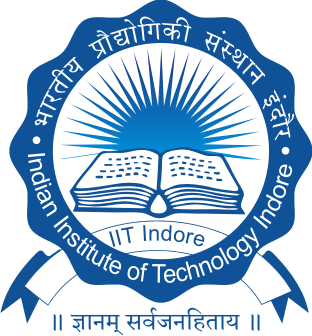
**B. TECH. PROJECT REPORT**

**On**

**TIRADS Based Thyroid Nodule Classification Using Texture Exploiting Descriptors**

BY

|  |  |
| --- | --- |
| **Sparsh Gupta** | **Neeraj Verma** |
| **170001049** | **170005020** |



**DISCIPLINE OF COMPUTER SCIENCE AND ENGINEERING**

**INDIAN INSTITUTE OF TECHNOLOGY INDORE**

**May 2021**

**TIRADS Based Thyroid Nodule Classification Using Texture Exploiting Descriptors**

**A PROJECT REPORT**

*Submitted in partial fulfillment of the*

*requirements for the award of the degrees*

***of***

**BACHELOR OF TECHNOLOGY**

**in**

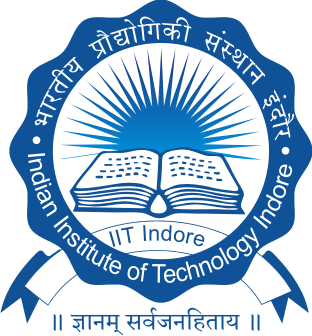
**COMPUTER SCIENCE AND ENGINEERING**

*Submitted by:*

|  |  |
| --- | --- |
| **Sparsh Gupta** | **Neeraj Verma** |
| **170001049** | **170005020** |

*Guided by:*

**Dr. Kapil Ahuja, Associate professor**



**INDIAN INSTITUTE OF TECHNOLOGY INDORE**

**May 2021**

# CANDIDATE’S DECLARATION

We hereby declare that the project entitled “**TIRADS Based Thyroid Nodule Classification Using Texture Exploiting Descriptors”** submittedin partial fulfillment for the award of the degree of Bachelor of Technology in ‘Computer Science and Engineering’ completed under the supervision of **Dr. Kapil Ahuja,** IIT Indore is an authentic work.

Further, we declare that we have not submitted this work for the award of any other degree elsewhere.

|  |  |
| --- | --- |
| Sparsh Gupta  26/05/2021 | Neeraj Verma  26/05/2021 |

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

# CERTIFICATE by BTP Guide

It is certified that the above statement made by the students is correct to the best of my/our knowledge.

|  |  |
| --- | --- |
|  | Dr. Kapil Ahuja  (Associate Professor) |

# Preface

This report on “TIRADS Based Thyroid Nodule Classification Using Texture Exploiting Descriptors” is prepared under the guidance of Dr. Kapil Ahuja.

*Through this report we have tried to give a detailed design of a classification system for thyroid nodules and try to cover every aspect of the system.*

*We have tried to the best of our abilities and knowledge to explain the content in a lucid manner. We have also added graphs, tables and figures to make it more illustrative.*

**Sparsh Gupta, Neeraj Verma**

B.Tech. IV Year

Discipline of Computer Science and Engineeering

IIT Indore

# Acknowledgements

We wish to thank Dr. Kapil Ahuja and Aditya Anand Shastri for their kind support and valuable guidance.

It is their help and support, due to which we became able to complete the design and technical report.

Without their support this report would not have been possible.

**Sparsh Gupta, Neeraj Verma**

B.Tech. IV Year

Discipline of Computer Science and Engineering

IIT Indore

# ABSTRACT

Researchers have developed Computer-Aided Diagnosis (CAD) systems to help doctors diagnose thyroid nodules to reduce traditional methods errors. Doctors' experiences are the basis of the traditional diagnosis methods. Therefore, such systems' performance plays a vital role in enhancing the quality of a diagnosing task. Although state-of-the-art studies regarding this problem are based on handcrafted features, deep features, or the two's combination, their performances are still limited. To overcome these problems, we propose an ultrasound image-based diagnosis of the malignant thyroid nodule method using artificial intelligence. We have used a two-stage classification to classify the thyroid nodules into its respective TIRADS score in this work. Our experiments with a popular open dataset, namely the Thyroid Digital Image Database (TDID), confirm our method's superiority compared to state-of-the-art methods.

# CONTENTS

[CANDIDATE’S DECLARATION 2](#_Toc72992422)

[CERTIFICATE by BTP Guide 2](#_Toc72992423)

[Preface 3](#_Toc72992424)

[Acknowledgements 4](#_Toc72992425)

[ABSTRACT 5](#_Toc72992426)

[CONTENTS 6](#_Toc72992427)

[List of Figures 8](#_Toc72992428)

[List of Tables 9](#_Toc72992429)

[INTRODUCTION 10](#_Toc72992430)

[TIRADS (Thyroid Imaging Reporting and Data Systems) 12](#_Toc72992431)

[LITERATURE REVIEW 14](#_Toc72992432)

[PROPOSED METHOD 17](#_Toc72992433)

[Preprocessing and Enhancement 18](#_Toc72992434)

[Image Binarization 18](#_Toc72992435)

[Image Normalization 20](#_Toc72992436)

[Image Enhancement 20](#_Toc72992437)

[Feature Extraction 21](#_Toc72992438)

[Histogram of Oriented Texture (HOT) 21](#_Toc72992439)

[Pass Band – Discrete Cosine Transform 26](#_Toc72992440)

[Feature Selection 27](#_Toc72992441)

[Minority Oversampling 28](#_Toc72992442)

[EXPERIMENTAL RESULTS 29](#_Toc72992443)

[Dataset and Experimental Setup 29](#_Toc72992444)

[Criteria for Classification Performance 30](#_Toc72992445)

[Result and Comparison of stage 1 32](#_Toc72992446)

[Results of Stage 2 33](#_Toc72992447)

[CONCLUSION and future work 36](#_Toc72992448)

[References 37](#_Toc72992449)

# List of Figures

[Figure 1 - TIRADS score with respective cancer risk % 13](#_Toc72994821)

[Figure 2 - Proposed work 13](#_Toc72994822)

[Figure 3 - Flow diagram of the proposed classification system; (a) training phase and (b) testing phase 18](#_Toc72994823)

[Figure 4 - Example result of thyroid region detection algorithm. (a) an input ultrasound thyroid image; (b) the binarized image; (c) the final detection result 19](#_Toc72994824)

[Figure 5 - Example of thyroid image preprocessing. (a) an input image; (b) cropped image; (c) enhanced image 21](#_Toc72994825)

[Figure 6 - HOG descriptor calculation 24](#_Toc72994826)

[Figure 7 - Performance of Benign-HOT 34](#_Toc72994827)

[Figure 8 - Performance of Benign-PB-DCT 34](#_Toc72994828)

[Figure 9 - Performance of Malignant-PB-DCT 35](#_Toc72994829)

[Figure 10 - Performance of Malignant-HOT 35](#_Toc72994830)

# List of Tables

[Table 1 - TIRADS scores and its associated suspicion 12](#_Toc72994584)

[Table 2 - Comparison of existing and proposed classification systems 15](#_Toc72994585)

[Table 3 - Distribution of benign and malignant images according to TIRADS scores 29](#_Toc72994586)

[Table 4 - Comparison of classification accuracies for various descriptors on the TDID 32](#_Toc72994587)

[Table 5 - Results of the second stage 33](#_Toc72994588)

# INTRODUCTION

A thyroid nodule (or a lump) that develops in the thyroid gland of a human being is a disease in which cells grow abnormally and are likely to spread to other parts of the body [1]. The presence of this nodule may or may not be an indication of thyroid cancer. When a thyroid nodule is found, an ultrasound of the thyroid region is done to confirm if this nodule is, in-fact, a cancerous/non-cancerous nodule, which in medical terms is called benign/malignant nodule, respectively. Favorably, most of the detected thyroid nodules are benign. However, the nodule's presence (whether benign or malignant) causes various health problems in patients, like difficulty in breathing and swallowing [2]. Moreover, malignant thyroid nodules can produce an additional hormone called thyroxine, which causes some critical problems with a patient's health and may result in his/her death. Hence, classifying these nodules in their early stage can reduce the chances of the patient's death.

The abnormalities like hoarseness, swollen glands in the neck, difficulty in swallowing, difficulty in breathing, pain in the throat or neck, a lump in the front of the (neck near Adam's apple), etc., are the symptoms of thyroid cancer [1]. There are several imaging techniques for examination of the thyroid, such as Computed Tomography (CT) scanning, Ultrasound (US) imaging, X-ray imaging, etc. [1]. Ultrasound imaging is the most effective tool for the early detection of thyroid cancer that uses high-frequency sound waves to create a picture of the internal organs [1] [2].

The traditional diagnostic technique, where doctors diagnose cancerous tumors from the images (CT, US, X-ray, etc.), sometimes may give false results as this diagnosis highly relies on the doctor's personal knowledge and experience. Furthermore, determining whether a thyroid nodule is benign or malignant is challenging for doctors based only on symptoms or experience. Hence, nowadays, researchers focus on developing Artificial Intelligence (AI) based imaging techniques [3] [2]. The development of image-based Computer-Aided Diagnosis (CAD) systems in medical research serve as the additional expertise to assist doctors for an accurate diagnosis.

Thyroid images can be better classified by using their texture properties [4]. We use the following two descriptors that capture the textural features for classifying thyroid nodules, i.e., the histogram of Oriented Texture (HOT) and Pass Band - Discrete Cosine Transform (PB-DCT). We use Discrimination Potentiality (DP) to select the appropriate features. These proposed descriptors are compared with the existing standard methods for thyroid nodule classification; image augmentation [5], VGG-16 [6], GoogLeNet [7], Circular Mask [2], and Convolutional Neural Network (CNN) [2]. Support Vector Machine (SVM) is the most suitable and widely used classifier for the two-class classification problem. Hence, we use SVM as a classifier.

CAD systems for thyroid nodule classification consist of basic modules: preprocessing thyroid ultrasound images, image enhancement, feature extraction, and classification. The preprocessing step helps remove background and artifacts (additional text or indicator made by the capturing system). Enhancement techniques are applied on the thyroid images to improve visualization of tissues or a tumor. Finally, a feature extraction technique is employed to obtain the features from images and a classifier to classify them.

CAD systems are usually tested on the Thyroid Digital Image Dataset (TDID), an open-access database of thyroid ultrasound images created by Universidad Nacional de Colombia.

## TIRADS (Thyroid Imaging Reporting and Data Systems)

TIRADS is a 5-point classification to determine the risk of cancer in thyroid nodules. The nodules are classified as not suspicious, probably benign, one or more suspicious features, and possible malignancy. These categories are represented by the TIRADS scores of 2,3,4, and 5, respectively. Based on this, we consider the ultrasound images with TIRADS scores of 2 or 3 as the benign cases, while the ultrasound images with TIRADS scores of 4 and 5 as the malignant cases.

Table - TIRADS scores and its associated suspicion

|  |  |
| --- | --- |
| **TIRADS 1** | Normal thyroid gland |
| **TIRADS 2** | Benign nodules |
| **TIRADS 3** | Probably benign nodules |
| **TIRADS 4** | With ultrasound features suspicious of malignancy |
| **TIRADS 5** | Nodules highly suggestive of malignancy |

Classifying the thyroid nodules as benign and malignant (which is performed by all previous studies) is not always useful because benign nodules (specifically of TIRADS score 3) may become malignant nodules (of TIRADS score 4) if not treated accurately. Similarly, patients having malignant nodules of TIRADS score 5 are at higher risk than the patients with TIRADS score 4. Hence, it is vital to classify *further* the thyroid nodules based on their TIRADS scores.

Figure - TIRADS score with respective cancer risk %

In this work, we propose a two-stage classification method based on TIRADS scores. In the first stage, the thyroid nodules are classified as benign and malignant. In the second stage, benign nodules are further classified as nodules with TIRADS scores 2 and 3, and malignant nodules are classified as nodules with TIRADS scores of 4 and 5.

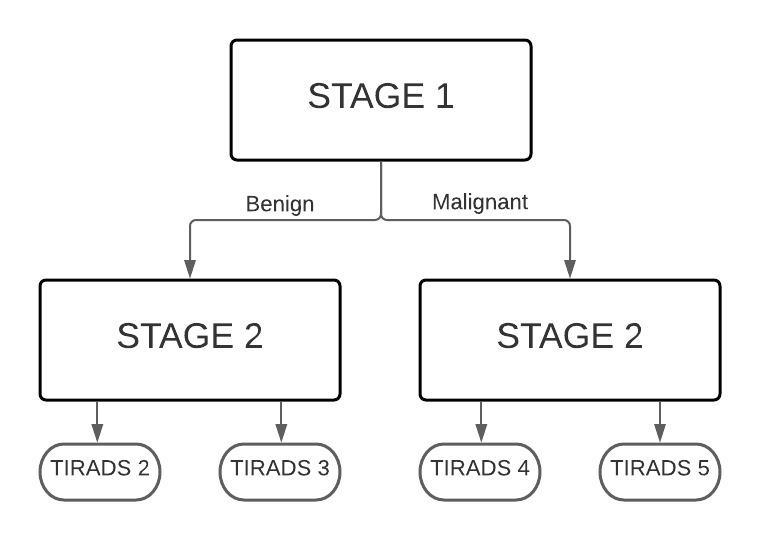


Figure - Proposed work

# LITERATURE REVIEW

Previous works on thyroid nodule classification can be grouped into two categories: deep learning-based and handcrafted-based methods [2]. Before the appearance of deep learning-based methods, handcrafted-based feature extraction methods have been widely used for a long time. This is because they can be easily implemented with simple image-based systems. We construct a deep learning model for feature extraction and thyroid nodules classification from captured ultrasound images in deep learning-based methods. In [7], the authors proposed a classification framework based on Convolutional Neural Network (CNN). Here, they extracted the features from the inputted ultrasound thyroid nodule image using a pre-trained CNN. Then, they used the Support Vector Machine (SVM) method to classify the images into benign and malignant. The authors also constructed another CNN for image classification, and the features were extracted using transfer learning techniques (like VGG16-Net [6] or Inception-Net). In [8], authors developed a multitask cascaded convolution neural network (MC-CNN) framework to exploit thyroid nodules' context information. Similarly, some authors used deep learning using the YOLOv2 neural network to classify the thyroid nodules.

Unlike the deep-learning-based methods mentioned above, handcrafted-based methods use several traditional feature extraction techniques to extract image features and a classifier to classify these features. However, these methods have low classification accuracy as they highly depend upon their feature extraction techniques. In [4], authors used textural features like Gray Level Co-occurrence Matrix (GLCM), Gray Level Run-Length Matrix (GLRLM), and Law's texture energy measures to obtain the features. These features are then classified using the standard SVM. Again, in [9], authors used the Discrete Wavelet Transform (DWT) to locate the tumor region and to extract subtle information from isolated tumor regions for classification. The authors in [10] performed an analysis of linear and nonlinear classifiers for ultrasound images. They showed that both the methods give comparable (almost similar) accuracy. Another study that employed a handcrafted-based method is done by Raghavendra et al. [11], where they used Segmentation-based Fractal Texture Analysis (SFTA) to extract the features.

|  |  |  |
| --- | --- | --- |
| **Category** | **Advantages** | **Disadvantages** |
| Handcrafted-based Methods | *•* Easy implementation  *•* High-performance hardware not required | *•* Low accuracy |
| Deep learning-based Methods | *•* Use deep learning and transfer learning methods  *•* Higher accuracy than handcrafted methods | *•* There is always room for  improvement  *•* High-performance hardware required.  *•* Require more processing time |
| Used Methods (DP-HOT and  DP-PB-DCT) | *•* Easy implementation  *•* High-performance hardware not required  *•* More accuracy due to texture exploiting descriptors | *•* DP-PB-DCT performance is low as compared to DP-HOT |

It is evident from the above works that most of the handcrafted-based methods used textural properties of the ultrasound images. The classification system's performance depends on the extraction of the features and very little on the classier used. The most recent work [12] has proposed the use of the two descriptors: Histogram of Texture (HOT) and Pass Band - Discrete Cosine Transform (PB-DCT) with a feature selection technique called Discrimination Potentiality (DP), which captures the textural information in a better way. These descriptors overcome the only disadvantage that the handcrafted-based methods have: low classification accuracy.

Table 2 - Comparison of existing and proposed classification systems

In this work, we propose to extend this work to perform a two-stage classification of thyroid nodules based on TIRADS scores. In the first stage, the thyroid nodules are classified as benign and malignant. In the second stage, benign nodules are further classified as nodules with TIRADS scores 2 and 3, and malignant nodules are classified as nodules with TIRADS scores of 4 and 5.

# PROPOSED METHOD

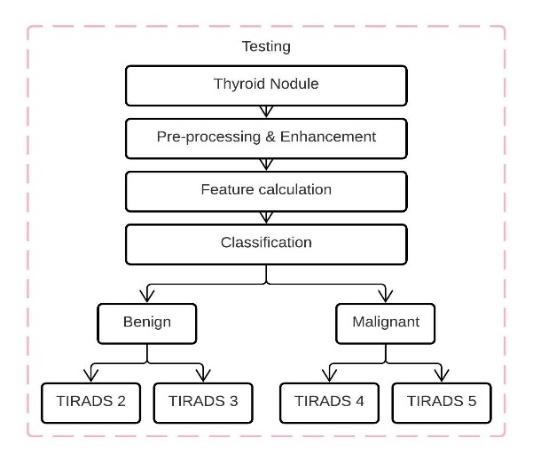
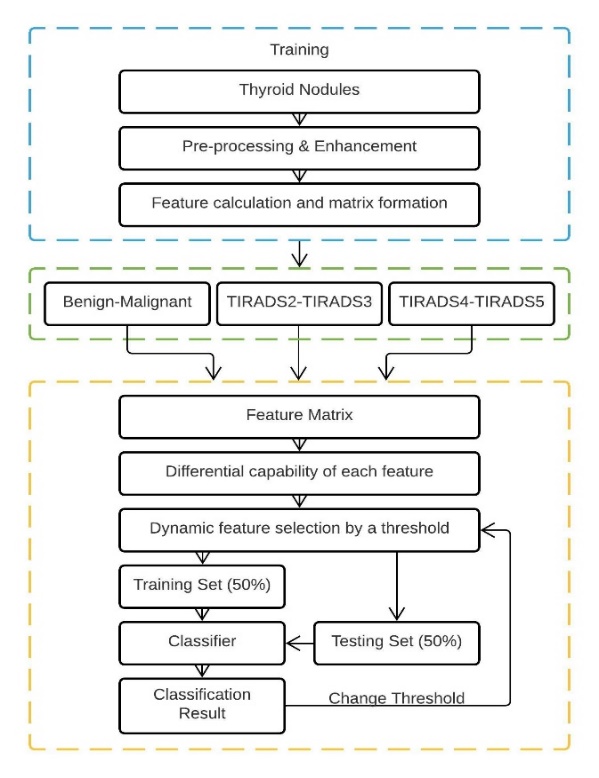
This work proposes a two-stage thyroid nodules classification system. In the first stage, thyroid nodules are classified as benign-malignant, and in the second stage, benign nodules are further classified in TIRADS 2/TIRADS 3, and malignant nodules are further classified in TIRADS 4/TIRADS 5. We are using SVM to classify the nodules in their respective classes.

The framework of the proposed work for the training and testing phase is shown in Figure 3. Here, we first discuss image preprocessing and enhancement techniques. Thyroid nodules are preprocessed for illumination normalization and visibility enhancement of tumors and tissues. A two-stage adaptive histogram equalization enhancement technique is used here for texture enhancement of thyroid nodules. Second, we discuss the two feature extraction techniques, where features of thyroid nodules are extracted from enhanced images. Then, we discuss the feature selection technique used by us, and finally, we discuss the minority oversampling technique.

Figure - Flow diagram of the proposed classification system; (a) training phase and (b) testing phase

(b)

(a)



## Preprocessing and Enhancement

### Image Binarization

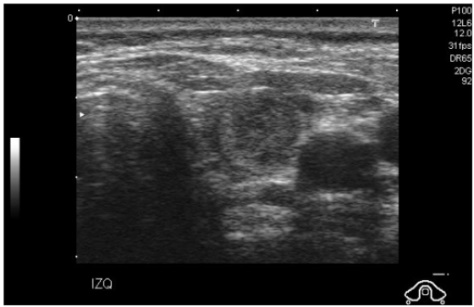
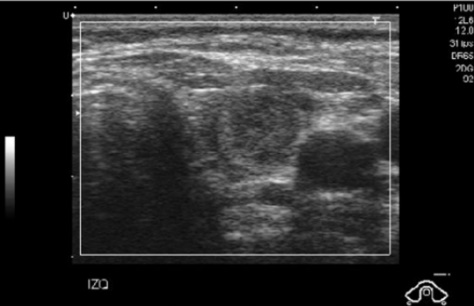
As shown in Figures 4(a), the captured ultrasound thyroid images contain two main parts, i.e., the background (boundary parts with low illumination and some additional artifacts) and the thyroid region (the inner brighter part that captures the details of the thyroid region). It is easy to see that the background regions contain no information about whether an image contains benign or malignant cases of thyroid nodules. It also contains some artifact information added to an image as indicators for the radiologist, such as the patient information or capturing system configuration during the image acquisition process. Due to this reason, the background region should be removed before passing images to the primary classification system. As shown in Figure 4(a), the thyroid region is typically displayed as the largest brighter region in the captured ultrasound thyroid image. Although several brighter regions exist in an ultrasound thyroid image, such as the illumination indicator and text for specifying capturing system configuration, these regions' size is much smaller than that of the thyroid region. We first performed an image binarization method to detect all brighter regions in the captured image using an optimal threshold value based on this observation. Our study used a binarization method proposed by Otsu et al. [13], which takes an input image and performs binarization adaptively by selecting the most suitable threshold value. The result of this binarization step is given in Figure 4(b) using the input image of Figure 4(a). As shown in Figure 4(b), although some brighter regions were detected, the thyroid region had the largest size. Finally, the detected thyroid region was determined by taking the bounding-box in the input image (in Figure 4(a)) based on the selected region of Figure 4(b). An example of a resultant image of this step is given in Figure 4(c) using the input image of Figure 4(a). As we can see from this example, the thyroid region was well localized using our localization method.

Figure - Example result of thyroid region detection algorithm. (a) an input ultrasound thyroid image; (b) the binarized image; (c) the final detection result

(a)

(b)

(c)



### Image Normalization

Then, we normalize the intensity of pixels of the thyroid nodule images because while capturing images, illumination conditions are usually not the same. So, the range of the gray level is different for different thyroid nodule images. Hence, we use a simple and the most commonly used normalization formula (eq.1), which normalizes pixels' intensity between 0 and 1 [14].

|  |  |  |
| --- | --- | --- |
|  |  | Eq.1 |

where is the pixel position, is the normalized pixel intensity, is the actual pixel intensity, is the minimum intensity over all the pixels, and is the maximum intensity over all the pixels.

### Image Enhancement

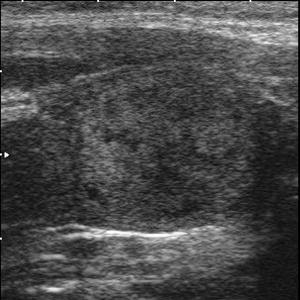
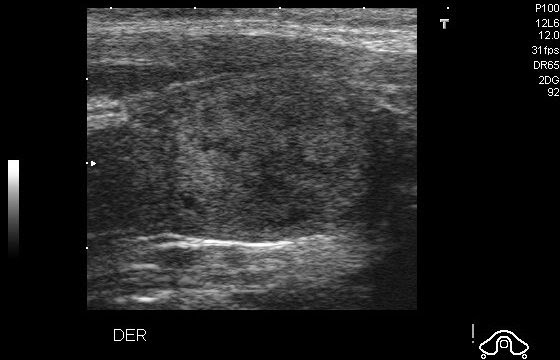
Next, we discuss tissue enhancement of thyroid nodules. Histogram equalization is one of the most basic techniques here, which stretches the contrast of the high histogram regions and compresses the low histogram regions' contrast. As a result, if the region of interest in an image occupies only a small portion, it will not be enhanced appropriately during histogram equalization. This leads to more advanced techniques for enhancement, e.g., Adaptive Histogram Equalization (AHE), Contrast Limited Adaptive Histogram Equalization (CLAHE), Unsharp Masking (UM), Non-Linear Unsharp Masking (NLUM), Two-Stage Adaptive Histogram Equalization (TSAHE), etc. [15]. CLAHE is more suitable for tissue enhancement in thyroid nodules [12]. Since cancerous cells mostly develop in tissues and the HOT and the PB-DCT descriptors are strongly tied to tissue texture, we use a combination of CLAHE and TSAHE. We apply two stages of CLAHE on thyroid nodules in a cascaded order. Firstly, histogram equalization is applied to 8×8 sized blocks, followed by an application to 4×4 sized blocks. Fig. 3 shows the normalized and the enhanced image of a thyroid nodule. It is observed that the tissues are clearly visible in the enhanced image.

Figure - Example of thyroid image preprocessing. (a) an input image; (b) cropped image; (c) enhanced image

(c)

(b)

(a)



## Feature Extraction

This work uses two descriptors (HOT and PB-DCT) for thyroid nodule classification. HOT is a modification of the HOG descriptor where a Gabor filter is used to calculate the angle and magnitude response of a thyroid nodule's texture. Selected PB-DCT coefficients-based features are used here to improve the classification accuracy for each TIRADS score. Next, we discuss these two techniques separately. To the best of our knowledge, these strategies have not been applied anywhere for thyroid nodule classification based on the TIRADS score.

### Histogram of Oriented Texture (HOT)

Here, we derive our HOT descriptor. Firstly, we discuss the calculations of gradient and orientation of an image and the HOG descriptor calculation from cells and blocks partitions [16]. Secondly, we describe a Gabor filter, which is used to extract the magnitude and orientation of tissue texture information, and finally, we discuss modifications to the HOG descriptor that involves a Gabor filter and parameter selection.

The gradient of an image in horizontal and vertical directions, for a pixel position is computed as:

|  |  |  |
| --- | --- | --- |
|  |  | Eq.2 |

for each pixel, the gradient magnitude and orientation are computed as below.

|  |  |  |
| --- | --- | --- |
|  |  | Eq.3 |

Orientation range is quantized into bins (i.e., with . The image is divided into non-overlapping cells, and cells are integrated as one block. Two adjacent blocks can overlap. The histogram of orientations of within cell is computed as

|  |  |  |
| --- | --- | --- |
|  |  | Eq.4 |

The histogram of jth block is obtained by integrating (Histogram of Cells) within this block as follows:

where, || denotes histograms concatenation into a vector. The vector of is finally normalized by -norm block normalization as below to obtain .

where, is a small constant to avoid the problem of division by zero. Histogram of Oriented Gradients (HOG) can be obtained by integrating normalized histograms of all blocks as below.

Where, is the number of possible blocks in an image, which is equal to . Figure 6 shows an example of cell partitions, the formation of overlapped blocks, and histograms' concatenation to get the HOG descriptor. Finally, the length of HOG

Different line-shape filters or tools are available in the literature to extract lines and orientation features of a texture image [17].

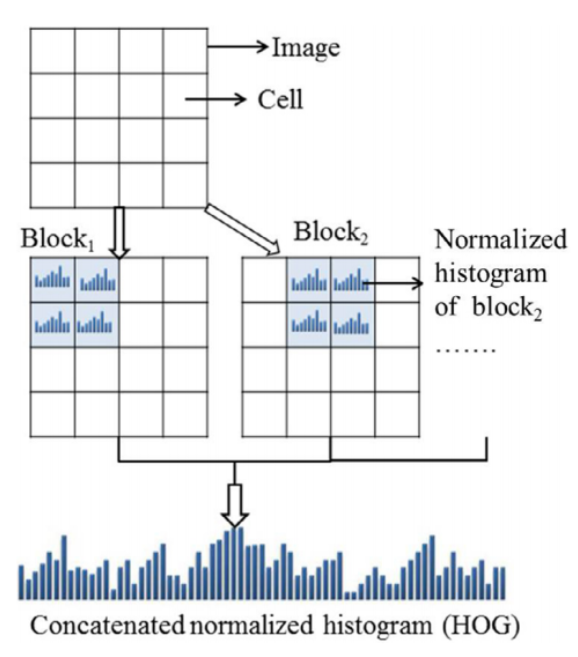


Figure - HOG descriptor calculation

2-D Gabor filters have been found more suitable filter bank to extract biological-like textural features of simple cells in the mammalian visual system [18]. Thus, a Gabor filter is ideal for calculating multi-orientation texture features of a thyroid nodule. A Gabor function is defined as follows:

|  |  |  |
| --- | --- | --- |
|  |  | Eq.5 |

where is the frequency of the sinusoidal wave, controls the function's orientation, and is the standard deviation of the Gaussian envelop. Based upon this Gabor function, a set of Gabor filters can be created for different scales and orientations. Here, texture feature extraction for a given thyroid nodule image is calculated by the real part of a Gabor filter bank with eight different orientations and a fixed scale. Gabor magnitude, and Gabor orientation, the response of each pixel are computed as

|  |  |  |
| --- | --- | --- |
|  |  | Eq.6 |

where, means the convolution operation. The direction is calculated as follows:

The features are calculated by varying the values of and . We combine HOG with a Gabor filter and name it as Histogram of Oriented Texture (HOT). HOT is computed in the same way as but and are used as magnitude and orientation of texture line instead of Eq.3.

Finally, the HOT descriptor's optimum parameters for both types of classification are chosen by experiments. The value of is varied from one to five to obtain an optimum number. The value of is computed as . In this work, the magnitude image is divided into equal-sized cells. The size of a block considered is therefore, overlapped blocks are formed. The orientation range is quantized into 8 bins, and therefore, the final length of the resultant HOT descriptor is The HOT descriptor's length is considerable, and all features do not have the same discrimination capability [19]. Feature selection schemes help to select more appropriate features.

### Pass Band – Discrete Cosine Transform

2D Discrete Cosine Transform (DCT) transforms images into frequency representation from the spatial form. It also provides energy compaction that helps to reduce the information redundancy by retaining only a few coefficients. DCT coefficients of image of size are calculated as follows:

|  |  |
| --- | --- |
|  |  |
| Eq.7 | |

where is defined by

DCT based various feature extraction and compression techniques have been proposed in the literature [20]. Usually, DCT features are formed by selecting the most prominent and discriminating coefficients based upon some criterion [20]. The DCT coefficients can be divided into three sets, low frequencies, middle frequencies, and high frequencies. Low frequencies are correlated with illumination conditions; middle frequencies represent texture features, while high frequencies represent small variance or noise. Illumination and texture properties are essential for thyroid nodule classification. Therefore, this work uses low and middle coefficients to form the descriptor (the Pass Band - Discrete Cosine descriptor or abbreviated as PB-DCT). Finally, the more discriminate DCT coefficients are selected based upon a discrimination criterion, discussed in the next section.

## Feature Selection

All features do not have the same ability to discriminate various classes (benign-malignant and TIRADS scores) [19], and they do not increase the accuracy based on available information for each class. Therefore, it is necessary to eliminate irrelevant features and select the most discriminative features among a given set of features. There exist many techniques for feature selection. Some of the common ones are as follows: PCA (Principal Component Analysis) method, Markov blanket method, wrapper methods (e.g., sequential selection algorithm, etc.), filter methods (e.g., Pearson correlation criteria, mutual information, etc.), embedded methods, and statistical measures-based methods (e.g., *T-*test, Kolmogorov–Smirnov test, etc.) [21].

Out of these, wrapper methods, filter methods, and statistical measures-based methods are usually used for thyroid nodule classification. Wrapper methods are computationally expensive since the number of steps required for obtaining the feature subset is very high. Filter methods sometimes lead to a redundant feature subset, and hence, are not optimal in this sense [21]. Thus, we go for statistical measures-based methods since they do not have the drawbacks mentioned above. These methods also have the advantage of reducing the feature space without significantly degrading the classification performance. The T-test method is one such method that gives a high score to features that capture the texture and the shape of thyroid nodules. As earlier, capturing texture is very important to us. Moreover, the *T*-test method is computationally light, and easy to implement. Thus, we use this feature selection method and show in the results section that this works very well with the two descriptors (DP-HOT and DP-PB-DCT). We term it as Discrimination Potentiality (DP) because of its capability in discriminating between the available features.

The discrimination potentiality of the feature between two classes (a and b) is computed from a given training set as follows:

|  |  |  |
| --- | --- | --- |
|  |  | Eq.8 |

where and are mean and standard deviation values of the feature for and classes, respectively. and are the number of thyroid nodules for and classes, respectively. A high value of DP means high discrimination ability of the corresponding feature.

All features (columns) of the feature matrix are arranged in descending order of their *DP* value. Initially, the first 100 features with the highest *DP* values are chosen for classification accuracy. Then, classification accuracy is calculated by adding features, with the next higher value of *DP*, one by one until we get the highest accuracy. The optimum subset of features corresponding to the highest accuracy is selected as the final descriptor.

## Minority Oversampling

If the instances for one class are relatively less than the instances for other class, then the dataset comes under an imbalanced dataset category. In this scenario, we have two classes: the majority class and the minority class. Thus, in this context, many classification algorithms have low accuracy for the minority class. The most common way to solve this problem is to use Synthetic Minority Over-sampling TEchnique (SMOTE) [22]. This technique resamples the original dataset, either by under-sampling the majority class and/or oversampling the minority class. Here, we perform the oversampling of the minority class, so that number of instances for both the classes are almost similar.

In stage 1, the minority class is 'Benign', and in stage 2, the minority class among the benign cases is of TIRADS 2, and the minority class among the malignant cases is of TIRADS 5.

# EXPERIMENTAL RESULTS

## Dataset and Experimental Setup

|  |  |  |
| --- | --- | --- |
| **TIRADS** | **No. of**  **Images** | **Classification (Total Images)** |
| 2 | 42 | Benign  (61) |
| 3 | 19 |
| 4 | 243 | Malignant  (288) |
| 5 | 45 |

As mentioned earlier, we use the TDID database for our experiments, which consists of 349 images. Each original image is of size 360 *×* 560, which becomes 300 *×* 300 after preprocessing. Out of these, 61 are benign, while 288 are malignant. Table 3 lists the number of images in TDID based on the TIRADS classification.

Table - Distribution of benign and malignant images according to TIRADS scores

Experiments are carried out in MATLABR2020 on a machine with an Intel i5 processor @2.5 GHz and 8GB RAM. Our system's performance (and comparative systems) is evaluated by standard metric of sensitivity, specificity, and accuracy.

We first use two-fold cross-validation, where the dataset is randomly divided into two equal parts. One part is used for training, and the other is used for testing. Then, the two parts are swapped, i.e., the one used for training earlier is now used for testing, and the one used for testing earlier is now used for training. At the end of this exercise, average performance is saved. Finally, we repeat two-fold cross-validation ten times to remove any bias related to the dataset's division.

## Criteria for Classification Performance

To measure a thyroid classification system's performance, three popular metrics have been in use, including sensitivity, specificity, and overall classification accuracy. Like the previous studies, we also used these three performance measurements in our experiments to measure our proposed method's performance.

Sensitivity is computed as the number of true positive cases over the number of actual positive cases. It is represented as follows:

|  |  |  |
| --- | --- | --- |
|  |  | Eq.9 |

where, TP means True Positive cases and FN means False Negative cases.

Specificity is computed as the number of true negative cases over the number of actual negative cases. It is represented as follows:

|  |  |  |
| --- | --- | --- |
|  |  | Eq.10 |

where, TN means True Negative cases and FP means False Positive cases.

Accuracy is computed as the number of correct classifications over the number of given cases. It is represented as follows:

|  |  |  |
| --- | --- | --- |
|  |  | Eq.11 |

## Result and Comparison of stage 1

Since CAD for a thyroid nodule classification system typically focuses on two different aspects of the classification problem: the correct classification of benign case images and correct classification of malign case images, the specificity, and sensitivity measurements were used to measure the accuracy of these aspects. First, the sensitivity was measured as the ratio between the true positive (TP) samples (samples that are malignant case images are correctly classified as the malignant ones) over the total number of the image of the malignant case (TP + false negative (FN)) in a test dataset as shown in Eq.9. Second, the specificity is the measurement of true negative (TN) samples (samples that are benign cases are correctly classified as benign ones) over the total number of benign case images (TN + false positive (FP)) in a test dataset, as shown in Eq.10. To access an overall (average) ability of the classification system, the third measurement (overall accuracy) was used and measured by the total number of correct classification/detection samples (true positive and true negative samples) over the total number of samples in a test dataset as shown in Eq.11.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sr. No.** | **Descriptors** | **Sensitivity** | **Specificity** | **Accuracy** |
| 1 | Image Augmentation [20] | 94% | 93% | 94% |
| 2 | VGG-16 [22] | 100% | 88% | 94% |
| 3 | GoogLeNet [22] | - | - | 79% |
| 4 | Circular Mask [11] | 95% | 64% | 91% |
| 5 | CNN [11] | 96% | 66% | 92% |
| 6 | DP-HOT | 100% | 90% | 96% |
| 7 | DP-PB-DCT | 90% | 70% | 90% |

We compare the performances of the descriptors with the five existing ones mentioned earlier for benign and malignant classification. The results for this are given in Table 4. From this table, it is evident that DP-HOT performs better than DP-PB-DCT.

Table - Comparison of classification accuracies for various descriptors on the TDID

## Results of Stage 2

The second stage of the thyroid nodule classification system operates on two parts, i.e., benign and malignant. Each part focuses on the binary classification between TIRADS, i.e., correct classification of TIRADS 2 and TIRADS 3 images in the benign part, and TIRADS 4 and TIRADS 5 images in the malignant part. For this purpose, specificity and sensitivity measurements were used to measure the accuracy of these aspects.

First, the sensitivity is measured for TIRADS 2 in the benign part and TIRADS 4 in the malignant part, as shown in Eq.9. Second, the specificity is measured for TIRADS 3 in the benign part and TIRADS 5 in the malignant part, as shown in Eq.10. The sensitivity reflects the ability of a classification system to correctly detect TIRADS 2 cases and TIRADS 4 cases. In comparison, the specificity reflects the ability of a classification system to correctly detect TIRADS 3 cases and TIRADS 5 cases. To access the overall (average) ability of the classification system, overall accuracy was used as shown in Equation (9).

The results for the 2nd stage are given in Table 6.3. From this table, it is evident that DP-HOT performs better than DP-PB-DCT.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Descriptor** | **Sensitivity** | **Specificity** | **Accuracy** |
| **Benign** | HOT | 100% | 90% | 96% |
| PB-DCT | 90% | 80% | 90% |
| **Malignant** | HOT | 89% | 88% | 91% |
| PB-DCT | 90% | 80% | 90% |

Table - Results of the second stage

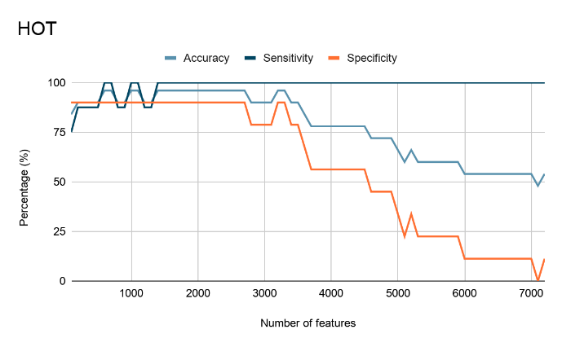
**Benign**

Figure - Performance of Benign-HOT

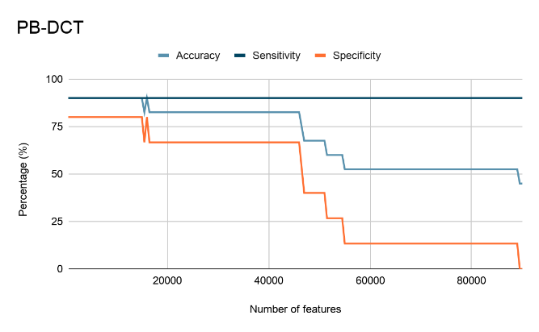


Figure - Performance of Benign-PB-DCT

**Malignant**

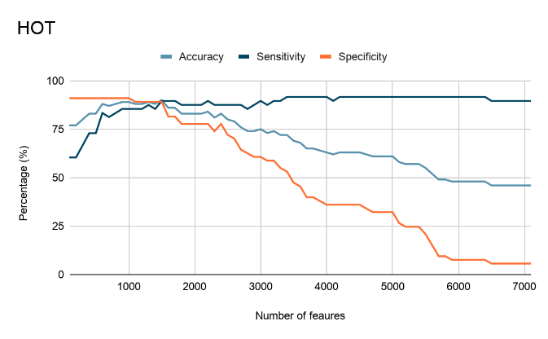
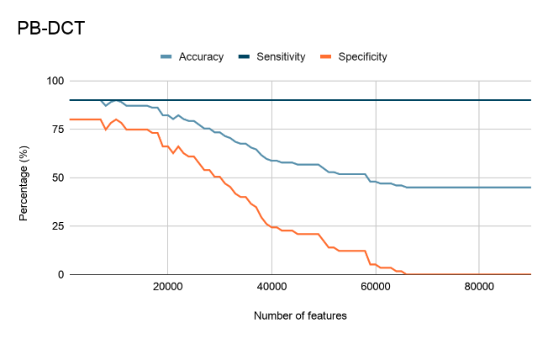


Figure - Performance of Malignant-PB-DCT

Figure - Performance of Malignant-HOT

# CONCLUSION and future work

We present the two-stage thyroid nodule classification method using texture exploiting descriptors (HOT and PB-DCT). In the first stage, the thyroid nodules are classified as benign and malignant. The benign and malignant nodules are further classified based on their TIRADS classes. In the second stage, benign nodules are further classified as nodules with TIRADS scores 2 and 3, and malignant nodules are classified as nodules with TIRADS scores of 4 and 5. This two-stage classification has not been done in any of the previous works. We achieve 96% accuracy for benign classification and 91% accuracy for malignant classification.

In the future, we plan to use the XML metadata associated with each thyroid nodule image to obtain more sensitive regions. It contains the coordinates of the region where the tumor might be present. Extracting features from this region may give more accurate features to classify the images. This can further improve the classification accuracy of our method.

# References

|  |  |
| --- | --- |
| [1] | “Thyroid cancer: Diagnosis,” American Society of Clinical Oncology, [Online]. Available: https://www.cancer.net/cancer-types/thyroid-cancer/diagnosis. [Accessed October 2020]. |
| [2] | D. Nguyen, J. Kang and T. Pham, “Ultrasound image-based diagnosis of malignant thyroid nodule using artificial intelligence,” *Sensors,* p. 20, 2020. |
| [3] | D. Nguyen, T. Pham and G. Batchuluun, “Articial intelligence-based thyroid nodule classication using information from spatial and frequency domains,” *Clinical Medicine,* p. 11, 2019. |
| [4] | C. Chang, S. Chen and M. Tsai, “Application of support-vector-machine-based method for feature selection and classication of thyroid nodules in ultrasound images,” *Pattern Recognition,* pp. 3494-3506, 2010. |
| [5] | Y. Zhu, Z. Fu and J. Fei, “An image augmentation method using convolutional network for thyroid nodule classification by transfer learning,” in *IEEE International Conference on Computer and Communications*, 2017. |
| [6] | K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv,* pp. 1409-1556, 2014. |
| [7] | K. Sundar, K. Rajamani and S. Sai, “Exploring image classification of thyroid ultrasound images using deep learning,” in *International Conference on ISMAC in Computational Vision and Bio-Engineering*, 2018. |
| [8] | W. Song, S. Li and J. Liu, “Multitask cascade convolution neural networks for automatic thyroid nodule detection and recognition,” *IEEE Journal of Biomedical and Health Informatics,* pp. 1215-1224, 2018. |
| [9] | V. Sudarshan, M. Mookiah and U. Acharya, “Application of wavelet techniques for cancer diagnosis using ultrasound images: A review,” *Computers in Biology and Medicine,* pp. 97-111, 2016. |
| [10] | F. Ouyang, B. Guo and L. Ouyang, “Comparison between linear and nonlinear machine-learning algorithms for the classication of thyroid nodules,” *European Journal of Radiology,* pp. 251-257, 2019. |
| [11] | U. Raghavendra, U. Acharya and A. Gudigar, “Fusion of spatial gray level dependency and fractal texture features for the characterization of thyroid lesions,” *Ultrasonics,* pp. 110-120, 2017. |
| [12] | A. Shastri, “Novel Statistical and Probabilistic Machine Learning Algorithms for Genotype Clustering and Cancer Classification,” *Ph.D. Thesis, IIT Indore,* 2020. |
| [13] | N. Otsu, “A threshold selection method from gray-level histogram,” *IEEE Trans. Syst. Man Cybern,* pp. 62-66, 1979. |
| [14] | A. Jain, K. Nandakumar and A. Ross, “Score normalization in multimodal bio- metric systems,” *Pattern recognition,* pp. 2270-2285, 2005. |
| [15] | S. Anand and S. Gayathri, “Mammogram image enhancement by two-stage adaptive histogram equalization,” *Optik-International Journal for Light and Electron Optics,* pp. 3150-3152, 2015. |
| [16] | S. Ergin and O. Kilinc, “A new feature extraction framework based on wavelets for breast cancer diagnosis,” *Computers in Biology and Medicine,* pp. 171-182, 2015. |
| [17] | S. Khan, M. Hussain, H. Aboalsamh, H. Mathkour, G. Bebis and M. Zakariah, “Optimized Gabor features for mass classification in mammography,” *Applied Soft Computing,* pp. 267-280, 2016. |
| [18] | Kamarainen, J. K., V. Kyrki and H. Kälviäinen, “Invariance properties of Gabor filter-based features - Overview and applications,” *IEEE Transactions on Image Processing,* pp. 1088-1099, 2006. |
| [19] | H. Liu, J. Li and L. Wong, “A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns,” in *Genome Informatics*, 2002, pp. 51-60. |
| [20] | S. Dabbaghchian, Ghaemmaghami, M. P. and A. Aghagolzadeh, “Feature extraction using discrete cosine transform and discrimination power analysis with a face recognition technology,” *Pattern Recognition,* pp. 1431-1440, 2010. |
| [21] | G. Chandrashekar and F. Sahin, “A survey on feature selection methods,” *Computers & Electrical Engineering,* pp. 16-28, 2014. |
| [22] | N. Chawla, K. Bowyer and L. Hall, “SMOTE: synthetic minority over-sampling technique,” *Articial Intelligence Research,* pp. 321-357, 2002. |