International Innovative Research Journal of Engineering and Technology
ISSN: 2456-1983 Vol: 5 No: 3 March 2020

Classification of Ultrasound Images for Thyroid Detection using Fuzzy C-Means and Artificial Neural Network (ANN) Classifier Technique

M. S. Anbarasi[1], B. Niketha[2], K. Lakshminarayanan[3], Nallandhula Rachana[4]
Department of Information Technology, Pondicherry Engineering College,
Puducherry, India
[1]anbarasims@pec.edu, [2]bniketha26@gmail.com, [3]klnarayanan21@gmail.com, [4]nrachana1998@gmail.com

Abstract— Nowadays the health issue is being worried a lot and so the workload of a doctor becomes comparatively huge. In current scenario, unidentified thyroid has lots of serious medical issues. So, to reduce the workload, the image segmentation and classification of thyroid ultrasound images is most necessary. In our proposed project, the ultrasound images first undergo a pre-processing stage, where it involves processes like grayscale conversion, intensity calculation and histogram equalization. Most commonly, K means clustering algorithm is being used for the segmentation of ultrasound images. It is effective when the clusters are not overlapped. In our work, thyroid tumor region is segmented and by using Fuzzy C Means (FCM) clustering it is extracted, which is advantageous compared to K means algorithm. After segmentation, Scale Invariant Feature Transform(SIFT) algorithm is used to detect local features in images. Since the thyroid ultrasound images are of different rotations and scales, SIFT is more efficient. The feature output that is obtained is then applied to the PCA (principal component analysis) and then to GLCM as input. The main work of PCA is for dimension reduction purpose. After the feature extraction, all features are extracted from each data and combined them into a single matrix for classification. ANN is a family of artificial intelligence. For classification purposes, ANN is chosen since it is self-learning, has high speed and parallelism, error tolerance and associative memory. Feedforward ANN that is trained with backpropagation algorithm is a model of ANN that is used here. The most commonly adopted classifier is SVM. In our proposal, classification is done by ANN because SVM is a binary classifier whereas ANN classifier is more efficient compared to SVM.

Keywords- ANN, Benign or Malignant thyroid, Gray level co-occurrence matrix(GLCM), Hyperthyroidism, Machine Learning, PCA, SIFT, Fuzzy C-Means clustering
.

1. INTRODUCTION

In our proposed project, the ultrasound images first undergo a pre-processing stage, where it involves processes like grayscale conversion, intensity calculation and histogram equalization. The ultrasound image dataset is collected and those are undergone to grayscale conversion. The output of the grayscale conversion is given for intensity calculation. The intensity calculation can be done by histogram equalization method. After the Image Enhancement, that enhanced image is segmented by FCM (fuzzy c means) clustering method for obtaining the thyroid region. After segmentation, to describe and detect local features in images, SIFT algorithm is used. In SIFT, essentially scale-space filtering is used, images are searched over scale and space, for local extrema. By identifying the nearest neighbors, key points between two images are matched. The feature output that is obtained is then applied to the PCA (principal component analysis) as input which is a tool used for reducing the dimension that can be employed to cut down a large set of variables to a small set that yet contains most of the data in the large set. The dimension reduced data is then applied to Gray Level Co-occurrence Matrix (GLCM). After the feature extraction, all features are extracted from each data and combined them into a single matrix. Based on that feature extracted data, predicted class/ label is applied to ANN Classifier. It will Analyze the data and predict the corresponding result/class based on feature extracted data. Classification is done to identify the types of tumor.

2. LITERATURE SURVEY

In this survey, review about the various pre-processing, clustering and machine learning techniques for classification and detection of thyroid in ultrasound images are given below. Thyroid tumor is also classified using those techniques.

"Dynamic Background Subtraction Using Histograms Based on Fuzzy C-Means Clustering and Fuzzy Nearness Degree"-Wei Lu, Tianming Yu, and Jianhua Yang, IEEE February 2019. Based on fuzzy c-means clustering and the fuzzy nearest degree, the study presents an approach for background subtraction using fuzzy histograms. In this method, the FCM algorithm is used to describe the temporal characteristics of the pixels by a fuzzy histogram. The proposed model is provided with the statistical characteristics of the previous samples of individual pixels. The main objective is to provide a fuzzy histogram background model, take up the fuzzy nearness degree to calculate the distance between the background model and the pixel, and then detect to which cluster the pixel belongs to.

"Application of Neural Network Based on SIFT Local Feature Extraction in Medical Image Classification" - Hong Jiang, Shuqi Cui, Zheng Wang, Chaomin Shen. Here, SIFT algorithm is integrated with sliding window and SVM classifier to obtain local features and ROI accurately in the image. Firstly, the image is de-noised, and SIFT is adopted to extract the local feature of ROI, this excludes the influence of the unwanted region. This local feature is taken as an input layer to enhance the generalization ability of the classifier. Here, SVM classifier is primarily used to detect the sliding window area is the breast area or not. The SIFT algorithm is used here since it has strong capacity of Image matching and the stable feature extraction, which helps in extracting the features of ROI.

"Rapid and Low-Cost Detection of Thyroid Dysfunction Using Raman Spectroscopy and an Improved Support Vector Machine" - Xiaoyi Lv, Zhengang Zhai, Guodong Lv, Guoli Du, Xiangxiang Zheng, and Jiaqing Mo, IEEE,2018. Using SVM integrated with serum Raman spectroscopy, the study presents a low-cost and rapid method to identify thyroid dysfunction. Principal component analysis (PCA) is used to extract features and minimize the dimension of high-dimension spectral data; an efficient discriminating model SVM was then used. The findings show that SVM combined with the Raman serum spectroscopy technique is highly capable of detecting thyroid dysfunction. This approach can be used to create low-cost and compact devices for the diagnosis of thyroid dysfunction in order to meet patient needs. The diagnostic models that they have used incorporated with serum Raman spectroscopy doesn't have ideal diagnostic accuracy of thyroid dysfunction.

"Multi-scale analysis of ulcer disease detection from WCE images" - Meryem Souaidi1., Mohamed El Ansari1, 29th July 2019. In this study, an automated, computer-aided design is proposed, in an aspect to discriminate between normal WCE and diseased ulcer images. A multi-scale analysis-based GLCM is used here to compute Laplacian pyramid decomposition from each sub-band, to obtain the Haralick features. GLCM for four directions (0°; 45°; 90° and 135°) around the central pixel is calculated. Consequently, to generate the final function matrix, all the selected features advanced a suitable concatenation strategy ahead of the classification. Whereas in this case the images reach an acceptable threshold to identify the peripheral area to prevent degeneration of the consistency of the ulcer recognition study.

"Thyroid Ultrasound Texture Classification Using Autoregressive Features in Conjunction with Machine Learning Approaches.", 2019. The thyroid texture classification and segmentation is done with three machine learning algorithms which are SVM, ANN and RFC. Volume measurement and thyroid segmentation are of particular significance when it comes to thyroid disease monitoring and treatment. In this proposal, they concentrated mainly on characterizing the texture of thyroids on a US picture using 3 machine learning techniques, including Random Forest Classifier, Artificial Neural Network and Support Vector Machine. The computed features in this work are entirely based on an algorithm for characterizing the texture. This method gives around 90 per cent accuracy. The key shortcomings of this approach are that it has only been tested from stable subjects with the thyroid images.

3. PROPOSED WORK

To reduce the clinical workload of the doctor, Classification and Image pattern recognition techniques for medical ultrasound images is necessary. The common image pattern recognition techniques are Support Vector Machine, Bayesian pattern classifier and neural network model. In our work, thyroid tumor region is segmented ant it's extracted by means of Fuzzy C Means (FCM) clustering. Where each data element can belong to more than one cluster and its membership level fuzzy value ranging between zero and one.

In our work, thyroid tumor region is segmented and it's extracted by means of FCM clustering. The textural and statistical information are retrieved from the extracted region using SIFT with PCA, GLCM. The features like standard deviation, mean and the entropy of the obtained sub-bands are determined and stored in a feature vector (in format of mat file). After that features are extracted, it is

International Innovative Research Journal of Engineering and Technology
ISSN: 2456-1983 Vol: 5 No: 3 March 2020

compared with the predefined datasets that are completely trained and classified by using the ANN classifier.

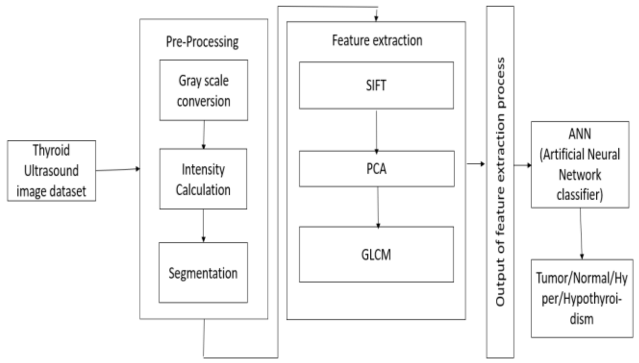

Figure 1.1. High level architecture diagram for Classification of Ultrasound images for Thyroid Detection using Fuzzy C-Means and Artificial Neural Network (ANN) Classifier technique

Pre-processing Module

The ultrasound image dataset is collected and those are undergone to grayscale conversion. Grayscale is a collection of monochromatic shades from white to black. Hence, a grayscale image consists of only shades of gray and no other color. To provide less information to each pixel, we discriminate grayscale images from another sort of color images. The output of the grayscale conversion is given for intensity calculation. Intensity refers to the amount of light or numerical value of a pixel. The intensity calculation can be done by histogram equalization method. After the Image Enhancement, that enhanced image is segmented by FCM (fuzzy c means) clustering method for obtaining the thyroid region.

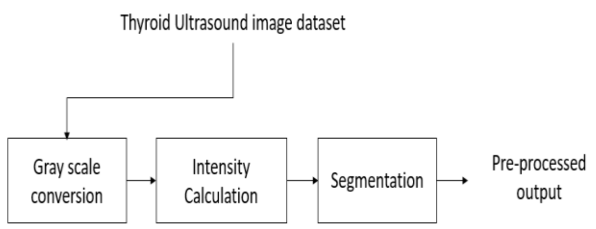

Figure 1.2. Pre-processing module of CUITDFANN

Grayscale conversion: In this process, the color images are converted into grayscale format to ensure that the input ultrasound image is in shades of gray. The bits no. specifies the maximum amount of colors a digital system supports. The picture on the grayscale is represented using 8-bit luminance valuesThe luminance of a grayscale pixel value varies between 0 and 255. The method of transforming a color image into a gray image is to transform RGB bit values into grayscale bit values.

Intensity calculation: The intensity calculation can be done by histogram equalization method. The enhancement of the image's contrast can be noticed by employing histogram equalization. For segmentation method, the thyroid region is segmented into two sides as left and right. The images are then converted into binary using im2bw function. For even further processing, the initialization mask and thyroid image are segmented using the Fuzzy C means Clustering.

Segmentation: There are several methods for segmenting grayscale images of which FCM is one of them. FCM is a method for clustering which allows a bit of data to fit into 2 or several clusters. Fuzzy logic is a multi-valued logic derived from the principle of fuzzy set. The FCM is utilized mainly for soft segmentation. In addition, FCM can produce good performance than most clustering algorithms. FCM depends on minimizing the generalized least-squared errors function. It splits the objects group V ¼ [v1 ; v2 ; …; vl] into k fuzzy clusters by lessening the total set of squared error of objective function. By this way, the thyroid tumor region can be segmented along with SIFT algorithm in the following process.

Feature Extraction module

After segmentation, to describe and detect local features in images, SIFT algorithm is used. In SIFT, essentially scale-space filtering is used, images are searched over scale and space, for local extrema. When the positions of possible focal points are defined, they must be optimized in order to achieve more accurate results. It therefore excludes any key points with low contrast and edge key points, and maintains only strong points of interest. Finally, to attain invariance to image rotation, an orientation is assigned for each main point. Based on the scale, a neighborhood is taken around the key point location, and the direction and gradient magnitude is determined in that region. It contributes to the stability of matching. By identifying the nearest neighbors, key points between two images are matched.

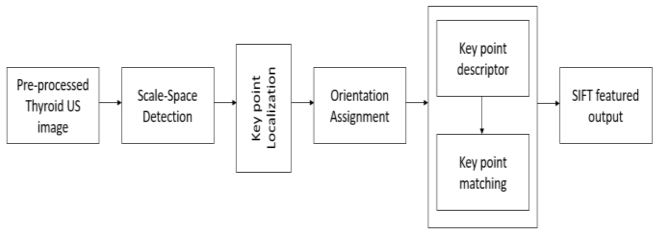

Figure 1.3. Feature Extraction module using SIFT

PCA and GLCM

The feature output that is obtained is then applied to the PCA (principal component analysis) as input which is a tool used for reducing the dimension that can be employed to cut down a large set of variables to a small set that yet contains most of the data in the large set. Each variable could be treated as a different dimension.

PCA is adopted to obtain the necessary data from a multivariate data table and to convey this information as a collection of fewer new variables called principal components. Which corresponds to a linear sequence of the original set. The data in a given data set corresponds to the total variation it encloses. The main goal of PCA is to determine directions (or) principal components along with the variation in the data which is maximal. In other words, the process of PCA is the reduction of the dimensionality of a huge data to two or three principal components, which can be visualized graphically, with minimal loss of data.

The dimension reduced data is then applied to Gray Level Co-occurrence Matrix (GLCM), and the following parameters listed below are calculated:

Energy/Uniformity/Angular Moment (done), Dissimilarity (done), Entropy (done), Inertia / Contrast (done), difference, Inverse, correlation, Homogeneity / Inverse difference moment, Auto-correlation, Cluster Prominence, Cluster Shade, Minimum probability, Sum Average, Sum of Squares, Sum Variance, Sum Entropy, Difference entropy, Difference variance, Information measures of correlation, Coefficient of correlation, Inverse difference moment normalized (IDN), and Inverse difference normalized (INN).

Even though, there is a function graycoprops () in the Image Processing Toolbox of Matlab that calculates four parameters Correlation, Contrast, Homogeneity, and Energy. The proposal by Haralick recommends a few more parameters that are also calculated here.

Classification module using ANN

After the feature extraction, all features are extracted from each data and combined them into a single matrix. While on preprocessing Predicted label/class are already separated, based on the class we represent the class to ANN Classifier. Based on that feature extracted data, predicted class/ label is applied to Classifier. It will Analyze the data and predict the corresponding result/class based on feature extracted data. In Neural network there are two process that has to be done which is training and testing. Here we are using Feed forward back propagation neural network. Feed-Forward Neural Network (FFNN) comprises of at least three layers of neurons: an input layer, minimal of one intermediate hidden layer, and an output layer.

The learning process in FFNN along with back propagation takes place during the training phase in which every input pattern from the training data set is given to the input layer and then propagates forward. The pattern of activation, which is arriving at the output layer is then matched with the correct output pattern to determine an error signal. The error signal for each of the target output pattern is then back propagated from the output layer to the input layer of neurons so that the weights in each layer of the network are adjusted. The training for the neural network tool involves the following steps :

Step1: import input data(h & g) from workspace

Step2: import target data(xx) from workspace

Step3: create new network by specifying the input and target

Step4: click and open the network give training datas and click start training

Step5: give input sample in the simulate option and click simulate

Step6: open the network output window and open it [if it is 1-true 2-false]

The NN learns the accurate classification for a set of inputs during the training phase, after which it can be tested on another set of samples to check how well it classifies new patterns. Hence, an important consideration is needed in applying back propagation learning is how well the network makes the generalization.

4. RESULT ANALYSIS

The output obtained involves the following: Sensitivity (also called the true positive rate, or probability of detection in some fields) determines the proportion of actual positives that are accurately analyzed as such (e.g., the percentage of infected people who are actually diagnosed as having the condition). Specificity (which is additionally referred to as true negative rate) determines the proportion of actual negatives that are accurately analyzed intrinsically (e.g., the share of healthy people that are actually diagnosed as not having the condition). And also analyze the Thyroid region in terms of pixels & area (in mm). In our proposed project, the segmentation process which is done by Fuzzy C Means algorithm which increases the efficiency by reducing time complexity. In K means clustering, one of the three outputs is used for further processing whereas in Fuzzy C Means the mean value of all the outputs is taken into consideration which proves to be advantageous. In our proposed project, the

International Innovative Research Journal of Engineering and Technology
ISSN: 2456-1983 Vol: 5 No: 3 March 2020

ANN classifier is adopted since it is a self-learning algorithm that alters its parameters based on internal and external information that moves through the network.

5. CONCLUSION

Thus in this paper we have proposed a model to identify thyroid tumor by using a neural network. The model classifies the tumor as benign and malignant based on the thyroid ultrasound images. In future, the CNN, RNN method can be used in place of ANN classification to further increase the accuracy of the thyroid classification in the ultrasound images.

REFERENCES

[1] Prabal Poudel , Alfredoillanes , Elmer J. G. Ataide, Nazila Esmaeill, Sathish Balakrishnan and Michael Friebe , "Thyroid Ultrasound Texture Classification Using Autoregressive Features in Conjunction With Machine Learning Approaches." Faculty of Medical Engineering, Otto-von-Guericke University, 39106 Magdeburg, Germany Corresponding *author: Prabal Poudel (prabal.poudel@ovgu.de),2019.*

[2] Mahdi Bayat , *Member, IEEE*, Mostafa Fatemi , *Fellow, IEEE*, and Azra Alizad , *Senior Member, "*Background Removal and Vessel Filtering of Noncontrast Ultrasound Images of Microvasculature", *IEEE transactions, March 2019.*

[3] D. China, A. Illanes, P. Poudel, M. Friebe, P. Mitra, and D. Sheet, ``Anatomical structure segmentation in ultrasound volumes using cross frame belief propagating iterative random walks," IEEE J. Biomed. Health Inform., *May 2019*, vol. 23, no. 3, pp. 1110_1118.

[4] Tianming Yu , Jianhua Yang, and Wei Lu, "Dynamic Background Subtraction Using Histograms Based on Fuzzy C-Means Clustering and Fuzzy Nearness Degree", IEEE transactions, *February 2019*

[5] A.Illanes, N. Esmaeili, P. Poudel, S. Balakrishnan, and M. Friebe, ``Parametrical modelling for texture characterization A novel approach applied to ultrasound thyroid segmentation," PLoS ONE, *Jan. 2019*, vol. 14, no. 1, Art. no. e0211215.

[6] Meryem Souaidi1 , Mohamed El Ansari1, "Multi-scale analysis of ulcer disease detection from WCE images" , *4th May 2019*, ISSN 1751-9659. doi: 10.1049/iet-ipr.2019.0415.

[7] Xiangxiang Zheng , Guodong Lv, Guoli Du, Zhengang Zhai, Jiaqing Mo, and Xiaoyi Lv , Khushboo Chandel, Veenita Kunwar , Sai Sabitha ,Tanupriya Choudhury "Rapid and Low-Cost Detection of Thyroid Dysfunction Using Raman Spectroscopy and an Improved Support Vector Machine",*2019*

[8] Anxiongwang,shuai Li Wenfengsong,HongQin,"Learning from weakly –labelled clinical data for automatic thyroid nodule classification in ultrasound images"-*2018.*

[9] Meriama Mahamdioua1 , Mohamed Benmohammed[2]1TLSI Department, NTIC Faculty, University of Constantine[2], Constantine, Algeria, ISSN 1751-96322. Ac doi: 10.1049/iet-cvi.2017.0190. "Automatic adaptation of SIFT for robust facial recognition in uncontrolled lighting conditions", *7th February 2018*

[10] Qin Yu, Tao Jiang, Aiyun Zhou, Lili Zhang, Cheng Zhang, Pan Xu,"Computer Aided Diagnosis of malignant or benign thyroid nodes based on ultrasound images", *Springer*, 2017.

[11] Shuqi Cui, Hong Jiang, Zheng Wang, Chaomin Shen, East China Normal University, "Application of Neural Network Based on SIFT Local Feature Extraction in Medical Image Classification" 978-1-5090-6238-6/17/$31.00 ©20 17 IEEE.

[12] P.Poudel, E. J. G. Ataide, A. Illanes, and M. Friebe, ``Linear discriminant analysis and K-means clustering for classification of thyroid texture in ultrasound images," in Proc. *July 2018*, 40th Int. Conf. IEEE Eng. Med. Biol. Soc.

BIOGRAPHIES

Dr. M.S. Anbarasi is currently working as Assistant Professor in the Department of Information Technology at Pondicherry Engineering College. Her area of interest mainly focuses on Data mining, Databases and Cloud Computing.

B. Niketha was born in Puducherry, currently pursuing her B.Tech degree in the stream of

Information Technology in Pondicherry Engineering College, Pondicherry, India.

K. Lakshminarayanan was born in Puducherry, currently pursuing his B.Tech degree in the stream of Information Technology in Pondicherry Engineering College, Pondicherry, India.

Nallandhula Rachana was born in Andhra Pradesh, currently pursuing her B.Tech degree in the stream of Information Technology in Pondicherry Engineering College, Pondicherry, India.