results. The improved KNN algorithm had better accuracy than traditional KNN algorithm with the average accuracy of more than 80%. However, according to the author, the tongue sample used during training process is typical sample without being interfered by much noise. Therefore, the robustness of this algorithm is questionable.

Chen Jingbo (2014) also built a classifier for heat syndrome (热证) and cold syndrome (寒证) based on tongue color features which involve using 8 color models (RGB, HSV, YIQ, YCbCr, XYZ, L*a*b, CIELuv and CMYK). Each pixel in the image will have total 25 parameters from 8 color models; for each parameter, mean, median, standard deviation and range of all pixels were calculated to generate a 100-dimensional feature vector. After that, SVM feature selection method was used for dimensionality reduction to improve the performance of the classifier. The algorithm could achieve an accuracy of 85.10% for the classification of normal tongue, tongue with heat syndrome, and with cold syndrome. However, the occurrence of these syndromes could be due to malfunction of different organs. Therefore, the classification of cold and heat syndrome solely cannot provide enough information for further diagnosis, it would be better if the malfunctioning organ that causes the syndrome could be identified.

Li Xiaoyu et al. (2006) proposed a method for classifying 15 types of tongue colour and coating colour with the combination of Directed Acyclic Graph (DAG) and Decision Tree. During the training process of the SVM classifier, different kernel functions and their parameters were adopted according to the linearly separable and linearly inseparable characteristics of tongue image sample. Compared to the direct use of DAG, the algorithm that combined DAG and decision tree had greatly reduced the number of classifiers to be passed and thus the classification process had been speeded up. The average accuracy of the classification model was 93.87%. However, the

accuracy depends upon the selection of SVM parameter. In this paper, the parameter was selected using k-fold cross validation method, which means several tests were carried out on the training dataset, the parameter value that resulted in highest accuracy will be selected, but this process may take a long time and maybe computationally expensive.

In addition to tongue colour and coating colour, the moisture of the tongue coating also provides important information for the tongue diagnosis. The moisture or dryness of the tongue coating reflects the amount of water in human body. Moist tongue coating indicates that the body fluid is sufficient; dry tongue coating indicates that the body fluid is depleted. Su Kaina et al. (1999) established a bisection light reflection model (二分光反射模型) and carried out the detection and identification of tongue moisture based on image bright spot feature analysis (图像亮斑特征分析). According to TCM, the thickness of tongue coating was judged based on the visibility of tongue surface (舌苔的厚薄以见底不见底为依据). The experimental result was then presented to the TCM physician and was approved by them.

After tongue body and tongue coating separation (舌质与舌苔分割), Shen Lanqi et al. (2003) quantified the degree of visibility of tongue surface, and used it to relate to the thickness of the tongue coating. Then, using the bisection light reflection model algorithm, the tongue coating moisture was automatically graded (dry, extremely dry, slightly moist, extremely moist, extremely wet, and wet). In addition, the moisture of the tongue coating was also reflected in the bright spot (亮斑) caused by the water film (水膜) on the tongue surface. However, the accuracy of the algorithm proposed was only close to 70%, and thus improvement is needed to increase the accuracy.

Xie Tao (2017) used the bisection light reflection model to analyse the distribution characteristics of pixel clusters, and then distinguished the white tongue coating (白苔) area from the bright spot area. After that, brightness gradient (亮度梯度) was calculated to screen out the qualified watery bright spot area, and then according to the size and brightness of bright spot, the degree of moisture of tongue coating was obtained. The experimental result differed from the result of manual screening by about 10%. The overall accuracy of this algorithm was 90%.

It is concluded that the identification of tongue moisture is about solving the problem of identifying the bright spot area, but there are some difficulties in the identification of the bright spot area. First, both the white tongue coating and the bright spots are bright white in colour, which is difficult to differentiate them by colour features alone. Second, the surface of the dry tongue coating is usually covered by a layer of secretory mucus; the water film covered on the mucus layer also can form bright spots. Therefore, the basis of the recognition of the tongue moisture is actually the recognition of correct bright spot area.

The spotted tongue (点刺舌) consists of spots (点) and thorns (刺). Spot is the spot that bulges on the tongue. According to TCM, the spotted tongue reflects a heat syndrome (热证), indicating a period in which the organs of the organs are vigorously heated (脏腑器官的阳热旺盛) or the blood is extremely hot (血液极度热). The characteristics of the tongue spot that can be investigated include its colour, distribution and shape. Wang Sheng (2016) proposed a method for identifying and extracting tongue spots (点刺) and petechiae (瘀斑). Firstly, spot detection algorithm was used, and the support vector machine was used to perform classification based on the spot features such as the number, size, and distribution. Then, the spot detection results were clustered into multiple small clusters by K-means clustering. Finally, the clustering results were compared with the spot detection results by defining a discriminant function based on the weighted color space distance (基于加权颜色空间距离的判别函数), and thus tongue spots and petechiae were extracted. This algorithm achieved low false positive rate of 6.0%, and at the same time high detection rate of 97.4%. However, the shortcoming of this algorithm is that it could not differentiate between spotted tongue (点刺舌) and tongue with petechiae (瘀斑舌); these two tongues will lead to different treatment according to TCM tongue diagnosis. Therefore, additional feature extraction algorithm is needed to distinguish tongue spots from petechiae.

Next, teeth mark (齿痕) refers to the trace of the tooth visible on the edge of the tongue. Due to the abnormal function of the spleen and stomach, the tongue becomes swollen and is squeezed by the teeth to form a tooth mark. Clinical studies have shown that the formation of teeth mark is closely related

to many diseases. The biggest difference between the scalloped tongue (齿痕舌) and the healthy tongue is that the scalloped tongue has a concave edge (凹边缘) at the tip of the tongue. Zhang Guangyu (2018) used the Canny algorithm to extract the edge of tongue tip, and then fitted the convex hull (凸包) at the edge of the tongue tip. He proposed two new geometric features (几何特征): convex hull degree (凸包度) and convex hull distance (凸包差). Experimental result showed that the use of convex hull degree allowed for better recognition of scalloped tongue, as compared to the use of circularity (圆形度) and convex hull distance. When the degree of convex hull was larger, it indicated that the degree of concavity of the edge of the tongue (舌尖边缘的凹陷) was small, which indicated a mild scalloped tongue; when the degree of convex hull was small, the degree of concavity of the edge of the tongue was high, indicating a severe scalloped tongue. This algorithm had a good recognition result on the tongue boundary with obvious concave edge; when the concave edge is not obvious, the result will be less accurate.

All the research papers mentioned above had performed feature extraction and then classification of the extracted features but there was a lack of the study on the link between the extracted features and diseases, until Yang (2018) proposed her work in feature recognition and disease prediction based on tongue samples of patients with chronic kidney disease (CKD).

Yang (2018) had proposed algorithms to perform feature extraction on tongue colour, teeth-mark, cracks, spots, coating colour and thickness of tongue coating. 12 color classification centers in CIExy chromaticity diagram were selected and were used together with the color histogram and color moment features to extract tongue colour and coating colour; based on the characteristics of different types of scalloped tongues, three boundary curves of the scalloped tongue were obtained, then the difference between them was calculated in order to identify teeth-mark; SLIC superpixel segmentation, seed node selection and region growth were implemented to segment the crack region; based on the shape and size characteristics of the spots, an algorithm was proposed that through setting Laplacian of Gaussian (LOG) operators of different sizes, the LOG cores of different scales were calculated to determine whether a pixel belongs to tongue spots; by extracting the gray level co-occurrence matrix

features (灰度共生矩阵特征) and the tongue coating ratio (舌苔比例) of the nine image blocks of the tongue, the thickness of the tongue coating was extracted. Then, random forest algorithm was used to train the recognition model for the severity degree of each tongue feature. The recognition accuracies for tongue colour, coating colour, teeth-mark, cracks, spots and thickness of tongue coating were 80.8%, 86.5%, 71.7%, 73.1%, 76.9% and 73.1% respectively. The downside was that although 12 color centers are selected in the CIExy chromaticity diagram, insufficient tongue samples resulted in a certain degree of bias in the selection process.

However, all the models above are only able to extract low-level features. According to Lai & Deng (2018), "these features lack representation ability for high-level problem domain concepts, and their generalization ability is rather poor." Deep learning model have been implemented in different field and have achieved excellent results in those application; but they have yet to be widely used in medical field. Deep learning models like neural network can "provide an effective way to construct an end-to-end model that can compute final classification labels with the raw pixels of medical images" (Lai & Deng, 2018).

Lai & Deng (2018) propose a deep learning model that integrates Coding Network with Multilayer Perceptron (CNMP), which combines high-level features that are extracted from a deep convolutional neural network and some selected traditional features. Inspired by deep convolutional neural network (CNN), Meng et al. (2017) propose a novel feature extraction framework called constrained high dispersal neural networks (CHDNet) to extract unbiased features. However, the tongue images in this paper are acquired under standard conditions with a box being designed to fix the camera position and illumination. Therefore, the performance of this network to extract features from phone-taking tongue images needs further justification.

2.6 Summary

Tongue diagnosis is all about the observation of the state of tongue. Through the observation, we can figure out the physiological changes of the human body, and then carry out the treatment. The existing tongue image acquisition system has strict requirements on the light source and the camera to reduce the colour difference between the captured tongue and real tongue.

However, the portability and popularity of these instruments are still poor, and thus improvement on current system is needed. Next, there is a variety of tongue segmentation algorithms: threshold method, edge detection method and clustering algorithm. With the development of artificial intelligence, artificial neural networks are becoming increasingly used in image segmentation. Existing algorithms are designed and proposed specifically to deal with tongues images that have been collected using specific tongue image acquisition system. Therefore, a tongue segmentation algorithm which is able to perform its job regardless of image brightness and the complexity of the background environment when an image is taken is needed. Although different colour correction algorithms and automatic feature extraction algorithms have been proposed, but most of them have been designed for purpose other than tongue diagnosis, therefore they have to be improved to be able to fit in tongue diagnosis application. Lastly, there is still a lack of complete system that comprises tongue image acquisition system, colour correction, automatic tongue segmentation using neural network and automatic feature extraction using neural network in a single system. Therefore, a complete system consists of all these components will be main focus of this project.

CHAPTER 3

METHODOLOGY AND WORK PLAN

3.1 Introduction

The proposed project is to design a predictive model for TCM tongue diagnosis using machine learning. The flowchart as shown in Figure 3.1.1 shows the design process of the project.

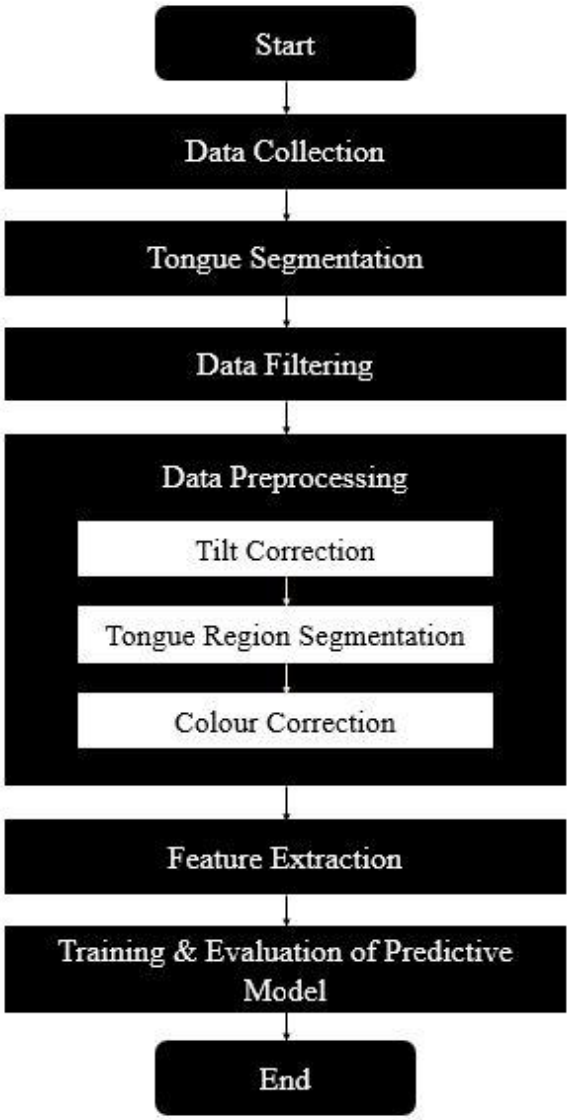


Figure 3.1.1 Design process of predictive model for TCM tongue diagnosis

3.2 Data Collection

Training data is the most important component of a machine learning project as it will be used to teach the model to learn patterns. There are a variety of open image datasets available online for machine learning research purpose. For example, MNIST which is used to classify handwritten digits and COCO which is widely used in object detection and image segmentation.

Unfortunately, there is no readily available dataset for this project. Therefore, tongue sample collection become one of the important part of this project. Sample collection is not an easy job but it is an imperative step for all supervised machine learning project. The dataset for this project should include images of both healthy tongues and unhealthy tongues. In machine learning, the performance of a trained model is also related to the size of dataset because a large dataset carries more information of each class, and thus the machine learning model can learn the features of each class better.

The performance of a machine learning model is also highly dependent upon the quality of dataset. Therefore, an excellent tongue image acquisition system has to be designed.

3.2.1 Design of Tongue Image Acquisition System

Among the computerized tongue diagnosis reports written by different researchers today, tongue images used to train and prove their algorithm are usually taken by the researcher himself through a tongue image acquisition system that had been designed for research purpose. Due to this, those who are interested to carry out computerized tongue diagnosis himself will lack the access to these tongue image acquisition system. Therefore, the improvements on popularity of the system is needed.

In this project, an easy way of taking tongue image by using mobile camera is proposed.

3.2.1.1 Justification of Suitability of Mobile Phone Camera in Taking Tongue Image

There are some skepticism towards the suitability of mobile phone camera alone as tongue image acquisition equipment. Therefore, the questions such as inconsistency in image size, brightness, tongue position and the loss of tongue

details due to low image resolution need to be justified to prove the qualification of phone-taking tongue image for subsequent computerized diagnosis

3.2.1.1.1 Inconsistency in Image Size

Different people have different mobile phone with different screen size. The design of a robust tongue diagnosis algorithm that could deal with tongue images of different image size is one of the important part of this project. Therefore, all tongue images have to be resized into same dimension before they go through subsequent operation. This could be easily solved by resizing the images programmatically.

OpenCV has cv2.resize() function which allows for 3 resize operations: with aspect ratio preserved, without aspect ratio preserved (resize only the width or height) and specific dimension (resize both width and height). For this project, all tongue images will be resized to 800×1000, with aspect ratio preserved, which means the resized image will be padded with zero either horizontally or vertically, so that the image size will all be 800×1000 before further processing.

3.2.1.1.2 Inconsistency in Brightness

The existing tongue image acquisition system has strict requirements on the light source and the camera. For example, Yiwu et al. (2002) uses a light with color temperature of 5300K and brightness of about 3100 Lux while Wei Baoguo et al. (2002) uses two Osram full-spectrum L18/72 Biolux D 6500 light with color rendering index, Ra = 96 and color temperature of 6500K.

The use of external light while taking tongue image is to ensure the consistency in image brightness and to reduce the colour difference between the captured tongue and real tongue. However, if the degree of colour difference between the captured tongue and real tongue for all tongue images taken using phone camera is the same, this will not cause any problem while training computerized tongue diagnosis model. For example, the real tongue looks red but it looks light red in the image; the real tongue looks purple but it looks red in the image. As long as the degree of colour difference is consistent, the colour

different between the captured tongue and real tongue will not affect much on the diagnosis result.

Today, most of the smartphones are equipped with LED flash. Taking tongue image with flash-on could ensure the consistency in image brightness. Figure 3.2.1 shows tongue image taken on the same person using iPhone 7 and Huawei Mate 10 Pro respectively. It can be clearly seen that the quality of phone-taking tongue image is not in the least inferior to that acquired by the system designed by Yang (2018), which is as shown in Figure 3.2.2. Although different mobile phones may equip with different flash, the degree of colour difference is very little and it could easily be corrected by colour correction algorithm.

Figure 3.2.1 Image taken using iPhone 7 (left) and Huawei Mate 10 Pro (right)

Figure 3.2.2 Image acquired by the system designed by Yang (2018)

3.2.1.1.3 Inconsistency in Tongue Position

There is a feature in most smartphone cameras that enables lines to be overlaid over the phone screen before taking photo, as shown in Figure 3.2.3. These lines

are known as grid lines which can be used as guidelines for taking phone, and they will not be shown on the photo.

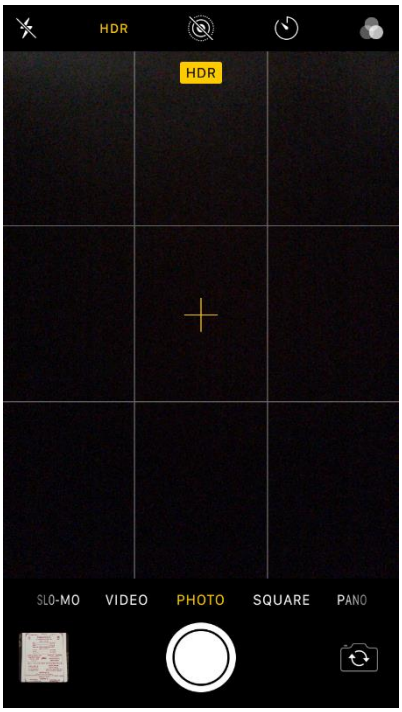


Figure 3.2.3 Grid lines on mobile phone camera

There is one important rule to follow while taking tongue image using mobile phone camera: the edge of the tongue have to be at about half of outer grids, as shown in Figure 3.2.4. By doing so, the inconsistency of tongue position and size in the image could be reduced.

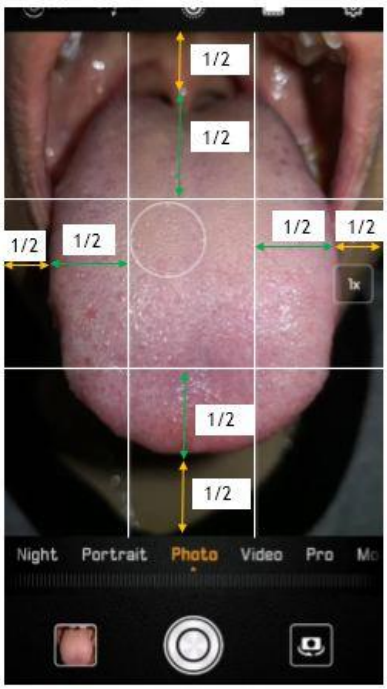


Figure 3.2.4 Grid lines as guidelines while taking tongue image

3.2.1.1.4 Loss of Tongue Details due to Low Image Resolution

Mobile phone camera technology has evolved rapidly, and it gets easier for us to get a high resolution photo with mobile phone camera. Figure 3.2.5 shows a tongue image taken using iPhone 5s rear camera. The details of the tongue, including tongue coating (舌苔) and texture (舌纹) can easily be observed from the image.

Figure 3.2.5 Tongue image taken using iPhone 5s

iPhone 5s was released by Apple Inc. in 2013. It is equipped with 8MP rear camera; today it is not rare to get a smartphone that is being equipped with camera of more than 8MP. Table 3.2.1 shows the report being released by Counterpoint on the best selling smartphones model for Q1 2018, Q3 2017 and Q2 2014 respectively.

Table 3.2.1 Counterpoint report on best selling smartphones model at different periods.

	Q1 2018		**Q3 2017**		**Q2 2014**	
	Model	**Camera resolution**	**Model**	**Camera resolution**	**Model**	**Camera resolution**
1	iPhone X	12 MP	iPhone X	12 MP	iPhone 5S	**8 MP**
2	iPhone 8 Plus	12 MP	iPhone 8	12 MP	Samsung Galaxy S5	16 MP
3	Redmi 5A	13 MP	iPhone 8 Plus	12 MP	Samsung Galaxy S4	13 MP
4	Oppo A83	13 MP	Samsung Galaxy Note8	12 MP	Samsung Galaxy Note3	13 MP
5	Samsung Galaxy S9	12 MP	iPhone 7	12 MP	iPhone 5C	**8 MP**
6	Samsung Galaxy S9 Plus	12 MP	Samsung Galaxy J7 Prime	13 MP	iPhone 4S	**8 MP**
7	iPhone 7	12 MP	iPhone 6	**8 MP**	Mi 3	13 MP
8	iPhone 8	12 MP	Vivo X20	12 MP	Samsung Galaxy S4 Mini	**8 MP**
9	Samsung Galaxy J7 Pro	13 MP	Oppo R11	20 MP	Xiaomi Redmi Note	13 MP
10	iPhone 6	**8 MP**	Galaxy S8 Plus	12 MP	Samsung Galaxy Grand 2	**8 MP**

As shown in the table, it can be seen that the smartphones are equipped with camera of at least 8 MP while most of them have camera of 12 MP. Therefore, the good quality of tongue image taken using mobile phone camera is unassailable.

3.2.1.2 Essential Rules to Follow When Taking Tongue Image

After the justification of the suitability of mobile phone camera in taking tongue image, there are five important rules (as shown in Figure 3.2.6) to follow when

taking a tongue image to ensure the quality of the image and to provide sufficient tongue information for further processing.

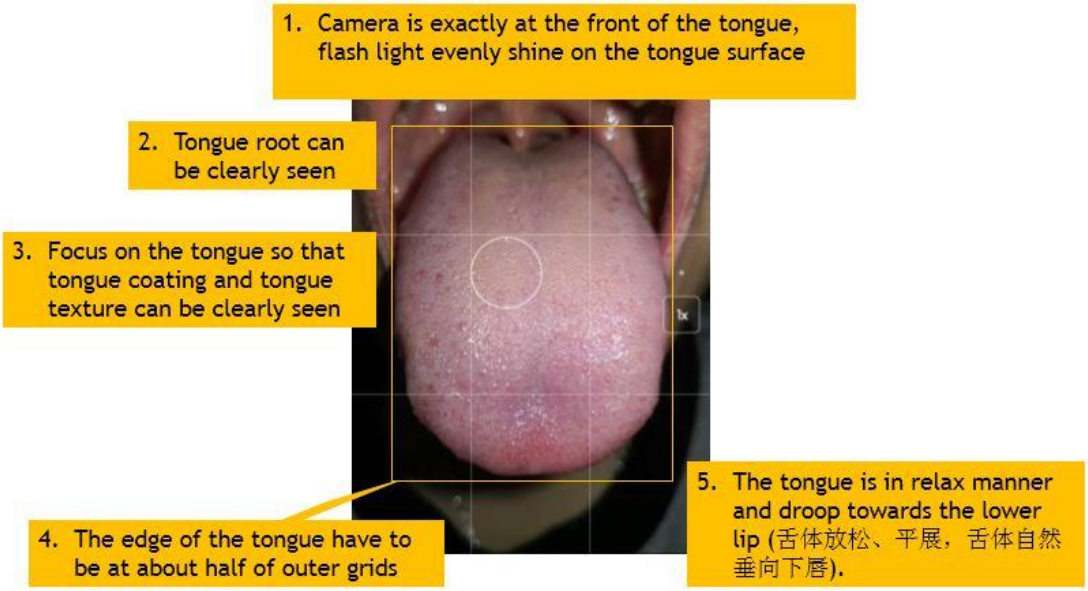


Figure 3.2.6 Five rules to follow when taking tongue image

In addition, there are some non-pathogenic factors that could lead to physical changes of tongue. For example, it is normal for someone who is just got up to have thick white coating on his/her tongue. Therefore, according to Wu (2011), in order to ensure the accuracy of diagnosis, there are few tips to take note:

- Don't take tongue image in one hour after someone is just got up
- Don't take tongue image in half an hour after a meal.
- Don't eat food that would colour the tongue
- Don't take tongue image in environment with colored lights
- Turn off beauty filter in mobile phone camera when taking tongue image
- Finish taking the image within one minute after someone has put his/her tongue out, otherwise the tongue color will change after some time

3.2.1.3 Record of Patient's Information

According to Liu Feilong (2014), the discoloration of a particular region of the tongue indicates a lesion in the human organ corresponding to that region. In the design process of computerized tongue diagnosis, the researchers who observe the tongue image are not able to have enough information about the patient's body conditions, so it was impossible to judge whether the

discoloration of the tongue is caused by non-pathogenic factors, and this could easily result in a wrong diagnosis.

Other than observation on the tongue, according to Wu (2011), there are other factors that are thought to be important influences in tongue diagnosis. Therefore, after taking the tongue image, the patient are required to complete a questionnaire comprising information as shown below:

- Age
- Gender
- Health condition

3.3 Tongue Segmentation

3.3.1 Types of Image Segmentation Algorithm

There are several types of image segmentation algorithm: traditional segmentation, semantic segmentation and instance segmentation.

"Before 2000, we used several methods in digital image processing: threshold segmentation, region segmentation, edge segmentation, texture features, clustering and so on. From 2000 to 2010, there are four main methods: graph theory, clustering, classification and combination of clustering and classification." (Eai Fund Official, 2018). These are some examples of traditional segmentation algorithm, which could separate the image pixels into different class without knowing what exactly each class is. In other words, if there is an image of cat and dog, these algorithm are able to group cat pixels into a cluster while grouping dog pixels into another cluster. However, they could not tell which cluster belongs to cat and which belongs to dog.

With the rise of machine learning and the development of artificial neural network, semantic and instance segmentation have been introduced. The main difference between semantic segmentation and image segmentation is that semantic segmentation algorithm are able to recognise the class of each cluster of image pixels, which means it is able to tell which cluster belongs to cat and which belongs to dog. One of the popular networks for semantic segmentation is Fully Convolutional Network (FCN).

Next, instance segmentation, as told by its name, is about instance recognition, instead of class recognition. In other words, if there are 2 dogs in an image, semantic segmentation algorithm will group them into a single cluster

as they belong to the same class; however, instance segmentation algorithm are able to distinguish one dog from another. The fundamental working principles of instance segmentation are to:

1. detect objects in the image and segment them into separate bounding boxes (object detection)
2. classify the class of object in each bounding box (semantic segmentation)

Therefore, the implementation of instance segmentation is much complex than semantic segmentation. However, an instance segmentation algorithm is needed for this project because there will be unpredictability about how complex is the background environment when a patient is taking tongue image using mobile phone camera. There is a possibility that the background contains something which looks like a human tongue. In order to avoid non-tongue region being segmented together with the tongue region, instance segmentation algorithm is used so that each instance will be grouped accordingly, and then the instance with highest recognition score will be selected as final detection and segmentation result.

3.3.2 Mask R-CNN

Mask RCNN is a neural network that is designed to perform instance segmentation and it will be implemented in this project to segment tongue from the image. “Mask R-CNN (Regional Convolutional Neural Network) has been the state-of-the-art model for object instance segmentation” which was proposed by He et al. (2017) and had won Best Paper at ICCV.

There are other algorithm such as DeepMask and Instance-sensitive FCN being proposed to perform instance segmentation; however, these algorithm is slow and with low accuracy as they try to implement segmentation first, and then followed by recognition/classification. Mask R-CNN is able to outperform all these algorithm with its powerful features as described below.

3.3.2.1 Backbone Architecture: ResNet-FPN

In Mask R-CNN, ResNet-FPN backbone is introduced. The backbone architecture is composed of Convolutional Neural Network (CNN), as shown in Figure 3.3.1, which is used for feature extraction. Each layer of the network is

responsible for the extraction of features at different level. For example, the preceding layers are responsible for the extraction of edges and corners which are referred to as low-level features.

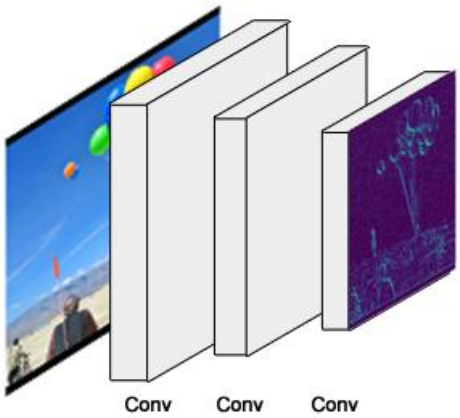


Figure 3.3.1 Illustration of backbone architecture (Zhang, 2019)

Mask R-CNN is said to be an extension of Faster R-CNN. ResNet backbone is used in Faster R-CNN as feature extractor. Additional Feature Pyramid Network (FPN) is introduced in Mask R-CNN as it is able to perform feature extraction for multi-scale objects by "adding a second pyramid that takes the high level features from the first pyramid and passes them down to lower layers" (Abdulla, 2018).

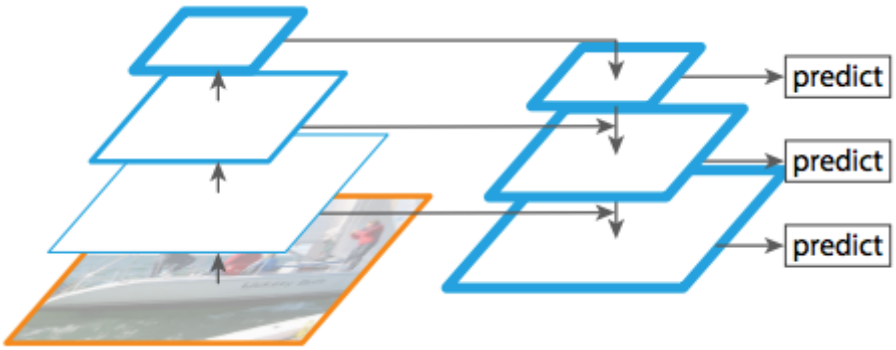


Figure 3.3.2 Illustration of FPN (Zhang, 2019)

Due to the structure of FPN (as shown in Figure 3.3.2), the features at each level are able to access to features at other levels. Therefore, the combination of ResNet and FPN as backbone makes Mask R-CNN to have better accuracy and speed than Faster R-CNN.

3.3.2.2 Head Architecture: Additional Mask Generation branch

Another important features that makes Mask R-CNN to outperform all other algorithms is that it performs prediction of segmentation masks and classes at the same time.

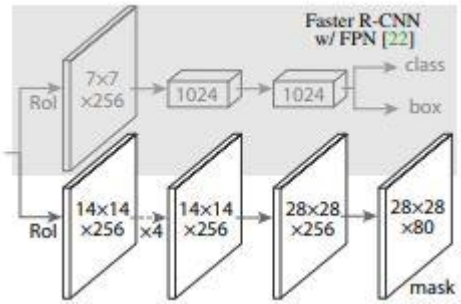


Figure 3.3.3 Head Architecture of Mask R-CNN (He, 2017)

Figure 3.3.3 shows the head architecture of Mask R-CNN. Grey area shows the head architecture for Faster R-CNN with FPN. Mask R-CNN extends Faster R-CNN with an additional fully convolutional branch (below the grey area) for segmentation mask prediction which is in parallel with the classification and bounding box regression branch in Faster R-CNN. According to He (2017), this additional mask branch is composed of Fully Convolutional Networks (FCN). Therefore, Mask R-CNN is able to perform classification and generation of segmentation mask at the same time. Furthermore, the additional branch does not cost much computational overhead; it is running at 5 fps.

3.3.2.3 RoIAlign

Next feature of Mask R-CNN would be the introduction of RoIAlign that has replaced the role of RoIPool in Faster R-CNN. According to He (2017), " RoIPool first quantizes a floating-number RoI to the discrete granularity of the feature map, this quantized RoI is then subdivided into spatial bins which are themselves quantized, and finally feature values covered by each bin are aggregated (usually by max pooling)".

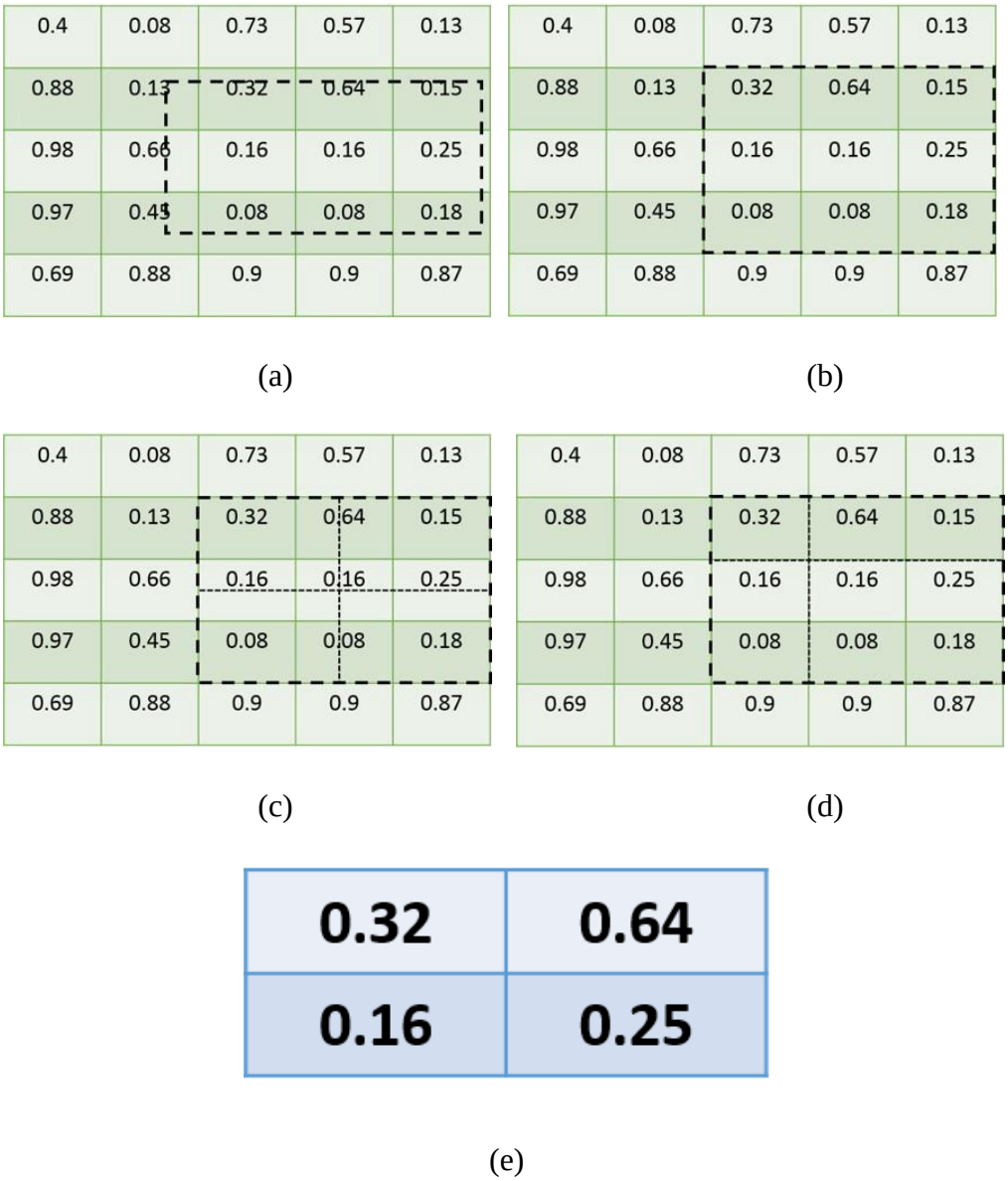
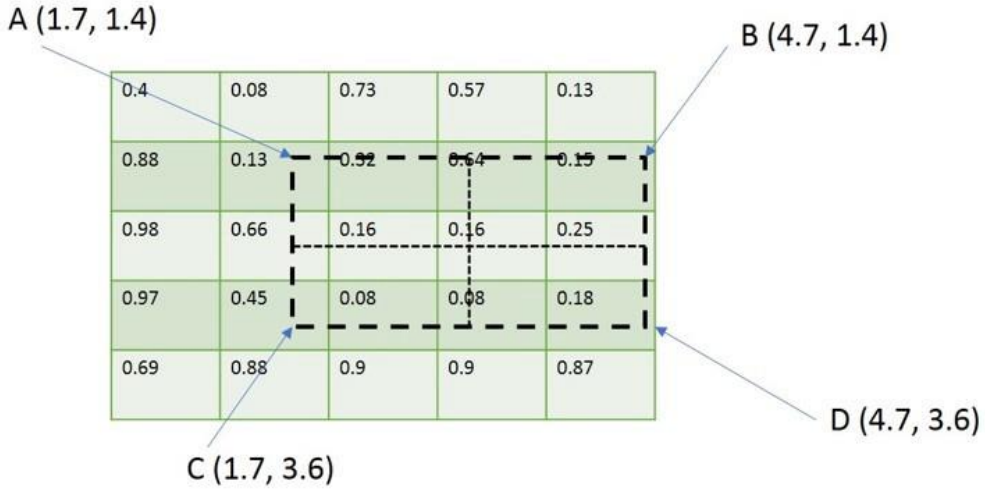


Figure 3.3.4 The working principle of RoIPool (Zhang, 2019)

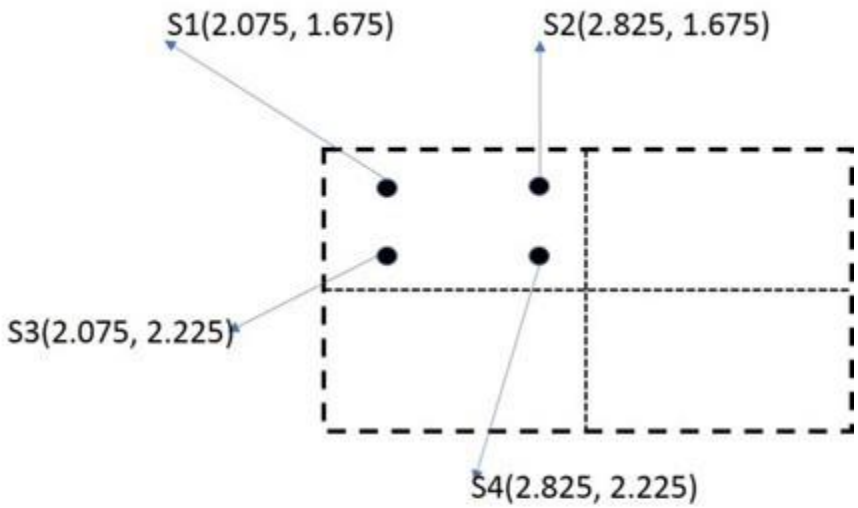
For example, as shown in Figure 3.3.4 (a), the boundaries of RoI doesn't match exactly with the boundaries of feature map box. In order to solve this, quantization is applied, as shown in Figure 3.3.4 (b). Next, after the subdivision of quantized RoI, the boundaries of each subdivision again doesn't match exactly with the boundaries of feature map box. Another quantization operation is applied and the final feature values obtained by RoIPool through max pooling is shown in Figure 3.3.4 (e).

The quantization causes the RoI to be misaligned with the extracted features; this misalignment can be simply ignored for classification task but not for instance segmentation that requires generation of segmentation masks.

Therefore, instead of quantization, bilinear interpolation is introduced in RoIAlign. The working principle of RoIAlign is explained below.



(a)

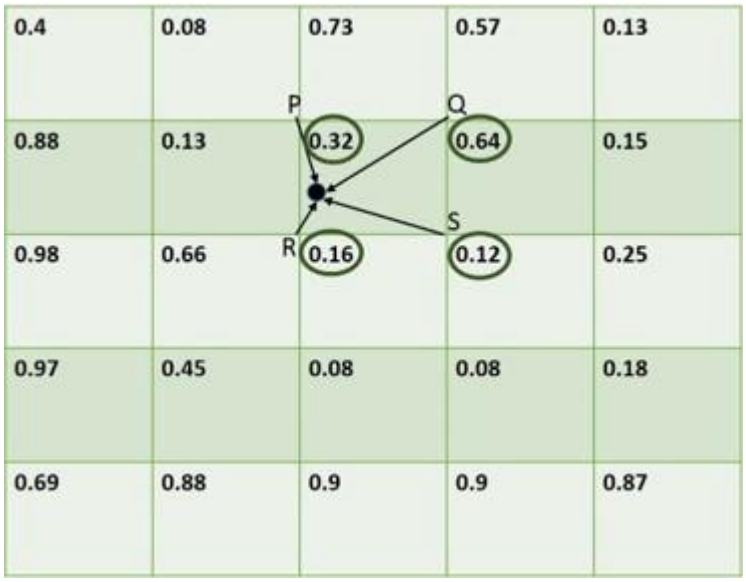


$$x = X_{low} + ((i + 0.5) * \frac{X_{high} - X_{low}}{numsamples}$$

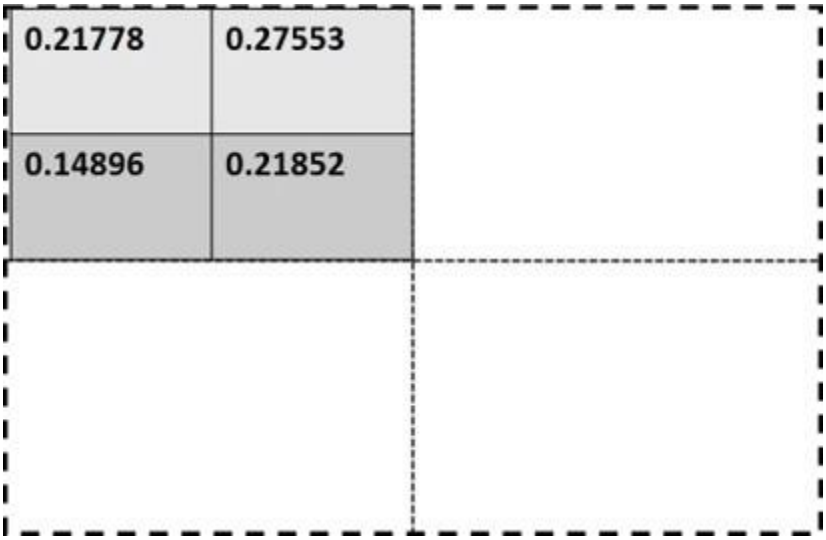
$$y = Y_{low} + ((j + 0.5) * \frac{Y_{high} - Y_{low}}{numsamples}$$

$$i, j \in [0,1 \ldots, num\ of\ samples)$$

(b)



(c)



(d)

0.2152	0.2335
0.3763	0.3562

(e)

Figure 3.3.5 The working principle of RoIAlign (Zhang, 2019)

Firstly, based on the location of RoI, four samples points are selected in each bin using the formula as shown in Figure 3.3.5 (b). After that, with the coordinates of four closest points on the feature map, bilinear interpolation is used to calculate the feature value of each sample point. As a result, there will be four values being generated in each bin as shown in Figure 3.3.5 (d), and then, the final feature values (as shown in Figure 3.3.5 (e)) will be obtained by performing average pooling on the values in each bin. Without the loss of precision due to quantization process, RoIAlign is able to perform prediction of pixel-accurate masks which is very important for instance segmentation tasks.

3.4 Data Filtering

The quality of the tongue image has direct impact on the result of tongue diagnosis; unqualified tongue images may lead to wrong diagnosis and cause unnecessary panic to the patient. Therefore, data filtering process is designed to reject those unqualified images and then request for a retake of qualified tongue image. Several tests have to be carried out before further tongue diagnosis process. The images that fail either one of the tests will be rejected immediately.

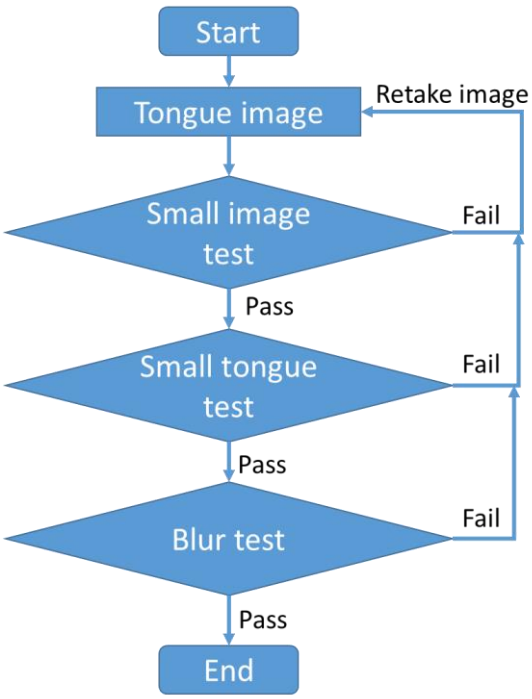


Figure 3.4.1 Flowchart of data filtering process

3.4.1 Small image test

Image resolution will determine the details that an image holds. If the image is compressed to a very small size, the details of the tongue will be lost, and thus there will be insufficient information for the subsequent process.

In section 3.2.1.1.4, it has been justified that the smartphone model used to take tongue image should have a camera of at least 8MP. The image resolution for 8MP camera is 3264×2448. By allowing the tolerance of 30%, the first test will be carried out to reject images with size smaller than 2200×1700.

3.4.2 Small tongue test

Mask R-CNN which is used as the tongue segmentation algorithm will perform the detection of tongue and generation of mask at the same time. Before segmenting the tongue from the image using the mask generated, a small tongue test should be carried out to evaluate the size of the tongue.

The tongue segmentation algorithm will output a bounding box containing tongue region, together with the mask of the tongue, as shown in Figure 3.4.2.

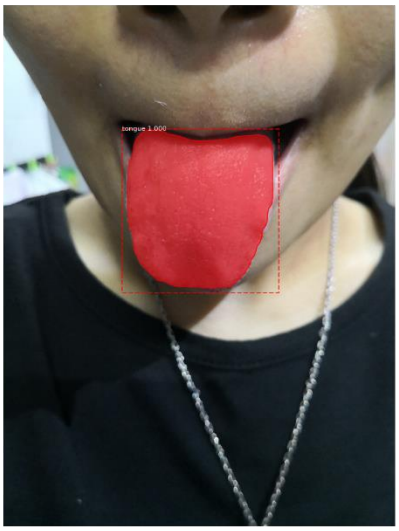


Figure 3.4.2 Result generated by tongue segmentation algorithm with bounding box (in dotted line) and red colour mask

With the bounding box, we could measure the size of the tongue in the image. As discussed in section 3.2.1.1.3, one of the rules for tongue image is that the edge of the tongue have to be at about half of outer grids. By allowing

some tolerance, a test will be carried out to reject the tongue which is smaller than $\frac{1}{3}$ of original image size. In other words, the tongue should be at least larger than the center grid (as shown in Figure 3.4.3).

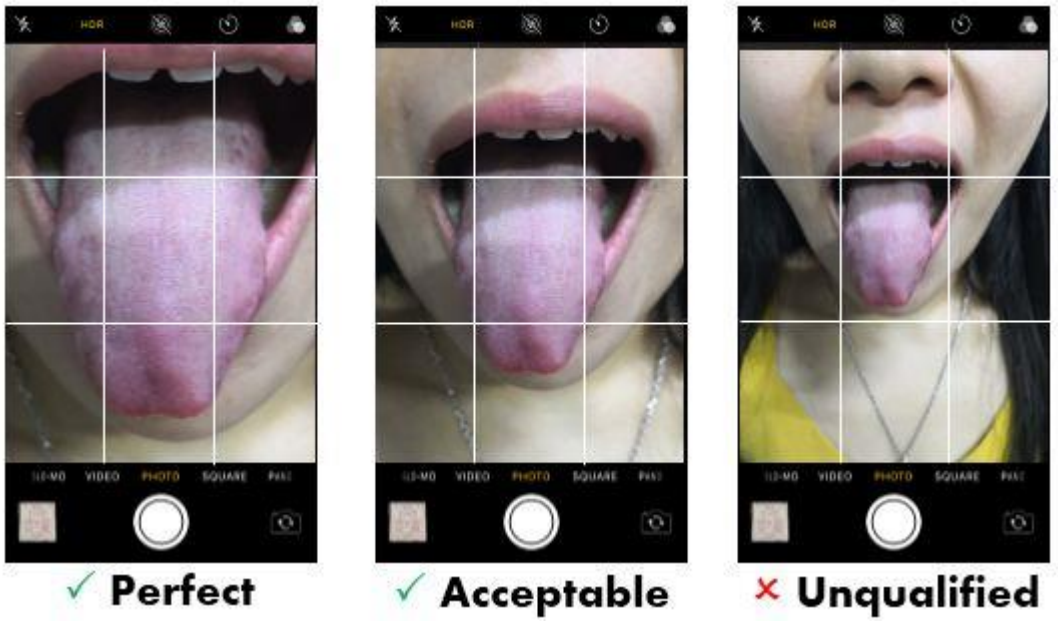


Figure 3.4.3 Qualified and Unqualified Tongue Size

3.4.3 Blur test

Due to the factors such as movement of the patient, shaking of the photographing device, or defocus, a blurry tongue image is obtained. Blurring is a kind of degradation of an image, which leads to the loss of image details, resulting in a decrease in the sharpness of the image, which has a serious impact on the quality of the image, as shown in Figure 3.4.4.

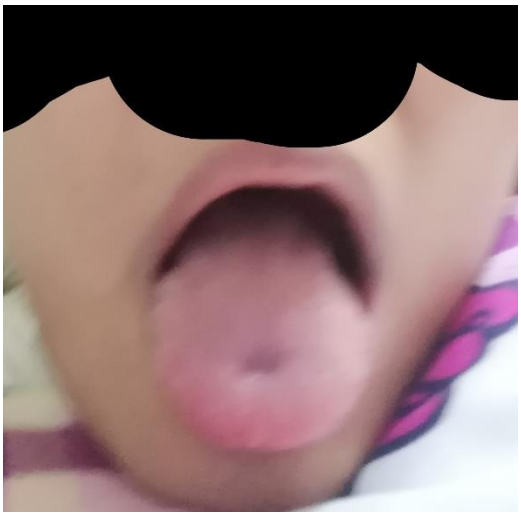


Figure 3.4.4 Example of unqualified blurry tongue image

Therefore, a test is carried out to reject blurry tongue images. Figure 3.4.5 shows the flowchart for blur test process.

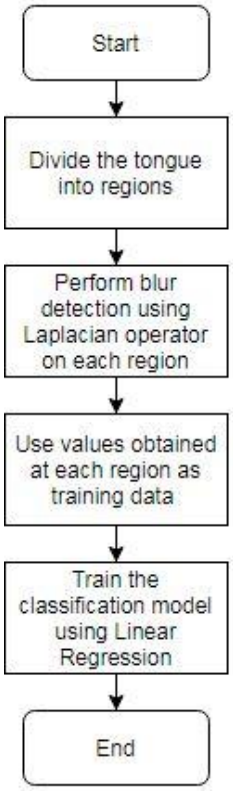


Figure 3.4.5 Flow chart for blur test process

There are two criteria being used to determine whether an image is clear:

1. The edge of the tongue is sharp
2. The tongue texture (舌纹) can be clearly seen

Therefore, instead of performing blur detection directly on the whole image, the blur test starts with dividing the image into 5 regions as shown in Figure 3.4.6. The first criteria will be evaluated on region 1-4 while the second criteria will be evaluated on region 5. After that, Laplacian operator will be used to detect blur for each region.

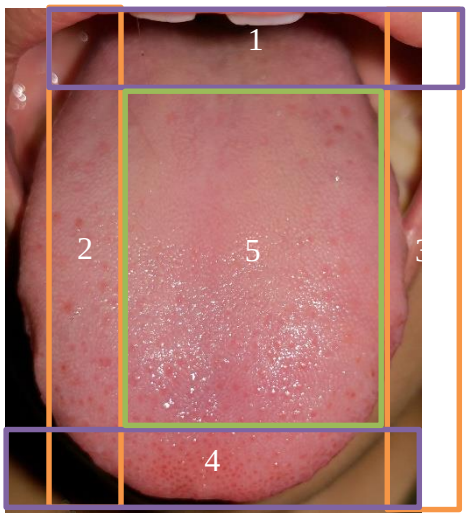


Figure 3.4.6 Division of tongue image into region to perform blur test

Laplacian operator is used in the measurement of second derivative of an image. It could determine whether intensity in a particular region changes rapidly; it is also used to detect edges. " The assumption here is that if an image contains high variance then there is a wide spread of responses, both edge-like and non-edge like, representative of a normal, in-focus image. But if there is very low variance, then there is a tiny spread of responses, indicating there are very little edges in the image" (Bhamidipati, 2018). A blurry image usually has very little edges. The variance can then be used as a measure to detect blurriness. Laplacian operator is applied to all five regions of the image; the values obtained from each region are combined as a feature vector.

The feature vectors are collected as training data to be fed to a machine learning model to classify blurry and clear images. There are several supervised machine learning algorithm: Linear Regression, Logistic Regression, K Nearest Neighbors (KNN), Decision Trees (DT), Support Vector Machine (SVM) and Random Forest. However, since the blur test involves only linear binomial classification problem and there are only five features in the training data, Linear Regression is chosen to carry out the blur test. Linear Regression is designed to deal with linear problem and it is the simplest machine learning algorithm. Therefore, Linear Regression is capable to perform the classification of blurry and clear image efficiently. In order to further justify the performance of Linear Regression, each machine learning algorithm mentioned above is trained for three times with their performance being tabulated in the table below.

Table 3.4.1 Comparison of Performance of Different Machine Learning Algorithms

Machine Learning Algorithm	Average Accuracy	Average Processing Time (ms)
Linear Regression	**1.0000**	**1.33**
Logistic Regression	**1.0000**	40.34
KNN	0.9667	16.67
Decision Trees	0.8278	3.67
SVM	**1.0000**	74.34
Random Forest	0.9556	29.33

According to Table 3.4.1, Linear Regression outperforms other algorithms in terms of accuracy and processing time. It has the shortest processing time because other algorithms have much complex structure as they are designed to solve much complex problem. However, blur test doesn't involve any complex problem, therefore Linear Regression is chosen to train the classification model for blur test. During blur test, only clear images are qualified for subsequent processing while blurry images will be rejected immediately.

3.5 Data Preprocessing

3.5.1 Tilt correction

According to Liu Feilong (2014), discoloration of a particular region of the tongue indicates a lesion in the human organ corresponding to that region. Tongue tip relates to lungs and heart; tongue center relates to spleen and stomach; tongue margin relates to liver; tongue root relates to kidney. Therefore, the tongue regions have to be identified and segmented.

A slanted tongue as shown in Figure 3.5.1 will make the tongue region segmentation work difficult. Therefore, tilt correction has to be carried out on slanted tongue before region segmentation is performed.

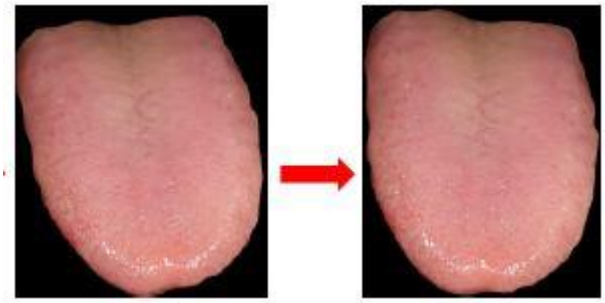


Figure 3.5.1 Tilt correction on the tongue

Figure 3.5.2 shows the flowchart for tilt correction process. The tongue center line has first to be determined, and then perform rotation according to the slope angle of the center line.

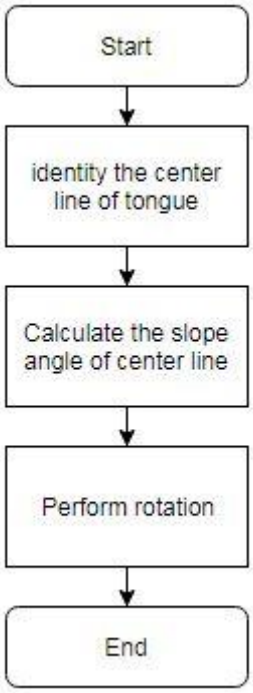


Figure 3.5.2 Flowchart for tilt correction process

In order to identify the center line of a tongue, first we extract parts of the image as shown in Figure 3.5.3 (a), so that the left and right edges of the tongue can be analyzed separately.

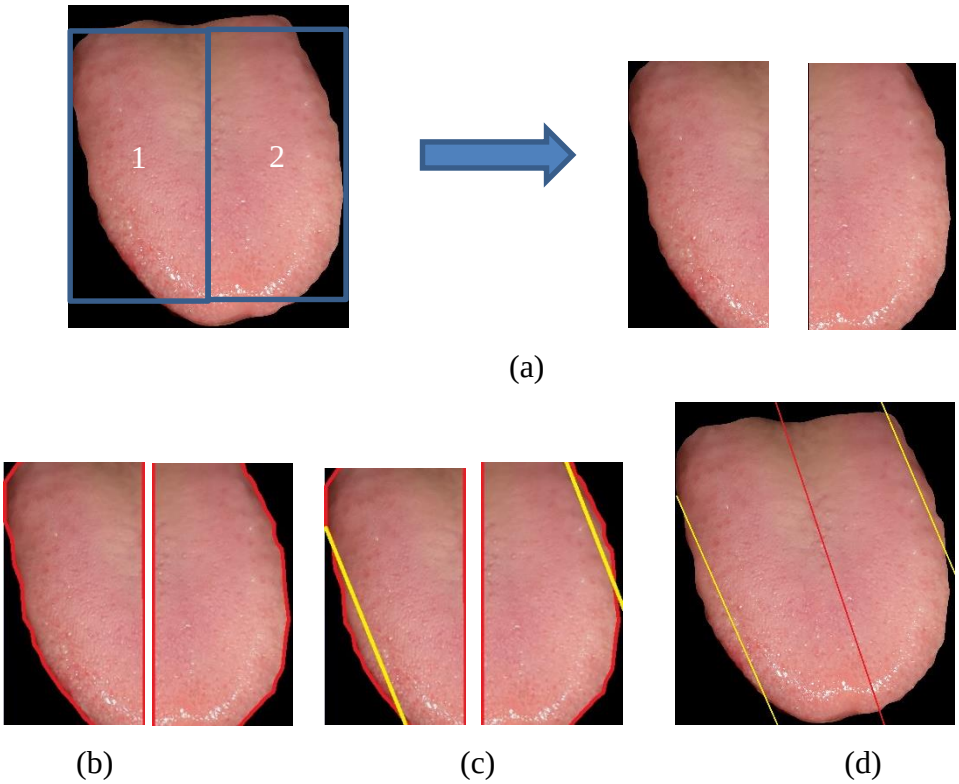


Figure 3.5.3 Process to identify center line of tongue

Next, Suzuki Boarder Following algorithm will be applied to find the contour (red lines in Figure 3.5.3 (b)) at both part of the tongue. According to Suzuki et al. (1985), there are several important definitions used in Suzuki Boarder Following algorithm (as illustrated in Figure 3.5.4) to find contour of the object:

1. The four outermost lines of pixels form the frame; assume the pixels of frame are '0', all the '0' pixels connected to the frame will be taken as background while those unconnected '0' pixels will be taken as hole.
2. If any pixels in neighborhood of pixel '1' is 0, then the pixel '1' will be taken as boarder point, B between two regions (region with '1' pixels, S1 and region with '0' pixels, S2).
3. Next, it is concluded that S1 surrounds S2 "if there exists a pixel belonging to S2 for any 4-path from a pixel in S1 to a pixel on the frame S1 and S2 are connected" (Suzuki, 1985); S1 surrounds S2 directly if there is border point between the two region.
4. An outer border is where the region with '1' pixels is surrounded directly by the region with '0' pixels while a hold border is where

the region with ‘0’ pixels is surrounded directly by the region with ‘1’ pixels.

5. Assume region A with pixels ‘0’ surrounds region B with pixels ‘1’, the border between region A and B is an outer border. The parent border of the outer border is the hole border between region A and region C with pixels ‘1’ which surrounds region A, or the parent border of the outer border is the frame if region A is background. For example, S4 in Figure 3.5.4 is surrounded by S3 which is a hole, then the parent border of B3 is B2; S2 in Figure 3.5.4 is surrounded by S1 which is background, then the parent border of B1 is the frame.

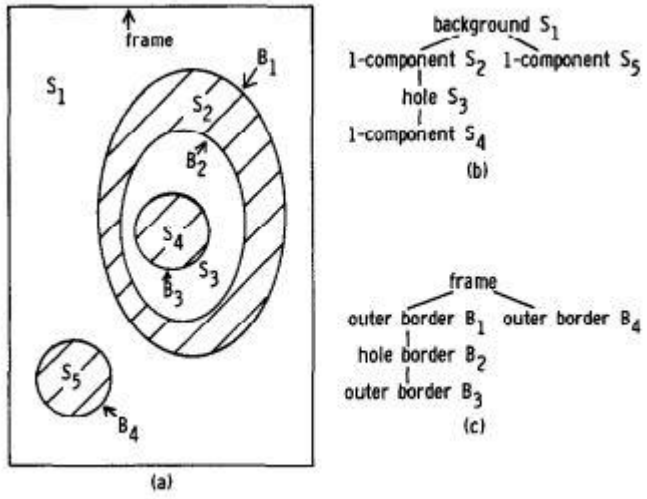


Figure 3.5.4 Surroundness among connected components (b) and among borders (c); plain areas refer to ‘0’ pixels while shaded areas refer to ‘1’ pixels. Adapted from Suzuki et al. (1985).

With the definitions above, different borders in the image can be identified easily. Figure 3.5.5 shows the requirements for a point *(i, j)* to be taken as starting point for outer border and hole border.

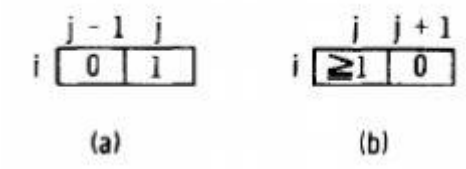


Figure 3.5.5 The conditions of the border following staring point *(i, j)* for an outer border (a) and a hole border (b). Adapted from Suzuki et al. (1985).

A unique index number will be assigned to each newly discovered border while a negative value of that index will indicate the last pixel of border

in a particular row. For example, the pixels on B1 in Figure 3.5.4 will be assigned index of '1' while the pixels on B2 will be assigned index of '2'; the rightmost pixel of B1 in each row will be assigned '-1' while the rightmost pixels of B2 in each row will be assigned '-2'. Lastly, the largest border will be taken as the contour of the tongue.

After that, total least-squares method will be used to fit a straight line (yellow lines in Figure 3.5.3 (c)) onto the contour. A straight line can be represented in the form of $ax + by + c = 0$. The working principle of total least-squares method is to get the minimum summation value of $|ax_i + by_2 + c|^2$, where x_i and y_i are the coordinates of each point along the contour and the selection of a and b values have to fulfil the equation $a^2 + b^2 = 1$. The line obtained using the total least-squares method is the line with minimum distance to all points on the contour, and thus the best fit line for the edge of tongue is obtained.

With the two straight line identified at the edge of tongue, the center line (red lines in Figure 3.5.3 (d)) of the tongue could be obtained by the calculating the midpoint of each point on the lines. The formula for calculating midpoint between a point on the straight line of left edge, (x_1, y_1) and a point on the straight line of right edge, (x_2, y_2) is shown below.

$$Midpoint = (\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2})$$

After that, with any two points on the center line of tongue, the slope angle can be easily obtained through the formula shown in Figure 3.7.4, where $length = y_2 - y_1$ and width$= x_2 - x_1$.

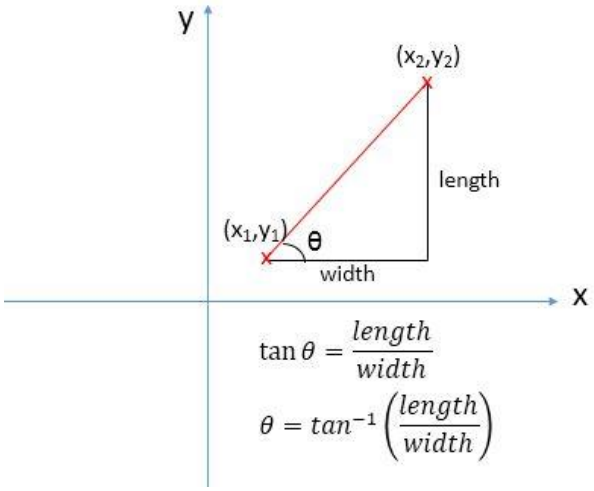


Figure 3.5.6 Calculation of slope angle

Lastly, the slanted tongue can be corrected by rotating the tongue image according to the slope angle obtained.

3.5.2 Tongue region segmentation

After the tilt correction, an upright tongue image is obtained. The tilt correction will eventually ease the subsequent tongue region segmentation process, as the tongue can be easily divided into regions based on the ratio as shown in Figure 3.5.7.

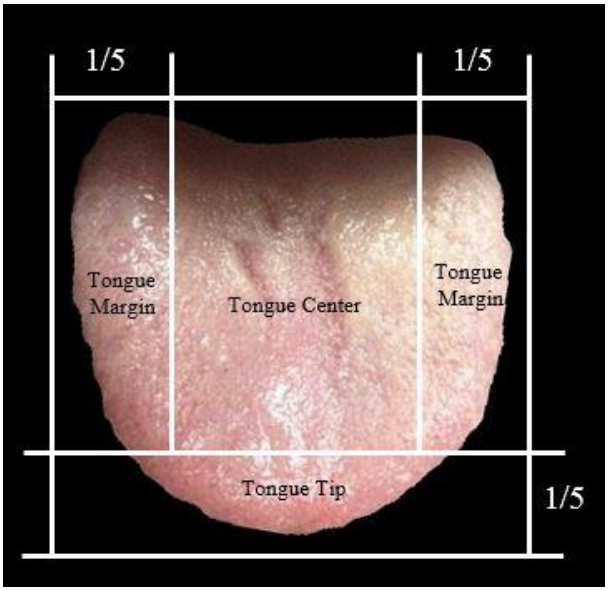


Figure 3.5.7 Tongue regions with their ratios. Adapted from Xu Jiayu (2009).

3.5.3 Colour Correction

Images taken using different mobile phone will have problem of device-dependent colour space rendering because the image colour information will depend on the imaging specification of the camera. Besides, some noises will be generated together with the images due to variation in environment lighting.

The objective of this section is to study on whether a colour correction stage is necessary in this tongue diagnosis system and whether it will help to improve the accuracy of the predictive model. There are 3 important criteria that a good colour correction algorithm is expected to achieve:

- colour constancy in tongue images under different illumination, where after colour correction is applied, images taken with flash and without flash should look the same in terms of contrast and brightness
- Tongue details should not be lost and also the corrected tongue colour should not deviate greatly from the original tongue colour, where a TCM physician should be able to recognise the same tongue features as in original tongue images from the corrected tongue images
- Corrected images should achieve higher accuracy in feature extraction (subsequent section), as compared to original images

Different colour correction techniques, such as Histogram Equalization and Retinex will be applied to the tongue images to find out whether a colour correction stage is necessary in this tongue diagnosis system and whether it will help to improve the accuracy of the predictive model.

3.5.3.1 Histogram Equalization

According to Xiao (2015), histogram equalization alters the brightness of colors so that the histogram of an image will be equalized, and so that the image has high dynamic range and shows more details.

Assume the pixel values of an image, *I* can go from *0 to L-1*, which *L* equals to *256* for an 8-bit image. The histogram of the image can be obtained through the formula below.

$$p_r(r_k) = \frac{n_k}{n}, k = 0,1, \dots, L-1,$$

where $p_r(r_k)$ is the probability of occurrence of pixel value r_k, n_k is the number of occurrence of pixel value r_k, and n is the total number of pixels in the image. After that, the cumulative frequency plot can be obtained using the formula below.

$$s_k = T(r_k) = \sum_{j=0}^{k} p_r(r_j) = \sum_{j=0}^{k} \frac{n_k}{n}, k = 0,1, \dots, L-1,$$

where $T(r_k)$ is the transformation function from r_k to new pixel values, s_k. So the new pixel value for output image of histogram equalization can be obtained using $T(r_k)$.

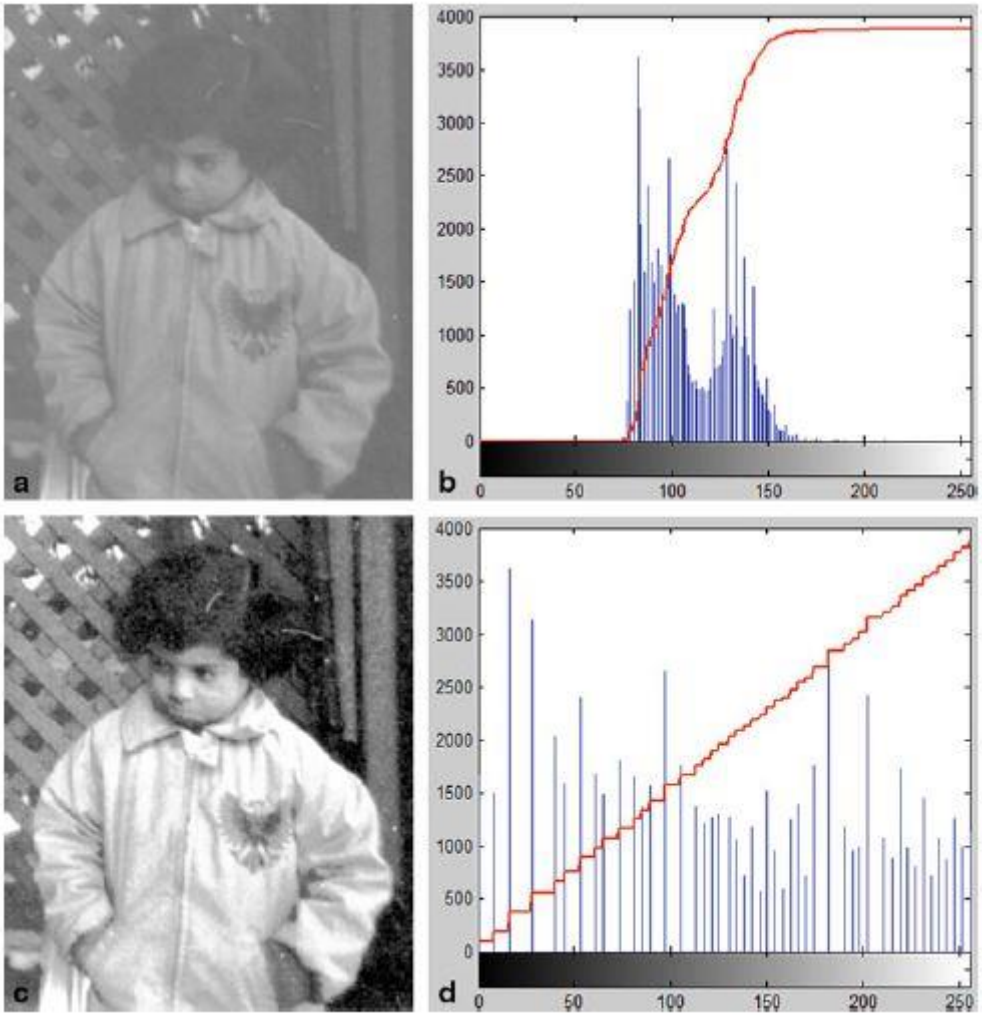


Figure 3.5.8 (a) original image, (b) corresponding histogram (blue) and cumulative frequency plot (red), (c) image corrected using HE, (d) corresponding histogram (blue) and cumulative frequency plot (red)

Figure 3.5.8(a) is a black-and-white image of a kid while Figure 3.5.8(b) shows its intensity histogram (blue) and its cumulative frequency plot (red). The

original image is a low contrast image and the pixels are concentrated in the middle of the histogram. After histogram equalization is applied, the histogram spreads out and is more evenly distributed and the cumulative frequency plot looks like a straight line, as shown in Figure 3.5.8(d). The output image is much brighter as compared to the original image.

Histogram equalization can be applied on colour image by first converting the image into HSV color space, then applying HE only on either the luminance or value channel.

3.5.3.2 Retinex

"The retinex theory was developed by Land and McCann (1971) to model how the human visual system perceives a scene. They established that the visual system does not perceive an absolute lightness but rather a relative lightness, namely the variations of lightness in local image regions." (Petro, et al., 2014)

According to Jobson et al. (1997), there are 3 important properties of color constancy algorithm:

a. dynamic range compression,
b. color independence from the spectral distribution of the scene illuminant
c. color and lightness rendition

Each image pixel can be described mathematically as below:

$$S(x, y) = R(x, y) * L(x, y),$$

where *L* represents illuminance, *R* represents reflectance and *S* represents the image pixel. According to Gonzalez & Woods (n.d.), the second property of color constancy algorithm can be achieved by removing the illuminance component while the first property can be achieved through logarithmic transformations.

Let $s = \log(S)$, $r1 = \log(R)$, $l = \log(L)$, equation above becomes

$$r1(x, y) = s(x, y) - l(x, y)$$

L can be obtained by convolving a low pass filter *F* with image *S*. The low pass filter function *F* proposed by Land (1986) doesn't satisfied the third property. Therefore, to solve this issue, Jobson et al. (1997) proposed Single Scale Retinex (SSR).

The general mathematical form of a Single-Scale Retinex (SSR) is given by

$$R_i(x,y) = \log\big(I_i(x,y)\big) - \log(I_i(x,y) * F(x,y)),$$

where I_i is the input image on the *i-th* color channel, R_i is the retinex output image on the *i-th* channel and *F* is the normalized surround function. However, Jobson et al. (1997) replaced the normalized surround function with a Gaussian function as shown below:

$$F(x,y) = Ce^{-\frac{(x^2+y^2)}{2\sigma^2}},$$

where σ, the filter standard deviation, controls the amount of spatial detail which is retained, and C is a normalization factor such that $\int F(x,y)\,dxdy = 1$.

Choosing the most suitable value for σ is important in SSR as it could affect both dynamic range compression and color rendition. Multiscale Retinex (MSR) was then proposed as it provides better trade-off between dynamic range compression and color rendition. MSR formula proposed by Jobson et al. (1997) is given by

$$R_{MSR_i} = \sum_{n=1}^{N} \omega_n R_{n_i} = \sum_{n=1}^{N} \omega_n [log\ I_i(x,y) - \log(F_n(x,y) * I_i(x,y))]$$

where *N* is the number of scales, ω_n is the weight of each scale and $F_n(x,y) = C_n e^{-\frac{(x^2+y^2)}{2\sigma^2}}$. Jobson et al. (1997) experimentally determined $N = 3$ and $\sigma_1 = 15, \sigma_2 = 80, \sigma_3 = 250$.

3.5.3.2.1 Multiscale Retinex with Color Restoration (MSRCR)

Petro et al. (2014) states that "given an image with sufficient amount of color variations, the average value of the red, green and blue components of the image should average out to a common gray value." This is known as gray-world assumption. For images which do not follow gray-world assumption, MSR output will be grayish images by reducing the saturation of dominant color in these images. Therefore, Jobson et al. (1997) proposed Multiscale Retinex with Color Restoration (MSRCR) which will multiply MSR output by a color restoration function of the chromaticity. MSRCR formula is given by

$$R_{MSRCR_i}(x,y) = G[C_i(x,y)R_{MSR_i}(x,y) + b],$$

where *G* and *b* are final gain and offset values while $C_i(x,y)$ is the *i*-th band of the color restoration function (CRF)

$$C_i(x,y) = \beta \log[\alpha I_i'(x,y)],$$

where β is a gain constant, α controls the strength of the nonlinearity, and $I_i'(x,y)$ is chromaticity coordinates for the i-th color band which is given by,

$$I_i'(x,y) = \frac{I_i(x,y)}{\sum_{j=1}^{S} I_j(x,y)},$$

where S is the number of spectral channels. Jobson et al. (1997) experimentally determined $\alpha = 125, \beta = 46, b = -30$ and $G = 192$.

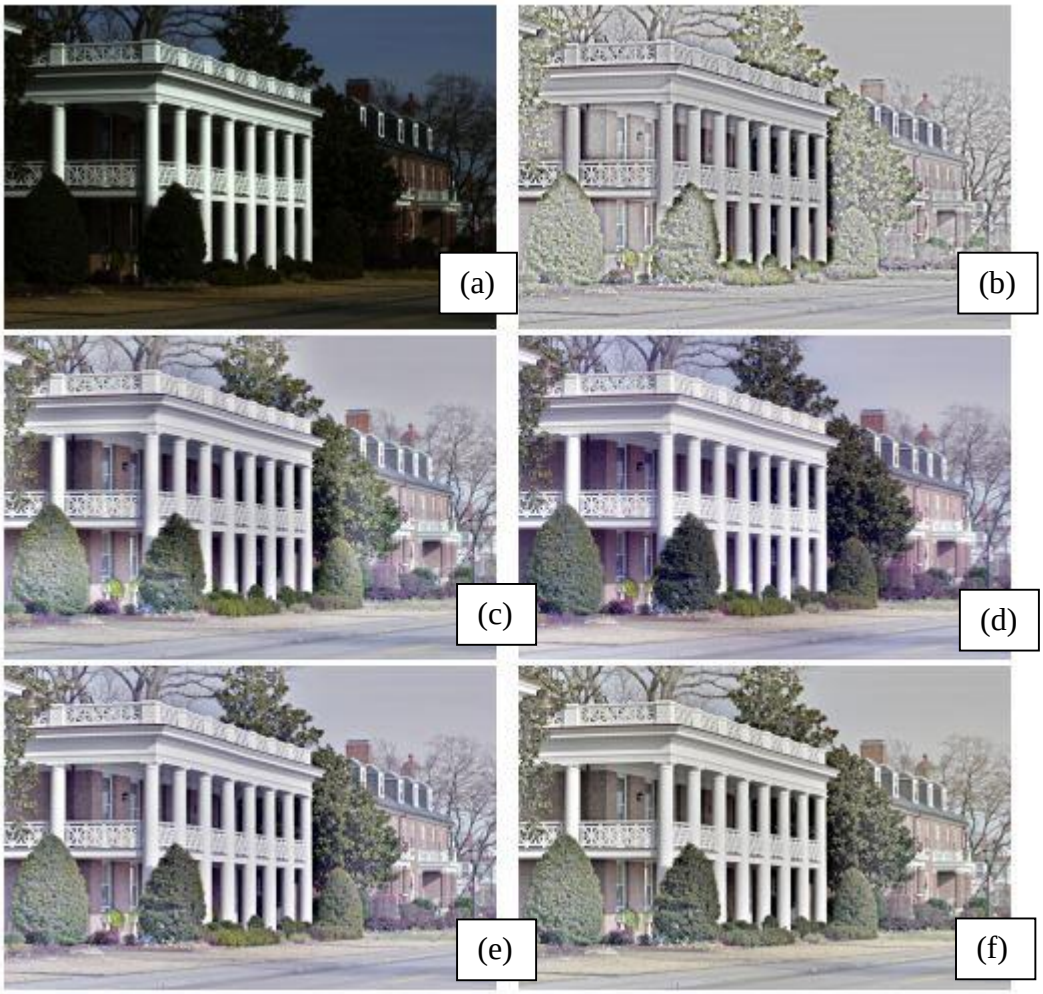


Figure 3.5.9 (a) original image (b) SSR with σ = 15 (c) SSR with σ = 80 (d) SSR with σ = 250 (e) MSR (f) MSRCR

3.5.3.2.2 Automated Multiscale Retinex with Color Restoration (Am-MSRCR)

However, it was found that even after MSRCR is applied, "greying -out" still happens in some output images. Jobson et al. (1997) proposed canonical gain/offset method to solve this problem on SSR.

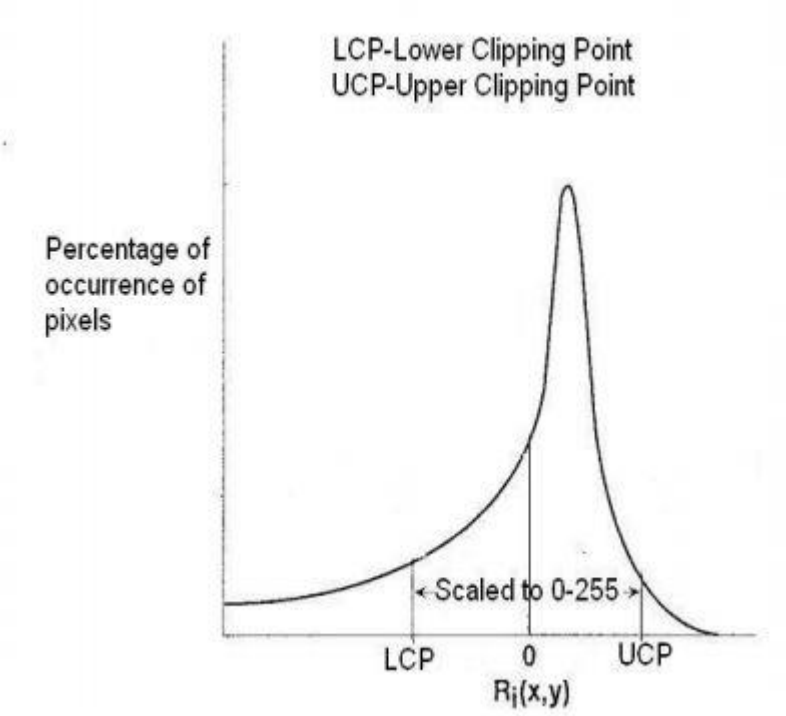


Figure 3.5.10 Histogram of SSR enhanced image

As shown in Figure 3.5.10, canonical gain/offset method causes some of the largest and smallest signal values being clipped, but fortunately these values do not carry much information. However, Jobson et al. (1997) in their paper, do not specify the method to decide the lower clipping and upper clipping point.

Parthasarathy (2012) propose an automated method to determine the lower and upper clipping points by "using the frequency of occurrence of pixels as a control measure."

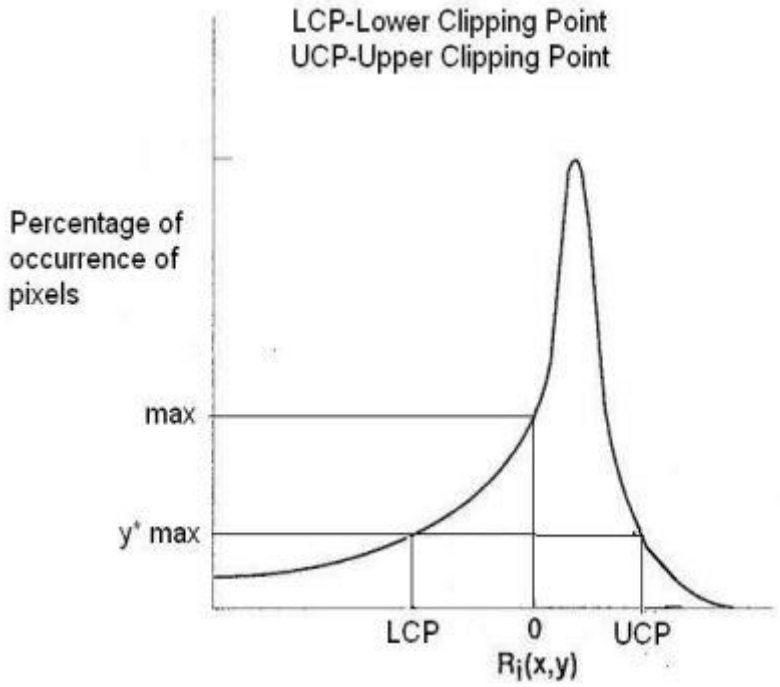


Figure 3.5.11 Histogram with clipping points chosen based on frequency of occurrence of pixels

First, let the frequency of occurrence of pixel value *'0'* be *'max'*, then the lower and upper clipping points can be obtained by multiplying *y* with *max*. Parthasarathy (2012) experimentally fixed y = 0.05, that is 5% of pixels on either end of the histogram are clipped. According to Parthasarathy (2012), "we found that better outputs are obtained if we apply this technique after MSR instead of applying at SSR stage itself, reason being that this approach is non-linear in nature." Besides, applying it once at MSR stage will improve computational speed, as compared to applying it at SSR stage for three times.

3.5.3.2.3 Multiscale Retinex with Chromaticity Preservation (MSRCP)

Also, it was found that some MSRCR output may have the problem of inverting color, that is, pixel values near 0 may go to values near 255 and vice versa.

Figure 3.5.12 (a) original image (b) MSR (c) MSRCR

For example, the blue ball at bottom left corner in Figure 3.5.12 (a) will have pixel value near '0' in red channel. The red channel of MSR for these pixels will be negative and CRF function will also be negative. Thus, "the red channel of MSRCR for these pixels becomes positive and their values are changed by the postprocessing step into a value higher than the image average" (Petro, et al., 2014). Therefore, the ball is magenta in colour in MSRCR output.

Multiscale Retinex with Chromaticity Preservation (MSRCP) was proposed to solve this problem by applying MSR only on intensity channel. The intensity channel formula is given by

$$I = \frac{\sum_{j=1}^{S} I_j}{S},$$

where S is the number of channels. MSR formula is then applied I to obtain R_I. "Then a linear transformation is applied to the intensity output to stretch the result to [0, 255]" (Petro, et al., 2014). While keeping the chromaticity the same as in the original image, the color channels are given by

$$R_i = I_i \frac{R_I}{I}$$

Figure 3.5.13 (a) MSRCR (b) MSRCP

In this project, the performance of Historgram Equalization, MSRCR, Automated MSRCR and MSRCP in correcting the colour of tongue images will be investigated and compared.

3.6 Feature Extraction

There are two types of feature extraction algorithms: manual and automatic feature extraction. Manual feature extraction refers to those algorithms designed specifically to extract a particular feature, such as using HSI (hue, saturation, intensity) model to distinguish tongue body (舌质) and tongue coating, establishing bisection light reflection model (二分光反射模型) to detect and identify tongue moisture, etc. Automatic feature extraction involves the use of deep learning models. Deep learning models like neural network can "provide an effective way to construct an end-to-end model that can compute final classification labels with the raw pixels of medical images" (Lai & Deng, 2018). Therefore, neural network will be used in this project to extract the tongue features.

There are several tongue features which a TCM practitioner usually observes during tongue diagnosis, as shown in Table 2.5.1. However, this project will only focus on 4 features which are greasy tongue coating (腻苔), teeth-marks (齿痕), cracks (裂纹), and spots (点刺). The reason for these features to be chosen is that these 4 features are the most common features among the tongue images collected, where we could easily get more than 50 images for each features which will then be used as training dataset for neural network.

There are a lot of neural networks that can be used as feature extraction algorithms, such as CNN, Region-based CNN (R-CNN), Fast R-CNN, Faster R-CNN, Mask R-CNN and YOLO (You only look once). In this section, two neural networks will be trained to extract tongue features and their performance will be compared.

CNN is designed to detect only one object at a time while R-CNN can detect multiple objects. However, R-CNN is very slow as it divides the image into 2000 regions and then perform feature extraction for each region. Fast R-CNN is then proposed to solve this problem by using a single deep ConvNet. Later on, Faster R-CNN further improves the processing speed by using Region Proposal Network (RPN) along with Fast R-CNN. Mask R-CNN extends Faster R-CNN with "an additional branch for predicting segmentation masks on each

Region of Interest (RoI) in a pixel-to pixel manner" (Khandelwal, 2019). Its architecture has been discussed in Section 3.4.2.

3.6.1 YOLO

Another popular neural network would be YOLO. According to Khandelwal (2019), "YOLO trains on full images and directly optimizes detection performance. This makes YOLO extremely fast and accurate."

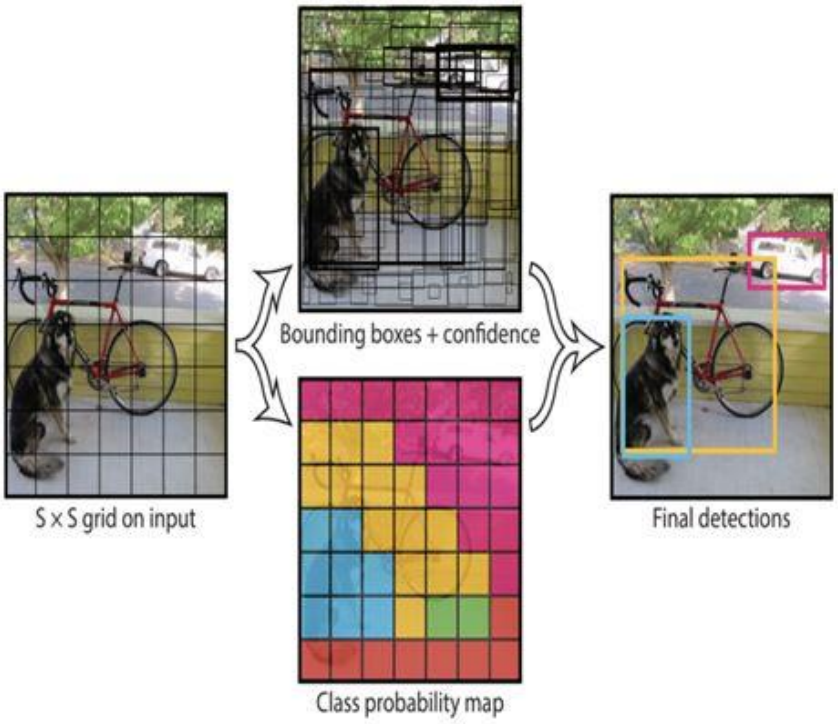


Figure 3.6.1 The working of YOLO

As shown in Figure 3.6.1, YOLO will first divide the input image into grids and each grid will be responsible to predict one object only. YOLO will then apply image classification and localization on each grid (Khandelwal, 2019). The grid where the center of an object falls onto will be responsible for object detection. Lastly, each grid will predict the bounding boxes with their corresponding confidence scores. Since each grid will carry out their jobs together, this makes YOLO able to perform prediction for several bounding boxes at the same time, resulting in a fast and accurate algorithm. Unlike the region-based techniques used in R-CNN and Fast R-CNN, YOLO is trained on entire image, therefore "it implicitly encodes contextual information about

classes as well as their appearances" (Khandelwal, 2019). This is why YOLO makes less background errors and is more accurate than other neural networks.

In this project, Mask R-CNN and YOLO will be trained for tongue feature extraction. Their performance will be compared and the better one will be implemented later. The YOLO version which will be used is YOLOv3 which employs Darknet-53 network to perform feature extraction. As the name implies, Darknet-53 has 53 convolutional layers.

3.7 Training and Evaluation of Predictive Model

Disease prediction is the last but also the most important part of this project. The disease to be predicted in this stage will be chosen after the sample data being collected is tabulated. This is to ensure that the number of sample data for a particular disease is enough to be fed into the machine learning algorithm for training.

The input feature vectors for the machine learning algorithm will have 7 elements, including patient's information and the tongue features extracted in Section 3.6, which are as below:

- Age
- Image taken with flash / without flash
- Gender
- Greasy tongue coating (腻苔) / no greasy tongue coating
- Spots (点刺) / no spots
- Teeth-marks (齿痕) / no teeth-marks
- Cracks (裂纹) / no cracks

The output of the machine learning model will be whether the person has that particular disease or not. The format of training data will be as shown in figure below.

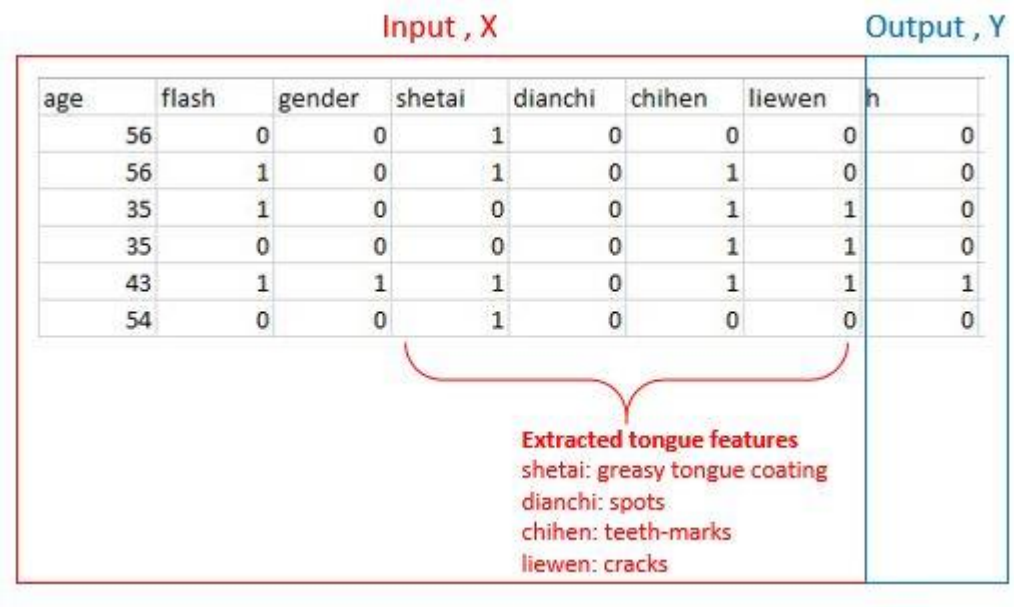


Figure 3.7.1 Format of disease prediction training data

3.7.1 Cross validation

There are several supervised machine learning algorithm: Linear Regression, Logistic Regression, K Nearest Neighbors (KNN), Decision Trees (DT), Support Vector Machine (SVM) and Random Forest. In order to get the most suitable machine learning algorithm for disease prediction, cross validation will be carried out. Cross validation allows us to compare different machine learning algorithms and get a sense of how well they will work in practice.

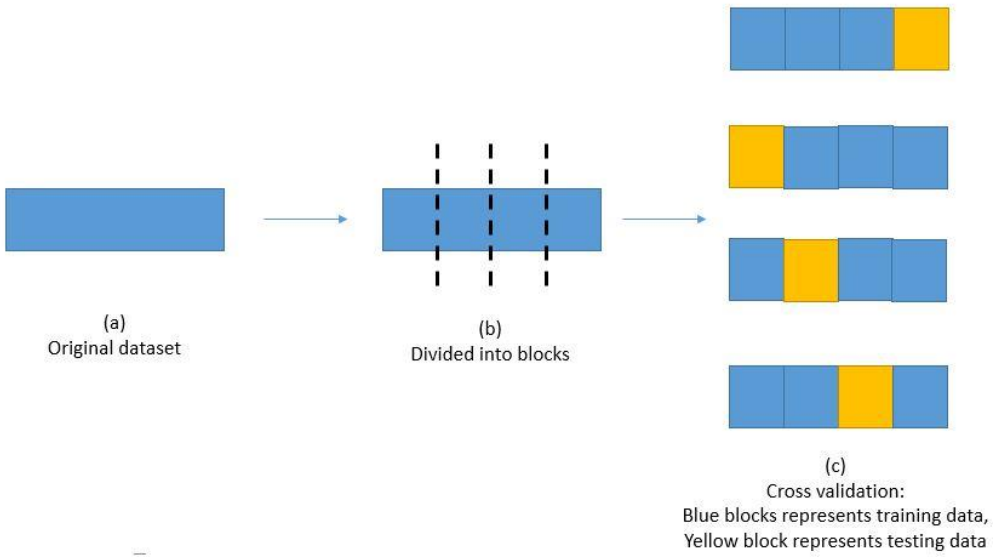


Figure 3.7.2 The working of cross validation

Figure 3.7.2 (a) shows the original data including samples with the chosen disease and without that disease. In order to carry out machine learning process, the data has to be first divided into two groups: training and testing data. As the name implies, the training data will be used to train the machine learning algorithm to perform the disease prediction while the testing data will be used to evaluate how well the machine learning algorithm work. To split the data into training and testing data, the data will first be divided into blocks, as shown in Figure 3.7.2 (b). Cross validation will use one block at a time as testing data while the remaining blocks as training data, and then summarizes the results at the end. For example, as shown in Figure 3.7.2 (c), cross validation would start by using the first 3 blocks to train the algorithm, then use the last block to test the algorithm, and then it keeps track of how well the model did with the testing data. After that, it will use another combination of blocks to train and test the algorithm. In the end, every block of data is used for testing, and the performance of these machine learning algorithms can be compared. The one with the highest accuracy will be chosen.

3.7.2 Bootstrap

Bootstrap is a resampling method by sampling the original data with replacement.

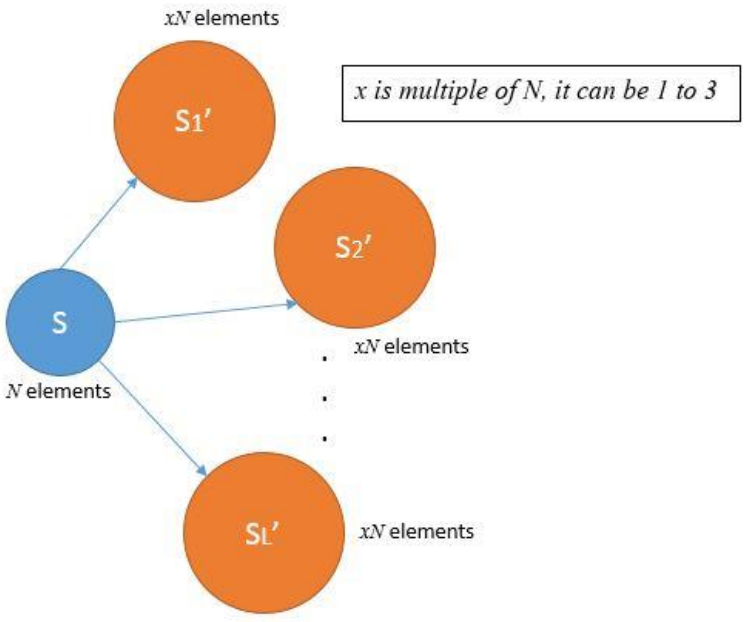


Figure 3.7.3 The ideas behind Bootstrap

As shown in Figure 3.7.3, assume dataset S has N elements, we can sample from it with replacement to obtain a new bootstrap sample, S' which can have multiple of N. Sampling without replacement is just duplication, but in bootstrap, sampling is carried out randomly. In other words, some data may be included more than once while some may be left out. In the case that the training data is imbalanced, which means positive data is much more than negative data or vice versa, bootstrap can be applied to obtain a balanced training data. To verify the robustness of the machine learning model, multiple bootstrap samples should be generated for training so that their accuracy.

Therefore, in this project, cross validation and bootstrap will be implemented to improve the accuracy and to ensure the machine learning model is robust.

3.7.3 Evaluation metrics

To evaluate the machine learning models better, classification evaluation metrics such as sensitivity, precision, specificity and f1-score have to be considered.

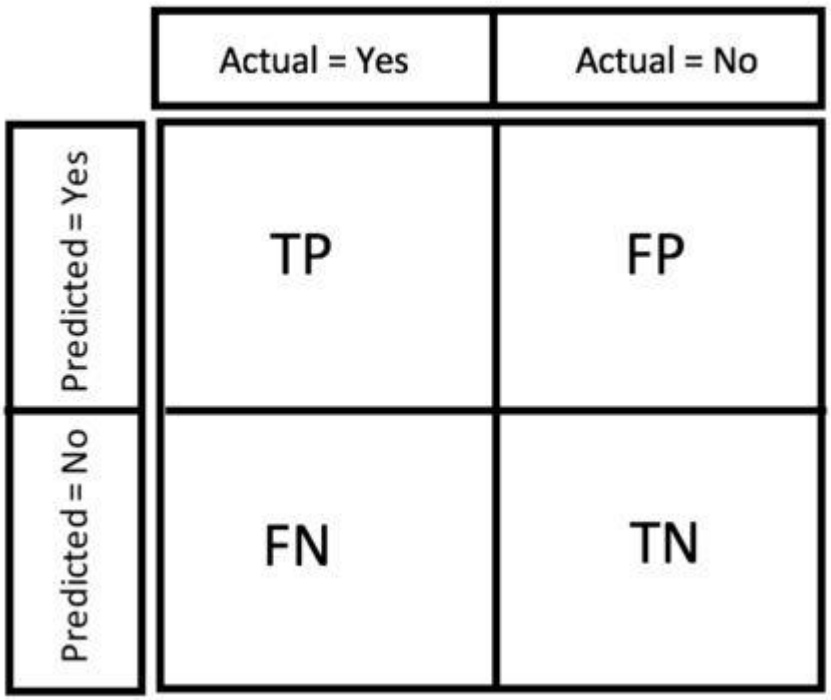


Figure 3.7.4 Confusion Matrix

Figure 3.7.4 shows a confusion matrix. "It is called 'confusion matrix' because it makes it easy to spot where your system is confusing two classes"

(Drakos, 2018). True Positive (TP) refers to when actual positive data is predicted as positive; True Negative (TN) refers to when actual negative data is predicted as negative; False Positive (FP) refers to when actual negative data is predicted as positive; False Negative (FN) refers to when actual positive data is predicted as negative.

The most ideal case would be 0 FP and FN, but this can be achieved only when the machine learning model has 100% accuracy, which is very difficult to achieve. Therefore, what we can do is to minimize FP and FN. However, minimizing FN is more important than minimizing FP. As in this project, disease prediction will be performed. FN represents a sick person being predicted as healthy while FP represents a healthy person being predicted as sick. Thus the cost of having FN is much higher than of having FP, as someone may miss the best treatment time due to the prediction results, but for FP, it can easily corrected after the person goes for further examination and diagnosis.

The formulas for sensitivity, precision, specificity and f1-score are as shown below:

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

$$Precision = \frac{TP}{TP + FP}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$F1 - score = \frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity}$$

Sensitivity is also known as true positive rate. It measures the performance of machine learning model with respect to FN while Precision measures the performance against FP. In other words, Sensitivity represents "how many did we miss" while Precision represents "how many did we caught" (Drakos, 2018). Therefore, if our focus is to minimize FN, sensitivity should be as high as possible without precision being too low. Next, Specificity is also known as true negative rate, it is the exact opposite of Sensitivity. Furthermore, if accuracy is the only evaluation metrics used in this project, then the value of accuracy can be used to compare the performance of different predictive models. However, sensitivity, precision, specificity were considered in this project, another evaluation metric is needed to tell us which model is better by

considering these metrics. Thus F1-score is used as it takes in precision and sensitivity to compute the score.

3.8 Summary

In this project, a predictive machine learning model for TCM tongue diagnosis will be built to predict a particular disease. The disease to be predicted in this stage will be chosen after the sample data being collected is tabulated. The major components of the system include tongue image acquisition system, colour correction algorithm, automatic tongue segmentation algorithm using neural network and automatic feature extraction algorithm using neural network. First, guideline for taking tongue images using mobile phone camera has been established. Next, colour correction algorithm will be applied to achieve colour constancy in tongue images under different illumination. Mask R-CNN is then used to segment the tongue from the image. Besides, data filtering process is designed to reject unqualified images and then request for a retake of qualified tongue image. There are three tests during data filtering process: small image test, small tongue test and blur test. The images that fail either one of the tests will be rejected immediately. After that, tilt correction is carried out on slanted tongue. According to Liu Feilong (2014), discoloration of a particular region of the tongue indicates a lesion in the human organ corresponding to that region. Therefore the tongue is further segmented into regions: margin, root, tip and center. Then Mask R-CNN and YOLO will be applied to perform the automatic feature extraction algorithm as neutral network is able to automatically select the features that contributes to accurate classification results. Lastly, the extracted features will be used to train a machine learning model to predict the chosen disease.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

The code to perform tasks described in Chapter 3 have been designed. The experimental results are as shown below.

4.2 Data Collection

On 19 December 2019, sample collection had been carried out during a visit to Annual General Meeting (AGM) of Persatuan Bimbingan Ajaran Confucian at Taman Connaught.

On that day, we were able to collect samples from 103 people with 206 tongue images, including images taken with flash and without flash. They were also asked to fill in a form to provide us their personal information (refer Appendix A for the form).

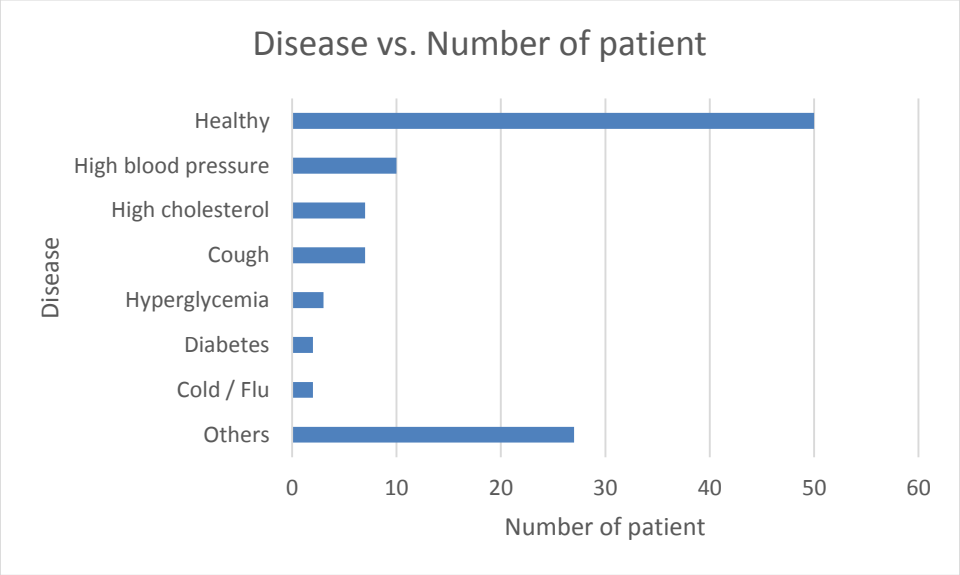


Figure 4.2.1 A chart of disease vs. number of patient

Figure 4.2.1 shows a chart of disease versus number of patient. Each patient provides 2 tongue images (flash and without flash), thus the number in the chart should be multiplied by 2. However, in subsequent training of machine learning model, 'flash/without flash' will be labelled for each image and it will be included in input training feature vector. There are 100 **healthy** samples (from 50 people) in the dataset. The second most common disease in the dataset

is high blood pressure (high blood pressure in Figure 4.2.1) with 20 samples collected (from 10 people). Therefore, **high blood pressure** will be chosen as the disease to be predicted in this project. 'Others' in the chart refers to those disease where only 2 samples (from 1 person) were collected, such as malignant tumour, heart disease, respiratory diseases, irregular menstruation, etc.

4.3 Data Filtering

Qualified and unqualified tongue images have been identified. Unqualified tongue images are rejected immediately while qualified images are allowed to go through subsequent process.

Table 4.3.1 Data filtering results

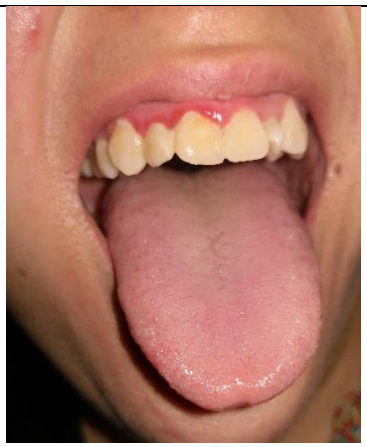
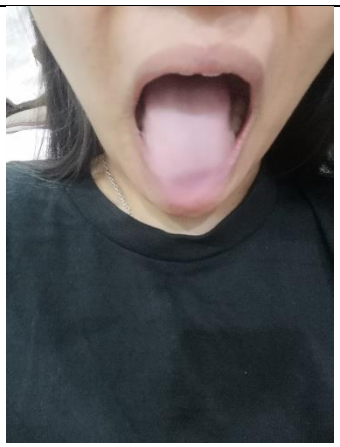
Qualified tongue images	Unqualified tongue images
	
	

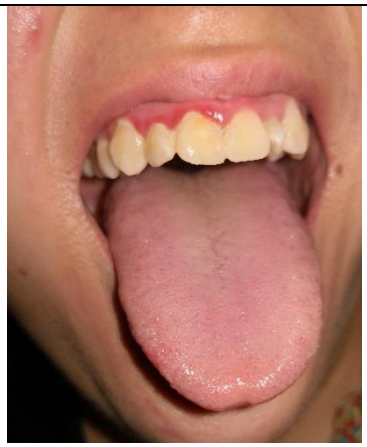
Table 4.3.1 shows the data filtering results. The left columns show the qualified tongue images while the right columns show the unqualified tongue images. The first unqualified image was rejected as it is not able to pass the

small tongue test while the second unqualified image was reject as it is not able to pass the blur test.

4.4 Tongue Segmentation

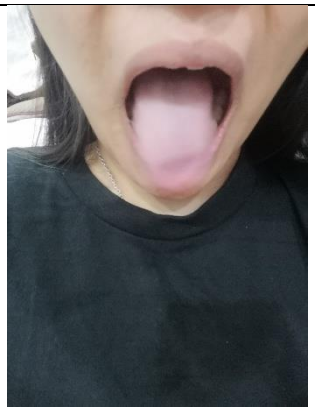
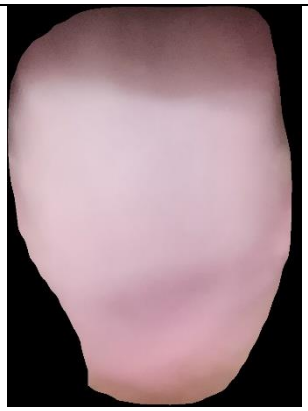
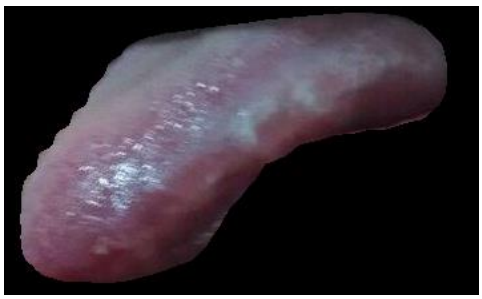
The tongue is then segmented from the images.

Table 4.4.1 Examples of segmentation results

Original Image	Segmented Image
	
	

In order to demonstrate the robustness of tongue segmentation algorithm, some unqualified tongue images are fed to the algorithm. The results below show that tongue segmentation algorithm proposed is able to segment the tongue under different illumination and even if it is blur or not captured exactly from the front of the tongue.

Table 4.4.2 Examples of segmentation results (extreme cases)

Original Image	Segmented Image
	
	
	
	

4.5 Data Preprocessing

4.5.1 Tilt Correction

The results below show that the algorithm is able to perform both clockwise and counter-clockwise rotation.

Table 4.5.1 Examples of tilt correction results

Original Image	Corrected Image
	
	

4.5.2 Tongue Region Segmentation

The tongue images is segmented into regions: margin, tip, root and center.

Table 4.5.2 Example of tongue region segmentation result

Original Image	Segmented Image
	 Left Margin  Right Margin  Tip  Root  Center

Although segmented tongue region is not used in later stages of this project, but this stage is included for future work, when the sample data for a certain disease is enough to study the relationship between each tongue region and human organ.

4.5.3 Colour Correction

In this section, the colour correction algorithm will be evaluated based on the 3 criteria mentioned in Section 3.6.3.

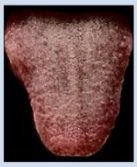
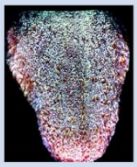
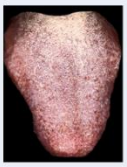
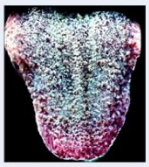
ORIGINAL	HISTOGRAM EQUALIZATION	MSRCR	MSRCP	AUTOMATED MSRCR
Without flash 				
With flash 				

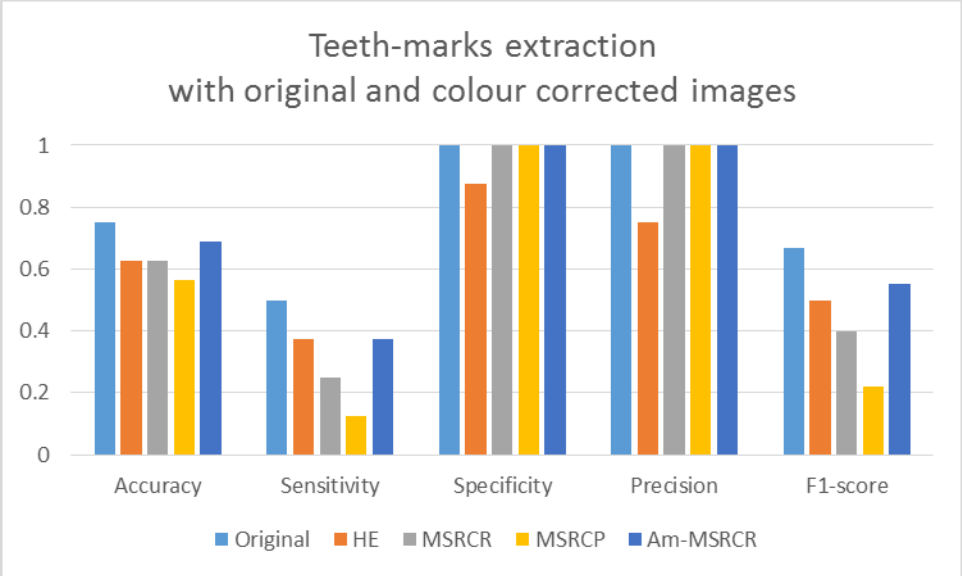
Table 4.5.3 Colour correction results using HE, MSRCR, MSRCP and Am-MSRCR

Table 4.5.3 shows the colour corrected images using HE, MSRCR, Am-MSRCR and MSRCP. According to Table 4.5.3, tongue images with and without flash corrected using MSRCR, Am-MSRCR and MSRCP achieve colour constancy while there is still huge colour difference in images corrected using HE.

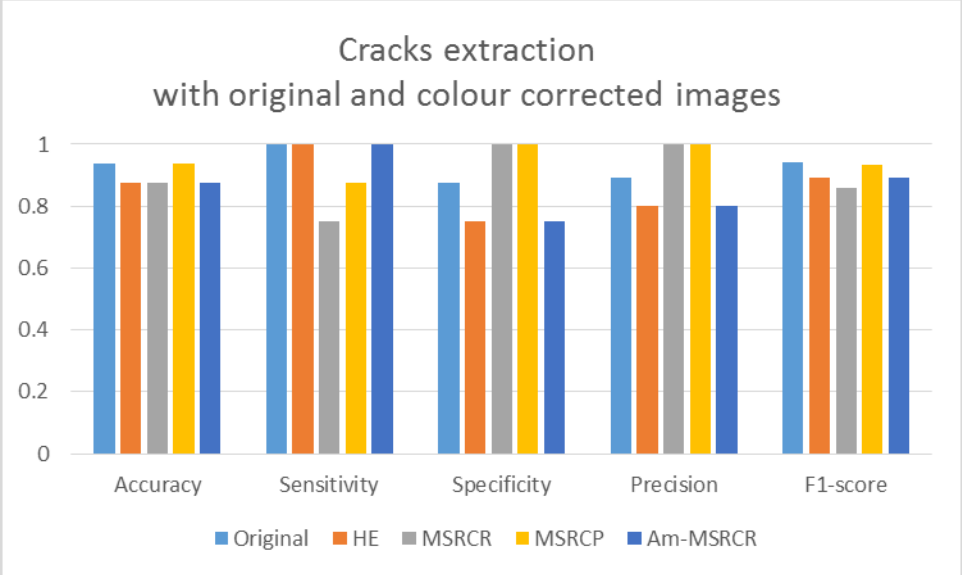
However, Am-MSRCR output has the problem of inverting color, as mentioned in Section 3.6.3.2.3. The red colour tongue may have pixel value near '0' in blue channel. The blue channel of MSR for these pixels will be negative and CRF function will also be negative. Thus, "the blue channel of MSRCR for these pixels becomes positive and their values are changed by the postprocessing step into a value higher than the image average" (Petro, et al., 2014). Therefore, the corrected tongue is bluish or greenish colour.

Another problem with tongue images corrected using Retinex algorithm is that the tongue details are lost and also the corrected tongue colour deviates greatly from the original tongue colour. The lost tongue details increase the difficulty in subsequent tongue feature extraction. Besides, the deviated colour will affect the TCM physician's decision and judgement while verifying our results.

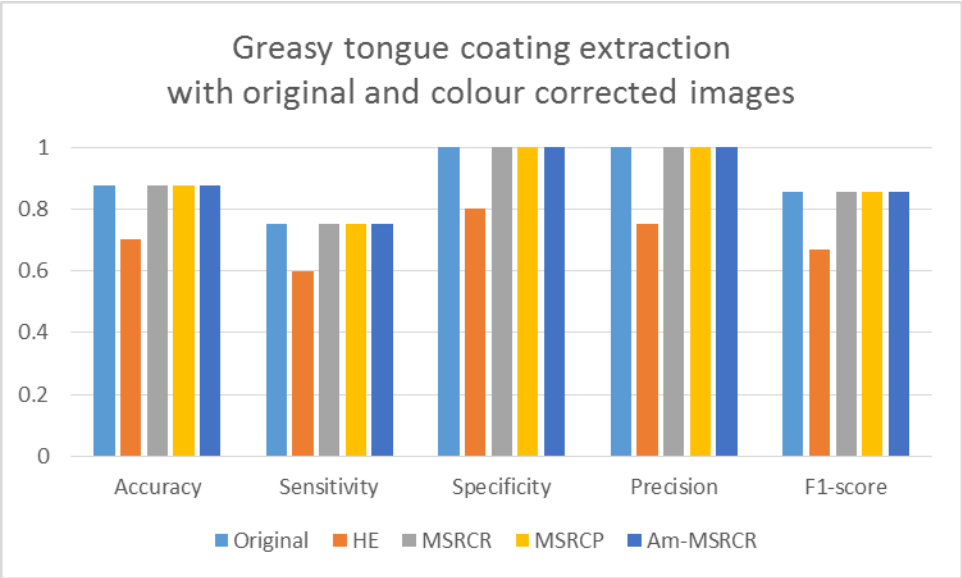
To verify the third criteria whether corrected images can achieve higher accuracy in feature extraction than original images, original and corrected images are fed to Mask RCNN feature extraction neural network (one of the feature extraction neural network; details in Section 4.6). The results are as shown in figures below.



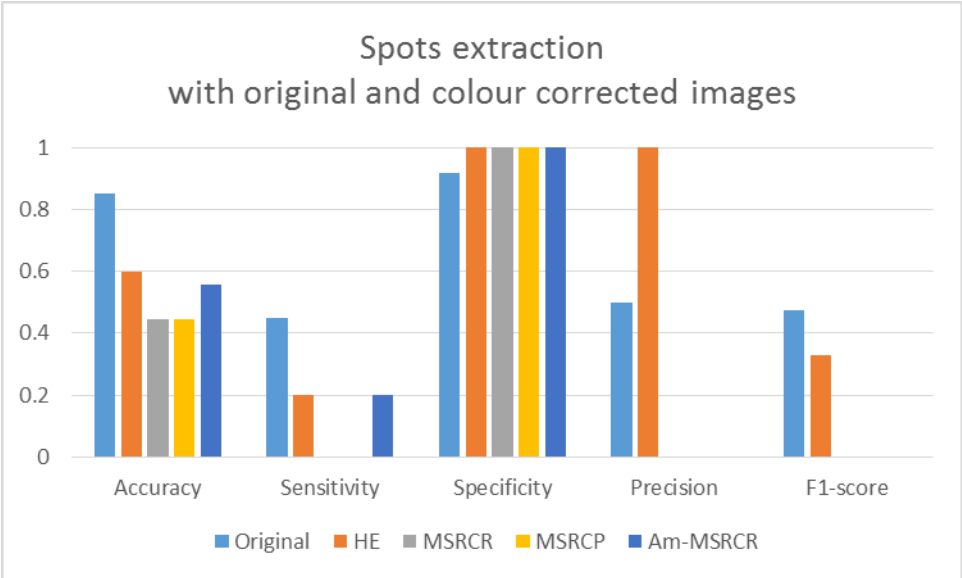
(a)



(b)



(c)



(d)

Figure 4.5.1 Comparison of performance of feature extraction model by feeding original and colour correction images as input

As shown in Figure 4.5.1, original images achieve higher accuracy, sensitivity and F1-score than corrected images in extracting all 4 features. Also, original images achieve higher specificity and precision in almost all features, except cracks and spots. MSRCR and MSRCP achieve higher specificity and precision in extracting cracks but their sensitivities are lower than using original images. Since this project is about disease prediction, our focus is to minimize

FN, sensitivity should be as high as possible without precision being too low. Therefore, it is concluded that corrected images are not able to help to achieve better feature extraction results than using original images.

Since all the 3 criteria could not be met, colour correction stage would be removed from this project. However, to make the tongue feature extraction model more robust, both images taken with and without flash will be included in training dataset of tongue feature extraction model. This is to make sure that the feature extraction model would be able to recognise the features under different illumination. Also, the input variable 'with/without flash' is included in the feature vector for disease prediction.

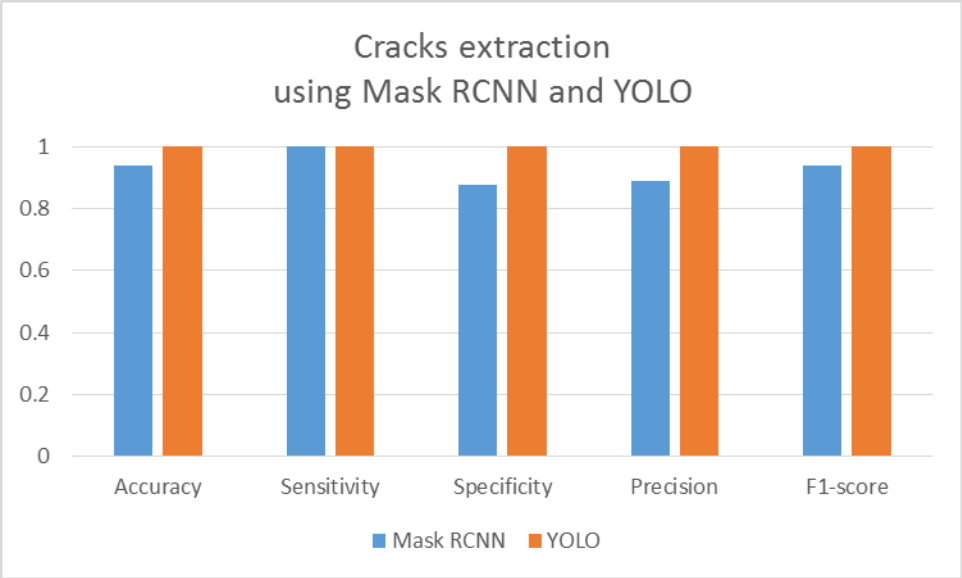
4.6 Feature Extraction

In this stage, Mask R-CNN and YOLO will be trained to extract tongue features. Before that, training data for each feature have to be prepared. A total of 86, 52, 82 and 60 tongue images had been collected for teeth-marks, spots, cracks and greasy tongue coating respectively. Since the dataset is quite small, it is split into 90/10 where 90% is used as training data while 10% is used as testing data. However, now the testing data contains only positive data, negative data needs to be included to test the performance of the neural network. For example, 10% of 86 teeth-marks images will be used as testing data, which is 8 images, then another 8 tongue images without teeth-marks have to be added to the testing data, thus the testing data for teeth-marks will have a total of 16 images. The number of images for training and testing is tabulated in the table below.

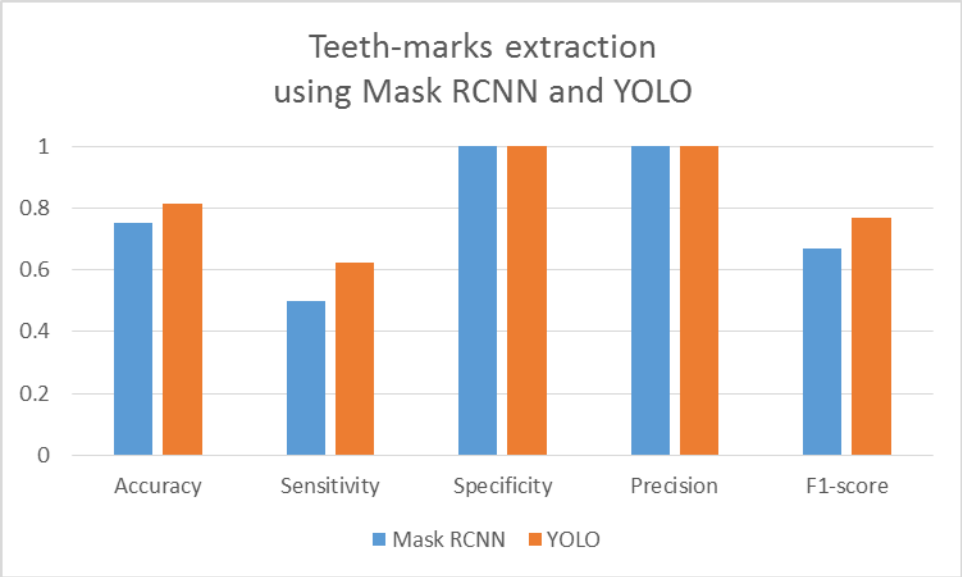
Tongue features	Number of training data	Number of testing data
Teeth-marks	78	16
Spots	47	10
Cracks	74	16
Greasy tongue coating	54	12

Table 4.6.1 Number of images for training and testing tongue feature extraction models

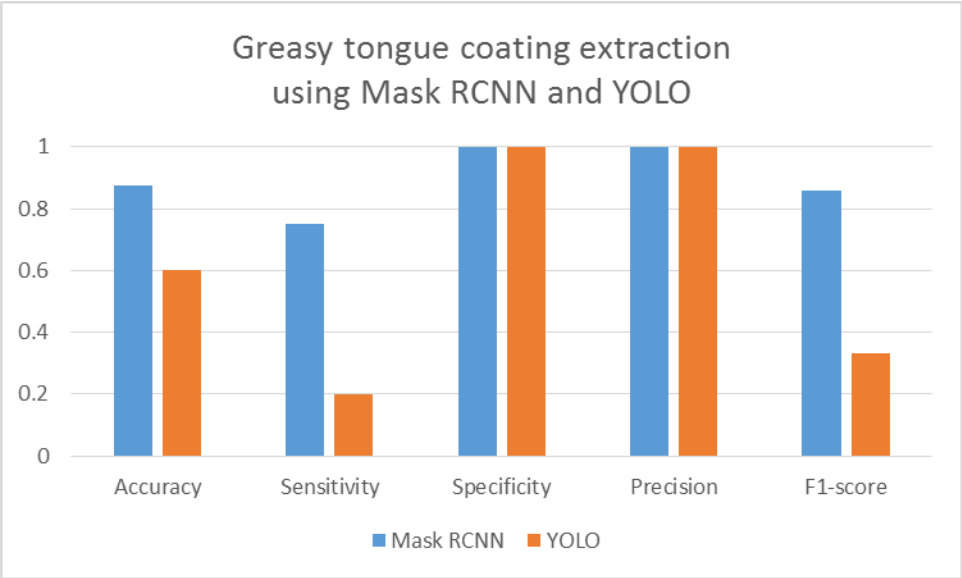
After that, the training data are labeled and are fed to Mask R-CNN and YOLO. The accuracy, sensitivity, precision, specificity and F1-score of Mask R-CNN and YOLO on extracting different features are as shown in the figures below.


Cracks extraction
using Mask RCNN and YOLO
1
0.8
0.6
0.4
0.2
0
Accuracy
Sensitivity
Specificity
Precision
F1-score
Mask RCNN
YOLO

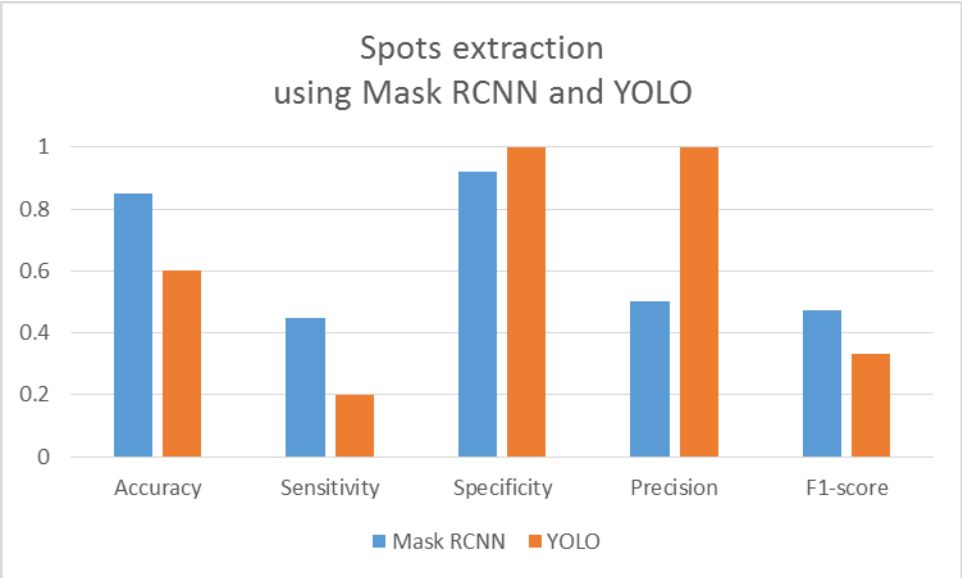
(a)


Teeth-marks extraction
using Mask RCNN and YOLO
1
0.8
0.6
0.4
0.2
0
Accuracy
Sensitivity
Specificity
Precision
F1-score
Mask RCNN
YOLO

(b)



(c)



(d)

Figure 4.6.1 The performance of Mask R-CNN and YOLO in extracting (a) cracks (b) teeth-marks (c) greasy tongue coating (d) spots

According to Figure 4.6.1, both Mask R-CNN and YOLO have excellent results in extracting cracks with YOLO achieved 100% accuracy. Next, both Mask R-CNN and YOLO achieve near 80% accuracy in extracting teeth-marks but YOLO is better in terms of accuracy, sensitivity and F1-score. Meanwhile, Mask R-CNN achieves 87.5% accuracy and 75% sensitivity in extracting greasy tongue coating, which is much higher than YOLO. However, both Mask R-CNN

and YOLO do not perform well in extracting spots. Although Mask R-CNN achieves 85% accuracy, its sensitivity and F1-score are just 45% and 47% respectively.

With small training dataset where there are only less than 90 training images for each features, these results are satisfying, especially for extraction of cracks and teeth-marks. The reason for the difference in results of different feature extraction is the difference in the number of training data. Teeth-marks and cracks have more than 70 training images while spots and greasy tongue coating have only 47 and 54 training images respectively. Therefore, by increasing the size of training dataset, the accuracy and other evaluation metrics should be improved.

Based on the result shown in Figure 4.6.1, YOLO are employed in this project to extract cracks and teeth-marks while Mask R-CNN are used to extract greasy tongue coating and spots. The extracted features are included in feature vectors which were then fed to disease prediction model.

4.7 Disease Prediction

As mentioned in Section 4.2, we were able to collect samples from 103 people with 206 tongue images. Among these images, 100 of them (collected from 50 people) were healthy tongue images, thus the remaining 106 images are unhealthy tongue images. Meanwhile, the second most common disease in the dataset is high blood pressure with 20 images collected (from 10 people). Therefore, in this stage, two predictive models will be built, which are to predict healthy/unhealthy and high blood pressure/no high blood pressure. These data are fed to 7 different machine learning algorithms for training, which were Random Forest (RF), Decision Tree (DT), K-nearest neighbor (KNN), Support Vector Machine (SVM), Logistic Regression (LOG_R) and Linear Regression (LIN_R). Then the performance of these algorithms will be compared to choose the best algorithm.

However, the performance of KNN is greatly affected by the selection of its parameter: n_neighbors. According to Fraj (2017), “n_neighbors represents the number of neighbors to use for kneighbors queries”. The most suitable n_neighbors is the one where training data accuracy is close to testing

data accuracy. Training data accuracy refers to the accuracy obtained when the same data used for training is used to test the machine learning model.

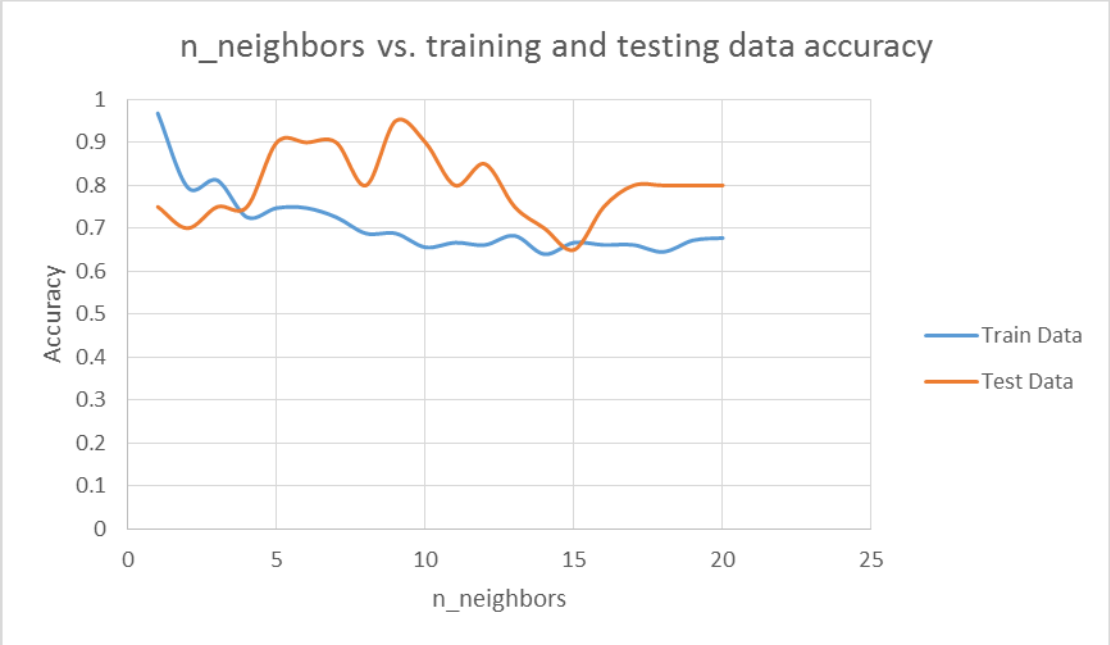


Figure 4.7.1 A graph of training and testing data accuracy with different n_neighbors

Figure 4.7.1 shows the training and testing data accuracy obtained when different n_neighbors were used. Training and testing data accuracy is the closest when n_neighbors = 4, but n_neighbors is usually an odd number because “in the case of a tie vote, the decision on which class to assign will be done randomly when weights is set to uniform” (Greenj, 2018). Also, the accuracy around small n_neighbors is too erratic due to small training size. Then next qualified closest point would be at n_neighbors = 15. Therefore, KNN with n_neighbors = 15 (KNN15) will be used.

4.7.1 Prediction of Healthy/Unhealthy

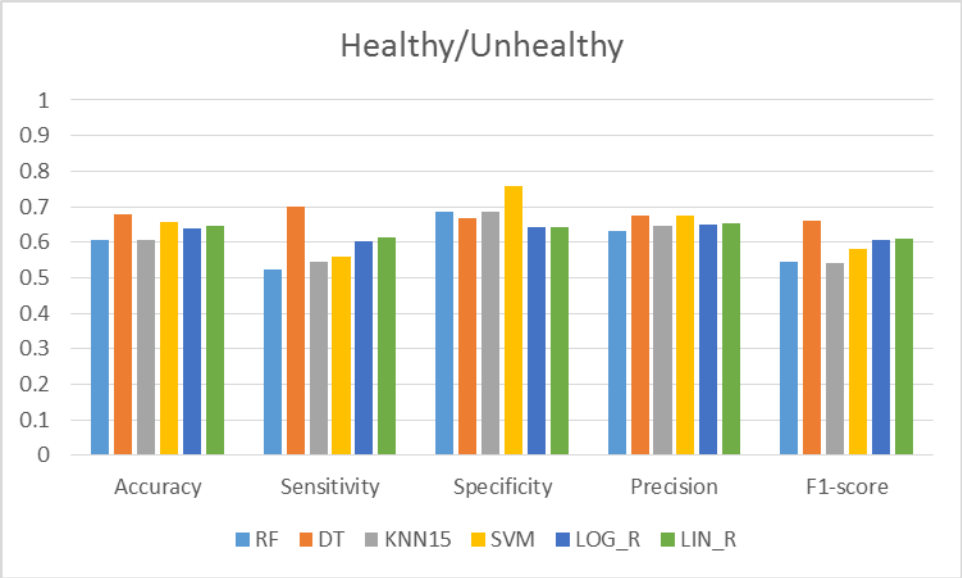


Figure 4.7.2 The performance of difference machine learning algorithms in predicting healthy/unhealthy

Figure 4.7.2 shows the performance of difference machine learning algorithms in predicting healthy/unhealthy. These values are the average values for 10-fold cross validations. In this case, the training data contains 100 healthy tongue images and 106 unhealthy tongue images, the data is nearly balanced, and thus bootstrap is not applied. According to Figure 4.7.2, it can be clearly seen that Decision Tree outperforms all other algorithms as it has the highest score in all evaluation metrics, except specificity. However, all algorithms perform poorly in predicting healthy/unhealthy as nearly all the scores are less than 70%, even the highest F1-score is just 66% which was obtained using Decision Tree. The reason for this poor accuracy can be due to inaccurate personal information provided by people. The labels of healthy/unhealthy on these training data are solely based on the information they filled in the Appendix A, where some may claim they are healthy but in fact, they are not. These inaccurate information can result in such inaccurate results. However, this problem can be easily overcome by collaborating with TCM practitioners or hospitals in future.

4.7.2 Prediction of High Blood Pressure / No High Blood Pressure

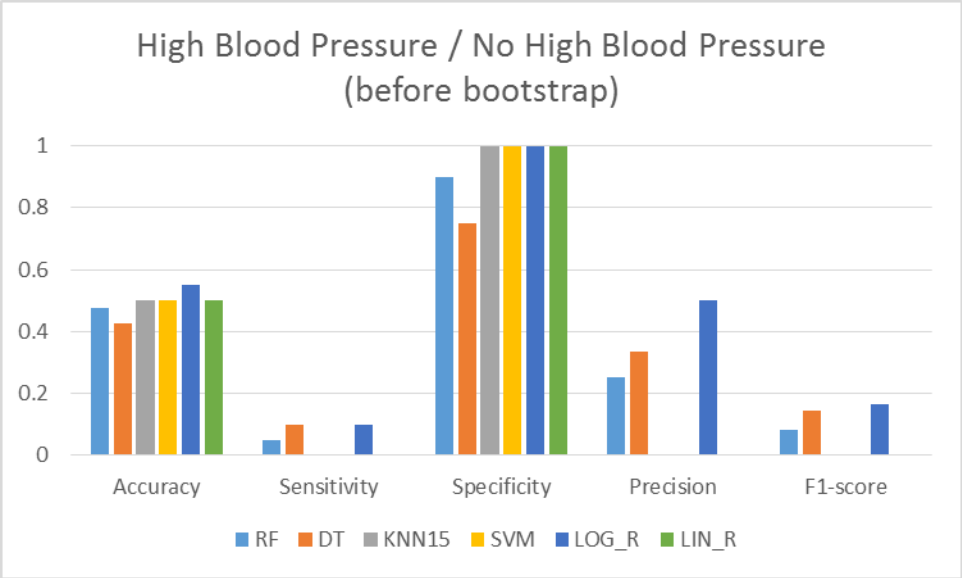


Figure 4.7.3 The performance of difference machine learning algorithms in predicting high blood pressure/no high blood pressure before bootstrap is applied

Figure 4.7.3 shows performance of difference machine learning algorithms in predicting high blood pressure/no high blood pressure before bootstrap is applied. These values are the average values for 5-fold cross validations. In this case, the training data contains 20 high blood pressure tongue images and 186 tongue images without high blood pressure, the data is greatly imbalanced, and there is a need to apply bootstrap. Without the application of bootstrap, it can be clearly seen that all algorithms have really bad performance, where their accuracies are all around 50% while their sensitivity and F1-score were less than 20%. Although they score very high specificity, this can be achieved by predicting all testing data as negative.

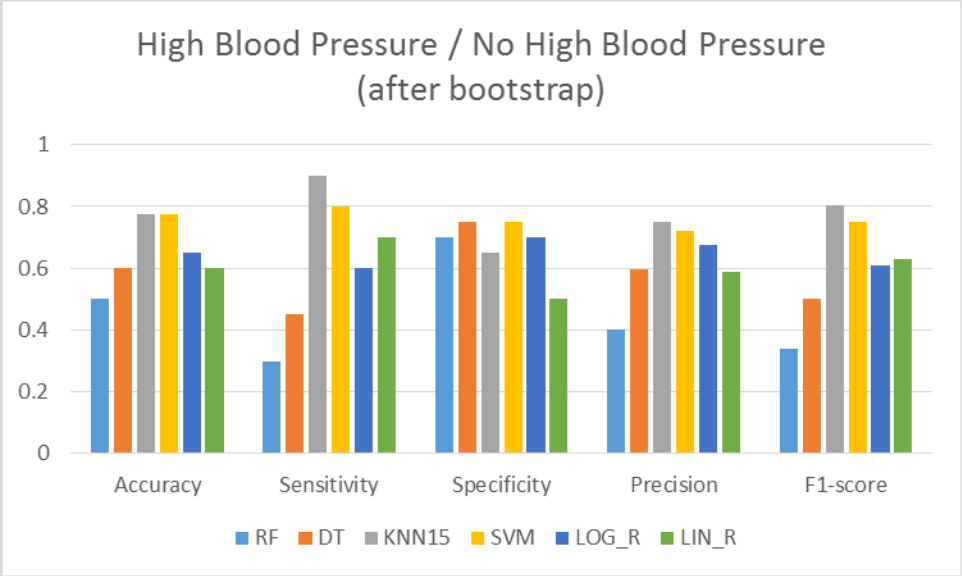


Figure 4.7.4 The performance of difference machine learning algorithms in predicting high blood pressure/no high blood pressure after bootstrap is applied

Figure 4.7.4 shows the results after bootstrap is applied. Out of 20 high blood pressure tongue images, 5 images are kept as testing data for each round of cross-validation. After bootstrapping, the new training data contains 45 high blood pressure tongue images and 45 tongue images without high blood pressure. It can be clearly seen that the results have been improved greatly, where KNN15 and SVM are able to achieve near 80% in accuracy, sensitivity and F1-score. KNN15 even achieves 90% sensitivity. In other words, KNN15 is able to catch most of the positive cases correctly.

Through the three experiments carried out above, Decision Tree model will be saved to predict healthy/unhealthy while KNN15 model will be used to predict high blood pressure/no high blood pressure. Also, it is very important to apply bootstrap when the data is greatly imbalanced, where the number of samples of a class is much less than another classes.

4.8 Summary

In a nutshell, 103 samples with 206 tongue images, including images taken with flash and without flash have been collected. There are 100 healthy samples (from 50 people) in the dataset. The second most common disease in the dataset is high blood pressure with 20 samples collected (from 10 people). Therefore, two predictive models are built in the last stage, which are to predict

healthy/unhealthy and high blood pressure/no high blood pressure. In data filtering process, qualified and unqualified tongue images have been identified. Unqualified images with small or blurry tongues are rejected. Next, the results prove that Mask R-CNN is a robust tongue segmentation algorithm and it is able to segment the tongue under different illumination and even if the image is blur or not captured exactly from the front of the tongue. Also, the tilt correction algorithm is able to perform both clockwise and counter-clockwise rotation. Next, the tongue is also successfully segmented into regions, but the segmented tongue will only be used in the future, when the sample data for a certain disease is enough to study the relationship between each tongue region and human organ. Meanwhile, colour correction stage is removed from the project as the corrected tongue has the problems of inverting color, loss of tongue details and deviation of colour from original tongue colour. However, the alternative would be to include both images taken with and without flash in later stage which is Tongue Feature Extraction to make sure that the feature extraction model would be able to recognise the features under different illumination. After comparing the performance of Mask R-CNN and YOLO in extracting 4 tongue features, YOLO were employed in this project to extract cracks and teeth-marks while Mask R-CNN were used to extract greasy tongue coating and spots. In the last stage, two predictive models are built. Decision Tree has the best performance among other algorithms in predicting healthy/unhealthy, but with just less than 70% accuracy, which can be caused by inaccurate personal information provided by people. However, KNN with n_neighbors = 15 achieves near 80% accuracy and 90% sensitivity in predicting high blood pressure/no high blood pressure after bootstrap is applied.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The objectives of this project have been achieved. First, according to the principles of TCM tongue diagnosis, observation on the color, texture, shape, state and coating of an individual's tongue can provide information of his or her overall health. In this project, the observation process was carried out through tongue image analysis, where the tongue was first segmented automatically from the image using Mask RCNN, then it was passed through data filtering and data preprocessing process, followed by automatic tongue feature extraction.

Next, in order to analyze the changes in healthy tongue and tongue of high blood pressure, four tongue features had been extracted from each tongue image. These extracted features were included in the input feature vector of the predictive model to preform prediction of healthy/unhealthy and of high blood pressure/no high blood pressure. Also, experiments had been carried out to study on whether color correction stage was necessary. However, the results showed that the color corrected images cannot achieve better results in feature extraction than using original images, and thus it was removed from this project.

Thirdly, throughout this project, different machine learning algorithms had been used in different stages. Mask RCNN was used to segment the tongue from the image while both Mask RCNN and YOLO were trained to extract four tongue features, then their performance were compared. Also, seven machine learning algorithms were trained to perform disease prediction, then their performance were compared.

Lastly, the predictive model was evaluated in terms of accuracy, precision, sensitivity, specificity and F1-score. For feature extraction, the results showed that YOLO was better in extracting teeth-marks and cracks while Mask RCNN was better in extracting greasy tongue coating and spots while for disease prediction, the results showed that Decision Tree had the best performance in predicting healthy/unhealthy while KNN with n_neighbors = 15

had the best performance in predicting high blood pressure/no high blood pressure.

5.2 Recommendations for future work

In this project, a complete framework for the design of predictive model for TCM Tongue Diagnosis using machine learning has been built. However, there are few things to be accomplished in the future.

First, with the results achieved in this project, we can ask for collaboration with hospital or TCM clinics to get more reliable data, so that prediction of more diseases can be carried out. The three most common diseases in our dataset are high blood pressure, high cholesterol and cough. Therefore, in the future, we can first collect enough data for these diseases to perform the predictions.

Next, all these functionalities have to be integrated into a mobile application. This will allow people to have a simple examination first before going for further diagnosis. Meanwhile, the mobile application can also help to boost the dataset as the users will provide their tongue images when they use the application.

Thirdly, when new data from hospital are obtained, the machine learning models have to be trained again with more data to improve the accuracy.

Lastly, data of different diseases have to be collected to relate these diseases to different tongue region. The segmented tongue regions can be used to study the relationship between each tongue region and human organ.

REFERENCES

Abdulla, W., 2018. *Splash of Color: Instance Segmentation with Mask R-CNN and TensorFlow*. [Online] Available at: https://engineering.matterport.com/splash-of-color-instance-segmentation-with-mask-r-cnn-and-tensorflow-7c761e238b46 [Accessed 30 June 2019].

Bhamidipati, S., 2018. *How I detect blurry images using OpenCV?*. [Online] Available at: https://blog.sandilya.com/how-i-detect-blurry-images-using-opencv/ [Accessed 3 July 2019].

Chen, R., Xie, J. & & Li, C., 2017. Research on color correction algorithm for mobile-end tongue images. *2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS),* pp. 795-800.

Cheung, V., Westland, S., Connah, D. & Ripamonti, C., 2004. A comparative study of the characterisation of colour cameras by means of neural networks and polynomial transforms. *Coloration Technology*, 120(1), pp. 19-25.

Comaniciu, D. & Meer, P., 2002. Mean Shift: A Robust Approach Toward Feature Space Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 24(5), pp. 603-619.

Ebner, M., 2007. *Color constancy.* s.l.:John Wiley & Sons.

Fraj, M. B., 2017. *In Depth: Parameter tuning for KNN*. [Online] Available at: https://medium.com/@mohtedibf/in-depth-parameter-tuning-for-knn-4c0de485baf6 [Accessed 25 March 2020].

Gonzalez, R. C. & Woods, R. E., n.d. *Digital Image Processing.* 3rd ed. s.l.:Pearson Publications.

Greenj, E., 2018. *k-Neighbors Classifier with GridSearchCV Basics.* [Online] Available at: https://medium.com/@erikgreenj/k-neighbors-classifier-with-gridsearchcv-basics-3c445ddeb657 [Accessed 25 March 2020].

He, K., Gkioxari, G., Dollar, P. & Girshick, R., 2017. Mask R-CNN. *The IEEE International Conference on Computer Vision (ICCV),* pp. 2961-2969.

Hu, J.-l., Ding, Y.-t. & Kan, H.-x., 2018. Tongue Body Constitution Classification Based on Machine Learning. *Journal of Jiamusi University (Natural Science Edition),* 36(5), pp. 709-713.

Hu, M. C., Cheng, M. H. & Lan, K. C., 2016. Color correction parameter estimation on the smartphone and its application to automatic tongue diagnosis. *Journal of Medical Systems*, 40(1), pp. 1-8.

Jobson, D. J., Zia-ur-Rahman & Woodell, G. A., 1997. Properties and performance of a Center/Surround Retinex. *IEEE Transactions on Image Processing,* Volume 6.

Jobson, D., Rahman, Z. & Woodell, G., 1997. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing,* Volume 6, pp. 965-976.

Jobson, D., Rahman, Z. & Woodell, G., 1997. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing,* Volume 6, pp. 965-976.

Kang, T., 2014. *Top 10 Smartphones in May 2014 – Galaxy S5 Fails To Displace iPhone 5s.* [Online]
Available at: https://www.counterpointresearch.com/top-10-smartphones-in-may-2014-galaxy-s5-fails-to-displace-iphone-5s/
[Accessed 3 July 2019].

Khandelwal, R., 2019. *Computer Vision — A journey from CNN to Mask R-CNN and YOLO -Part 2.* [Online]
Available at: https://towardsdatascience.com/computer-vision-a-journey-from-cnn-to-mask-r-cnn-and-yolo-part-2-b0b9e67762b1
[Accessed 3 March 2020].

Khandelwal, R., 2019. *Computer Vision: Instance Segmentation with Mask R-CNN.* [Online]
Available at: https://towardsdatascience.com/computer-vision-instance-segmentation-with-mask-r-cnn-7983502fcad1
[Accessed 20 March 2020].

Lai, Z. & Deng, H., 2018. Medical Image Classification Based on Deep Features Extracted by Deep Model and Statistic Feature Fusion with Multilayer Perceptron. *Computational Intelligence and Neuroscience,* pp. 1-13.

Land, E., 1986. An alternative technique for the computation of the designator in the Retinex theory of color vision. *Proceedings of Natural Academy of Science,* Volume 83, pp. 3078-3080.

Land, E. & McCann, J., 1971. Lightness and retinex theory. *Journal of the Optical Society of America,* Volume 61, pp. 1-11.

Luo, M. R., Hong, G. & Rhodes, P. A., 2001. A study of digital camera colorimetric characterization based on polynomial modeling. *Color: Research and applications*, 26(1), pp. 76-74.

Meng, D. et al., 2017. Tongue Images Classification Based on Constrained High Dispersal Network. *Evidence-Based Complementary and Alternative Medicine.*

Naiya, P., 2018. *Global Top Selling Smartphones – November 2017.* [Online] Available at: https://www.counterpointresearch.com/global-top-selling-smartphones-november-2017/

[Accessed 3 July 2019].

Official, E. F., 2018. *History of image segmentation.* [Online] Available at: https://medium.com/@eaifundoffical/history-of-image-segmentation-655eb793559a

[Accessed 30 June 2019].

Parthasarathy, S. & Sankaran, P., 2012. An automated multi Scale Retinex with Color Restoration for image enhancement. *2012 National Conference on Communications (NCC)*, pp. 1-5.

Petro, A. B., Sbert, C. & Morel, J.-M., 2014. Multiscale Retinex. *Image Processing On Line,* p. 71–88.

Raschka, S., n.d. *How can I know if Deep Learning works better for a specific problem than SVM or random forest?.* [Online] Available at: https://sebastianraschka.com/faq/docs/deeplearn-vs-svm-randomforest.html

[Accessed 3 July 2019].

Srivastava, S., 2018. *Apple iPhone X Remains the Best-Selling Smartphone; Xiaomi Enters Top 3 For the First Time.* [Online] Available at: https://www.counterpointresearch.com/apple-iphone-x-remains-best-selling-smartphone-xiaomi-enters-top-3-first-time/

[Accessed 3 July 2019].

Suzuki, S. & Abe, K., 1985. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, 30(1), pp. 32-46.

Xiao, L., Wang, L. & Wei, Z., 2015. Color Equalization and Retinex. In: M. E. C. B. S. M. Emre Celebi, ed. *Color Image and Video Enhancement.* Switzerland: Springer International Publishing Switzerland, pp. 253-289.

Xu, W. et al., 2011. An Automatic Tongue Detection and Segmentation Framework for Computer-Aided Tongue Image Analysis. *2011 IEEE 13th International Conference on e-Health Networking, Applications and Services*, pp. 189-192.

Yang, S., 2018. *Research on Automatic Capture and Feature Recognition for TCM Tongue Image in Chronic Kidney Disease*, s.l.: School of Information and Software Engineering.

Yu, X.-l.et al., 1994. Study on Method of Automatic Diagnosis of Tongue Feature in Traditional Chinese Medicine. *Chinese Journal of Biomedical Engineering*, 13(4), pp. 336-344.

Zhang, D., Zhang, H. & & Zhang, B., 2017. Introduction to Tongue Image Analysis.. *Tongue Image Analysis*, pp. 3-18.

Zhang, H., Wang, K. J. X. & Zhang, D., 2005. SVR based color calibration for tongue image. *IEEE*, pp. 5065-5070.

Zhang, M., 2019. *New SOTA on Instance Segmentation: Mask Scoring R-CNN Tops Mask R-CNN on COCO.* [Online] Available at: https://syncedreview.com/2019/03/12/new-sota-on-instance-segmentation-mask-scoring-r-cnn-tops-mask-r-cnn-on-coco/ [Accessed 30 June 2019].

Zhuo, L. et al., 2015. A K-PLSR-based color correction method for TCM tongue images under different illumination conditions. *Neurocomputing*, 174(9), pp. 815-821.

刘飞龙, 2014. *The Auxiliary system of Tongue Diagnosis*, Zhejiang: Zhejiang Sci-Tech University.

卫保国, 沈兰荪 & 王艳清, 2002. 数字化中医舌象分析仪. *中国医疗器械杂志*, 26(3), pp. 164-166.

吴祖春, 2011. *舌诊客观化研究中舌象图片采集方法与应用的探讨*, 广州: 广州中医药大学.

周越, 沈利 & 杨杰, 2002. 基于图像处理的中医舌像特征分析方法. *红外与激光工程*, 31(6), pp. 490-494.

张广宇, 2018. *基于中医舌诊的舌像分割与特征提取算法研究*, 2018: 江西理工大学.

张广宇, 2018. *基于中医舌诊的舌像分割与特征提取算法研究*, s.l.: 江西理工大学.

戚进；彭颖红；, 姜. 夏., 2017. 基于 Otsu 阈值法与形态学自适应修正的舌像分割方法. *高技术通讯*, 27(2), pp. 150-155.

李晓宇, 张新峰 & 沈兰荪, 2006. 基于支撑向量机的中医舌色苔色识别算法研究. *北京生物医学工程*, 25(1), pp. 43-46.

梁金鹏, 杨浩 & 张海英, 2017. 基于颜色特征的常见舌质舌苔分类识别. *微型机与应用*, 36(17), pp. 102-105.

沈兰荪, 蔡轶行 & 国, 卫., 2003. 中医舌象分析技术的研究. *世界科学技术—中医药现代化*, 5(1), pp. 15-20.

王昇, 2016. *中医舌诊图像分割与识别方法研究*, 天津: 天津大学.

瞿婷婷, 2016. *基于图像处理和模式识别的舌苔分析研究*, s.l.: 华东理工大学.

苏开娜 & 卢翔飞, 1999. 基于图象处理的舌苔润燥分析方法的研究. *中国图象图形学报*, 4(4), pp. 345-348.

蒋依吾, 建仲, 陈. & 鸿, 张., 2000. 电脑化中医舌诊系统. *中 国中西医结合杂 志*, 20(2), pp. 145-147.

许家佗, 2009. *中医舌诊彩色图谱*. 上海: 浦江教育出版社有限公司.

谢涛, 2017. *基于图像处理的舌像分割及润燥识别研究*, s.l.: 华东理工大学.

陈竟博, 2014. *基于颜色特征的计算机辅助舌诊分类模型*, s.l.: 吉林大学.

陈飞飞, 2018. *舌象瘀斑识别与舌象采集装置改进的研究*, s.l.: 华东理工大学.

APPENDICES

APPENDIX A: TCM Tongue & Eye Diagnosis Data Collection Form

中医舌诊、眼诊数据采集表格
TCM Tongue & Eye Diagnosis Data Collection Form

尊敬的先生/女士，

我们是来自 **拉曼大学（UTAR）**的电子工程系学生。我们目前正在进行"**人工智能中医舌诊和眼诊"**的研究项目。因此，我们需要收集大量舌头和眼睛的照片数据样本以进行研究。我们致力于保护您所提供的数据以及个人资料，并承诺所有资料仅作为研究用途。非常感谢您的帮助。谢谢！

Dear Sir/Madam,

We are electronic engineering students from **Universiti Tunku Abdul Rahman (UTAR)**. We are currently doing a project entitled **"DESIGN OF PREDICTIVE MODEL FOR TCM TONGUE AND EYE DIAGNOSIS IN MALAYSIA USING MACHINE LEARNING"**. Therefore, we need to collect image samples of tongue and eyes for research purpose. We are committed to protect the data samples and personal information you provide, and we promise that all data will be used for research purposes only. Your help is greatly appreciated. Thank you!

No.____

年龄 Age：_____

性别 Gender: ☐ 男 Male ☐ 女 Female

身体健康状况 Health Condition:

☐ 健康 Healthy
☐ 发烧 Fever
☐ 咳嗽 Cough
☐ 伤风感冒 Cold / Flu
☐ 恶性肿瘤 Malignant tumor
☐ 心脏病 Heart disease
☐ 高血压 High blood pressure
☐ 高血糖 Hyperglycemia
☐ 高胆固醇 High cholesterol
☐ 呼吸系统疾病 Respiratory diseases
☐ 癫痫 Epilepsy
☐ 怀孕 Pregnancy
☐ 糖尿病 Diabetes
☐ 月经失调 Irregular menstruation

☐ 其他 Other：_____

签名
Signature :_____