Tongue diagnosis can be an effective, noninvasive method to perform an auxiliary diagnosis any time anywhere, which can support the global need in the primary healthcare system. This work reviews the recent advances in tongue diagnosis, which is a significant constituent of traditional oriental medicinal technology, and explores the literature to evaluate the works done on the various aspects of computerized tongue diagnosis, namely preprocessing, tongue detection, segmentation, feature extraction, tongue analysis, especially in traditional Chinese medicine (TCM). In spite of huge volume of work done on automatic tongue diagnosis (ATD), there is a lack of adequate survey, especially to combine it with the current diagnosis trends. This paper studies the merits, capabilities, and associated research gaps in current works on ATD systems. After exploring the algorithms used in tongue diagnosis, the current trend and global requirements in health domain motivates us to propose a conceptual framework for the automated tongue diagnostic system on mobile enabled platform. This framework will be able to connect tongue diagnosis with the future point-of-care health system.

The global demand for primary healthcare support and the advancement of technology enable the platforms for point-of-care (POC) diagnostics. Despite the recent advances in automated disease diagnosis tools, the requirement of blood serum via nonprofessionals, accuracy, reliability, detection time, and requirement of second confirmatory test are the challenges yet to overcome. Thus, temperature, tongue diagnosis, retinopathy, facial expressions, skin color, and surface could be few vital parameters for future smartphone-based clinical expert systems to attain simplicity, immediacy, noninvasiveness, and automatic analysis.

Tongue diagnosis is an effective noninvasive technique to evaluate the condition of a patient’s internal organ in oriental medicine, for example, traditional Chinese medicine (TCM), Japanese traditional herbal medicine, traditional Korean medicine (TKM). The diagnosis process relies on expert’s opinion based on visual inspection comprising color, substance, coating, form, and motion of the tongue. Instead of tongue’s abnormal appearance and disease, traditional tongue diagnosis is more inclined to recognize the syndrome. For example, the tongue coating white-greasy and yellow-dense appearance signifies cold syndrome and hot syndrome, respectively, which are tied with health conditions such as infection, inflammation, stress, immune, or endocrine disorders. These are two parallel, yet correlated syndromes in TCM. Eliminating the dependency on subjective and experience-based assessment of tongue diagnosis may greatly increase the scope for wider use of tongue diagnosis in the world including Western medicine. Computerized tongue inspection implicating light estimation, color correction, tongue segmentation, image analysis, geometry analysis, etc. could be an effective tool for disease diagnosis aiming to address these concerns.

Tongue diagnosis is not a widespread perception in western medical practice, which leads to one of the key challenges, that is, availability of the authentic classified dataset. On the other hand, Cheng et al questioned about the reliability and validity of practitioners of western medical system using TCM. They emphasized on the importance of the degree of agreement and criticized the adaptation of ‘proportion of agreement’. Belonging to oriental medicine, significant amount of work on tongue diagnosis is published in Chinese. To the authors’ best knowledge, the only detailed survey conducted up to 2012 was motivated by hardware and software of tongue diagnosis. There are recent advancement of mobile phone sensors, dynamics of its application, machine learning techniques, and expert systems, which has the potential to be implemented for automatic tongue diagnosis (ATD). Thus, there is a need to cover the insufficiency of focus on algorithms, machine learning techniques, and clinical decision support system.
(CDSS) related to tongue diagnosis. This work aims to investigate the scope of combining tongue diagnosis with the current research practice, making it work as an expert system without human intervention. The merits and capabilities of existing systems are explored to identify the research gap to achieve this goal. This work will provide an insight into the achievements so far and the mechanism to achieve the ultimate goal of automated tongue diagnosis.

Though few techniques used in case of tongue diagnosis find its applicability in different fields such as speech mechanism,19 tongue-operated system for Tetraplegia patients, this paper does not focus on such cases. This work considers the work published in English only. Manual, intended for digital diagnosis, semiautomatic, and ATD are considered to be reviewed. This work explores the state-of-the-art works done on tongue diagnosis and presents a direction toward a standard and complete work so that tongue diagnosis can be included as an auxiliary diagnosis to support the primary healthcare system. In Section 2, tongue diagnosis systems in general, especially hardware components associated with them are overviewed. In Section 3, tongue image segmentation and feature extraction techniques are critically reviewed. The consequential works done on specific disease detections are explored in later section (Section 4) followed by limited work done on mobile-enabled platform for tongue diagnosis (Section 5). The main focus of the work is algorithms used in different steps of ATD. On the basis of the research gap in the existing works, an overall conceptual diagram is presented in the last section (Section 6) to address the concerns undertaking the trend of the demand.

2. Tongue diagnosis systems

Because of advancement in image sensing modules, various image acquisition and analysis devices for tongue inspection commonly known as tongue diagnosis systems (TDSs) are surveyed and shaped. Jung et al 23 reviewed both type of TDSs:

i) hardware consisting image sensing module, illumination module, compute, and

ii) control module and software for color correction, tongue image segmentation, and tongue classification in literature up to 2012.

To obtain quantitative and objective diagnostic results, the TDS needs to deliver reproducible tongue images under varying environmental conditions. In addition to commercially available digital cameras, many works used charge-coupled devices (CCD) with different configuration as the image sensing module.12,24−28 In recent literatures, both Cihin et al24 and Zhang and Zhang23 used 8-bit resolution 3-chip CCD cameras. The use of tongue image capturing device22,23 is similar to the use of conventional digital camera. Kim et al,25 Zhang et al,21 and Jiang et al20 used CCD cameras to outline tongues, whereas Zhang et al22 placed D65 fluorescent tubes around the 8-bit resolution CCD camera to maintain uniform illumination. Different imaging schemes are described in Section 2.1. The work performed on color correction and light source estimation is explored in Section 2.2.

1 The tongue image capturing devices are defined as the computerized tongue image analysis system (CTIS). ISO/TC 249 specified the requirement and scope of the CTIS. A CTIS can be any device or a system that acquires tongue images to be analyzed by a computer, for example, digital camera, bayer CCD camera, even hyperspectral camera. However, a CTIS by definition does not incorporate interpretation of tongue image.
particular pathologies. In TCM, the shapes of tongues are classified into seven types: ellipse, square, rectangular, round, acute triangular, obtuse tri-angular, and hammer. The ellipse tongue represents the normal tongue.

An alternative view of tongue diagnosis is thus 3D imaging. The thickness measurement, impulsively changing the curvature of the tongue surface angle could be better represented with the 3D tongue modeling scheme. This is supported by Liu et al., who described a framework to reconstruct bending human tongue on the basis of multiview geometry. They used two light sources and a frame to fix the face-location. With the help of four cameras, they adjusted the exposure time and high speed frame rate. To locate the tongue, a laser device was utilized. The artificial features were created on a tongue surface through an adjustable laser beam. Finally, a finite-element-based representation of the tongue was attained.

Many of these devices, utilized by the existing systems, are capable of compensating lighting environmental condition (discussed in Section 2.2), but none of these systems is standalone (i.e., the system cannot provide the diagnosis or syndrome itself). Most systems are portable but they lack a collective intelligence to provide the final diagnosis.

2.2. Light source estimation and color correction

Lighting condition is considered to be an important prerequisite for tongue diagnosis. Different approaches have been taken to solve the overwhelming problem of lighting condition estimation. These solutions include controlled environment,27,29 color correction,41 hardware based implementation, for example, tongue image acquisition device,20 and ColorChecker.42 An image taken with the same camera may provide different tongue diagnosis due to the variation in the color properties at different lighting conditions. The imperfection in camera operation may cause some consequences.28 The inconsistency imaging with different devices may affect as an added disadvantage making it inconvenient to share or interchange.

Jang et al.26 and Zhang et al.20 used halogen tungsten lamp as the light source to capture tongue images. Both of them used low color temperatures. The color temperature of a lamp is the temperature of an ideal black-body radiator that radiates light of comparable hue to that of the lamp. It is a way to describe the light appearance provided by a light source, measured in degrees of Kelvin (K) on a scale from 1000 to 10,000. Typically, the light appearance of 2000–3000 K is referred to warm white, producing orange to yellow-white light; 3100–4500 K is cool or bright white emitting more neutral white light with probable blue tint and 4600–6500 K and above tries to imitate the daylight, producing a blue-white light and creating a crisp and invigorating ambiance. However, the use of inadequate color temperature in the mentioned work failed to render color in high fidelity. Because of work done with reddish-biased images produced under halogen tungsten lamp, this is no longer a conventional method.

Research on tongue diagnosis includes placement of fluorescent lamp,25 standard light source installed in a dark chest,43 Köhler illumination,44 built-in LED lighting for tongue diagnosis. In general, Köhler illumination is used in optical microscopy as a specimen illumination method for transmitted and reflected light. It is useful to generate an incredibly even illumination of the sample. It has the ability to repress illumination source in the resulting image. However, the requirement of additional optical elements makes it more expensive. Another passing fad is taking images in public office environment.46 Hu et al.41 analyzed tongue features observed under different lighting conditions such as fluorescent, incandescent, and halogen illuminant. They used Support Vector Machine (SVM) to make the system independent of lighting environment. To nullify inconsistent or varying lighting conditions, another current trend is hyperspectral imaging.33,35 Another SVM-based work is recently reported to attain 94% accuracy to classify four tongue-related colors, that is, red, light red, and deep red.46 This work encompassed classification in two stages. The first stage involved the unsupervised machine learning, k-means. The background separation was partly involved at this stage. The later stage was based on SVM. The execution time was only 48 seconds, which shows the potential to be implemented as a real time tongue image segmentation technique. Because of the 2-layer segmentation process, SVM improved the color detection, that is, classification by 41.2%.

Even after using low color temperature halogen lamps, some work26 did not involve any color correction method. As a color correction scheme, the printed color card has been used in the earlier work.20,25,43,45,47 Some45,47 used Munsell Color Checker as well. Munsell was considered to be the first color order system to separate hue, value, and chroma into perceptually uniform and independent dimensions and to systematically illustrate the colors in 3D space. Hu et al.48 used SVM-based color correction matrix nullifying the dependency on controlled lighting environments and tongue positions. The lighting environment included with and without flashlights. They further improved their work and acquired better result.31 To the best of authors’ knowledge, this is the best work done so far in the field of tongue diagnosis for light estimation and color correction. Another popular convention is transforming the image into standard color space which could be achieved by both hardware and software manipulation.

On the other hand, recent advances of mobile phones are endeavoring to address lighting condition and color correction issue on its own. The camera lens, resolution, aperture, stabilizer, filter, multiframe imaging, and strong image processing software are able to manage the ambient condition better than ever. The back camera-front camera issue is also resolved for many phone brands. However, as chromatic features of tongue images hold the disease information, light estimation can still play a crucial role.

3. Image segmentation and feature extraction

The most challenging part of tongue diagnosis is an adequate segmentation and optimum feature extraction. Before starting the image segmentation, few works in literature preferred tongue detection and positioning. At the next stage, both color and geometric features need to be carefully extracted to feed into the subsequent steps.

3.1. Tongue detection and positioning

Zhong et al.49 used mouth location method to locate dark hole’s position in the mouth and active appearance model (AAM) for automatic tongue segmentation. The method required a different initial contour for tongue body segmentation which made it technically infeasible.

In order to fix the tongue position, a guidance scheme for image acquisition system which outline the frontal and feedback gridlines, was sketched by Jung et al.4 The study included three categories, namely, nil, frontal, and frontal and profile gridline and 120 images were assessed based on profile angle and tongue area ratio. Jung et al. discussed the relationship between diagnosis accuracy and repeatability of tongue features, for example, color and shape. The intra-class correlation coefficient was computed to examine repeatability. The result suggested an improvement in color and shape repeatability. Their study involved only healthy tongues which could be further used as a calibration model for related works.
3.2. Color models and chromatic features

The tongue diagnosis comprises recognizing the disease-related colors on the tongue surface and analyzing the texture. The tongue image segmentation steps such as edge detection and background–foreground separation are commonly processed in gray scale. However, the overall tongue image analysis needs to be processed in color format. The color gamut of a tongue image needs to be transformed into L*a*b* from RGB.

Table 1

<table>
<thead>
<tr>
<th>Color model</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGB</td>
<td>Most of the devices take the image in RGB. Then they are transformed into other color spaces. Tongue image segmentation⁴⁸,⁵⁰ Color correction³⁵</td>
</tr>
<tr>
<td>HSI</td>
<td>Fusion of color and space information⁵¹,⁵² Tongue image extraction⁵¹–⁵³ Morphology⁵⁴,⁵⁵</td>
</tr>
<tr>
<td>HSV</td>
<td>Color matching⁴⁷,⁵⁶ Contour shape¹⁵</td>
</tr>
<tr>
<td>LAB</td>
<td>Redness measurement⁵⁸ Color matching⁵⁶,⁵⁷,⁶⁰ Automatic tongue image segmentation⁵¹</td>
</tr>
</tbody>
</table>

Figure 1. Tongue image in various color spaces.

The RGB (Red, Green, Blue) is the most popular model for tongue diagnosis describing what kind of light needs to be emitted to produce a given color.⁴⁸,⁵⁰,⁶² The RGB space was utilized to extract tongue color features, for example, means and standard deviations of the colors per pixel within the region of interest.⁵⁰ It is seen as true color in MATLAB, but the model is device dependent, which means there are many variations of RGB color space for the same image using different devices. HSI (Hue, Saturation, Intensity) and HSV (Hue, Saturation, Value or Brightness) are two cylindrical-coordinate representations of points in a RGB model. HSI is closer to human vision. Zhu et al⁵¹ favored HSI due to its specificity of the application in tongue image extraction. With the help of HSI color model, the greedy rules were utilized to develop the template. Then, the template was processed. After successful separation of background from the raw image, the template was found to be effective even for thick tongue coating. It can be useful for the quantitative evaluation of tongue.⁵² The hue and intensity were exploited by Jian-Qiang et al⁵⁵ for tongue image segmentation which worked well for regular tongue surfaces, but not for the irregular ones. Li et al⁵⁶ used hue and brightness of HSV to decide the initial position of the tongue, whereas Jiang et al⁴⁰ reported that HSV is unsuitable for tongue diagnosis as it has discontinuities in hue value around red, making it sensitive to noise. The color information is separated from the energy or image intensity in HSV, making the color space less noise-prone than in RGB. HSV being popular in the library like OpenCV for Android platform makes it a strong candidate to consider.

In the recently reported article on tongue diagnosis, researchers have found to prefer LAB over the other color space as it is more perceptually linear giving the advantage of effective visual inspection and comparison. L stands for lightness and ‘a’ and ‘b’ are the color opponent dimensions. Analyzing the tongue indices, the mean a* of the whole tongue area was found to be distinguishably different between people with and without Yin-deficiency.⁵⁸ Thus, the color channel values can provide significant information about the patient’s condition.

Continuing with LAB color space, Jung et al⁵⁷ explored diagnosis accuracy and considered the intraclass correlation coefficient to inspect tongue feature’s repeatability. The median value of tongue region, that is, tongue body was taken into account for color feature computation.⁶³ Kawanabe et al⁵⁹ examined tongue color, its body, and coating by analyzing tongue images of 1080 subjects. They cataloged tongue color and tongue coating into five and six categories, respectively in the L*a*b* color space. To compute tongue body and coating information, k-means clustering (an unsupervised machine learning technique) was employed. The authors suggested the outcome as a globally suitable tongue color diagnostic standard. However, the reliability was not quantified, especially for the tongue of unhealthy subjects. From the a* value difference, Kim et al⁶⁸ separated the tongue coating from the tongue body area. The tongue coating was quantified by tongue coating percentage index. As the image was captured by a CTIS, the generated image needed to be transformed into L*a*b* from RGB.

Although tongue image capturing devices are useful, in literature, more works are done using a CCD camera. Hsu et al⁵⁴ used a head mounted tongue capture device. Tongue image segmentation based on chromatic information is performed in the literature using all standard color models. However, sRGB (standard Red Green Blue) is found to be more popular choice to fuse color and space information.

Extracting primary core colors by an iterative method, Chen et al⁵⁷ used Generalized Lloyd Algorithm (GLA) to obtain the color histogram. Because of the inadequacy of chromatic features, textural features were taken into account by an edge histogram descriptor. This ‘content-based image retrieval technology’ produced improved result, considering different weights for two features in retrieval. The work focused on feature extractions only, and disease classification was not considered. Inclusion of disease classification with same dataset of 268 images would have enlightened the readers more about the result obtained from feature extraction.
Table 2
Neural Network for Chromatic Analysis

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Camera</th>
<th>Input</th>
<th>Layer/Structure</th>
<th>Performance</th>
<th>Computing cost (s)</th>
<th>Dataset</th>
<th>Context</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP/BP</td>
<td>Watec WAT-202D CCD camera</td>
<td>R, G, B histogram</td>
<td>Multistructure 192-32-84</td>
<td>Training set: 100%, recall set: 58.7%</td>
<td>Learning rate: 0.1 s</td>
<td>31</td>
<td>Color analysis</td>
<td>26</td>
</tr>
<tr>
<td>SONY 3-CCD video camera (Sony DXC-390P)</td>
<td>3 unit for camera response</td>
<td>–</td>
<td>–</td>
<td>ANN (sigmoid) Training time: 3.3 Testing time: 0.0277 ANN (linear) Training time: 2.8511 Testing time: 0.0257 RBF (SVR) Training time: 0.01 Testing time: 0.0009</td>
<td>&gt;9000</td>
<td>Color correction</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>n/m</td>
<td>R, G, B, H, S, I of the pixels</td>
<td>Multi (9) Self-adaptive network structure</td>
<td>Accuracy: 88%</td>
<td>–</td>
<td>200</td>
<td>Color recognition</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>RFBNN</td>
<td>Hyperspectral camera</td>
<td>120-band hyperspectral image</td>
<td>N/m</td>
<td>Accuracy: &gt;86.74</td>
<td>844</td>
<td>375</td>
<td>Hyperspectral imaging</td>
<td>44</td>
</tr>
<tr>
<td>KNN LDL</td>
<td>Digital camera (CANON1100D)</td>
<td>Tongue body image</td>
<td>Single and multi</td>
<td>Accuracy: &gt;85.93 68.41%</td>
<td>691</td>
<td>N/m</td>
<td>Labeling tongue color</td>
<td>70</td>
</tr>
<tr>
<td>AA-kNN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>86.47%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The artificial neural network (ANN) is not an exceptional choice for chromatic analysis. ANN has been successfully used in medical diagnosis due to its ability to process large volume of data, reduced probability of ignoring relevant information, and faster computational time.\(^\text{64}\) For example, the output of a wireless capsule endoscopy for evaluating the gastrointestinal tract is an 8 hours long video. To reduce the inconvenience of already over burden health system, Barbosa et al\(^\text{65}\) utilized an ANN to diagnose and classify small bowel tumors analyzing the visual information of wireless capsule endoscopy. An enhanced HSV color model was utilized for the segmentation purpose, using the convolutional neural network (NN).\(^\text{66}\) However, Cheung et al\(^\text{67}\) argued that the neural network is complex and time consuming. Table 2 demonstrates the practice of neural network for color-based analysis. However, Zhuo et al\(^\text{68}\) discouraged NN for online color correction in the current context due to the requirement of large sample size to train the NN-based mapping model. The polynomial-based correction method was favored for its lower computational complexity.

In the field of medical image processing, gray scale images are more prevalent than color images. The tongue diagnosis comprises recognizing the disease-related colors on the tongue surface and analyzes the texture. Therefore, one needs to address the issue of volume of colors to be managed. The complex part is the resemblance of the involved colors. Most of the disease-reflection features are reported to be located in the 2% region between the 100% and 98% of the full tongue color range.\(^\text{47}\) Thus, J measure based segmentation (JSEG) algorithm is performed to test the homogeneity of a given color-texture pattern. The method involved unsupervised machine learning techniques such as k-means clustering and a hierarchical clustering (called agglomerative clustering). This segmentation method is computationally more feasible than model parameter estimation. The method proposed by Deng and Manjunath\(^\text{71}\) is illustrated in Fig. 2. The color quantization starts with the peer group filtering,\(^\text{72}\) which provides a smooth image by removing the noise. It also provides a smoothness indicator value. The weight to each pixel is assigned in such a way that pixel at textured area's weight is less than smooth area's weight. The implementation of agglomerative clustering algorithm increases the pixel of approximate same color, decreasing the global distribution. The color class-map is applied to local window and then J-image is obtained using Fisher's multiclass linear discrimination.

The implementation of JSEG, proposed by Deng and Manjunath,\(^\text{71}\) is shown in Fig. 3 to illustrate the effective color reduction keeping the useful information intact. Although chromatic features are essential to analyze, it is not sufficient enough to provide the complete information. Thus, this work will continue to discuss about different image processing techniques in the next section.

### 3.3. Edge detection techniques

Edge detectors are essential tools for tongue diagnosis and one of the most popular techniques. The prewitt, sobel, and Laplacian filters are commonly used edge detectors for medical image processing. To separate the tongue from face and mouth (i.e., background–foreground separation), different versions of canny filter\(^\text{73}\) are used\(^\text{21,56,74,75}\). Shi et al\(^\text{75}\) used canny to find the tongue tip point. Zhang et al\(^\text{21}\) used CCD camera to outline tongue via canny algorithm and then utilized feature extraction to diagnose diseases. The authors transformed the image into gray scale and an optimized canny algorithm to obtain the edge image. Kanawong and Xu\(^\text{76}\) presented a complete automatic tongue detection and tongue segmentation framework method excluding any parameter adjustment and initialization. They proposed a Principal Component Analysis (PCA)-based tongue detection algorithm and hybrid tongue segmentation utilizing mean shift algorithm, canny edge detection algorithm, and Tensor Voting algorithm. However, the use of only one sample makes the system incompetent. The Sobel is another popular filter to be found in the literature.\(^\text{56,59}\) Few works have been done using mathematical morphology. Li et al\(^\text{56}\) compared different tongue image extraction methods such as geodesic active contour, HSI color model (mathematical morphology and sequential algorithm), and presented a method that involves multijective greedy rules to employ fusion of color and space information. Jian-Qiang et al\(^\text{55}\) utilized morphological opening to eliminate the thin link of lower lip attached with the tongue image and morphological closing to get rid of the small holes on tongue.

### 3.4. Contour shape

The contour detection is quite prevalent for tongue region segmentation. The snake model finds its usefulness on tongue image segmentation with modifications.\(^\text{50,75,77–79}\) The dual snake was used for tongue segmentation with 92.89% accuracy.\(^\text{78}\) Noise removal and HIS color model transformation were conducted before implementing the algorithm to achieve an accurate contour. However, the method was not feasible due to the hurdle to attain the initial contour, both inside and outside. Shi et al\(^\text{79}\)

---

**Fig. 2.** J measure based segmentation (JSEG).
used snake geodesic active contour (GAC) model for tongue image segmentation. GAC is also known as geometric active contour or conformal active contours. The model is inspired by the Euclidean curve shortening evolution. On the basis of the object of the image, the contour splits and merges. Shi et al.\textsuperscript{79} categorized the images of the database by tongue colors and it contained four classes and ten branches. The double geodesic flow was applied after dividing the tongue image into two parts based on previous knowledge of tongue shape and location. The result showed low error and good true positive volume fraction (98%) compared to other works. But because of expensive computational time, it is not feasible for real time solution. Moreover, the system is not full-proof for background separation in case of irregular tongue images. Later, the work continued with color control-geometric and gradient flow snake (C2G2FSnake) algorithm for fully automated tongue segmentation, keeping the number of classifications and branches same.\textsuperscript{75} The dataset included 30 images. This method enhanced the curve velocity. Though the work showed good accuracy and less complexity, it needed four points to be specified for initial contour formation. Chen et al.\textsuperscript{61} performed tongue image segmentation utilizing the same dataset via MATLAB and the test result showed 98.4% accuracy taking 11.46 seconds to compute with C2G2FSnake. Testing on 40 images, recent work has been reported to achieve 75% accuracy.\textsuperscript{80}

3.5. Watershed transformation

The word watershed signifies a ridge that divides areas drained by different river systems. The geographical area draining into a water source, for example, river or reservoir is called the catchment basin. The watershed algorithm is adaptable to various kinds of images, for example, blood cells, stars, toner spots on a printed page, DNA microarray elements. This segmentation algorithm is capable to differentiate incredibly complex images. However, the performance depends on correct choice of markers to avoid oversegmentation. The watershed transform is also utilized as a segmentation technique for tongue diagnosis.\textsuperscript{50,61,81}

A possible route for watershed transform on tongue could be contrast-limited adaptive histogram equalization (after preprocessing), thresholding, filtering, and then overlaying the perimeter as shown in Fig. 4. The last image of Fig. 4 is where watershed algorithm was employed in this work and this was introduced by Eddins\textsuperscript{82} for cell segmentation with microscopic images. The image does not contain the anticipated textural information. The region of interest is cropped and the rest of the image information is not useful.

Snake models are often combined with watershed transform to produce better edge detection. Wu et al.\textsuperscript{81} used such combination on 120 images, where the initial contour was generated using watershed transform whereas the active contour model was used to converge to the exact edge. However, there may not be a clear edge border for tongue images in natural setting like the ones used in this work. Taking RGB tongue images as input, Ning et al.\textsuperscript{50} sequentially used gradient vector flow (GVF), watershed transformation, and then an interactive region merging method called maximal similarity based region merging (MSRM) and finally snake model for refinement. The idea was to avert oversegmentation of watershed as GVF can well preserve main edge structures while removing trivial structures.

3.6. Tongue shape based analysis

In order to analyze the shape of the tongue, one may require contour extraction and tongue shape correction. Zhang and Zhang\textsuperscript{23} used decision tree to classify five different types of tongue shapes among 130 healthy and 542 diseased people. They included 13 geometric features. Their work outline is sketched in Fig. 5. Huang et al.\textsuperscript{40} corrected the tongue deflection based on three geometric categories and analyzed tongue shape based on seven geometric features. The Analytic Hierarchy Process (AHP) was utilized for tongue classification by computing relative influences of different parameters such as length, area, and angle factor. Each AHP module represented each tongue shape class. The ambiguity and impression between the quantitative features and the shape classes were presented by the fuzzy fusion framework. This framework combined seven AHP modules. The study included 362 tongue samples and the result showed success to decrease tongue shape deflection. The classification accuracy was 90.3%. The authors compared the result with kNN\textsuperscript{83} and linear discriminant analysis,\textsuperscript{29} where the accuracy was 78.2% and 82.9%, respectively. Although
their study includes unhealthy tongues, the correlations between disease types and tongue shapes are not taken into account.

4. Disease classification

Tongue diagnosis deals with syndrome, rather than disease itself. Xu et al.\textsuperscript{84} classified syndromes using machine learning techniques. Another syndrome classification-based work, conducted by Zhao et al.\textsuperscript{85} focused on searching for its correlation with viral hepatitis. Kim et al.\textsuperscript{86} utilized statistical analysis to prove that the self-reported Yin deficient patients may have more redness in tongue and less tongue coating in comparison to non-Yin-deficient control group. Statistical analysis was further utilized to determine the relationship between ischemic stroke with and without retroflex tongue among 308 patients. A statistical analysis was utilized by Kim et al.\textsuperscript{87} to track the treatment of functional dyspepsia. This work analyzed the change in the Nepean Dyspepsia Index scores to indicate the primary outcome. The baseline and the images taken after four weeks were examined to achieve this goal. More parameters (such as tongue coating thickness) were added to find out a few secondary outcomes.

Effective CDSS may be yet to be established, but there is no denial that CDSS has the potential to improve healthcare by linking health observations with health knowledge to influence health choices by clinicians. The diagnosis decision support systems (DDSS) is a specialized CDSS for diagnosis that requests for the patients data, for example, questionnaires, images, previous medical record. In response to the request, the system provides a disease decision or proposes a set of appropriate diagnoses. DDSS are becoming popular day by day and is being tested in dynamic scenario.\textsuperscript{88,89}

Zhang et al.\textsuperscript{22} used tongue color gamut on 1045 images to differentiate between healthy tongue and tongue with disease and to diagnose different diseases including chronic kidney disease, diabetes nephritis, hypertension, verrucous and erosive gastritis, pneumonia, nephritic syndrome, chronic cerebral circulation insufficiency, upper respiratory tract infection, coronary heart disease, chronic bronchitis, and mixed hemorrhoid. They used k-NN and SVM to find whether the subject is healthy or not. Both of these techniques provided 91.99% classification accuracy. Later on, clustering was performed on 11 diseases with 70% accuracy. Though SVM and k-NN both show similar outcome for clustering, SVM performs better. For both cases, diabetes and hypertension classification result are poor.

Zhang and Zhang\textsuperscript{23} conducted quantitative analysis on 672 images to distinguish between healthy and diseased tongue and to diagnose diabetes mellitus, different gastritis, nephritis, and coronary heart disease. Samples were collected from Traditional Chinese Medicine Hospital but the images were classified based on their color gamut of the tongue coating.
on western medical practice. The decision tree was used to classify among five different types of tongue shapes, SVM was used to classify diseases for each tongue shape utilizing 13 geometric features, and number of features were optimized by Sequential Forward Selection (SFS). They also used k-NN but it gave poor result. The average accuracy of disease classification is 76.24%. In addition to tongue shape and geometric features, fusion of other features such as tongue color, fur, and fissure could be explored for better diagnosis of the diseases.

With an aim to diagnose diabetes, tongue image information of total 827 subjects, both diabetic and non-diabetic obtained from TDA-1 digital tongue instrument were analyzed.91 Tongue body and coating separation, color, and textural features were examined by division merging, chrominance-threshold, SVM, and few other algorithms. After sample equalization and feature normalization, 23 input parameters including information of the subject, tongue color, and texture PCA were used to reduce the dimensions of the feature. Optimizing the kernel parameter by GA, SVM showed 78.77% accuracy for diabetes prediction. Though the prediction accuracy is low, the work carries clear indication about the potential of tongue diagnosis for diabetes detection. In future, inclusion of other tongue features may increase the accuracy.

Inspired by deep convolutional neural network (CNN), a feature extraction technique named constrained high dispersal neural networks was implemented on tongue images of 315 people.92 The objective of this feature extraction was to separate healthy and unhealthy class. A weighted LIBLINEAR SVM was trained to perform the task, which showed 91.14% accuracy. The authors claimed the system to be faster than CNN; however, to be implemented in real-time, one has to be careful about using deep neural network, especially when the accuracy is still low and the system cannot provide decision on specific disease.

In recent years, researchers have been working on the ATD. But most of them are concerned only up to feature extraction stage. On the other hand, CDSS-based works paid an inadequate attention on the prior steps involved, which raises the issue of compatible standardization. Besides, the limited work done on tongue diagnosis-based disease detection is not on the same diseases. Thus, it becomes hard to compare the accuracy and repeatability of the best decision support system found in this field. Moreover, the reported works suggest the accuracy slightly under 80%, which may not be acceptable in many cases of CDSSs.

There are few works93 done on early stage breast cancer (BC) detection. Nine tongue features of both breast cancer patient (stage 0 and 1) and healthy persons were investigated in a hope to diagnose BC at an early stage to ensure timely treatment to increase the likelihood of recovery, lowering the relapse rate. The considered features are tongue color, quality, fur, fissure, red dot, ecchymosis, tooth mark, saliva, and tongue shape. The subdivision of the extracted features according to the corresponding organs such as spleen–stomach, liver–gall-left, liver–gall-right, kidney, and heart–lung area are illustrated in Fig. 6.

The data of 57 early-stage BC patients and 60 healthy people were utilized for training, whereas the data of 10 early-stage BC patients and 10 healthy people were applied for testing. The tongue features along with Mann–Whitney test were exploited to perform the logistic regression. Though the result is not good for the early stage or potential cancer patient, more feature extraction techniques as well as different features could be explored to attain better accuracy. The accuracy for healthy tongue was varied from 80% to 90%. Higher accuracy was attained using lesser number of features. The work was further upgraded without any significant change in the result.93

A summary of algorithms used in disease diagnosis is provided in Table 3.

5. Mobile-enabled tongue diagnosis schemes

In 2009, American Recovery and Reinvestment Act encouraged widespread adoption of health information technology (HIT) via Health Information Technology for Economic and Clinical Health Act (HITECH). Gradually these HITs are becoming more robust and diverse. Disease diagnosis on a mobile platform has gained the attention of researchers and developers in the last few years.95

Though available smartphone technology offers enormous prospect to employ tongue diagnosis for disease classification, very few work has been done exploiting existing means. To authors' best knowledge, only one mobile application is available in iOS and android platform called i-Tongue © University of Missouri.96 Because of openness, total interface customization, universality, security, few more tongue diagnosis applications are initiated focusing android platform than iOS.21,41,48

The research work conducted by Otsu97 and Hsu et al54 was taken under consideration by Hu et al41 for tongue region segmentation and tongue diagnosis, and the algorithm was tested in smartphones such as Samsung Galaxy S2 and S3, HTC Sensation, and J Z32 providing similar outputs and server running in PC is able to perform analysis of each image taking below 3 seconds. Hu et al pointed out that most of the mobile phone camera other than tongue with any kind of deformity this work is promising for fur and fissure detection. Though based on this work’s output, the team is working on liver disease diagnosis, and this work does not include any decision support system and thus predominantly relies on a physician’s opinion. Zhang et al21 used CCD camera to outline tongue via canny algorithm and then utilized feature extraction to diagnose diseases.

Taking teeth color as a standard, Lin98 presented a smartphone-based tongue color calibration method. 20 female subjects were chosen for the study, continued for a month. K-means algorithm was used to detect tongue coating differentiating from other body parts. However, the distinguishing method itself was unfitting to start with.

Taking lighting condition into account, Hu et al48 used a smartphone to calculate SVM-based color correction matrix and to automatically detect fur and fissure of tongue. The lighting environment included with and without flashlights. This work focused on detection accuracy, but robustness was not explored.

Like most diagnosis-based health and fitness mobile applications, iTongue © University of Missouri96 is available for both guest and registered users in google play and apple app store. The height (cm) and weight (kg) of the user has to be provided during registration. There is a series of questions users must provide answer to; otherwise no result will be generated. There are some additional questions for registered users; few of them are gender-specific. Using these answers, the application generates the result and recommendations including constitute of qi asthenia, yang asthenia,
<table>
<thead>
<tr>
<th>Algorithm/Technique</th>
<th>Ref.</th>
<th>Dataset (no. of images (I)/subject (S))</th>
<th>No. of diseases</th>
<th>Diseases</th>
<th>Other techniques used</th>
<th>Features/Context</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian</td>
<td>20</td>
<td>14,696 and S-600 56</td>
<td>544</td>
<td>9</td>
<td>Leukemia; pulmonary heart disease; abdominal tuberculosis; appendicitis; Nephritis; gastroduodenal perforation; gastritis; Pancreatitis; bronchitis</td>
<td>SVR</td>
<td>Color correction</td>
</tr>
<tr>
<td>SVM</td>
<td>22</td>
<td>&gt;9000 and S-5222a 440</td>
<td>2780</td>
<td>200</td>
<td>Chronic kidney; disease diabetes; nephritis; hypertension; verrucous gastritis; pneumonia; nephritic syndrome; chronic cerebral circulation insufficiency; upper respiratory tract infection; erosive gastritis; coronary heart disease; chronic bronchitis; mixed hemorrhoid; miscellaneous</td>
<td>k-NN</td>
<td>Tongue color gamut; Tongue color gamut; 12 tongue color categories; tongue foreground pixels</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>1-1045 143</td>
<td>902</td>
<td>13</td>
<td>Diabetes mellitus; nephritis; gastritis verrucosa; nephritic syndrome; erosive gastritis; chronic gastritis; coronary heart disease</td>
<td>k-NN: poor result</td>
<td>13 geometric features</td>
</tr>
<tr>
<td></td>
<td>91</td>
<td>827 531</td>
<td>296</td>
<td>1</td>
<td>Diabetes</td>
<td>k-NN, Naive Bayes, BP-NN, GA-SVM</td>
<td>Classification based on color and textural features</td>
</tr>
</tbody>
</table>

*Note: SVR = Support Vector Regression, k-NN = K-Nearest Neighbors, Naive Bayes, BP-NN = Backpropagation Neural Network, GA-SVM = Genetic Algorithm Support Vector Machine, Decision tree, SFS = Sequential Forward Selection, 5 tongue shape features, To reduce no of features.*
<table>
<thead>
<tr>
<th>Algorithm/Technique</th>
<th>Ref.</th>
<th>Dataset (no. of images (I)/subject (S))</th>
<th>No. of diseases</th>
<th>Diseases</th>
<th>Other techniques used</th>
<th>Features/Context</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPSS software; mothur software</td>
<td>[8]</td>
<td>S-386</td>
<td>100</td>
<td>286</td>
<td>3</td>
<td>Colorectal cancer Lung cancer Gastric cancer Breast cancer</td>
<td>Tongue coating thickness Fisher's exact test Tongue color, quality, fur, fissure, red dot, ecchymosis, tooth mark, saliva, and tongue shape. Mann-Whitney test</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>[93]</td>
<td>137</td>
<td>–</td>
<td>–</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical analysis</td>
<td>[90]</td>
<td>71</td>
<td>–</td>
<td>–</td>
<td>1</td>
<td>Aspiration pneumonia</td>
<td>Tongue plaque index (TP): tongue-coating oral bacteria Tongue color and gist features</td>
</tr>
<tr>
<td>ANN (SOM Kohonen Classifier)</td>
<td>[94]</td>
<td>100</td>
<td>–</td>
<td>–</td>
<td>1</td>
<td>Diabetes</td>
<td></td>
</tr>
</tbody>
</table>

- 2002 people were grouped in the sub-health category.
- The study included analysis of diseases tongue, and the diagnosis itself was not conducted. Diseased class diagnosed using western medicine, not TCM.
yin asthenia, phlegm-dampness, and damp-heat. Each constitute can recommend environment, exercise, emotion, and diet. The guest version cannot store any images. The application removes the image when the user tries to take another picture. There is \((x,y)\) axis to guide the user about tongue exposure in front of the phone camera. The registered users can save the history and use stored images from the phone. Image segmentation, feature extraction, and machine learning techniques are used to process the result. The user can request for manual assessment by TCM professionals with a fee. The current version 2.0.2 is an update released on Sep 14, 2016 requires minimum android 2.2 or iOS 5.1.1 and can be operated in English, Simplified Chinese, and Traditional Chinese. The update includes sharing features as well. The application is intended for an auxiliary diagnosis with questionable accuracy and reliability.

A mobile enabled system for tongue diagnosis without any human intervention is outlined in Tania et al\(^2\) (Fig. 7). Such a system would comprise image processing and machine learning algorithms to act as a standalone system. However, the implementation on mobile platform was not included in the paper.

6. Proposed framework and discussion

On the basis of our study, we have identified the research gaps to be addressed to develop an intelligent tongue diagnosis system. They are summarized as below.

- Unlike the western research convention, it is not a common practice to deposit the classified authentic tongue images in an open access database, thereby limiting the scope of improvement. There should be benchmark datasets of classified tongue images in the open access database to broaden the horizon of tongue diagnosis. Without a benchmark dataset, it is hard to perform comparable experiments, verify, validate, or improve the results.
- There is a lack of overall consistency in the ATD systems including characteristics of the input image, for example, image quality, design guideline, image segmentation, and feature extraction. Thus, the accuracy, robustness, and reliability are inevitably improbable.
- There is an infrequency of interchangeable scheme, making it incomprehensible to utilize the best found result from the prior step as an input for the subsequent step for different work.
The tongue diagnosis requires more autonomy to react to its full potential and ensure its widespread application in the digital health system. On the basis of the work done in the literature, a conceptual diagram is illustrated in Fig. 8 as a completely ATD system using CDSS, independent of human expert involvement. The idea is to incorporate disease detection with extracted features from tongue diagnosis in a mobile enabled platform (Fig. 7).

The commercial health applications, for example, iPhone health app usually provides provision for manual input regarding symptoms. The manual input of disease symptoms can be based on selected diseases. Any of the standard questionnaire set, for example, TCM, Kampo can be utilized. A polythetic approach on the mobile platform should consider the symptoms provided by the user with the color, texture, and geometric features.

With the improvement of hardware, preprocessing requirements are reducing day by day. The commercially available mobile device possesses camera resolution as high as 12–16 megapixel, with F1.7 lens. Higher focus and built-in automatic adjustment features are narrowing down, and the necessity of light estimation and color correction is abridging or nullifying. Moreover, a strong image processing algorithm can often manage the lighting condition by itself. The extracted features (e.g., fur, texture, shape), subsequent to image segmentation, can be utilized for classification. All these information, including the symptoms, need to be fused together by a fusion algorithm to provide the final result. The conceptual diagram can be trained offline and the whole diagnosis system can be deployed as a smartphone application to avail the system as a homecare or mobile setting (Fig. 7).

7. Conclusion

This paper explored the standardized and automated tongue diagnosis systems to demonstrate its potential. The investigation is in favor of tongue diagnosis to make it an integrated part of auxiliary diagnosis system within the primary healthcare system and an early diagnostic tool. Because the chromatic, textural and geometric shape based analysis of tongue projects indispensable information about health status. Remarkable work has already been done on different steps, for example, tongue image segmentation, background–foreground separation, healthy–unhealthy classification, and few specific disease-related works. This work explored each of these steps in the literature, especially focusing on the algorithms. Identifying some general trends, this section broadly summarizes the research gap.

While evaluating the prospect of tongue diagnosis, the absence of CDSS was noticeable. Our study indicated that the ATD embracing a CDSS can open the door for standalone systems, rather than relying on limited TCM experts only. The recent advancement in the field of camera of smart devices, classifiers, and CDSS algorithms present the opportunity to develop a real-time framework and standard protocol for everyone to be followed so that collective knowledge can be built on existing one. However, it was also noticeable from our survey that the transition from one step to another is incomprehensible due to ‘casual’ diversity in practice. Thus, the need for standardization is inevitable from the literature, as everyone is defining their own system for various steps but in most cases, there is no way to put them together. To proceed with the research on tongue diagnosis, undeniably there is a need for an open access authentic dataset with classified images of tongue.

The volume of work done in the literature does not combine the overall system to make it human intervention-free. The literature reflected the lack of a complete standalone system from capturing the tongue picture to attain the disease diagnosis. On the basis of the literature review, this work proposed an integrated framework to develop such a system. The derivation of general recommendations may help those who intend to bring tongue diagnosis into modern technology enabled practice.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments

The research work has been funded by Erasmus Mundus Partnerships Action 2 “FUSION” (Featured eEurope and South Asia mObility Network) project of Erasmus Mundus. Grant reference number: 2013-32541/001001.

References


