Article

Novel Image Processing Method for Detecting Strep Throat (Streptococcal Pharyngitis) Using Smartphone

Behnam Askarian [1], Seung-Chul Yoo [2],* and Jo Woon Chong [1],*

[1]   Department of Electrical and Computer Engineering, Texas Tech University, Lubbock, TX 79409, USA
[2]   School of Communication & Media, Ewha Womans University, Seoul 03760, Korea
*   Correspondence: communication@ewha.ac.kr (S.-C.Y.); J.Chong@ttu.edu (J.W.C.);
     Tel.: +82-2-3277-2240 (S.-C.Y.); +1-806-834-8392 (J.W.C.)

check for updates

Abstract: In this paper, we propose a novel strep throat detection method using a smartphone with an add-on gadget. Our smartphone-based strep throat detection method is based on the use of camera and flashlight embedded in a smartphone. The proposed algorithm acquires throat image using a smartphone with a gadget, processes the acquired images using color transformation and color correction algorithms, and finally classifies streptococcal pharyngitis (or strep) throat from healthy throat using machine learning techniques. Our developed gadget was designed to minimize the reflection of light entering the camera sensor. The scope of this paper is confined to binary classification between strep and healthy throats. Specifically, we adopted *k*-fold validation technique for classification, which finds the best decision boundary from training and validation sets and applies the acquired best decision boundary to the test sets. Experimental results show that our proposed detection method detects strep throats with 93.75% accuracy, 88% specificity, and 87.5% sensitivity on average.

Keywords: strep throat; image processing; color space; classification; smartphone

1. Introduction

According to the U.S. National Health Statistics Report, strep throat (*streptococcal pharyngitis*) is one of the main reasons for patient visits to hospital emergency departments in the U.S. [1]. Strep throat is an infection that is caused by bacteria [2]. Specifically, Group A beta-hemolytic streptococcus is the main cause of streptococcal pharyngitis in children and adults [3,4]. One of the risks of late strep throat diagnosis is rheumatic fever, which may lead to chronic rheumatic heart disease [5]. Rheumatic fever is the cause of death for approximately 320,000 patients a year globally [6,7]. Hence, early diagnosis of strep throat is crucial for preventing deaths related to rheumatic heart disease, especially in remote areas with a medical shortage. Moreover, a false diagnosis of strep throat may cause inappropriate treatment using antibiotics that would lead to bacterial resistance [8,9].

The common diagnosis method is the clinical decision utilizing the Centor score that is calculated from a set of criteria which includes coughing, fever, etc. [2,3,5,7,8,10]. However, its accuracy is less than 86% [10,11]. Throat culture is another clinical diagnosis method detecting streptococcal pharyngitis [9,11–16], which adds a sample of cells from the throat to a substance to promote the growth of the bacteria and diagnoses the disease. If bacteria grows (positive), it indicates that the patient has a bacterial infection [15]. Otherwise, the patient does not have a bacterial infection. The accuracy of this culture method for strep detection is 98% [15]. Strep throat was also diagnosed with the help of touch spray ionization mass spectrometry [14]. However, these diagnosis methods need trained physicians or specialists. Hence, timely and accessible diagnosis for all patients is still a challenge.

There have been studies which use color intensity values to detect diseases like diabetes [17,18], internal-organ diseases [19–21], or heart and kidney diseases [17,18,22–29]. These color intensity value-based methods have been combined with machine learning techniques such as naive Bayes, Bayes net, and sequential minimal optimization (SMO) [30–32]. In these studies, 21 properties were extracted from tongue color intensity values to diagnose 23 different types of diseases. Despite the capability of diagnosing different diseases using tongue color features, there exist some limitations identifying syndromes, distinguishing color features, and classifying the diseases [17,22–24]. For example, Zhang and Kim et al., concluded that different light conditions, color spaces, and devices can make the fore-mentioned methods to be less reliable in diagnosing corresponding diseases [17,33,34]. Even though there have been studies on smartphone-based tongue color analysis for medical diagnosis [34,35] as mentioned above, to the best of the authors' knowledge, there has been no research on smartphone-based strep throat detection using color analysis.

In this paper, we propose a novel and robust throat color analysis technique using YCbCr color space and least square estimation-based color correction method with images obtained from the smartphone camera to detect strep throat. Our proposed method uses an add-on gadget which helps to acquire throat images in an accurate manner. The YCbCr color space separates the luminance factor from the color space and makes it independent of luminance changes to detect the region of interest (ROI). The novel color correction method copes with different sensors and chroma variations to provide a unified color space. For classification, the *k*-NN classifier was adopted to distinguish healthy and diseased throat. As a result, the proposed method provides detection of strep throat with the images captured by the smartphone camera. The rest of this paper is organized as follows: Section 2 describes data collection and feature extraction. Section 3 describes the results from our proposed method, and finally Section 4 concludes the paper.

2. Materials and Methods

Strep throat symptoms are inflammations, red spots on the back of the throat, and enlarged tonsils, which are shown in Figure 1b [36]. In this paper, we propose a smartphone-based strep throat detection method, which classifies strep throats from healthy throats using the image features shown in Figure 1. The classification of our proposed method is confined to binary classification between strep and healthy throats. Data acquisition required for testing the proposed method is explained in Section 2.1 while the proposed strep detection method consisting of (1) preprocessing, (2) feature extraction, and (3) classification is described in Sections 2.2–2.4, respectively.

(a)   (b)
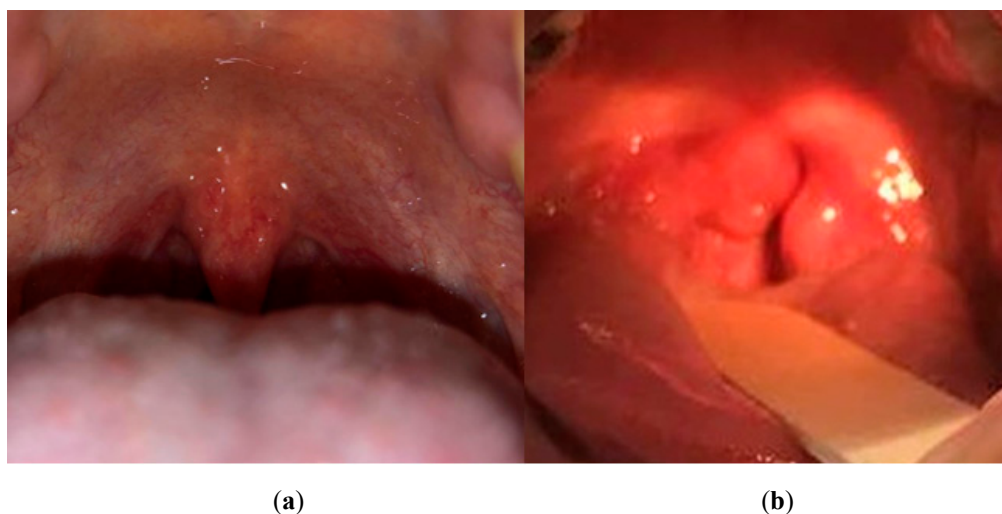
Figure 1. Example of a healthy and strep throat. (**a**) Healthy throat where there is no sign of any infection, and (**b**) strep throat where there are red swollen uvula and tonsils, and whitish spots on the throat.

2.1. Data Acquisition

We recruited 56 subjects following the Texas Tech University Institutional Review Board (IRB) (IRB#: IRB 2018-701). The subjects (56) consisted of 28 healthy and 28 strep throat-diagnosed subjects whose ages were in the range of 20 to 38 years old. Among 56 subjects, 31 were male and 25 were female. Subjects were asked to sit in a relaxed position without any movement and instructed to open their mouths widely. At that moment, experimenters captured subjects' throat images using a smartphone camera. We used the iPhone X rear camera and set the resolution of the camera to its maximum resolution at 12-megapixels (4032 x 3024 pixels). We used the autofocus function of the iPhone X and turned the light emitted diode (LED) flashlight on during the image acquisition.

Figure 2 shows our developed add-on gadget and its usage with the iPhone X. We designed and manufactured this add-on gadget customized to iPhone X using a 3-D printer. This gadget made the smartphone's flashlight shine on the throat in a bright and uniform way. Moreover, it eliminated the effect of ambient light, minimized tongue movement, and prevented the tongue from blocking the throat, Figure 2.

(a)   (b)

Figure 2. Our developed add-on gadget and its usage for data acquisition. (**a**) Add-on gadget designed and manufactured by 3D printing, and (**b**) image acquisition setup using the iPhone X with the add-on gadget.

2.2. Preprocessing

The preprocessing step is needed for accurate and effective feature extraction in throat images. Two main parts of the preprocessing steps are (1) color correction and (2) image segmentation. Color correction is required to derive the output image independent from the color space since each smartphone camera has its own color space parameters [37]. On the other hand, image segmentation is required to extract a region of interest (ROI) from the input raw image since images taken by the smartphone camera may include other parts of the inner mouth (soft palate and teeth, lips, etc.).
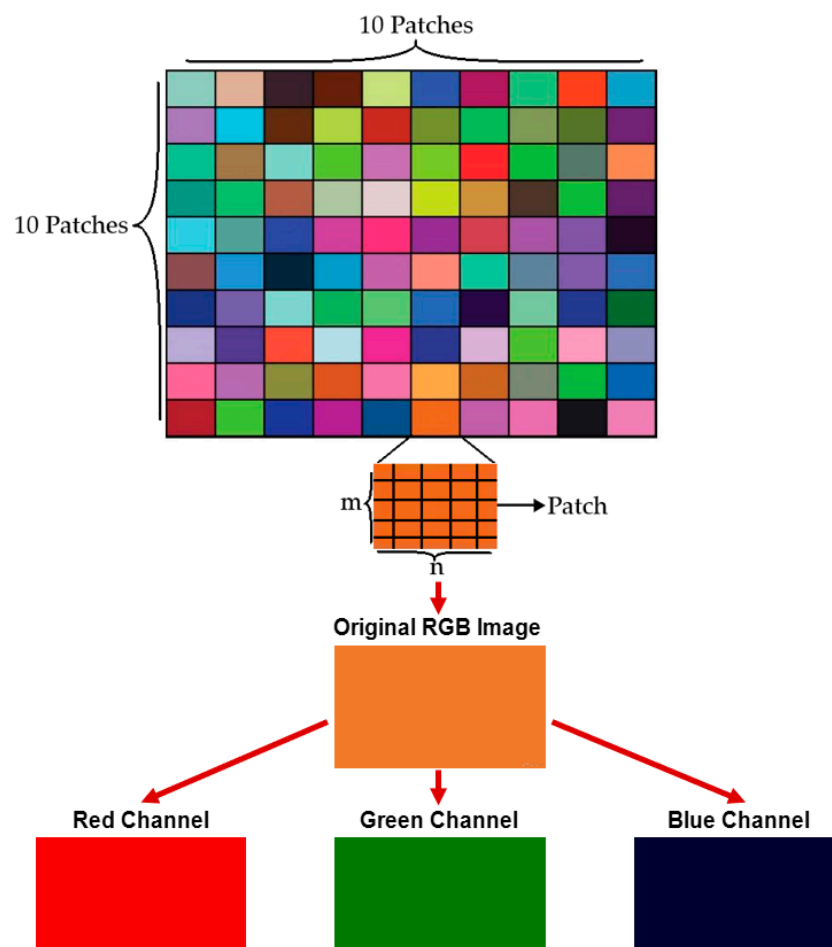
2.2.1. Color Correction

For color correction, we adopted the least square estimation-based color correction method [38], which calculates color correction matrix A based on least-square estimation toward the reference color. We generated the color chart having 100 color patches (10×10 color patches) using MATLAB as shown in Figure 3 [39], and took a picture of the color chart using a smartphone. The two-dimensional original image and its processed image are represented by O and P matrices, respectively, which are $i \times 3$ matrices where i is the number of patches and 3 comes from the number of color channels containing

R, G, B (red, green, blue) color channels (see Equation (1) below). Here, each patch consists of *m* rows (height) × *n* columns (width) pixels as shown in Figure 3.

$$O = \begin{bmatrix} O_{1R} & O_{1G} & O_{1B} \\ O_{2R} & O_{2G} & O_{2B} \\ \vdots & \vdots & \vdots \\ O_{iR} & O_{iG} & O_{iB} \end{bmatrix}, \quad P = \begin{bmatrix} P_{1R} & P_{1G} & P_{1B} \\ P_{2R} & P_{2G} & P_{2B} \\ \vdots & \vdots & \vdots \\ P_{iR} & P_{iG} & P_{iB} \end{bmatrix}. \tag{1}$$

Here, the individual terms in the $i \times 3$ image matrices O and P are denoted by O_{xy} and P_{xy}, respectively, where x varies in the range from 1 to i and y may be R, G, or B. O_{xR}, O_{xG}, and O_{xB} are the red, green, and blue intensities of the xth original image patches, and P_{xR}, P_{xG}, and P_{xB} are the red, green, and blue intensities of the processed image patches, respectively.

Figure 3. Color correction step. Color chart with 100 patches (10 × 10 patches) were generated. Each patch inside the color chart is an image with $m \times n$ pixels. Each patch is presented in its R, G, B color channel components.

Denoting by A the color correction matrix, O can be expressed by A and P as follows:

$$O = \begin{bmatrix} O_{1R} & O_{1G} & O_{1B} \\ O_{2R} & O_{2G} & O_{2B} \\ \vdots & \vdots & \vdots \\ O_{iR} & O_{iG} & O_{iB} \end{bmatrix} = \begin{bmatrix} 1 & P \end{bmatrix} A = \begin{bmatrix} 1 & P_{1R} & P_{1G} & P_{1B} \\ 1 & P_{2R} & P_{2G} & P_{2B} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & P_{iR} & P_{iG} & P_{iB} \end{bmatrix} \begin{bmatrix} A11 & A12 & A13 \\ A21 & A22 & A23 \\ A31 & A32 & A33 \\ A41 & A42 & A43 \end{bmatrix}, \tag{2}$$

where **1** denotes the column vector consisting of *i* rows of 1s. By adding column **1** to *P*, a DC offset is added. Due to the appended **1** column vector with the matrix *P*, A_{11}, A_{12}, and A_{13} were added in *A* to determine the optimal color offset. The product of xth row of the processed image $(1, P_{xR}, P_{xG}, P_{xB})$ and the first column of matrix *A* $(A_{11}, A_{21}, A_{31}, A_{41})$ becomes O_{xR}. Similarly, O_{xB} (or O_{xG}) is can be expressed by the product of xth row of matrix *P* and the second (or the third column) of matrix *A*. Color correction matrix *A* is calculated using the following equation [38]:

$$A = \left(\begin{bmatrix} 1 & P \end{bmatrix}^T \begin{bmatrix} 1 & P \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & P \end{bmatrix}^T O, \tag{3}$$

where $[\cdot]^T$ stands for the transpose of a matrix. The color correction of 10 patches are presented in Figure 4. In Figure 4, $(\cdot, \cdot)$ below each tick label on the *x*-axis indicates the location of the patch. e.g., (1,2) indicates the patch located at the 1st row and 2nd column. The corrected color values (gray bar) from the iPhone X color value (orange bar) became similar to the reference values (blue bar) after the color correction step as shown in Figure 4. The output examples obtained by this color correction step of our proposed method are shown in Figure 5.
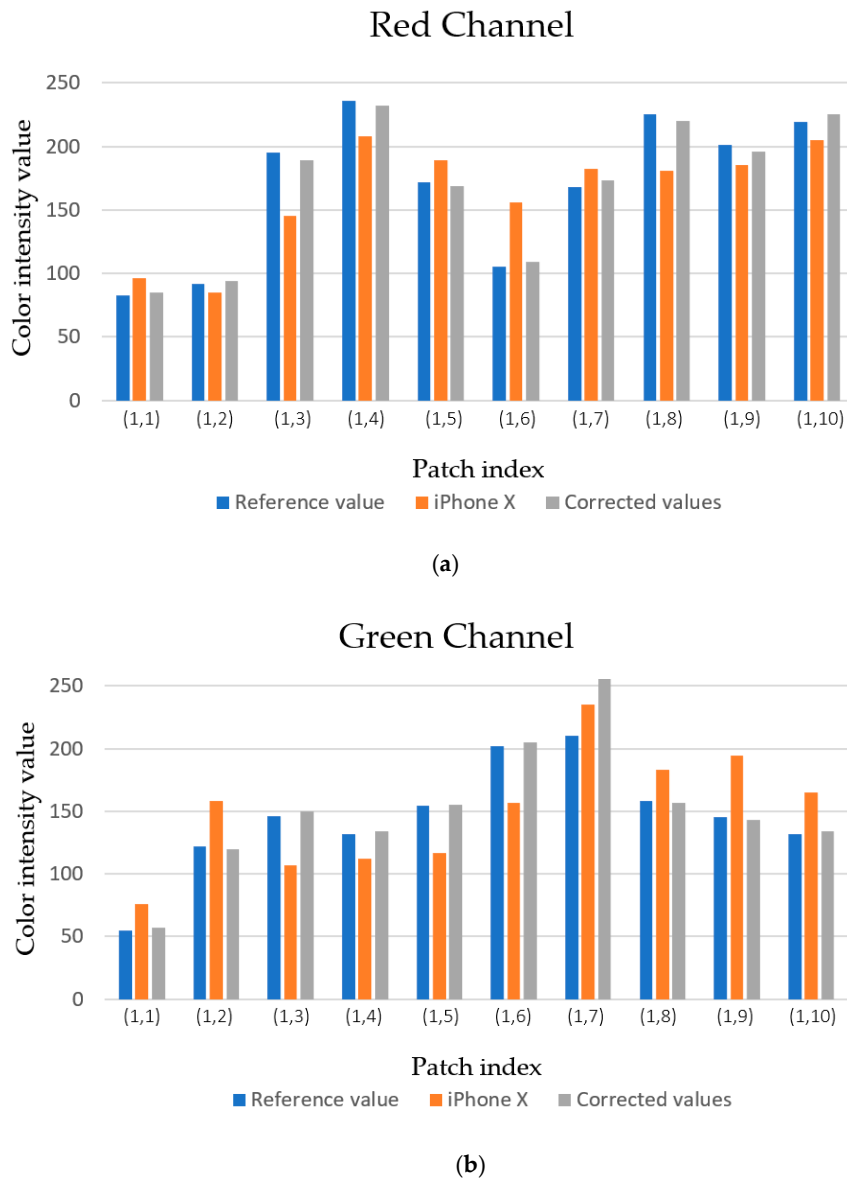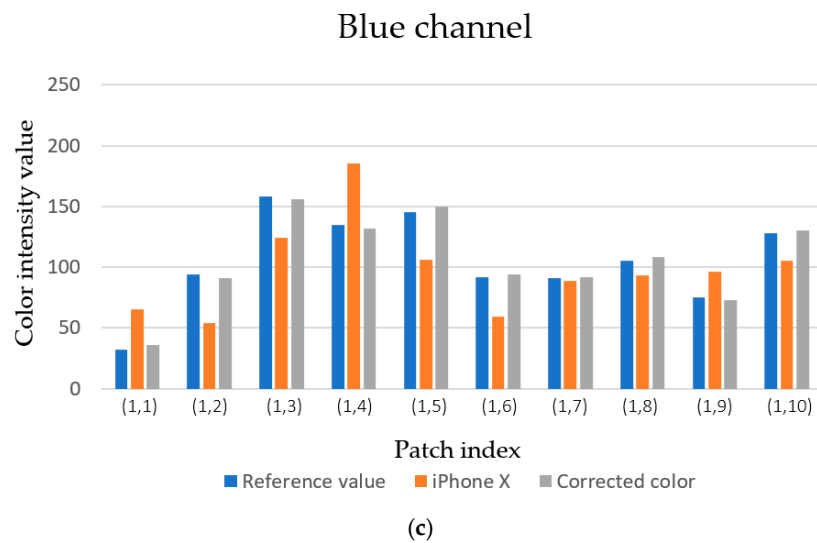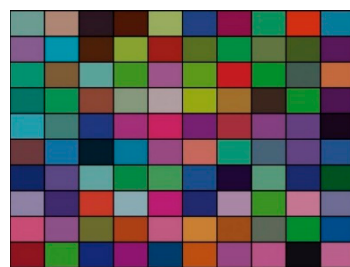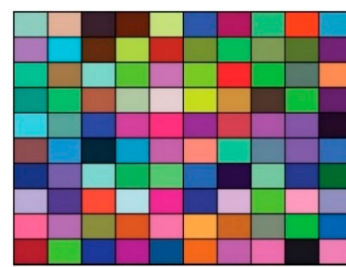

(**a**)

(**b**)

Figure 4. *Cont.*

Blue channel

(**c**)

Figure 4. Comparison of R, G, and B values among reference, iPhone X, and color-corrected patches. Here, 10 color chart patches are chosen from 100 patches. (**a**) Red channel values, (**b**) green channel values, and (**c**) blue channel values.

(**a**)

(**b**)

(**c**)

(**d**)

Figure 5. Output examples obtained by the color correction step of our proposed method. (**a**) Original image of the color chart, (**b**) corrected image of the color chart after applying our color corrected method, (**c**) original throat image which is presented by the color space in (**a**), and (**d**) color corrected throat image.
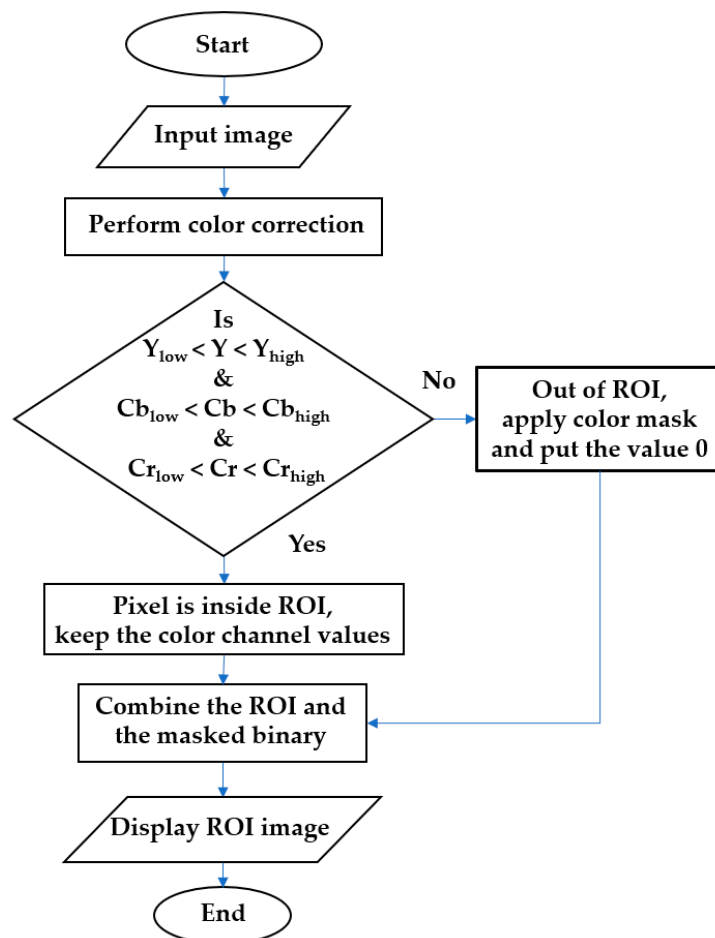
2.2.2. Image Segmentation

In the throat images acquired by the smartphone, there were five regions: (1) tongue, (2) palate, (3) lip, (4) teeth, and (5) throat tissue of the inner mouth. The image segmentation step is aimed at acquiring only the throat tissue region, which is the ROI in this paper, among the five regions in the input image. Since the color of the ROI was different from the other regions, we used the color intensity

thresholding algorithm to find the ROI [40]. Specifically, we converted a raw RGB image obtained from the smartphone into a YCbCr image. Next, we extracted Y, Cb, and Cr channels, and finally, applied threshold values into each channel to find the ROI. Figure 6 shows the flowchart of the proposed color intensity thresholding algorithm to extract the ROI. The color intensity values of Y, Cb, and Cr channels were extracted from the color corrected image obtained in Section 2.2.1. We set the color intensity threshold values of Y, Cb, and Cr channels considering the ranges of color intensity values of ROI's Y, Cb, and Cr channels. Specifically, the minimum and the maximum values of ROI's Y, Cb, and Cr color intensity values were extracted to determine the corresponding threshold values of each channel. Denoting by Y_{low}, Cb_{low}, and Cr_{low} low threshold values of ROI's Y, Cb, and Cr channels and denoting by Y_{high}, Cb_{high}, and Cr_{high} high threshold ones, the pixels which satisfied the following conditions are considered to constitute the ROI. Otherwise, the other pixels were considered to constitute non-ROI region as shown in Figure 6.

$$R_a(r,c) = \begin{cases} R_b(r,c) & \text{if } Y_{low} < Y < Y_{high}, \ Cb_{low} < Cb < Cb_{high}, \ Cr_{low} < Cr < Cr_{high} \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where $R_b(r,c)$ and $R_a(r,c)$ are color intensity values at the pixel location at r^{th} row and c^{th} column before and after the image segmentation step, respectively. Figure 7b shows an example of the ROI selection obtained by the image segmentation step of our proposed method on the throat image of Figure 7a.

Figure 6. Flowchart of the image segmentation step of our proposed method. The image segmentation step extracts the region of interest (ROI) from the original image.

(**a**)

(**b**)

Figure 7. The example output obtained by the image segmentation step of our proposed method. (**a**) Original image, and (**b**) the ROI extracted from the original image in (**a**).

2.3. Feature Extraction

Strep throat symptoms include red spots on the roof of the mouth, red and swollen tonsils, and white and yellow dots on the tonsils and the back of the mouth. These symptoms are the indications of bacterial inflammation. Hence, our proposed method extracts these features to detect strep throat symptoms [12,13,41]. Our method was designed and implemented to only distinguish strep throats from healthy ones. We first introduced throat color gamut and throat color features. We then used these color features to distinguish the strep throat images from healthy ones. All possible colors representing the throat surface are mainly distributed in the red and blue boundaries of Figure 8 [42]. The blue one provides the tighter boundary which covers almost 98% of the points of the throat surface. The colors that exist inside the blue boundary are the colors in the YCbCr range of the ROI mentioned in Section 2.2.
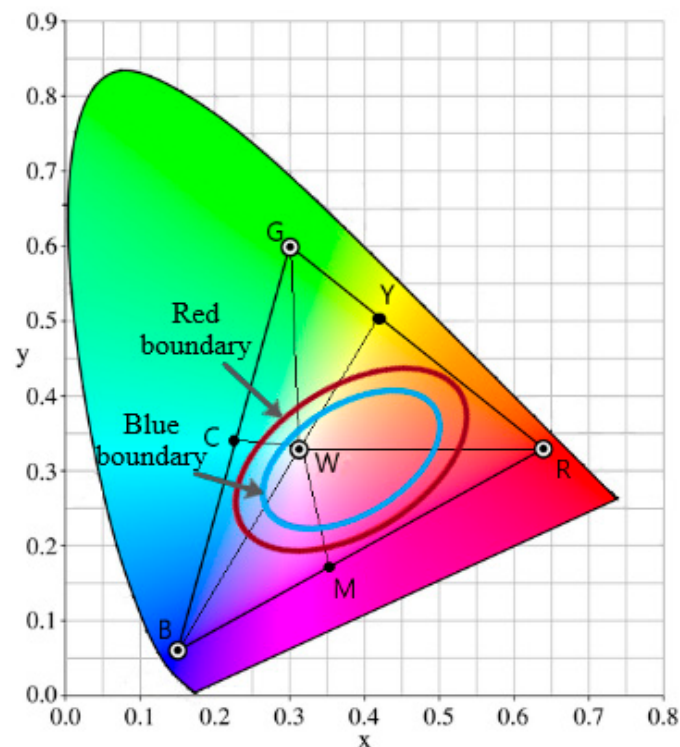

Figure 8. Throat color gamut. Over 98% of the throat color is in the blue boundary. The R, G, B, Y, C, M, and W stands for pure red, green, blue, yellow, cyan, magenta, and white colors, respectively [42].

2.4. Classification

We applied the *k*-NN classifier to distinguish strep throats from healthy throats since it is widely used in various fields such as medical imaging for brain tissue segmentation, MRI (magnetic resonance imaging) image classification, skin and breast cancer cell classification, and tongue image classifications due to its accuracy, fastness, and simplicity [43–46]. The *k*-NN classifier has also been shown to be compatible with running on smartphones [47]. We divided 56 data sets into 40 training and 16 test sets. This division was done in a random way to avoid bias [48,49]. Forty training sets consisted of 20 healthy subject images and 20 strep throat images. For the validation step, we adopted a *k*-fold cross-validation technique to prevent over-fitting. Specifically, we adopted 10-fold cross-validation which divided the data set into ten subsets and iteratively trained the algorithm on 9 folds while using the remaining fold as the validation set. Hence, the algorithm was trained on 9 folds (36 subjects) and the remaining set (four subjects) was left out for validation. This step was repeated for 10 turns (iterations) as shown in Figure 9. As a result of the 10-fold validation, we found the optimal parameter value *k* of the *k*-NN classification algorithm. As mentioned, 16 subjects (eight from healthy class and eight from diseased class) were left out for the test data set. We applied the decision boundary determined by this optimal parameter to the 16-test data set shown in Figure 9.
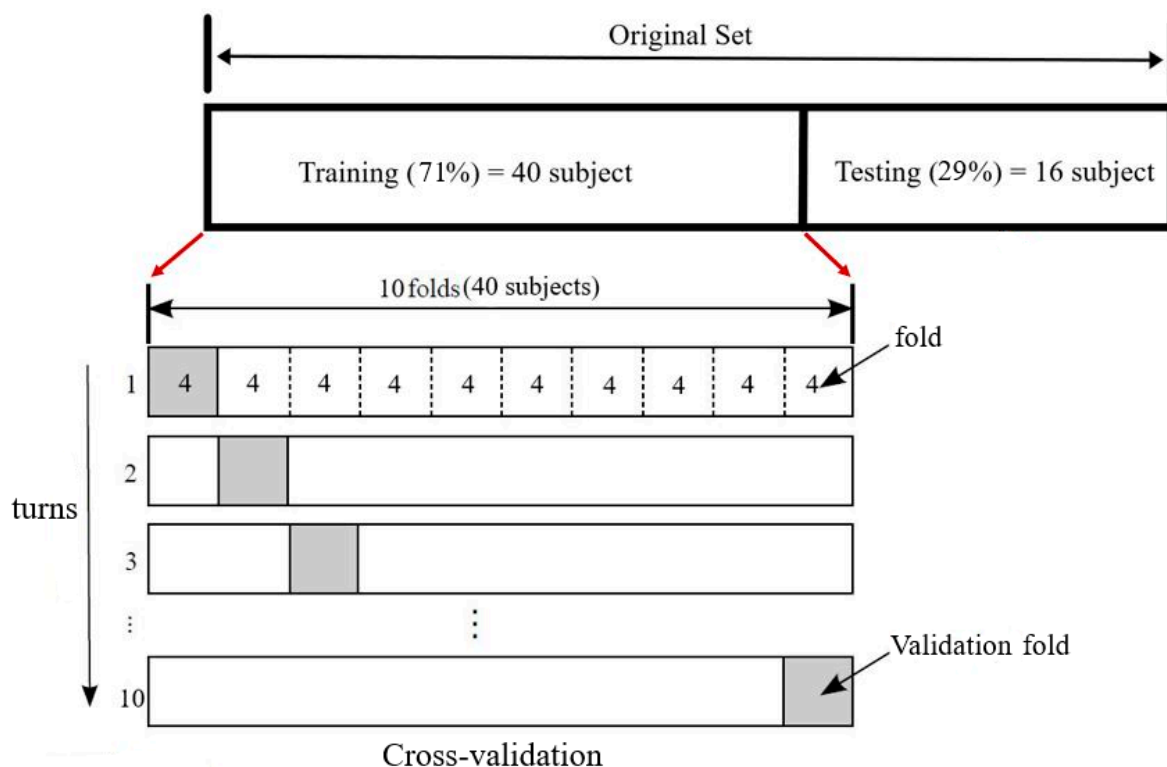

Figure 9. 10-fold cross-validation technique used in our proposed method. The original data set was split into training (71%) and testing (29%). We applied 10-fold cross-validation to the training data set by dividing it into 10 folds (each fold contained four subjects). Specifically, 9 folds were used for training and the remaining 1-fold was used for validation. The cross-validation step was repeated 10 turns, rotating the training and validation folds.

3. Results

We evaluated the performance of our proposed smartphone-based strep throat detection method by calculating accuracy, sensitivity, and specificity when the detection algorithm was applied to throat images of 56 subjects. We derived the color gamut of the throat area where three color features Y, Cb and Cr were extracted. The histograms of Y, Cb and Cr components values of healthy and strep throats

are shown in Figure 10a,b, respectively. The mean values of the color components (channels) for the healthy throat and strep throats were derived and represented in Table 1. Figure 11 shows the color distribution of the Y, Cb, and Cr color channels. The distribution of Y-Cb color channels is shown in Figure 11a while the color distribution of the Cb-Cr channels is shown in Figure 11b. As shown in Table 1 and Figure 11, Cb values are similar between healthy and strep throats while Y and Cr values were noticeably different between healthy and strep throats.
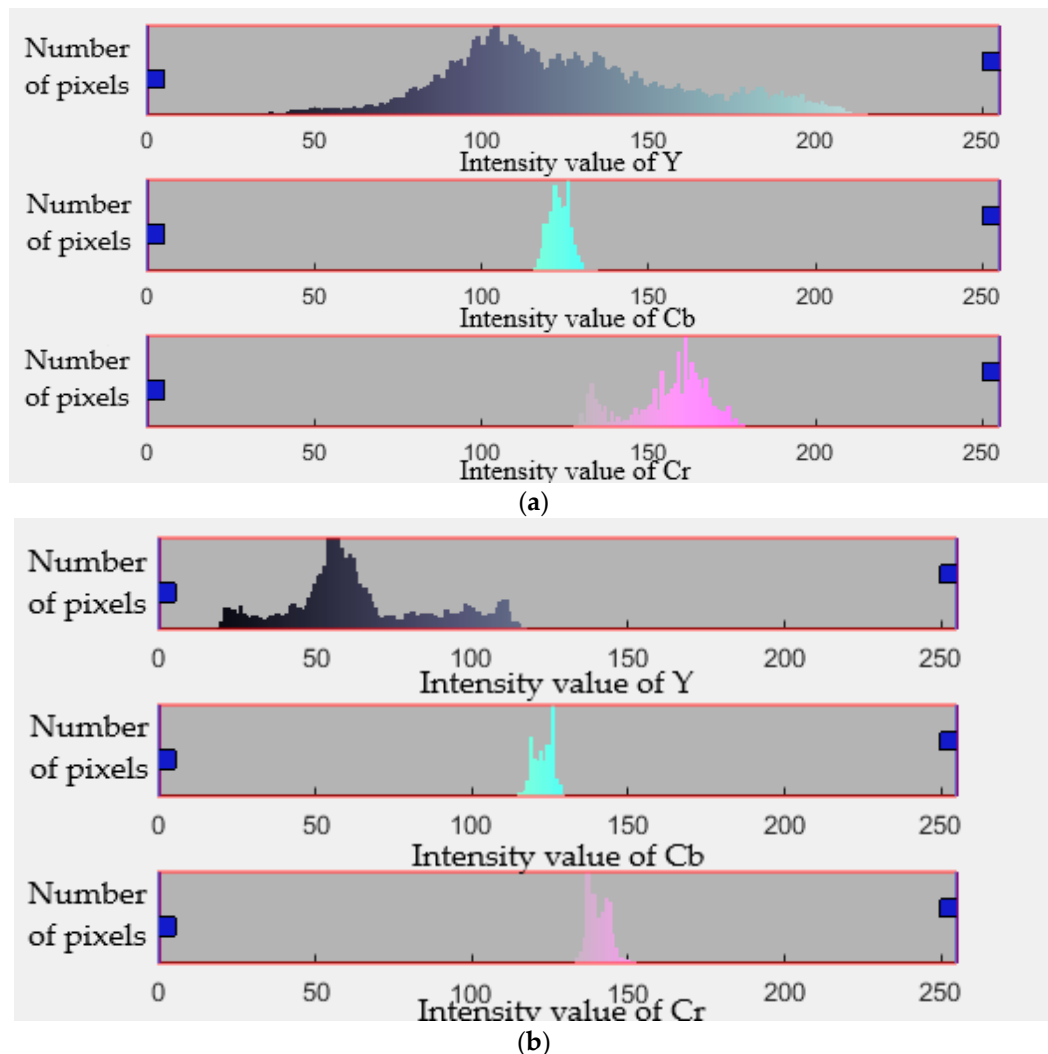
Figure 10. Histograms of Y, Cb and Cr color features of the acquired images. (**a**) Healthy throat color component histograms, and (**b**) diseased throat color component histograms. Here, the x-axis shows the intensity value of each color channel while the y-axis shows the number of pixels.

Table 1. Mean, standard deviation, and the range of the color intensity values of healthy and strep throats.

Color Channel	Y	Cb	Cr
Healthy (Mean ± STD)	133.5 ± 12	127 ± 5	168.5 ± 11
Diseased (Mean ± STD)	97 ± 5	137 ± 6	141 ± 8
Healthy (range)	122–145	112–142	155–185
Diseased (range)	92–103	118–132	135–147

(a)

(b)
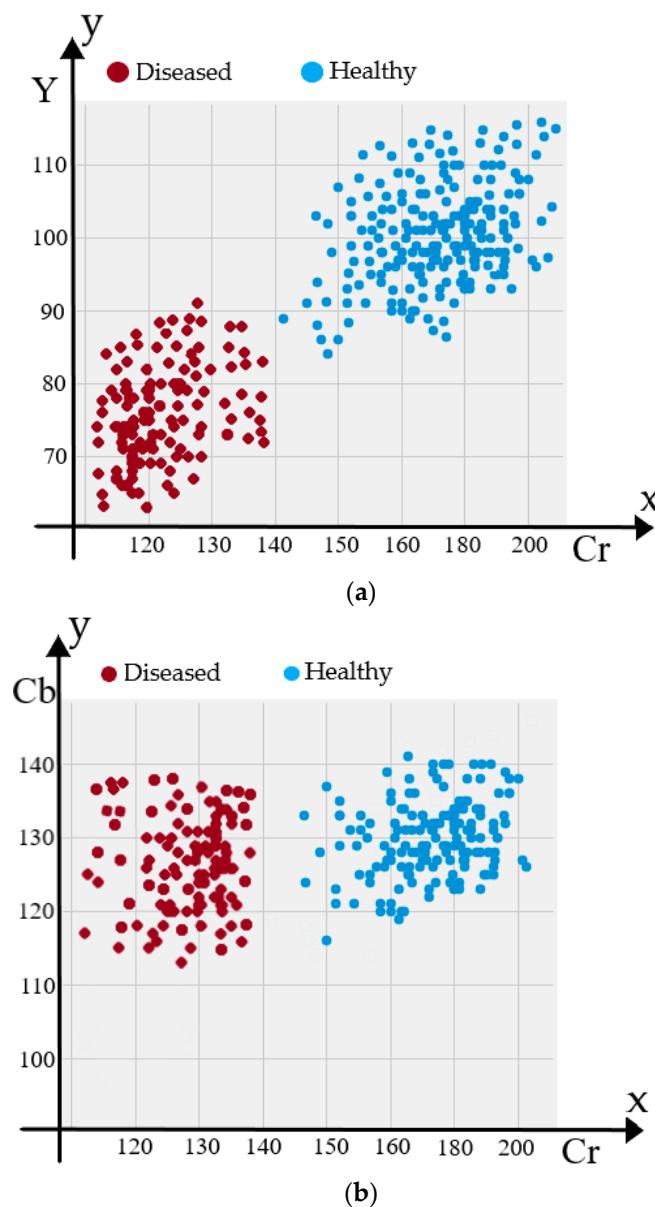
Figure 11. Color distribution of different color channels in healthy and diseased throats: (**a**) Y and Cr color intensity distribution of healthy and strep throats, and (**b**) Cb and Cr intensity distribution of healthy and strep throats.

Figure 12 shows an example of the strep detection procedure. The acquired RGB image is shown in Figure 12a. Figure 12b shows the YCbCr image converted from RGB image in Figure 12a. Figure 12c,d show the infected tissue detected in Figure 12b and in white colors, respectively. The colors that we were seeking as symptoms of the strep throat have been in Figure 12. The strep tissue are indicated by A, B, C, and D symbols in Figure 12 and the color intensity values of the infected tissue have been represented in Table 2. A paired-*t* test was performed to compare the average Y, Cb, and Cr values of healthy and diseased throats. The significant difference test was performed on the parameter value $YCbCr_{avg} = \frac{Y+Cb+Cr}{3}$ which has been proven to be effective in distinguishing healthy and diseased tissue with bacterial infection [17,32,34]. The paired-*t* test indicated that the $YCbCr_{avg} = \frac{Y+Cb+Cr}{3}$ from the healthy throat (mean = 146.3, STD = 6.8) was significantly higher than diseased ones (mean = 124.4, STD = 5.1) with p=0.04. Specifically, the values of mean difference and standard deviation of difference were 21.9 and 5.6, respectively.
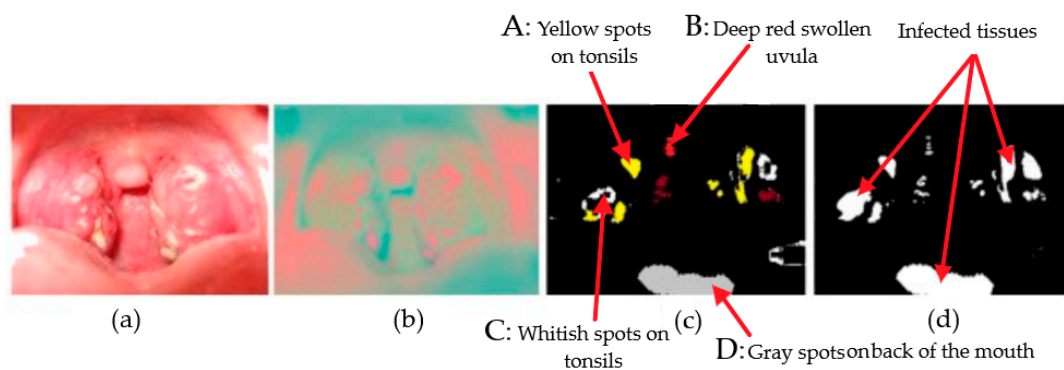
Figure 12. Strep detection procedure based on YCbCr components of a throat image: (**a**) Original RGB image, (**b**) YCbCr image converted from the RGB image in (**a**), (**c**) infected tissue detected in (**b**), and (**d**) infected tissue marked in white color.

Table 2. Mean and standard deviation of $YCbCr_{avg}$ values in the A, B, C, and D regions from all healthy and diseased throats.

Strep Throat Symptoms	Healthy $YCbCr_{avg}$ (Mean ± STD)	Disease $YCbCr_{avg}$ (Mean ± STD)
A in Figure 12	154 ± 6.8	141 ± 4.3
B in Figure 12	165 ± 7.6	143 ± 5.1
C in Figure 12	136.2 ± 4.4	152.6 ± 6.7
D in Figure 12	151.2 ± 6.6	134.6 ± 5.4

We divided the data (56 subjects) into a training and validation set (40 subjects), and a test set (16 subjects). Here, for the training and validation set (40 subjects), 20 healthy and 20 strep subjects were randomly chosen from the total 56 subject data to avoid biasing [48]. As a result of 10-fold validation, we found the optimal k value for the k-NN classifier is 13 since it gives the highest accuracy as shown in Figure 13. We applied the decision boundary determined by this optimal k value ($k = 13$) to the test data set (16 subjects).
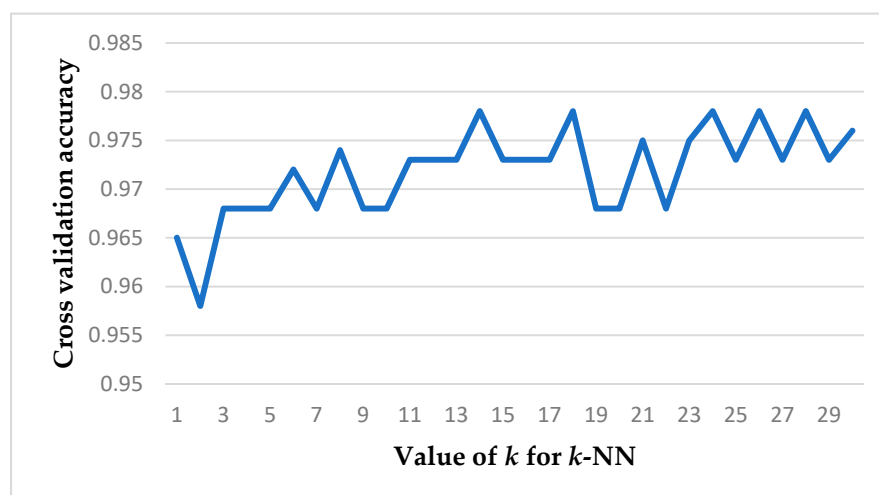
Figure 13. Cross-validation accuracy for varying k values of the k-NN classifier from 1 to 30. As k value increases, the accuracy value of cross validation increases while the processing takes more time. The optimal k value was achieved at $k = 13$ in terms of cross validation accuracy and processing time.

As performance metrics, we considered accuracy, sensitivity, and specificity which were calculated using true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values as follows:

$$Sensitivity = \frac{TP}{TP + FN} \times 100\%, \tag{5}$$

$$Specificity = \frac{TN}{TN + FN} \times 100\%, \tag{6}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%, \tag{7}$$

where *TP*, *FP*, *TN*, and *FN* were counted in terms of the number of images. Since the scope of this paper was confined to binary classification between strep and healthy throats as mentioned in Section 2, *TP*, *FP*, *TN*, and *FN* were calculated considering this binary classification. That is, *TP* is the number of images which were correctly determined to be strep given that they are strep, and *FP* is the number of images which were incorrectly determined to be strep given that they are healthy. On the other hand, *TN* is the number of images which were correctly determined to be healthy given that they are healthy, and *FN* is the number of images which were incorrectly determined to be healthy given that they are strep.

The average accuracy of the 10-fold cross-validation was calculated by averaging the accuracy values of all turns (iterations) of the cross-validation. Table 3 shows the average accuracy, sensitivity, and specificity values of the proposed algorithm. The average and standard deviation value of the cross-validation accuracy was 97.8% ± 0.014% as shown in Table 3. We applied the decision boundary obtained from this 10-fold cross-validation into the test data set (8 healthy and 8 strep throat images). As a result, we obtained 93.75% accuracy, 87.5% sensitivity, and 88% specificity, for the test dataset as shown in Table 3.

Table 3. Average accuracy, sensitivity, and specificity values of the proposed method.

Cross Validation Accuracy (Mean± STD)	Average Test Accuracy	Average Test Sensitivity	Average Test Specificity
0.978 ± 0.014	0.9375	0.875	0.88

Figure 14 shows example outputs of our proposed method on one healthy throat and one strep throat. Figure 14a is the original image from the healthy throat and Figure 14b is the result of our method on the healthy throat. Figure 14c is the original image from strep throat and Figure 14d is the result of our method on the strep throat. Infected tissue are detected in the strep throat as shown in Figure 14d while those are not detected in the healthy throat as shown in Figure 14b.
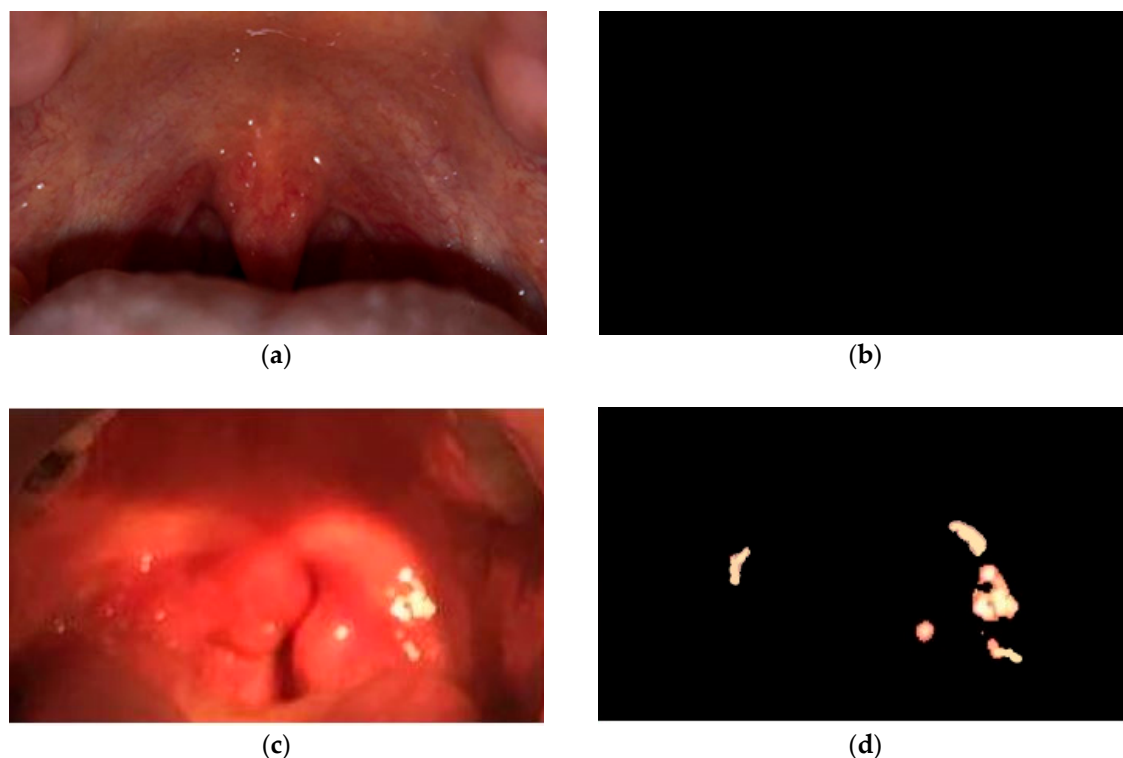
Figure 14. Detection of strep throat tissue using the proposed method. (**a**) Original image of a healthy throat, (**b**) the throat in (a) is diagnosed to be healthy (no infected tissue is detected in the image), (**c**) original image of a diseased throat, and (**d**) the throat in (c) is diagnosed to be infected (infected tissue is marked in bright color).

4. Conclusion and Discussion

In this paper, we have investigated the plausibility of using a smartphone to detect strep throat by evaluating our developed smartphone-based strep throat detection method on subjects' throat images taken by a smartphone camera. We recruited 56 subjects consisting of 28 strep and 28 healthy subjects, acquired subjects' throat images using an iPhone X, and tested our method on them. The aim of the proposed method was to find symptoms (color features) that indicate the signs of streptococcal pharyngitis in the throat. To improve the performance of our proposed method, we designed and manufactured an add-on gadget to control the lighting conditions and avoid ambient light and reflection. We proposed the use of color intensity thresholding techniques to segment throat tissue from a throat image. In this paper, a novel least square color correction method and YCbCr color space that is luminance-independent (by extracting Y channel) has been proposed. The color intensity thresholding technique has been applied and evaluated in detecting tongue color as well [50]. However, they had different approaches in evaluating their color intensity-based techniques. For example, a support vector machine (SVM) was adopted as a classifier to distinguish diseased subjects from healthy ones in Refs. [17,31–34,44]. We adopted a *k*-NN classifier as in Refs. [31,44] and evaluated the performance using *k*-fold validation approach as in Refs. [17,32–34]. The experimental results have shown that the proposed color intensity thresholding system could segment throat image tissue in a throat image. We have simplified the categories of throat images into strep and healthy throats since the scope of this paper was not the multiclass classification of different degrees of strep (or *streptococcal pharyngitis*) but was confined to binary classification between strep and healthy throats. Cross-validation was performed to prevent overfitting. Here, 10-fold cross-validation was specifically adopted. After running 10-fold cross-validation on a range *k* from 1 to 30 for the *k*-NN classifier, the highest validation accuracy 97.8% was achieved at $k = 13$. The experimental results have shown that the proposed method detects strep throat with 97.8% average accuracy (validation score) for the 10-fold

cross-validation training data set. Using the *k*-NN classifier, the proposed strep detection method can detect strep from the throat tissue with 93.75% accuracy, 87.5% sensitivity, and 88% specificity for the testing dataset. This method can be implemented using any smartphone, including iOS or Android phones with an appropriate add-on gadget using a retargetable application platform [51]. Extending this result into classifying different degrees of strep throat and differentiating bacterial from viral infections can be considered in future work.

Author Contributions: B.A. collected the data, conceived and designed the analysis, wrote the original and revised manuscript, and conducted most details of the work. S.-C.Y. set the direction of the revised paper based on reviewers' comments; re-designed the research experiment and analysis; verified data analysis and statistical analysis; wrote the revised draft based on reviewers' comments. J.W.C. wrote the original/revised drafts; designed and re-designed the analysis; verified image data analysis, and guided direction of the work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Niska, R.; Bhuiya, F.; Xu, J. National hospital ambulatory medical care survey: 2007 emergency department summary. *Natl. Health Stat. Rep.* **2010**, *26*, 358.
2. Kalra, M.G.; Higgins, K.E.; Perez, E.D. Common Questions About Streptococcal Pharyngitis. *Am. Fam. Physician* **2016**, *94*, 24–31.
3. Choby, B.A. Diagnosis and treatment of streptococcal pharyngitis. *Am. Fam. Physician* **2009**, *79*, 383–390. [PubMed]
4. Hing, E.; Cherry, D.K.; Woodwell, D.A. National Ambulatory Medical Care Survey: 2004 summary. *Adv. Data* **2006**, *374*, 1–33.
5. Dajani, A.; Taubert, K.; Ferrieri, P.; Peter, G.; Shulman, S.; Association, A.H. Treatment of acute streptococcal pharyngitis and prevention of rheumatic fever: A statement for health professionals. *Pediatrics* **1995**, *96*, 758–764. [PubMed]
6. Watkins, D.A.; Johnson, C.O.; Colquhoun, S.M.; Karthikeyan, G.; Beaton, A.; Bukhman, G.; Forouzanfar, M.H.; Longenecker, C.T.; Mayosi, B.M.; Mensah, G.A. Global, regional, and national burden of rheumatic heart disease, 1990–2015. *N. Engl. J. Med.* **2017**, *377*, 713–722. [CrossRef] [PubMed]
7. Carapetis, J.R.; Steer, A.C.; Mulholland, E.K.; Weber, M. The global burden of group A streptococcal diseases. *Lancet Infect. Dis.* **2005**, *5*, 685–694. [CrossRef]
8. Klepser, D.G.; Klepser, M.E.; Dering-Anderson, A.M.; Morse, J.A.; Smith, J.K.; Klepser, S.A. Community pharmacist-physician collaborative streptococcal pharyngitis management program. *J. Am. Pharm. Assoc.* **2016**, *56*, 323–329. [CrossRef]
9. Spellerberg, B.; Brandt, C. Streptococcus. In *Manual of Clinical Microbiology*, 11th ed.; American Society of Microbiology: Washington, DC, USA, 2015; pp. 383–402.
10. Fine, A.M.; Nizet, V.; Mandl, K.D. Large-scale validation of the Centor and McIsaac scores to predict group A streptococcal pharyngitis. *Arch. Intern. Med.* **2012**, *172*, 847–852. [CrossRef]
11. Aalbers, J.; O'Brien, K.K.; Chan, W.-S.; Falk, G.A.; Teljeur, C.; Dimitrov, B.D.; Fahey, T. Predicting streptococcal pharyngitis in adults in primary care: A systematic review of the diagnostic accuracy of symptoms and signs and validation of the Centor score. *BMC Med.* **2011**, *9*, 67. [CrossRef]
12. Bisno, A.L. Diagnosing strep throat in the adult patient: Do clinical criteria really suffice? *Ann. Intern. Med.* **2003**, *139*, 150–151. [CrossRef]
13. Ebell, M.H. Strep throat: Point of Care Guides. *Am. Fam. Physician* **2003**, *68*, 937–938. [PubMed]
14. Jarmusch, A.K.; Pirro, V.; Kerian, K.S.; Cooks, R.G. Detection of strep throat causing bacterium directly from medical swabs by touch spray-mass spectrometry. *Analyst* **2014**, *139*, 4785–4789. [CrossRef] [PubMed]
15. Kellogg, J.A. Suitability of throat culture procedures for detection of group A streptococci and as reference standards for evaluation of streptococcal antigen detection kits. *J. Clin. Microbiol.* **1990**, *28*, 165. [PubMed]

16. Ebell, M.H.; Smith, M.A.; Barry, H.C.; Ives, K.; Carey, M. Does this patient have strep throat? *JAMA* **2000**, *284*, 2912–2918. [CrossRef] [PubMed]

17. Zhang, D.; Zhang, H.; Zhang, B. *Tongue Image Analysis*; Springer: Berlin/Heidelberg, Germany, 2017.

18. Seo, S.E.; Tabei, F.; Park, S.J.; Askarian, B.; Kim, K.H.; Moallem, G.; Chong, J.W.; Kwon, O.S. Smartphone with Optical, Physical, and Electrochemical Nanobiosensors. *J. Ind. Eng. Chem.* **2019**, *77*, 1–11. [CrossRef]

19. Gong, Y.-P.; Lian, Y.-S.; Chen, S.-Z. Research and Analysis of Relationship between Colour of Tongue Fix Quantity, Disease and Syndrome. *Chin. J. Inf. Tcm* **2005**, *7*, 45–52.

20. Li, C.H.; Yuen, P.C. Tongue image matching using color content. *Pattern Recognit.* **2002**, *35*, 407–419. [CrossRef]

21. Li, Q.; Liu, Z. Tongue color analysis and discrimination based on hyperspectral images. *Comput. Med. Imaging Graph.* **2009**, *33*, 217–221. [CrossRef]

22. Tang, J.-L.; Liu, B.-Y.; Ma, K.-W. Traditional chinese medicine. *Lancet* **2008**, *372*, 1938–1940. [CrossRef]

23. Lo, L.-C.; Chen, Y.-F.; Chen, W.-J.; Cheng, T.-L.; Chiang, J.Y. The study on the agreement between automatic tongue diagnosis system and traditional chinese medicine practitioners. *Evid.-Based Complement. Altern. Med.* **2012**, *2012*, 505063. [CrossRef] [PubMed]

24. Kim, M.; Cobbin, D.; Zaslawski, C. Traditional Chinese medicine tongue inspection: An examination of the inter-and intrapractitioner reliability for specific tongue characteristics. *J. Altern. Complement. Med.* **2008**, *14*, 527–536. [CrossRef] [PubMed]

25. Askarian, B.; Tabei, F.; Askarian, A.; Chong, J.W. An affordable and easy-to-use diagnostic method for keratoconus detection using a smartphone. In Proceedings of the Medical Imaging 2018: Computer-Aided Diagnosis, Houston, TX, USA, 10–15 February 2018; p. 1057512.

26. Chong, J.W.; Cho, C.H.; Tabei, F.; Le-Anh, D.; Esa, N.; McManus, D.D.; Chon, K.H. Motion and Noise Artifact-Resilient Atrial Fibrillation Detection using a Smartphone. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2018**. [CrossRef] [PubMed]

27. Tabei, F.; Kumar, R.; Phan, T.N.; McManus, D.D.; Chong, J.W. A Novel Personalized Motion and Noise Artifact (MNA) Detection Method for Smartphone Photoplethysmograph (PPG) Signals. *IEEE Access* **2018**, *6*, 60498–60512. [CrossRef] [PubMed]

28. Tabei, F.; Zaman, R.; Foysal, K.H.; Kumar, R.; Kim, Y.; Chong, J.W. A novel diversity method for smartphone camera-based heart rhythm signals in the presence of motion and noise artifacts. *PLoS ONE* **2019**, *14*, e0218248. [CrossRef] [PubMed]

29. Askarian, B.; Jung, K.; Chong, J.W. Monitoring of Heart Rate from Photoplethysmographic Signals Using a Samsung Galaxy Note8 in Underwater Environments. *Sensors* **2019**, *19*, 2846. [CrossRef] [PubMed]

30. Hui, S.C.; He, Y.; Thach, D.T.C. Machine learning for tongue diagnosis. In Proceedings of the 2007 6th International Conference on Information, Communications & Signal Processing, Singapore, 10–13 December 2007; pp. 1–5.

31. Pang, B.; Zhang, D.; Li, N.; Wang, K. Computerized tongue diagnosis based on Bayesian networks. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 1803–1810. [CrossRef]

32. Wang, K.; Zhang, D.; Li, N.; Pang, B. Tongue diagnosis based on biometric pattern recognition technology. In *Pattern Recognition: From Classical to Modern Approaches*; World Scientific: Singapore, 2001; pp. 575–598.

33. Zhang, H.-Z.; Wang, K.-Q.; Jin, X.-S.; Zhang, D. SVR based color calibration for tongue image. In Proceedings of the 2005 International Conference on Machine Learning and Cybernetics, Guangzhou, China, 18–21 August 2005; pp. 5065–5070.

34. Zhang, B.; Wang, X.; You, J.; Zhang, D. Tongue color analysis for medical application. *Evid.-Based Complement. Altern. Med.* **2013**, *2013*, 264742. [CrossRef]

35. Wang, Y.-G.; Yang, J.; Zhou, Y.; Wang, Y.-Z. Region partition and feature matching based color recognition of tongue image. *Pattern Recognit. Lett.* **2007**, *28*, 11–19. [CrossRef]

36. Wessels, M.R. Streptococcal pharyngitis. *N. E. J. Med.* **2011**, *364*, 648–655. [CrossRef]

37. Dang, D.; Cho, C.H.; Kim, D.; Kwon, O.S.; Chong, J.W. Efficient color correction method for smartphone camera-based health monitoring application. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, Korea, 11–15 July 2017; pp. 799–802.

38. Wolf, S. *Color Correction Matrix for Digital Still and Video Imaging Systems*; National Telecommunications and Information Administration: Washington, DC, USA, 2003.

39. MathWorks. MATLAB 2017. Available online: https://www.mathworks.com/products/new_products/release2017b.html (accessed on 19 December 2017).

40. Bhandari, A.K.; Kumar, A.; Chaudhary, S.; Singh, G.K. A novel color image multilevel thresholding based segmentation using nature inspired optimization algorithms. *Expert Syst. Appl.* **2016**, *63*, 112–133. [CrossRef]

41. Schachtel, B.P.; Fillingim, J.M.; Beiter, D.J.; Lane, A.C.; Schwartz, L.A. Subjective and objective features of sore throat. *Arch. Intern. Med.* **1984**, *144*, 497–500. [CrossRef] [PubMed]

42. File:CIExy1931.png. Available online: https://commons.wikimedia.org/wiki/File:CIExy1931.png (accessed on 24 March 2019).

43. Tsai, C.-F.; Hsu, Y.-F.; Lin, C.-Y.; Lin, W.-Y. Intrusion detection by machine learning: A review. *Expert Syst. Appl.* **2009**, *36*, 11994–12000. [CrossRef]

44. Deng, Z.; Zhu, X.; Cheng, D.; Zong, M.; Zhang, S. Efficient kNN classification algorithm for big data. *Neurocomputing* **2016**, *195*, 143–148. [CrossRef]

45. Vrooman, H.A.; Cocosco, C.A.; van der Lijn, F.; Stokking, R.; Ikram, M.A.; Vernooij, M.W.; Breteler, M.M.; Niessen, W.J. Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification. *Neuroimage* **2007**, *37*, 71–81. [CrossRef] [PubMed]

46. Rajini, N.H.; Bhavani, R. Classification of MRI brain images using k-nearest neighbor and artificial neural network. In Proceedings of the 2011 International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, India, 3–5 June 2011; pp. 563–568.

47. Medrano, C.; Igual, R.; Plaza, I.; Castro, M.; Fardoun, H.M. Personalizable smartphone application for detecting falls. In Proceedings of the 2014 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Valencia, Spain, 1–4 June 2014; pp. 169–172.

48. Borovicka, T.; Jirina, M., Jr.; Kordik, P.; Jirina, M. Selecting representative data sets. In *Advances in Data Mining Knowledge Discovery and Applications*; IntechOpen: London, UK, 2012.

49. Scott, D.W.; Terrell, G.R. Biased and unbiased cross-validation in density estimation. *J. Am. Stat. Assoc.* **1987**, *82*, 1131–1146. [CrossRef]

50. Pang, B.; Zhang, D.; Wang, K. Tongue image analysis for appendicitis diagnosis. *Inf. Sci.* **2005**, *175*, 160–176. [CrossRef]

51. Cho, C.H.; Tabei, F.; Phan, T.N.; Kim, Y.; Chong, J.W. A Novel Re-Targetable Application Development Platform for Healthcare Mobile Applications. *Int. J. Comput. Sci. Softw. Eng.* **2017**, *6*, 196–201, arXiv:1903.05783.