J. Appl. Environ. Biol. Sci., 6(3)72-82, 2016
© 2016, TextRoad Publication
ISSN: 2090-4274
Journal of Applied Environmental and Biological Sciences
www.textroad.com

Intelligent Diagnostic System for Papillary Thyroid Carcinoma

Jamil Ahmed[1], M. Abdul Rahman[2]

[1,2]Department of Computer Science, Sukkur Institute of Business Administration
Airport road Sukkur, Pakistan

Received: October 28, 2015
Accepted: January 31, 2016

ABSTRACT

Due to exponential growth and high diversity of DICOM images in healthcare industry, medical image mining has gained prominent attention because extraction of useful and effective patterns is one of the major problems in DICOM images. For instance, differentiating between the mimic, mix and complex patterns of thyroid papillary carcinoma (PTC) and other cancers in FNAC images is really challenging since it requires in-depth study of cells and tissues. In order to reduce the chances of misdiagnosis of thyroid cancer and to discernment the mimic lesions of papillary thyroid carcinoma (Small cell carcinoma may mimic insular carcinoma). This article proposes a framework, so called (IDSPTC) Intelligent Diagnostic System for papillary thyroid carcinoma, which offers a systematic way to classify papillary and non-papillary structures by using AI base techniques. In first phase, we prepared own dataset due to unavailability of training and testing datasets in literature. We applied our proposed algorithm "DICOM_Segs graph-Sag" to extract the useful movement and coordinate based patterns of nuclei and constructed decision model using ANN (artificial neural network) to identify the papillary structures by analysing nuclei coordinates; finally we performed test model and performance evaluation to measure the classification accuracy using confusion matrix, precision and recall measures and visualized the ROC curve for papillary and non-papillary classes. The measured accuracy of our proposed system is 90.32% with 10-k fold cross validation.

KEYWORDS: Medical Image mining, Decision support system, Pre-process, DICOM, Papillary Thyroid Carcinoma

1. INTRODUCTION

Recently, Medical image analytics have become one of the well-established research areas of ML (Machine Learning) and shown significant impact upon the diagnosis and prognosis of different diseases such as retinopathy, diabetes, lung cancer, breast cancer, thyroid cancer and so on. Specially; classification problem of PTC (papillary thyroid carcinoma) using FNAB (Fine needle aspiration biopsy) images have become one of the difficult classification problems due to the existence of mimic, mix and complex cytological cancerous material, because in many tumours mimic morphological features may be presented with confused encapsulated nuclear patterns containing the potential evidences of papillary thyroid carcinoma. Even an experienced cytologist may be deceived to discriminate the mimic morphological features belonging to the distinct categories of well differentiated, un-differential and poorly differentiated cancers. A wide-range of CAD (Computer added diagnostic) systems came into existence to assist the doctors and many related approaches have been seen in recent past, i.e. [1, 3, 4, 5, 8]. Some of proposed approaches are offering very nice services by considering cell segmentation at the abstract level but it needs in-depth level of cell segmentation and careful analysis of cell structures because DICOM (Digital Imaging and Communications in Medicine) images are heterogamous in nature, always found with different shapes, sizes and structures depends upon the stage of tumours [Figure 1]. Thus; doctors would be assisted in more precise way by offering a system which may covert the traditional diagnostic methods into automated digital decision support system. In-order to mitigate all above stated problems, this article proposes a system so called IDSPTC: (Intelligent Diagnostic System for papillary thyroid carcinoma).The proposed system provides more meticulous assistance to doctors to avoid the confusion of deceiving morphological features of thyroid cancer variants presented with mimic, mix and complex patterns of human tissues. The system is a systematic framework which helps to classify the papillary and non-papillary structures. Our methodology comprises upon four main stages. In first stage we prepare the data sets by using**"DICOM_Segs graph-Sag"**algorithm which uses graph-cut segmentation, grey scale segmentation and extracts the movement features of nuclei. In second and third stage, we extract coordinates of the images objects and constructed a classification model by using ANN (artificial neural network). In final stage we performed test and performance evaluation. In this article we segmented 1539 nuclei from 40 DICOM images to evaluate and to confirm the papillary structures. The used real-world dataset was received from SMBBMU, (Shaheed Muhtarma Be Nazir Bhutto Medical University) Pakistan. The measured classification accuracy is about 90.32% with 10-k fold cross validation.

* **Corresponding Author:** Jamil Ahmed, M. Abdul Rehman Soomrani, Department of Computer Science, Sukkur Institute of Business Administration (SUKKUR IBA), Pakistan jameelahmed.phdcs@iba-suk.edu.pk

This paper is organized in six sections. section one is used to describe introduction of this paper, related works are presented in section two, back-ground information is described in section three, methodology is defined in section four, results are presented in section five and discussion &conclusion is discussed in section six.

2. RELATED WORKS

Basically; our approach is based upon the predictive modelling to identify papillary thyroid structure from DICOM (Digital imaging and communication in medicine) images of FNAB (Fine needle aspiration biopsy). Many related approaches were seen in the past and some of them are presented below.

2.1 A system [1] was proposed to identify the thyroid papillary cancer by using bio-markers. A SVM (Support vector machine) based classifier was trained to construct the decision model from cytological material such as proteins from DICOM images. The reported measured sensitivity and specify were about 95.14% and 93.97% respectively. We use a pixel level segmentation algorithms to auto-detect the homogenous objects from the DICOM images based upon the colour properties such as graph-cut segmentation and applied greyscale feature extraction technique to absolve the effects of cytological material such like H & E stains and biomarkers.

2.2 A compression [2] of three machine learning techniques was presented for the prediction of follicular thyroid cancer, i.e. artificial neural network, Decision Tree and logistic regression. The measured best classification accuracy was 80%. We obtained 90.32% percent overall accuracy of our proposed system.

2.3 A system [3] based on SVM technique was proposed for detection of thyroid cancer. The obtained P value for non-cancerous class was between 0.97 verses 80, P<.0001. Our system offers an automated method for prediction of papillary thyroid cancer by extracting movement based features and coordinates of nuclei from the images of FNAB after performing the process of noise reduction to declare the evidences of papillary structures.

2.4 An approach [4] proposed for thyroid cancer using ultrasound images. Benign and malignant DICOM texture features were taken by employing Fuzzy based local binary patterns (LBP) and fuzzy grey-level histograms. A region selection procedure was applied to extract the textures at global levels from ultrasound images whereas, our proposed approach offers in-depth level of cell segmentation and efficient feature extraction technique to acquire nuclei sequences.

2.5 An approach [5] with 78.00% accuracy was proposed for the classification of ultrasonic images of thyroid gland, six SVMs were used in feature extraction step and entropy of image was recorded to measure the differences of contrast and homogeneity between the extracted properties. We extract the movement features by using histograms for further classification operations.

2.6 A system [6] with 90% accuracy was proposed to classify the thyroid FNAC images. Otsu threshold method was used in image pre-processing stage and grey level features of segmented nuclei were visualized. Our proposed technique offers a systematic solution fully compatible with the traditional cytological experiments.

2.7 A comparative study [7] of two machine learning techniques was conducted for diagnosis of thyroid disease i.e. ‘SVM’ and ‘Multi-scale edge detection’. The reported classification accuracy measured as 91% and 89% respectively. Our proposed system provides more precise assistance to doctors by providing in-depth level of nuclei segmentation to discernment the mimic features to-words the diagnosis of papillary thyroid carcinoma because the system is designed as a substitute method for doctor’s traditional FNAB glass slide observing procedure without damaging the colour information of raw DICOM images.

As the matter of fact Papillary thyroid carcinoma (PTC) is one of the unique classification problems because of having mimic and confused features with the other types of malignant diseases (such as small cell carcinoma may mimic insular carcinoma) even an experienced doctor may be deceived with these confused properties. Due to the unavailability of datasets of papillary thyroid carcinoma in literature, we prepared our own datasets to train and test the performance of classifiers with the assistance of domain expert doctor.

3. BACKGROUND INFORMATION

Cancer is one of the most dangerous and fatal disease that can occur in any part of body except nail and hair, it grows rapidly and does not allow any physician to treat it at higher stages. Thus, cancer must be diagnosed at early stages to maximize the survival rate of patients; because timely, systematically and accurately diagnosis would support the physicians to restore the human health. Doctors are always found interested to investigate the state of internal parts of human-body by observing the X-Rays, CTs (Computed Tomography), Ultrasounds, FNAB (Fine needle aspiration biopsy) and other DICOM (Digital image and communication in medicine) images. FNAB (Fine needle aspiration biopsy) is a very useful tool to diagnose the various kinds of cancerous diseases such as thyroid papillary carcinoma and its variants, but due to the involvement of mimic, complex and mix features of cancerous material on FNAB slides [Figure 2] produces maximum chances for misdiagnosis. Thus it becomes more difficult for doctors to declare a specific malignancy type.

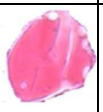
Features	Cell wall & plasma	Capsular invasion	Nuclei shape	Nuclear size	Nuclei appearance	Nuclear structure	Chromatin	Collide follicular matter outside	Collide follicular matter inside
Data type	Percentage of Cell wall & plasma	Percentage of Capsular invasion	Oval, Cubic, Tall, small, Columnar, Overlapping	No, Small, Large, Multiple	Grass ground appearance, Thick, Thin	Papillary, follicular and so on	Percentage of Chromatin in nuclei	Percentage of Collide follicular matter	Percentage of Collide follicular matter
Complex geometries of different components of human tissues									

Figure 1: Taxonomy of DICOM images

A. Fine Needle Aspiration Biopsy (FNAB)

FNAB (Fine needle aspiration Biopsy) is diagnostic tool used to determine a malignancy type. Traditionally; cytologists prepare the slides from the received tissues as specimen for a careful investigation of cancer. Recently surgeons use to conduct the FNAB (Fine needle aspiration biopsy) test before applying surgical procedures. This is a very nice approach because most of the times surgical procedures redone for the removal of neoplasm. Thus FNAB is beneficial tool to avoid un-necessary treatment, time and money of a patient. A FNAB observation[Figure 2, A, B, C, D] is a tedious job to declare a diagnosis and to mention cancer type with proper stage, because FNAB consists upon micro-architectural components of tissue as stated in [Figure 1] and it is difficult to examine but an expert cytologist is the best observer. In case of malignancy tumour areas are determined and removed surgically. Traditionally, cytologists use different stains for example H &E to label a colour to cells and to identify the micro-architectural components of different tissues. There are some types of thyroid cancer from them papillary carcinomas are more aggressive than the other cancers. Few of them are presented in [figure 2].

B. Papillary carcinoma

Papillary carcinoma is commonly found with gross, changeable and infiltrative behaviours. In these types of cytological studies, the white, grey and solid material with considerable average indicates presence of tumours and palpability cystic material. Usually; primary site of a tumour may exist at other organ of human body but cystic metastasis may be seen in non-palpable lymph nodes. Basically; papillary carcinoma is seen as finger like architecture so called papillae that are surrounded by micro-architectural material known as nuclei,cytoplasm and so on. There are some types of papillary carcinoma can be distinguished based on several varying attribute that are stated as bellow.

C. Fullicular carcinoma

This category of cancers are determined by considering the follicular neoplasm on the abhorrently observation. Usually; this type of material consists on the variable length of follicles such as, either darker colloid looks a 'bubble-gum' or hyper eosinophilic is judged against colloid material. F.C diagnosis is based an important multiple nuclei visibility in gigantic cells situated follicles, whereas, papillary carcinoma is very difficult, complex and real challenge for cytologists to declare because is a material which is derived from multiple nodular or follicular like lesions, sometimes the features are found aggressive in clinical observations. Importantly; features of follicular carcinoma which are involved with thecombination of its variantssuch as follicular neoplasm become more complicated and could not be differentiated with the effortless findings. It is a big deal to confirm the diagnosis and to differentiate the differences between these features for example papillary micro-carcinoma is rarely found completely encapsulated tumours whereas Tall cell variant tumours are found with abundant cytoplasm, nuclear features.

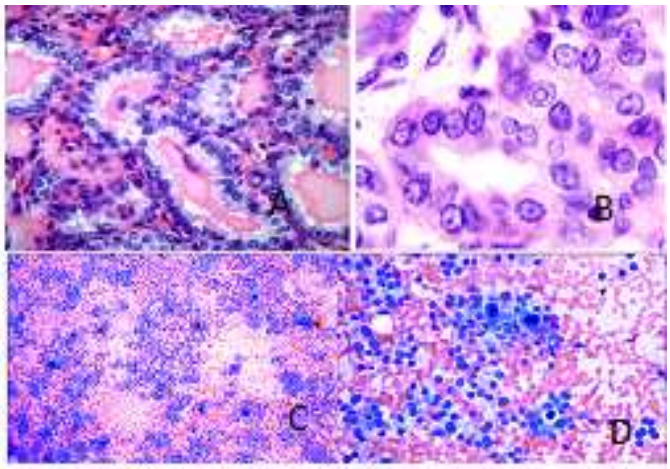


Figure 2: A. Thyroid Papillary Small size carcinoma B. Thyroid Papillary Carcinoma with glass ground appearance C. Thyroid Papillary Small cell carcinomas of thyroid cancer and D. Papillary carcinoma of lung cancer

4. METHODOLOGY

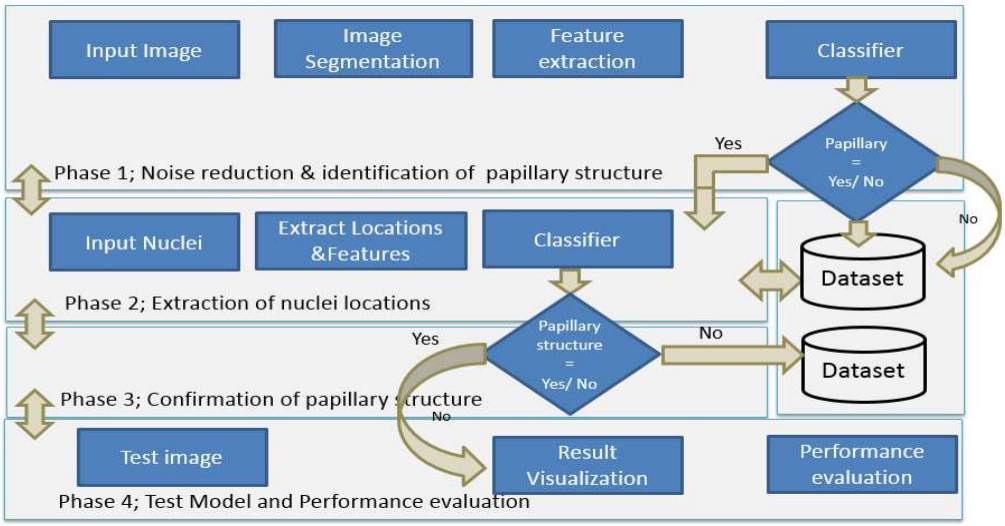


Figure 3: IDSTPC: Intelligent Diagnostic System for thyroid Papillary carcinoma workflow

Overall methodology of our proposed system comprises upon four major stages as presented in [Figure 3]. In first stage; we prepare the datasets by extracting the histogram basedmovement featuresafter performing noisereduction technique such like graph-cut segmentation.In second and third stages; we extract the nuclei structures using Euclidian distance algorithm and trained a classifier with the help of doctor.The constructed decision model has to predict the potential candidate having maximum properties to become the strongest candidate for declaration of papillary structure.Infourthstage we evaluated performance evaluation using confusion matrix, precision and recall measures. A DICOM (Digital imaging and communication in medicine) image comprises upon set of 'object of interests' (OOIs) such as nuclei structures, shapes, sizes, behaviours and other properties as shown in [Figure 1]. As these heterogeneous micro-architectural components are difficult to classify under the normal microscopic views, therefore we used a pixel level algorithm that is very useful in noise reduction [Figure 4] to perform the in-depth level of cell segmentation of DICOM images such like graph cut segmentation as presented in following sub-section.

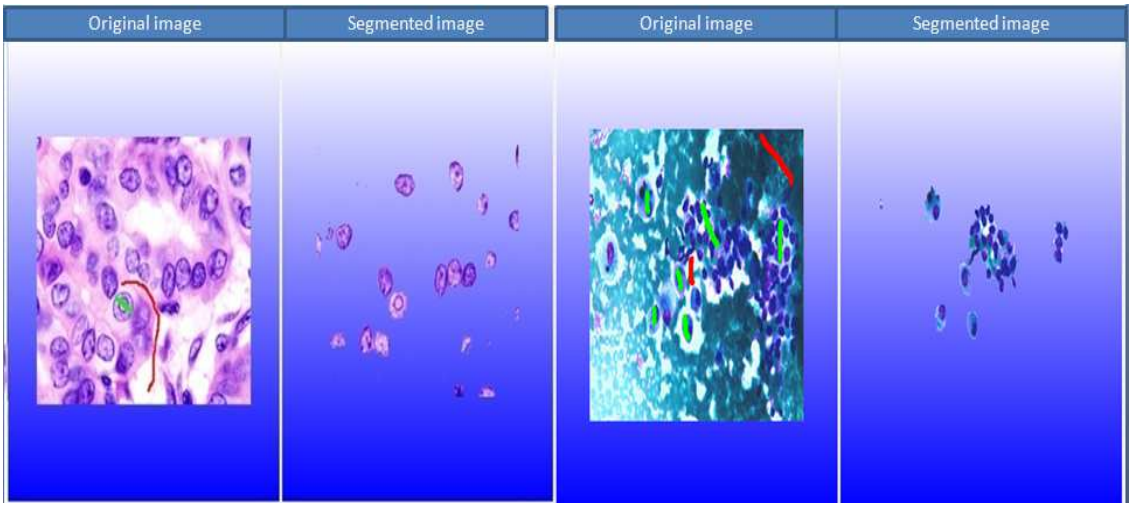


Figure 4: Noise reduction and identification of papillary structure

A. Noise reduction and Identification of papillary structure

Specially; DICOM images are found with lots of noise and with respect to noise reduction techniques many approaches have been seen to reduce the noise from DICOM images and to extract useful features with automated procedures. In our experiment, we used graph-cut segmentation algorithm due to some solidadvantages observed during the experiment of DICOM object detection from the slides of FNAB (Fine needle aspiration biopsy). The graph-cut segmentation exploits pixel level colour information to segment and detects the objects from user provided DICOM image as shown in [Figure 4]. Traditionally, doctors uses different colour stains (such as H & E) and biomarker to label the cytological material for careful examination of FNAB slide. Our proposed method provides substitute computable method with doctor's traditional method as it uses colourlabelling information known as fore-ground and back-ground functions. A doctor can use min-cut max-flow functions by marking the digital lines designated as foreground and back-ground to collect the information from DICOM images, it is an automated process for the detection of homogeneous nuclei of human cells based upon the colour properties. The fore-ground selected nuclei will remain prominent in out-put window and back-ground selection will be considered as noise. In our experiment we found that colour intensity information contains sufficient information to discriminate the mimic and complex structures, shapes of heterogamous micro-architectural components. Our proposed "DICOM_Segs graph-Sag" [algorithm 1] is presented bellow which is very useful to detect the set of nuclei by considering the colour propertiesand to extract the features such as movements from DICOM image by using the histograms.

Algorithm 1:DICOM_Segs graph-Sag

Input: DICOM_Graph, $X_{(V)Foreground,Background}$ as $HSV_{Values\ (E)}$
Output: Seg_objects
Step 1:
H: = get_Proximity ()
For DICOM _Feature (HSV) =1 to N do
$XDICOM_{(S),}$ = $X_{(v)Foreground,Background}$ as $HSV_{Values\ (E)}$ $\leftarrow \cos(H)(i) \leftarrow \mu, \sigma, \chi$ //Calculate user inputs from in N by N matrix← Repeat 1 to n until $X_{(v)} \notin (S)$
Return H values:

Step 2:
V: get comparison ()
For DICOM_VACTOR (y) =1 to M do
Segment each pixel of DICOM_VACTOR (y) (i) =: $XDICOM_{(\acute{S}),}$ $\leftarrow X_{Foreground(HSV),}X_{Background(HSV)}$
DICOM_VACTOR * h ←μ, σ, Ϫ images create set of connected graphs

Step 3:
Split_A (1 to n): (Graph) :($XDICOM_{(i),}$ / DICOM_VACTOR (i) ≤ 0 //Construct Graph A
Else
Split_B (1 to m): $XDICOM_{(i),}$ / DICOM_VACTOR (i) ≥0 // Construct Graph B

Step 4:
DICOM_VACTOR (y) = $DICOM\_VACTOR\ (y)^{T}(XDICOM - W)_{y}/$
$DICOM\_VACTOR\ (y)^{T}DICOM\_VACTOR\ (y)$ //select best split

Step 5: Return Seg_objects

Let consider DICOM_Segs graph-Sag is an iterative method to segment the objects on the basis of user selected colour code defined as$X_{(V)Foreground\ ,Background}\ as\ HSV_{Values\ (E)}$. The get_Proximity calculates user selected colour code in $XDICOM_{(S),}$ function until $X_{(v)} \notin (S)$ where$X_{(v)}$ is the pattern to be traversed between the lines of foreground and background as selected by a doctor. Get_comparison function is used to create the best splits of graphs as satisfy the condition $\notin (S)$, than the DICOM $DICOM\_VACTOR\ (y)^{T}$ and $(XDICOM - W)_{y}$ generates the objects by qualifying the criteria.

B. Extraction of nuclei locations

Before extraction of locations, we need to convert the colour DICOM images into the gray scale images to absolve effects of stains as used by doctors for labelling of nuclei as discussed in previous sections. We converted colour nuclei to gray scale nuclei to extract the locations in uniform way to acquire better results. Mathematically above experiment can expressed as Let DICOM is an array of pixels which contains nuclei and other micro-architectural material intoMxN matrix, where every object of interest is a point, line and polygon V and scalar S is the exact location of every object as unique place designated as $G = (E, V)$. In binary segmentation method, a color gradient of image is converted gray scale or binary image, where the gradient places represented as intensity. Consider [figure 5] and [figure 2, [A, B, C, D]], where [figure 2, B] nuclei are segmented converted into gray level intensity.

$$\nabla f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right] \quad (1)$$

Let consider derivatives as presented in equation (1) that plays central role to seek the direction of connected components of every cell and micro architectural components as stated in [figure 2], such that following equation is presented to traverse directions of various gradients.

$$\theta = \tan^{-1}\left(\frac{\partial f}{\partial y} / \frac{\partial f}{\partial x}\right) \quad (2)$$

However, strength of edges is also acquired through the degree of gradient and its intensity is recorded by measuring the norms of derivatives assigned to equation (3) as presented below.

$$||\nabla f|| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} \quad (3)$$

Same procedure may represented to differentiate by considering a digital image as a function f=(x, y), where an image is reconstructed as a continuous image, followed by a gradient is acquired by using the discrete derivate, such as consider the following equation, the image segmentation is done by applying the first derivative.

$$\frac{\partial f}{\partial x}[x, y] \approx f[x + 1, y] - f[x, y] \quad (4)$$

Effects of noise may also be reduced by implementing cross-correlation as stated in theorem of convolution for derivatives. Equation (5) is presented to demonstrate the noise reduction by multiplying the function of f, where h represents every pixel on x-axis by taking derivatives.

$$\frac{\partial}{\partial x}(h \star f) = \left(\frac{\partial}{\partial x}h\right) \star f \quad (5)$$

Suppose every segmented image is visualized by means of threshold as stated in equation (7) and second derivatives (x, y) image, where every row or column of the image is described as per following equation (6).

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (6)$$

$$g(x) = \begin{cases} \geq 1 \\ < 0\ otherwise \end{cases} \quad (7)$$

After acquiring the gray scale set of nuclei we applied the Euclidian distance algorithm to extract the coordinates of nuclei and applied watershed algorithm to label the points of coordinates [Figure 6]. Finally

doctors included class label attribute for every image, in this way, we recorded 1539 observations for training and testing dataset.

C. *Confirmation of papillary structure*

We used ANN (Artificial Neural Networks) algorithm to train the classifier to construct the decision model because ANNs are highly inspired from human brain and work like human brain neurons in shape of weighs [Figure 6]. With respect to our experiment we found significant advantages of ANNs in clinical cyto-pathology and the model predicted the classes such as papillary and non-papillary with good accuracy. ANNs are interconnected with three different layers 'input layer', 'hidden layer' and 'output layer' where input-layer is responsible to collect and forward input-values of image features to neurons (number of neurons or hidden layers depends upon the complexity of problem). Supplied data to neurons has to be calculated with some hidden 'bias' weights for further processing. Due to the neuron like structure one input is connected to another input vectors such like bias values, the predicted output becomes reliable. Overall measured classification accuracy of our system is about 90.31%.

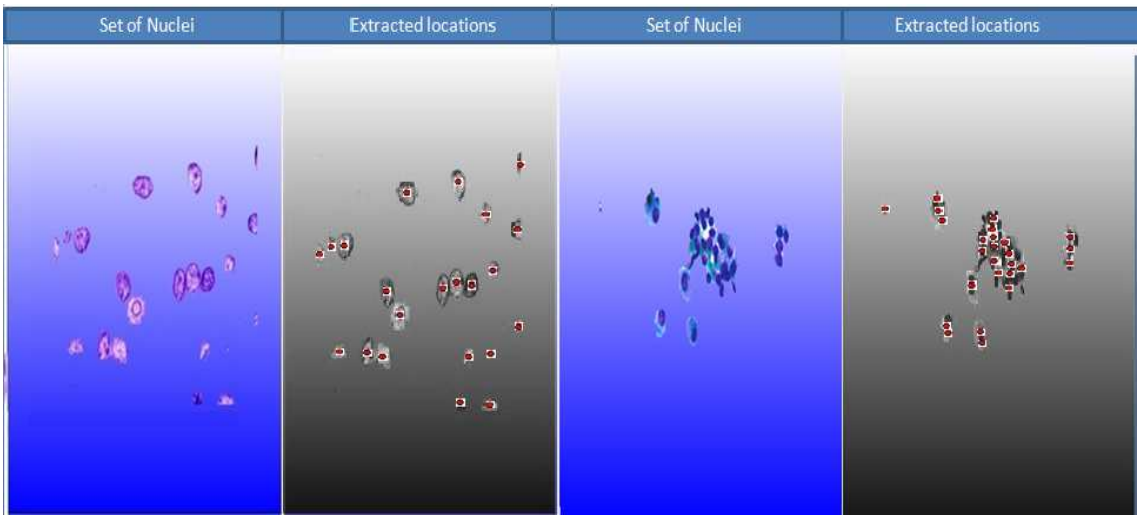


Figure 5: Extractions of longitude and latitude locations of nuclei from segmented image

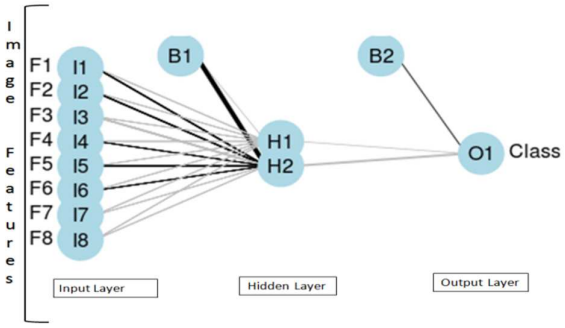


Figure 6: ANN classifier model representation

5. RESULTS

The complex DICOM (Digital imaging and communication in medicine) images of biopsy [Figure 2 (A, B, C, D)] show that noise reduction and feature selection is one of the most difficult task to quantify since it requires significant method to select the appropriate features as presented in our proposed pre-processing algorithm [Algorithm 1]. Results of the algorithm are shown in figure [Figure 4] where a set of segmented nuclei are separated on the basis of colour and selected as feature of interest. The properties of each nuclei are acquired by using colour movements [Figure 7], colour histograms [Figure 8], grey scale intensities and coordinate level features [Figure 9]. These properties of nuclei provided assistance to doctors during the phase of confirmation of papillary structure identification and each quantified observation was recorded in dataset preparation phase. Distance measuring results are presented in [Figure 9], [Figure 10] where distances of each location are measured on a central line and [Figure 5] is used to visualize the segmented nuclei with seeds where coordinate relationships are shown in [Figure 11]. Doctors are provided in-depth assistance at the time of class

label declaration by providing potential candidates for linear relationships between the nuclei since doctors required threshold was more than two nuclei with linear spatial locations to qualify the condition at the time of training classifier. The illustration of these associations are shown in [Figure 12] and [Figure 13] where evidence of papillary structure is represented in shapes of graphs. In the final phase, we evaluated performance evaluation by using confusion matrix, precession and recall measures. The measured overall classification accuracy of our system is about 90.32% with 10-k fold cross validation [Table 1] and estimated ROC curve is presented in [Figure 14].

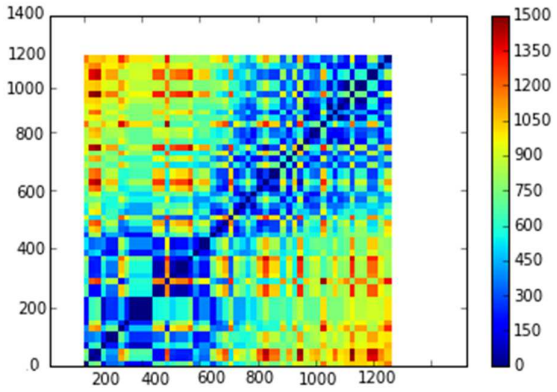


Figure 7: Colour feature movements shown in colour spectrum

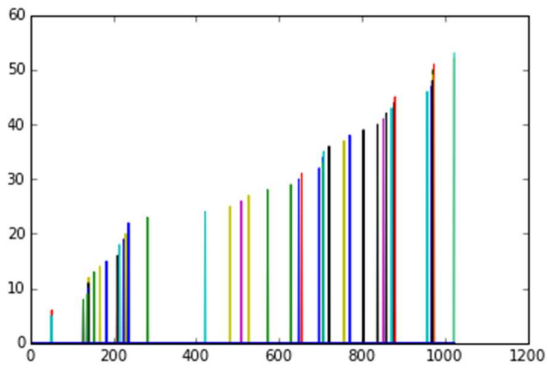


Figure 8: Colour Histogram for coordinates of nuclei

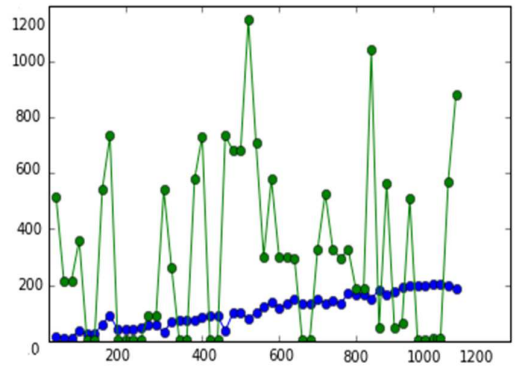


Figure 9: Nuclei Coordinate distances

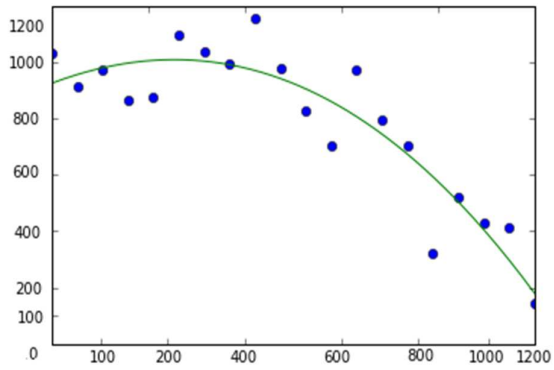


Figure10: Nuclei locations on a central line

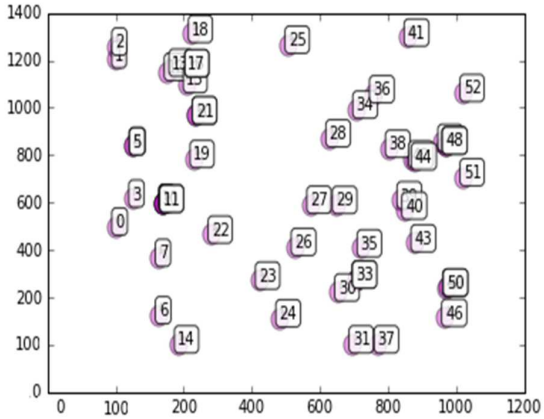


Figure 11: Nuclei spatial points with total count

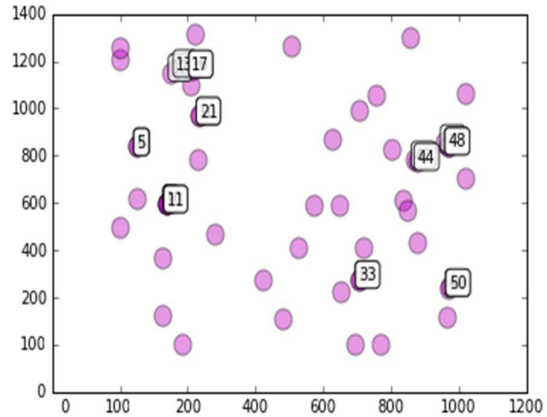


Figure 12: visualization of segmented nuclei with spatial position

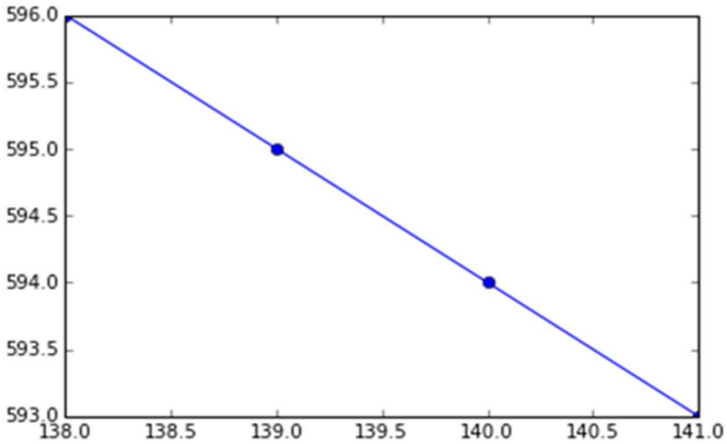


Figure 13: Linear relation of nuclei

In [Figure 7] colour movements are visualized by acquiring the dot product of colours which shows that those colours which have blue type combinations are the nearest colours to each location of nuclei. The quantification of these colours is presented in histogram [Figure 8] where lowest distance movements have maximum blue type combinations and yellowish colour lines have maximum distances. The purpose of this quantification analysis is to obtain minimum distances of objects with linear locations. Thus linear distances of every nuclei are shown in [Figure 10] and theses distance can be analysed in [Figure 9] where each nuclei is presented in shape of points on a matrix having 1400 by 1200 units and the range of maximum distance y is in 1200 units on the x axis variables designated as unique location of each nuclei since the [Figure 10] have a central line like hyperplane where unique locations are significant when it is most linearly nearer to the line. In [Figure 11] all nuclei locations are plotted in shape of spatial class labels and total count of nuclei is visualized. our proposed algorithm [Algorithm 1] have separated the sub-graphs of objects and further analysis is shown in [Figure 12] where clusters of most nearer distances of objects are presented to doctors showing the potential candidates for papillary structure whereas [Figure 13] have shown linear lines for most probable set of nuclei. We have performed performance evaluation in [Table 1, Table2]. The precision for papillary class is measured as 90% and recall is estimated as 68%. Precision for non-papillary class is measured as 90% and recall is estimated as 90.31% as shown in [Table 2]. [Table 3] is presented to describe the comparison of our proposed approach with literature, which shows that our novel approach has more 90% classification accuracy. All other approaches use DICOM images of FNAC (Fine needle aspiration cytology) whereas in our use case we presented FNAB (Fine needle aspiration biopsy) images.

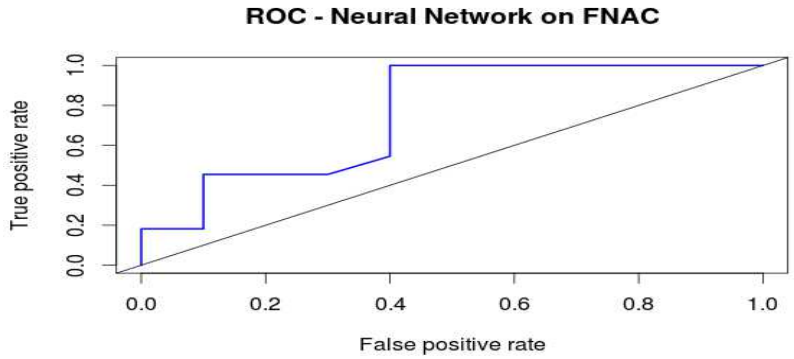


Figure14: ROC Curve estimated for papillary and non-papillary class

Table 1. Confusion matrix

	Papillary	Non-Papillary
Papillary	354	54
Non-Papillary	1053	78

Table 2.Overall performance of proposed methodology

	Raw DICOM images	No of Extracted nuclei	No of classified Nucleis	No of miss-classified Nuclies	Precession	Recall
Papillary	20	408	354	**54**	90%	**68.00%**
Non-Pappilary	20	1131	860	**78**	90%	90.31%

Table 3.Comparison of our system with literature

Approaches	Image Type (1-Ultrasound, 2- FNAC)	Cancer Type (A-Thyroid cancer, B- others) + Accuracy
1	1	A ,97.50 %
2	1	A, 78.00%
3	2	A, 90%
4	2	A, 98.00%
5	2	A, 98.80%
Our Proposed approach	2	A, 90.32%

6. DISCUSION AND CONCLUSION

FNAB (Fine needle aspiration biopsy) is useful traditional tool used for medical diagnosis and it is considered as gold standard to investigate about the malignancy but it is very difficult to differentiate between the mimic, mix and complex patterns of thyroid papillary carcinoma (PTC), because some of features are hard to discernment from each other such as small cell carcinoma may mimic insular carcinoma.We prepared our own datasets for training and testing purposes from the real-world dataset of FNAB consisting upon the 40 raw images because we could not found any related dataset in literature. The used datasets were received from department of pathology, SMBBMU (Shaheed Muhtarma Be Nazir Bhutto Medical University), Pakistan. We segmented1539 nuclei and extracted (colour and grey level) 36 features for each object. Our proposed system, so called (IDSPTC) Intelligent Diagnostic System for thyroid papillary carcinoma consists upon four major stages. In first stage we prepare the dataset, in second and third stages, we extract the coordinates of nuclei and build a decision model by using ANN (artificial neural network), finally, estimated performance evaluation and model testing. The measured accuracy of our proposed system is 90.32% with 10-k fold cross validation. We conclude that chances of misdiagnosis may be avoided by making a careful investigation conducted through machine learning techniques; since nuclei sequences and its properties on DICOM images containsufficient colour information to discernment the mimic behavioursof papillary thyroid carcinoma. A pixel level (in-depth level) cell segmentation would provide more precise assistanceto doctors in their diagnostic process.

ACKNOWLEDGEMENTS

The author's activity was supported by Higher Education commission, Indigenous PhD. fellowship programme, HEC, Pakistan. I would like to say special thanks to Dr. Siraj Ahmed and Dr. Manzoor Ahmed, for providing environment to test and validate the proposed system.

REFERENCES

[1] Fan, Y. Shi, L. Liu, Q. Dong, R. Zhang, Q. Yang, S & Wang, J, 2009. Discovery and identification of potential biomarkers of papillary thyroid carcinoma. Mol Cancer, 8(1):79-93.

[2] Pourahmad, S. Azad, M. Paydar, S. & Reza, H, 2012. Prediction of malignancy in suspected thyroid tumour patients by three different methods of classification in data mining. In First International Conference on advanced information technologies and applications: 1-8.

[3] Dimitris K. Iakovidis, L,2010. Fusion of fuzzy statistical distributions for classification of thyroid ultrasound patterns. Journal of Artificial Intelligence in Medicine 50 (2010): 33–41.

[4] Ding, J. Cheng, H. Ning, C. Huang, J. & Zhang, Y, 2011. Quantitative measurement for thyroid cancer characterization based on elastography.Journal of Ultrasound in Medicine, 30(9):1259-1266.

[5] Iakovidis, D. K., Keramidas, E. G. &Maroulis, D, 2010. Fusion of fuzzy statistical distributions for classification of thyroid ultrasound patterns. Journal of Artificial Intelligence in Medicine 50 (2010): 33–41.
[6] Chuan, Y, 2010. Application of support-vector-machine-based method for feature selection and classification of thyroid nodules in ultrasound image. Journal of Pattern Recognition 43 (2010): 3494–3506.
[7] Bell A.A. Kaftan, J.N. Schneider, T. E, 2006. Imaging and Image Processing for Early Cancer Diagnosis on Cytopathological Microscopy Images towards Fully Automatic AgNOR Detection, WrithArchin science journal.
[8] Dimitris G. Panagiota S. Stavros T. Giannis K. Nikos D. George N.Dionisis C, 2004. Unsupervised segmentation of fine needle aspiration nuclei images of thyroid cancer using a support vector machine clustering methodology, 1st IC-SCCE Athens, 8-10 September, 2004.
[9] Suzuki, H. Saita, S. Kubo, M. Kawata, Y. Niki, N.Nishitani, H. & Moriyama, N, 2009. An automated distinction of DICOM images for lung cancer CAD system. In SPIE Medical Imaging. International Society for Optics and Photonics: 72640Z-72640Z.
[10] Glotsos, D, Tsantis, S. Kybic J, Daskalakis, A. (2013). Pattern recognition based segmentation versus wavelet maxima chain edge representation for nuclei detectionin microscopy images of thyroid nodules. Euromedica medical center, Department of Medical Imaging, Athens, Greece 2013.
[11] Gopinath, B, Shanthi, N. (2013). Support Vector Machine Based Diagnostic System for Thyroid Cancer using Statistical Texture Features, Asian Pacific J Cancer Prev, 14 (1): 97-102.
[12] Gopinath, B, Shanthi, N. (2013). Computer-aided diagnosis system for classifying benign and malignant thyroid nodules in multi-stained FNAB cytological images, Australas Phys EngSci Med (2013) 36:219–230.
[13] Stavros T., (2004). Improving diagnostic accuracy in the classification of thyroid cancer by combining quantitative information extracted from both ultrasound and cytological images. 1st IC-SCCE-Athens, 8-10 September, 2004.
[14] http://www.thyroidmanager.org/chapter/fine-needle-aspiration-biopsy-of-the-thyroid-gland/ accessed on 15-01-2014.
[15] Thyroid Cancer: Web Site, http://www.medicinenet.com/thyroid_cancer/article.htm accessed on 15-01-2014
[16] Han J., Kamber M. & Pei P. Data Mining: Concepts and Techniques", (2011). 3rd (ed.) the Morgan Kaufmann Series in Data Management Systems.
[17] Rafeal, C. G. Richard, E. W. 2007. Digital Image Processing, 3rd (ed.). Prentice Hall. 2007.
[18] Hussein, Mohamed, Amitabh Varshney, and Larry Davis,2007. On implementing graph cuts on cuda. First Workshop on General Purpose Processing on Graphics Processing Units. Vol. 2007.
[19] Boykov, Y. &Funka-Lea, G, 2006. Graph cuts and efficient ND image segmentation. International journal of computer vision, 70(2): 109-131.
[20] Malik, J.Belongie, S. Leung, T. & Shi, J. 2000. Contour and texture analysis for image segmentation. In Perceptual Organization for artificial vision systems, Springer US:139-172.
[21] Al-Kofahi, Y. Lassoued, W. Lee, W. &Roysam, B, 2010. Improved automatic detection and segmentation of cell nuclei in histopathology images.Biomedical Engineering, IEEE Transactions on, 57(4): 841-852.
[22] Liu, X.Huo, Z. & Zhang, J, 2006. Automated segmentation of breast lesions in ultrasound images. In Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of theIEEE: 7433-7435.
[23] Al-Kofahi, Y. Lassoued, W. Lee, W. &Roysam, B, 2010. Improved automatic detection and segmentation of cell nuclei in histopathology images.Biomedical Engineering, IEEE Transactions on, 57(4): 841-852.
[24] Xu, N. Ahuja, N. & Bansal, R, 2007. Object segmentation using graph cuts based active contours. Computer Vision and Image Understanding, 107(3): 210-224.
[25] Freedman, D.& Zhang, T, 2005. Interactive graph cut based segmentation with shape priors. In Computer Vision and Pattern Recognition, 2005.CVPR 2005. IEEE Computer Society Conference on IEEE Vol. 1: 755-762.