International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
Impact Factor (2018): 5.426

A Novel Approach on Tongue Images for Early Diagnosis of Breast Cancer and Lung Tumor

B. Saritha

Jay Shriram Engineering College, Avinashipalyam

Abstract: ***Metastatic carcinoma may be a sickness of early carcinoma that typically happens many years when the first carcinoma. carcinoma is that the most typical cancer among Iranian girls. consistent with the new statistics in Persia 6160 carcinomas square measure diagnosed within the country annually and 1063 cases result in death.Among all cancers, respiratory organ and breast may be attributed to an oversized variety of deaths worldwide. Given the assorted inconveniences and risks related to ancient diagnostic strategies, economical non-invasive detection approaches victimisation computerised strategies square measure required. With recent advances in medical biometry, notably specializing in the analysis of facial and tongue pictures to notice varied diseases, there's a scarcity of studies in tongue sub-lingual veins. Therefore, this paper can analyze tongue sub-lingual veins so as to tell apart people who square measure healthy from people who have cancer (either respiratory organ or breast). Tongue sub-lingual veins square measure initial captured employing a unambiguously designed device. Segmentation is then disbursed to separate the vein foreground pixels from its background. This facilitates feature extraction within the variety of color and pure mathematics. As for the last step, classification is performed. Experimental results on a dataset consisting of 628 healthy samples, eighty one samples with carcinoma, and 147 with carcinoma yielded a mean accuracy of eighty two.07% at healthy vs. carcinoma, and 79.23% at healthy vs. breast cancer, proving the effectiveness of the projected technique.***

Keywords: tongue sub-lingual veins; color and geometry features; cancer detection; statistical pattern recognition; medical biometrics

1. Introduction

According to studies, ten to thirty percentages of women having breast cancer and undergoing mammography has negative mammograms, i.e. are misdiagnosed. Furthermore, only twenty to forty percentages of the women who undergo biopsy has cancer. Biopsies are expensive, invasive and traumatic to the patient. The high rate of false positives motivate research aimed to enhance the mammogram images, to provide computer aided diagnostics tools that can alert the radiologist to potentially malignant regions in the mammograms and to develop tools for automated classification of mammograms into benign and malignant classes. Radiologists basically look for two types of patterns in mammography. They are micro-calcification and masses. The diagnosis result of tissue is classified into three categories: normal which represents mammogram without any cancerous cell, benign which represents mammogram showing a tumor, but not formed by cancerous cells and malign which represents mammogram showing a tumor with cancerous cells In 2015, carcinoma caused 1.69 million deaths worldwide [1] per the planet Health Organization (WHO). A similar report expressed that the amount of deaths caused by carcinoma was 571,000. each cancers stratified 1st and fifth, severally, where 8.8 million deaths were led to by cancer. this suggests one person died from carcinoma each eighteen.66 seconds and fifty five.23 seconds for carcinoma. moreover, WHO rumored [2] that China (the most thickly settled country in 2014) had 459,495 cases of carcinoma in males (ranked 1st in cancers) and in females there have been 380,560 cases of respiratory organ (193,347) and breast (187,213) cancers combined (ranked 1st and second, severally in cancers).

The common ways that to diagnose carcinoma [3] embody a body fluid microscopic anatomy, a chest X-ray, and a CT (CT) scan. A body fluid microscopic anatomy check involves analyzing a sample of secretion below a magnifier to appear for cancer cells. However, the method is time overwhelming because it is usually recommended to gather the secretion for 3 consecutive days (in the first morning). A chest X-ray is generally the primary medical examination conducted if carcinoma symptoms arise. If there area unit signs of cancer, additional in depth testing are going to be disbursed. A CT scan produces a additional elaborate read than a chest X-ray. These scans will show the position, size, and form of tumors within the lungs. That being same, the procedure emits a additional powerful dose of radiation than X-ray. A diagnostic technique [4] is usually wont to screen for carcinoma and involves employing a low energy X-ray to look at the human breast. However, there area unit some drawbacks with this procedure as well as the exposure to low-dose radiation, and therefore the accuracy of the diagnostic result, that depends on the expertise and talent of the specialist.

Given the importance of the 2 cancers and its current diagnostic problems, there's associate pressing got to develop non-invasive, reliable, and economical strategies to find these 2 diseases. Recently, researchers have developed computerised techniques to non-invasively find various diseases (diabetes, liver disease, rubor, and heart disease) [5-10]. This was disbursed by analyzing the options extracted from facial [7-10] and tongue [5-6] pictures returning from morbid people. the thought comes from medical bioscience [11], wherever distinctive and distinguishable options is extracted from a specific health problem. The notion of analyzing the face and tongue was borrowed from ancient medicines [12-13], that believed the standing of the interior organs is mirrored on the face and tongue. despite the fact that important results were achieved via the analysis of facial and tongue options, there exists very little to none within the literature once it involves analyzing tongue sub-lingual veins (also a part of the tongue) to find cancers, specifically respiratory organ and breast.

Therefore, this paper proposes to non-invasively find carcinoma and carcinoma via the analysis of options extracted from tongue sub-lingual vein pictures. Segmentation is 1st applied to a tongue sub-lingual vein image so as to separate the left vein and right vein foreground pixels from its background. Next, 2 form of options area unit extracted from every vein and forms a feature vector. These area unit color options and pure mathematics options. Sample of mucous secretion underneath a magnifier to seem for cancer cells. However, the method is time intense because it is suggested to gather the mucous secretion for 3 consecutive days (in the first morning). A chest X-ray is generally the primary medical conducted if carcinoma symptoms arise. If there area unit signs of cancer, a lot of intensive testing are administrated. A CT scan produces a a lot of elaborated read than a chest X-ray. These scans will show the position, size, and form of tumors within the lungs. That being aforesaid, the procedure emits a a lot of powerful indefinite quantity of radiation than X-ray. A diagnostic procedure [4] is usually accustomed screen for carcinoma and involves employing a low energy X-ray to watch the human breast. However, there area unit some drawbacks with this procedure together with the exposure to low-dose radiation, and also the accuracy of the diagnostic result, that depends on the expertise and skill of the specialist.

Given the importance of the 2 cancers and its current diagnostic problems, there's AN pressing ought to develop non-invasive, reliable, and economical ways to notice these 2 diseases. Recently, researchers have developed processed techniques to non-invasively notice varied diseases (diabetes, liver disease, rubor, and heart disease) [5-10]. This was administrated by analyzing the options extracted from facial [7-10] and tongue [5-6] pictures coming back from morbid people. the concept comes from medical bioscience [11], wherever distinctive and distinguishable options is extracted from a selected health problem. The notion of analyzing the face and tongue was borrowed from ancient medicines [12-13], that believed the standing of the inner organs is mirrored on the face and tongue. although vital results were achieved via the analysis of facial and tongue options, there exists very little to none within the literature once it involves analyzing tongue sub-lingual veins (also a part of the tongue) to notice cancers, specifically respiratory organ and breast.

Therefore, this paper proposes to non-invasively notice carcinoma and carcinoma via the analysis of options extracted from tongue sub-lingual vein pictures. Segmentation is 1st applied to a tongue sub-lingual vein image so as to separate the left vein and right vein foreground pixels from its background. Next, 2 form of options area unit extracted from every vein and forms a feature vector. These area unit color options and pure mathematics options.

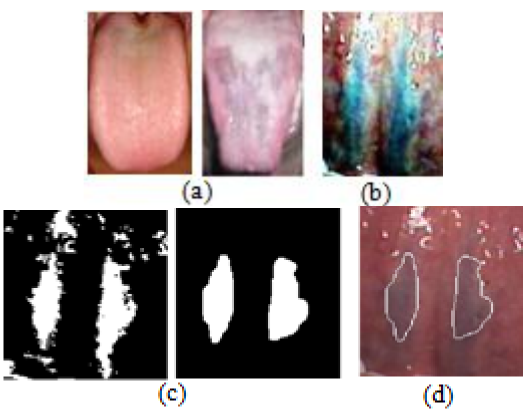
Finally, the feature vectors of the left and right veins area unit combined to represent a sample and accustomed perform classification. A. Tongue identification for TCM and Western medication

Tongue identification plays an important role in ancient Chinese medication as a result of it'll assist doctors in distinctive varied pathophysiological changes inside the body [8]. The apply is also a significant component of the scrutiny section of TCM examination, and has connectedness to most syndrome differentiation ways that. traditionally, TCM tongue identification is performed through AN exhaustive analysis of surface choices of the tongue body and coating, at the side of color, shape, moisture, texture, thickness of coating, etc. In Western medication, abnormal choices on the tongue area unit coupled to dehydration, allergies or plant life infestation [7].

Recently, machine power-assisted tongue identification has been used in CM to produce associate objective record and analysis of tongue choices. With this, some studies have investigated the link between tongue choices and differing kinds of diseases. In [4] the first comprehensive study investigation the association between ancient tongue identification and thus the tongue coating microbiome practice next-generation rRNA sequencing is given. The study renowned 123 and 258 species-level operational compartmentalization units, that were prevailing in patients with Cold/Hot Syndromes, severally, representing "Cold Microbiota" and "Hot Microbiota" which may be associated to one or the alternative syndrome. Recently, [5] used tongue identification as a predictor of Metabolic Syndrome, showing but tongue coating was significantly all completely different in healthy patients if compared to those choked with Metabolic Syndrome. In patients with type-2 polygenic disorder (DM) and diagnosed with “blood stasis”, [3] shows that patients with a light-blue tongue would possibly want succeeding likelihood of developing vessel stiffness.

2. Tongue Sub-Lingual Veins Segmentation

A region of interest (ROI) is initially established encompassing the tongue sub-lingual veins (refer to Fig. 2(a)). Next, the ROI is enhanced as illustrated in Fig. 2(b). Afterwards, based on the enhanced image, a coarse representation of the sub-lingual veins is created in the form of two masks (left and right, see Fig. 2(c)). Finally, operations are applied to fine tune the masks as shown in Fig. 2(d) to cover only the foreground pixels of the tongue sub-lingual veins. The ROI with the segmented tongue sub-lingual veins can be seen in Fig. 2(e).



(e)

Figure 2: Tongue sub -lingual veins segmentation steps: (a) ROI, (b) enhanced ROI, (c) coarse level representation, (d) fine level representation, and (e) segmented veins.

3. Tongue Sub-Lingual Veins Feature Extraction

After locating and segmenting the tongue sub-lingual veins from an image, feature extraction is performed. Two types of features, color and geometry are extracted from each vein (left and right) to form a feature vector. In the following sub-sections each feature type will be explained in more detail.

A. Color

Three groups of color features are extracted from each vein. They include RGB, HSV, and Lab. These three are chosen given its popularity and its use in facial [7-10] as well as tongue [5 -6] image analyses. Given a segmented sub-lingual vein, its R, G, and B values are defined as the average pixel intensities. The same is true for the H, S, and V values, which are computed by converting RGB to HSV. As for the Lab color features, the sub -lingual vein is first transformed from RBG to XYZ [15], followed by XYZ to Lab [15]. This forms a feature vector sub-lingual veins (from the same individual) with its corresponding color features (below) of the left vein [0.46 0.31 0.34 0.96 0.33 117.52 38.05 17.71 2.49] of the right vein [0.47 0.31 0.34 0.94 0.33 118.96 38.50 17.81 2.72]

B. Geometry

Table 1: Tongue Sub-Lingual Vein Geometry Features

Tongue Sub-lingual Vein	Geometry Features
Left	[980 1038 0.94 0.69 35.32 60.57 21.22 0.94]
Right	[1172 1327 0.88 0.66 38.63 62.83 26.36 0.91]

The following experiments were carried out on a dataset collected from the Guangdong Provincial Hospital of Traditional Chinese Medicine. The dataset consisted of sub-lingual vein images from 628 healthy samples, 81 suffering from lung cancer, and 147 with breast cancer. The patients with cancer were diagnosed using the raditional methods mention above (in Section I), while a blood test and other commonly used medical examinations were used to classify the healthy individuals. It should be noted that all procedures performed involving human participants were in accordance with the ethical standards of the institution and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

In the experiments, we performed two binary class classifications based on the extracted features: healthy vs. breast cancer and healthy vs. lung cancer. The classifier used was the Support Vector Machine (SVM) [16] with a linear kernel, given its performance in many classification problems. To measure the classification results, accuracy, which is defined as the number of correctly classified samples divided by the total number of samples is applied. This means given a person's sub-lingual vein image, we can determine if s/he is healthy or has breast/lung cancer.

4. Experimental Results and Discussion

A. Healthy vs. Breast Cancer

Since there significantly more healthy samples (628) than breast cancer (147) samples, directly performing classification using the extracted features (from Section II) would produce unreliable results. To resolve this issue, we randomly selected 147 samples from the healthy class. This meant the number of samples in both classes would be equal. Next, half the tongue sub-lingual vein images from both classes were randomly selected, where its extracted features represented the training set, while the remaining images and its extracted features were designated the test set. To fully utilize the dataset, each sample will be represented by its two sub-lingual veins (left and right). Therefore, in total a given sample from either class will have a feature vector (+) with 34 elements (left vein features numbered from 1 to 17; right vein features numbered from 18 to 34). To optimize the classification result, feature selection in the form of Sequential Forward Selection (SFS) was applied. This meant the best feature combination was chosen from the two feature types (color and geometry) by maximizing the accuracy of SVM for each selected feature. After achieving a result, the training and test sets are switched, and the classification process is performed again, where the final accuracy is the average of the two rounds. The first row of Table II shows this classification result. As can be seen, 12 features (4, 5, 15, 3, 7, 1, 31, 34, 10, 13, 32, and 28) out of 34 were selected achieving an average accuracy of 79.23%. Given that 8 (4, 5, 15, 3, 7, 1, 10, and 13) features from the 12 belong to the left sub-lingual vein, we can conclude that the left side plays a bigger part in healthy vs. breast cancer detection. Fig. 4 shows three typical samples of tongue sub-lingual veins from healthy (top) and breast cancer (bottom). To be thorough, the *k*-Nearest Neighbor (*k*-NN) classifier [17] was also applied with SFS. Its highest classification was 77.89% using $k = 1$.

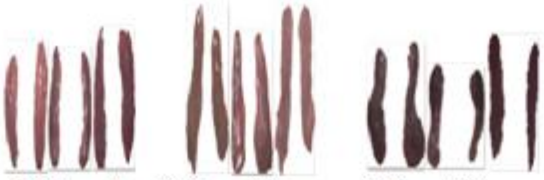


Figure 4: Three typical healthy (top), lung cancer (middle), and breast cancer (bottom) tongue sub-lingual veins.

B. Healthy vs. Lung Cancer

The classification of these two classes followed the same procedure as Section IV.A. First, the classes were balanced and had 81 samples each. Then, half of the samples were randomly assigned to the training set represented by its extracted sub-lingual vein features. The remaining samples and its corresponding features made-up the test set. Classification was performed twice (by switching the training and test sets around) using SFS and SVM. Finally, its result of 82.07% using 8 features (28, 21, 2, 33, 34, 25, 10, and 1) is reported in the last row of Table II. This compares to an average accuracy of 77.16% using *k*-NN, where $k = 1$. In this case 5 (28, 21, 33, 34, and 25) features from the 8 belong to the right sub-lingual vein, where we can conclude that the right side plays a bigger part in healthy vs. lung cancer detection. Three typical samples of tongue sub -lingual veins from lung cancer patients can be seen in Fig. 4 (middle).

5. Conclusion

Give the severity and prevalence of each carcinoma and carcinoma, and also the indisputable fact that medically identification either comes with its own risks, this paper proposes a non-invasive detection technique through the analysis of tongue sub-lingual veins. The tongue sub-lingual veins square measure 1st captured exploitation our unambiguously designed imaging device. Afterwards, the sub-lingual veins square measure placed and segmental from its background. within the following step, 2 kinds of options (color and geometry) square measure extracted and combined to create a feature vector consisting of thirty four components

Experimentation was conducted on a dataset containing two hundred healthy samples, 70lung cancer patients, and one hundred twenty carcinoma patients. 2 sets of experiments were dispensed within the type of healthy vs. carcinoma and healthy vs. carcinoma. The results showed that a mean accuracy of eighty two.07% was achieved at healthy vs. carcinoma employing a kind of options. As for the opposite classification of healthy vs. carcinoma, the ultimate result was seventy nine.23% exploitation twelve options. This demonstrates the effectiveness of the projected technique and its potential as a future medical tool to non-invasively discover respiratory organ or breast cancers.

References

[1] Y. Dai, Y. Long, X. Xing, and G. Wang, Exploring Body Constitution inTraditional Chinese Medicine with K-Means Clustering. Cham: Springer, International Publishing, 2016, pp. 52–64.

[2] World Health Organization Cancer Country Profiles 2014, http://www.who.int/cancer/country-profiles/chn_en.pdf?ua=1, 2014.

[3] American Cancer Society, Exams and Tests That Look for Lung Cancer, https://www.cancer.org/cancer/lung-cancer/prevention-and-early-detection/exams-and-tests.html, 2016.

[4] L. Tabár, A. Gad, L.H. Holmberg, et al., “REDUCTION IN MORTALITY FROM BREAST CANCER AFTER MASS SCREENING WITH MAMMOGRAPHY: Randomised Trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare,” *The Lancet*, vol. 325, no. 8433, pp. 829-832, 1985.

[5] B. Zhang, B.V.K. Vijaya Kumar, and D. Zhang, “Detecting Diabetes Mellitus and Nonproliferative Diabetic Retinopathy using Tongue Color, Texture, and Geometry Features,” *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 2, pp. 491-501, 2014.

[6] B. Pang, D. Zhang, and K. Wang, “Tongue image analysis for appendicitis diagnosis,” *Information Sciences*, vol. 17, pp. 160-176, 2005.

[7] B. Zhang, B.V.K. Vijaya Kumar, and D. Zhang, “Noninvasive Diabetes Mellitus Detection using Facial Block Color with a Sparse Representation Classifier,” *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 4, pp. 1027-1033, 2014.

[8] J.X. Li, D. Zhang, Y.C Li, J. Wu, and B. Zhang, “Joint similar and specific learning for diabetes mellitus and impaired glucose regulation detection,” *Information Sciences*, vol. 384, pp. 191-204, 2017.

[9] J.X. Li, B. Zhang, and D. Zhang, “Joint discriminative and collaborative representation for fatty liver disease diagnosis,” *Expert Systems with Applications*, vol. 89, pp. 31-40, 2017.

[10] T. Shu, B. Zhang, and Y.Y. Tang, “Effective heart disease detection based on quantitative computerized traditional chinese medicine using representation based classifiers,” *Evidence-Based Complementary and Alternative Medicine*, vol. 2017, Article ID 7483639, 2017.

[11] G. Zhao, X. Wang, Y. Niu, L. Tan, and S.-X. Zhang, “Segmenting brain tissues from chinese visible human dataset by deep-learned features with stacked autoencoder,” BioMed Research International, vol. 2016, no. Article ID 5284586, pp. 1–12, 2016.

[12] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L. Mega, and D. R. Webster,

[13] X. Wang, and D. Zhang, “An optimized tongue image color correction scheme,” *IEEE Trans. on Information Technology in Biomedicine*, vol. 14, pp. 1355-1364, 2010.

[14] H.Z. Zhang, K.Q. Wang, X.S. Jin, and D. Zhang, “SVR based color calibration for tongue image,” in Proc. *International Conference on Machine Learning and Cybernetics*, pp. 5065-5070, 2005.