



# Extreme Learning Machine for Thyroid Nodule Classification with Graph Cluster Ant Colony Optimization Based Feature Selection

Sayyad Rasheeduddin, Kurra Rajasekhar Rao

**Abstract:** *Thyroid nodule is defined as an endocrine malignancy that occurs in humans due to abnormal growth of cells. Recently, an increasing level of thyroid incidence has been identified worldwide. Thus, it is necessary to detect the nodules at an early stage. Ultrasonography is an important tool that is utilized for the detection as well as differentiation of malignant thyroid nodules from benign nodules. The nodules in ultrasound appear in different heterogenic forms, which are difficult to differentiate by the physicians. Further, large number of features available in US characteristics increases the computation time as well as complexity of classification. In this paper, Graph-Clustering Ant Colony Optimization based Extreme Learning Machine approach is proposed to achieve efficient diagnosis of thyroid nodules. It will enhance thyroid nodule classification by selecting only the optimal features and further using it for improving the function of classifier. The main goal of this technique is to differentiate the malignant nodules from the benign nodules. The performance of both feature selection and classification are evaluated through parameters such as accuracy, AUC, sensitivity and specificity. From the experimental results, it is revealed that the proposed method is significantly better than the existing methods. Thus, it is considered to be an effective tool for diagnosing the thyroid nodules with less complexity and reduced computation time.*

**Keywords.:** *Thyroid nodule, ultrasound image, diagnosis, feature extraction, nodules classification.*

## I. INTRODUCTION

Thyroid is a butterfly shaped small gland situated in the lower region of neck under the layers of skin and muscle. The abnormal growth of cells in the thyroid glands referred to as thyroid nodules. These nodules may be either benign or malignant commonly called as non-cancerous or cancerous cells respectively (Acharya et al, 2016). Thyroid nodules are the most common search criteria in thyroid gland as it is present in almost 40% of the population among world-wide and about 5-10% is found to be malignant. Thus, radiologists are involved in diagnosing the thyroid gland to identify the risk of malignancy with respect to the guidelines provided by Thyroid Imaging, Reporting and Data System (TI-RADS). In general, thyroid nodules affect both men and women whereas it is severe in case of women and its formation depends upon different characteristics like gender, age and population (Erdem et al, 2010).

**Revised Manuscript Received on 30 July 2019.**

\* Correspondence Author

**Sayyad Rasheed Uddin\***, Research Scholar, Dept. of CSE, Nagarjuna University, Guntur, AP

**Dr.Kurra Raja Rajasekhar Rao**, Prof. Dept of CSE, Usha Rama College Of Engg And Technology, Vijayawada

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](#) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

The thyroid disease analytics have revealed that thyroid is a severe disorder which increases the mortality rate in humans.

Therefore, it is necessary to produce an accurate tool for malignancy risk detection in order to increase the survival rate of thyroid patients. Moreover, early identification of the symptoms of thyroid disorders can improve the survival rate thereby initiating the treatment at initial stage (Koundal et al, 2018). However, the diagnosing process as well as treatment of thyroid disease remains difficult and the main challenge in this field is differentiation between the nodules. It is necessary to accurately classify the thyroid nodule because of high prevalence of the nodules as well as less prevalence of the malignancy.

Fine needle aspiration biopsy (FNAB) is the standard treatment utilized for diagnosing the thyroid diseases but it is reported that it can mimic other kind of diseases (Bakshi et al, 2003). Several thyroid treatment plans use FNABs as reference since it is labor-dependent and expensive under large scale diagnosis. Likewise, unwanted biopsies cause anxiety, irritation and increase the treatment-expense to thyroid patients (Ma et al, 2017). Even though massive growth is achieved in the field of thyroid diagnostics with sources such as CT imaging, radionuclide and MRIs still it is necessary to select an appropriate and stable material for effective differentiation between the nodules (Wu et al, 2013). Furthermore, clinical procedures do not obtain better diagnosis and so, non-invasive imagery tools like Ultrasonography is identified as a best choice for distinguishing among the nodules. The American Thyroid Association (ATA) stated that ultrasound images are the primary choice of any radiologists for examining the thyroid nodules. Furthermore, if a nodule is identified on other kind of image modalities, detailed diagnostics are performed on US images. Thus, Ultrasonography is the initial stage diagnosing modality for thyroid disease identification. (Cooper et al, 2006) defined that US images are sensitive and suitable for examining the nodularity of thyroid compared to other images such as MRI and CT. Sonography visualize the different characteristics of thyroid glands like dimension, structure, echogenicity, availability of calcification, etc. In literature, more number of research works has been carried out to distinguish between benign and malignant nodules as it is necessary to provide proper and effective treatment to thyroid patients.

Thyroid nodules are comprised of different kind of textural features. Ultrasound images also resemble numerous features like electrographic, textural and morphologic which are important for the purpose of nodule classification.



In the last decade, medical assistants experimented different sonographic features to prove its efficiency in diagnosing the risk of malignancy of thyroid disorders. However, feature selection is the main task of many machine learning based disease classification techniques (Chang et al, 2010). A number of approaches have been developed to extract relevant features from US images. It is identified that the process of feature extraction and using them to train a classifier consumes more amount of time. Therefore, certain features are neglected during diagnosis in order to reduce time and to improve the reliability of classification. A proper method is required to select appropriate features and to neglect irrelevant ones. Classification and prediction achieves accurate result with limited features than processing with all the available features. Thus, proper feature extraction and classification techniques are needed to attain better results in disease prediction. (Tsantis et al, 2009) presented CAD based diagnosis system that used morphologic and wavelet-based features for classifying the thyroid nodules in US images. These features are extracted based on malignancy related characteristics like calcification, irregular shape, uniformity, echogenicity, etc. The efficiency of using extracted features in classification is evaluated through two different pattern recognition algorithms such as probabilistic neural network and support vector machine. It showed that the extracted features can improve the accuracy of classifier and lowers the faults in thyroid disease identification.

In general, computation time and prediction accuracy are the important aspects that are taken into consideration during thyroid nodules differentiation. Therefore, in this research, an improved fast learning based pattern recognition tool called as ELM is utilized for the prediction of thyroid disease with US characteristics. Extreme Learning Machine (ELM) is a new learning based approach that supports single hidden layer feed-forward neural networks (SLFNs). Compared to gradient-based methodologies that iteratively adjust the parameters of neural network, ELM randomly selects the input weights as well as hidden biases for the determination of output weights by adopting the generalized inverse of Moore-Penrose (MP) analytical method. Further, it learns faster with highly generalized performance and also keeps the parameter tuning-free. Due to these properties, ELMs are widely used on classification areas like predicting patient outcomes (Liu et al, 2011), sales forecasting (Chen & Ou, 2011) and so on. Moreover, ELM proved its reliability on number of disease diagnosing tasks over other learning based classification algorithms.

In this paper, a machine learning model is proposed to achieve efficient classification of the thyroid nodules. Existing classification techniques have recommended that using an optimal feature selection process will enhance the accuracy of classifier used for nodule differentiation task. Hence, a novel feature selection method called as graph-clustering based ant colony optimization is adopted in this proposed work. It will select the discriminant features thereby making it effective for classifying the nodules within limited time. Further, Extreme learning machine based classifier differentiates the benign and malignant nodules. The main contribution of this paper is defined as follows:

- A hybrid methodology is proposed to enhance the diagnostics of thyroid disease using ultrasound characteristics
- An optimal feature selection approach called as graph clustering based ant colony optimization tool is applied to extract the relevant features from the raw dataset
- To increase the efficiency as well as accuracy in differentiation of thyroid nodules, a computer aided diagnosis system based on extreme learning machine is also proposed. The remainder of this paper is organized as follows. Section 2 provides a literature review on different feature selection and machine learning based classification approaches. Section 3 presents a background review on graph clustering based ant colony optimization and extreme learning machine classifier. The detailed implementation of proposed methodology is given under Section 4. The experimental results and discussion of the proposed method is visualized in Section 5. Finally, the conclusion of this paper is defined in Section 6.

## **II. LITERATURE REVIEW**

Recently, a number of thyroid disease diagnostic systems were introduced to analyze the severity of thyroid disorders using ultrasound characteristics. It includes computer aided diagnosis (CAD) (Sollini et al, 2018), deep convolutional neural networks (Li et al, 2019; Li et al, 2019), machine learning and so on. As CADs works on the principle of machine learning algorithms, it is mostly preferred by radiologists for identifying the risk of malignancy in thyroid glands. It is identified that machine learning based thyroid disease diagnostic systems would increase the accuracy of analysis with ultrasound imaging.(Ardakani et al, 2015) identified a new approach to analyze the texture of US images based on computer aided diagnosis (CAD) to distinguish the thyroid nodule as benign or malignant. The Receiver Operating characteristic Curve (ROC) analysis showed that Texture Analysis (TA) is a reliable approach which provides useful information to identify and to classify the nodules. Furthermore, this technique consumes low cost and does not need any human intervention as the entire diagnosis is performed on computers. But is tested on small datasets along with highly sensitive FNAB approach. FNABs need the help of surgical pathology to obtain more definitive results and it excludes certain data due to an indeterminate operation.

In classification based applications, features are the important factors that impact the discriminatory functioning of the classifier. Generally, it is effective to consider all the features during classification but it is redundant due to mutual correlation between them. Due to this, it is enough to select the relevant features from the available dataset which will then increase the classification accuracy of classifiers. However, the main task is to choose the suitable feature selection approach for the particular classification algorithm. On the other hand, selecting the optimal feature selection method can improve the classification accuracy but increases the time as well as computation complexity. In the last decade, number of feature selection algorithms was introduced that include random searches, heuristic, greedy and exhaustive.

However, these techniques are computationally very expensive and get trapped into local optima. To overcome such situations, different kind of feature selection methods like Ant Colony Optimization (ACO) (Tabakhi et al, 2014), Genetic Algorithm (GA) (Kabir et al, 2011) and Particle Swarm Optimization (PSO) (Yong et al, 2016) are presented. Of these, ACO seems to be very affective as it is multi-agent based selection methodology. The advantages of ACO over swarm intelligence based techniques include its local and global optima ability, availability of long-term distributed storage, and utility of reinforcement based machine learning concept.

(Choi et al, 2015) provided a systematic approach to quantitatively observe the features of US images on calcified pathological thyroid nodule dataset. The features responsible for tumor malignancy were identified by means of a univariate algorithm and the nodules are differentiated by using neural network approach. The diagnostic ability of both the neural network and feature estimation algorithm sarederived from ROCs and AUCs (Area Under ROCs) respectively. However, the application of this technique is limited to descriptive 2D-calcified datasets as it is necessary to perform retrospective review on those datasets. In addition, only the dataset with surgical cases are taken for analysis and it does not consider clinically visible benign nodule dataset. Conversely, existing CADs cannot provide promising results and several radiologists have reported that its clinical usage is limited on certain practices.

(Ouyang et al, 2019) estimated the performance of linear as well as non-linear machine learning approaches in thyroid nodule malignancy identification with reference to a standard approach. The diagnostic performance analyzed through AUC showed similar AUCs on non-linear techniques compared to linear techniques. Particularly, Kernel SVM and Random Forest algorithms attained moderately larger AUCs compared to other algorithms taken for observation. As these analyses does not perform any pre-processing or feature extraction tasks on image datasets, it is found to be an enhanced approach than the CAD system. The echographic view of thyroid nodule in thyroid imaging is referred to as texture. In mathematical model based diagnosing applications, these textures are analyzed by means of quantitative parameters. This is very helpful for CAD based disease diagnosing applications.

The heterogeneous nature of thyroid nodules, presence of different internal substances and large number of echo patterns in US images confuse the physicians and radiologists to identify the appropriate textures. Thus, textural feature extraction techniques are introduced to distinguish suitable texture patterns thereby decreasing the misdiagnosis rate. (Chang et al, 2010) tested six kinds of SVMs in using important textures and to increase the efficacy of thyroid lesion classification. The experimental outcomes proved the reliability of their method in extracting the important features from thyroid imagery. It is then compared with an existing approach called as sequential-floating-forward-selection (SFFS). This comparison showed that the performance of SVMs in feature extraction is similar to SFFS but the execution time is 3-37 times faster than SFFS.

(Shankar et al, 2018) established a kernel-based classification model to classify the thyroid nodules after selecting appropriate features from thyroid dataset. Grey wolf optimization based feature selection algorithm is

adopted to improve the dataset classification. Their technique showed improved performance on dataset classification but it consumes large amount of time. (Han et al, 2006) discovered an ELM model to predict how long a non-small cell lung cancer postoperative patient can survive. This method showed accurate results in prediction and the convergence rate is faster than the ANN framework. (Zhang et al, 2007) determined the functioning of ELM model on multi-categorical classification of microarray dataset of cancer patients. It was observed that the classification accuracy, training time and the computation complexity of ELM are better than ANN and SVM classifiers. (Helmy&Rasheed, 2009) utilized ELMs to diagnose five types of diseases, and the classification accuracy and computational complexity for this observationis effective with reduced training dataset. (Gomathi & Thangaraj, 2010) suggested computer aided ELM lung cancer diagnostic system. The experimental outcome of this system is compared with SVM approach where ELM produce more accurate results in classification task.

(Li et al, 2012) recommended computer operated diagnostic system that works on the basis of principal component analysis and extreme learning machine. It performed thyroid diagnosis by taking into account different characteristics of ELM like simplicity, less complexity, generalized behavior, faster learning capability and less time for computation. Further, feature extraction is performed through PCA which ignores irrelevant data and consider only the appropriate features for ELM classification. This technique is efficient in classifying three different forms of thyroid disorders like hyperthyroid, hypothyroid and normal thyroid. PCA-ELM classifier is precise and it provides accurate categorization of thyroid malignancy. However, still is a challenging task to provide timely efficient feature selection approach for selecting appropriate features and using it for differentiating the nodules. Further, existing classifiers are highly complex which performs larger calculation throughout the entire operation. To overcome these issues, a hybrid approach is introduced in this paper by combining graph-clustering ant colony optimized feature extraction with extreme learning machine classifier.

### III. BACKGROUND METHODOLOGY

#### A Ant Colony Optimization

(Sivagaminathan & Ramakrishnan, 2007) explained that ant colony optimization is similar to the real-life behaviour of ants while travelling through the same path to reach their nest after collecting the food. It should be noted that it does not make any visual contact with the path itis travelling. This could be achieved through an indirect mode of communication known as "stigmery" by an odorous chemical substance named as pheromone. The quantity of the pheromone substance depends upon different properties of the food source like quantity, quality and distance of availability. All the ants follow a path that contains more pheromone thereby making it a positive feedback loop. The pheromone starts vanishing and evaporating after certain time period, and finally result in reduced pheromone in less demand path. Because of pheromone evaporation, the ants search other available paths and finalize the most optimal path for travelling.

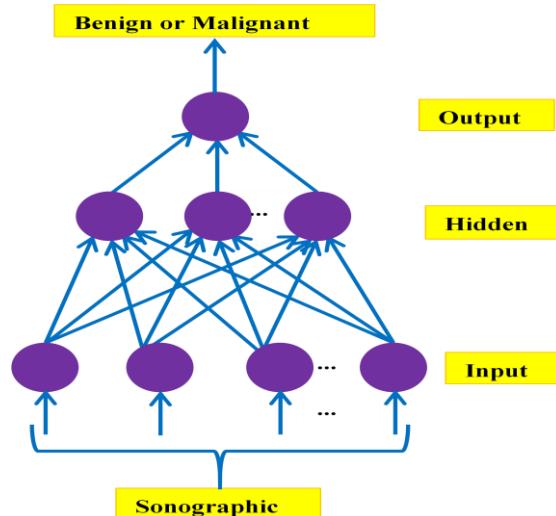


By following the same procedure, the optimal features are selected from the thyroid dataset using ant colony optimization algorithm. Due to its simplicity, it is found to be a better method for machine learning based classification designs. Nevertheless, the use of present ACO algorithms in recent application suffer from several drawbacks like computation complexity due to fully connected graph structure drawn from all the available features, need for a learning model to create feature subsets, highly correlated features are not excluded which inturn affects the classification accuracy and reliability of classifier depends upon the size of the selected feature subset.

To address the above issues, (Moradi&Rostami, 2015) introduced a new feature selection approach by integrating ant colony optimization with graph-clustering technique. Graph Clustering based Ant Colony Optimization is a filter based multivariate feature selection process that represent the features in the form of undirected graph with nodes and edges of the graph as features and similarities between features respectively. Similar to data clustering algorithms like Fuzzy C Means (FCM) and k-means, graph clustering initially identifies the similarity among pair of data points and form an undirected graph. This graph is then divided into clusters according to the linear/non-linear boundaries estimated through an optimal objective function. In GCACO algorithm, feature clustering is performed by means of a community detection algorithm, which produces a subset of features with each feature containing minimal redundancy with other features available in the subset. The key role of this technique over other techniques such as F-Score, L-Score, ReliefF and UFSACO is that it considers both redundancy and relevance analyses while performing the feature selection task. Likewise, GCACO identify the relationship between features before performing selection whereas univariate approaches rank the features without considering the dependency between each feature. Since graph clustering finds advantages by integrating with ant colony optimization, it is extensively used for feature selection in many applications. Graph Clustering based Ant Colony Optimization feature selection approach is adopted in the proposed work for selecting necessary features from the thyroid dataset. This is an effective approach to select discriminant features that are very essential to differentiate the thyroid nodules from the dataset.

### **B. Extreme Learning Machine**

The use of extreme learning machine (ELM) as novel machine learning algorithm for single layer feed forward neural networks (SLFNs) displayed in Fig. 1 was first initiated by Huang et al (2004). It overcomes the drawbacks of conventional SLFNs related to slow learning speed, tuning of trivial components and improper generalization ability. Therefore, ELM possesses different properties such as fast learning capability, highly generalized performance and free parameter tuning. ELM is designed in such a manner to function well with enhanced generalization capability for performing better classification and regression. On comparing the learning processes of both ELM as well as SLFNs gradient based iterative and back propagation methods, ELM learns faster than SLFNs.



**Fig 1. Structure of Extreme Learning Machine**

### **IV. PROPOSED METHOD**

The proposed method is designed to predict the thyroid disease from dataset by classifying the thyroid nodules using US features. Firstly, the discriminate features are partitioned from the dataset using Graph Clustering based Ant Colony Optimization feature selection method. Secondly, each of the selected features is experimented to differentiate the type of nodule using Extreme Learning Machine algorithm.

#### **A. Feature Extraction using Graph Cluster Ant Colony Optimization**

GCACO is a multivariate feature selection strategy that selects the optimal features by performing dependency analysis on features structured as an undirected graph. In order to select the discriminant features, the relevance analysis is performed by means of the Fisher score (F-Score) and the multiple discriminant analysis (MDA).

*i. Graph formation.* Initially, the features are analyzed one-by-one to know about its redundancy. This analysis is achieved by creating a weighted undirected graph with all the available features. The graph representation is defined as:  $G_w = \{f, E, W\}$ , where  $f = \{f_1, f_2, \dots, f_n\}$  depicts the nodes of the graph,  $E = \{E_{ab}, a, b = 1, 2, \dots, n, a \neq b\}$  represents the edges between the graph nodes, and  $W = \{W_{ab}\}$  denotes the weight of edges of the graph. The weight between nodes  $a$  and  $b$  is calculated as:

$$W_{ab} = \left| \frac{(f_a - \bar{f}_a)(f_b - \bar{f}_b)}{\sqrt{(f_a - \bar{f}_a)^2(f_b - \bar{f}_b)^2}} \right|$$

Where,  $f_a$  and  $f_b$  are the features and  $\bar{f}_a$  and  $\bar{f}_b$  are the mean value of the feature vectors. The feature vectors  $f_a$  and  $f_b$  is either extremely correlated or uncorrelated when the weight of their edges produce a value one or zero respectively. The weight value in GCACO is normalized by means of softmax scaling technique inorder to get rid of the impact of outliers. The procedure of softmax scaling is defined as follows:

$$W_{ab}^c = \frac{1}{1 + \exp \left( \frac{W_{ab} - \bar{W}}{\sigma} \right)}$$



Where  $\bar{W}$  and  $\sigma$  are the mean and standard deviation of all values of  $W_{ab}$ .

**A. Feature clustering.** Redundancy analysis in GCACO is performed by an effective algorithm called as Louvain community detection. In this, the weighted undirected graph is partitioned into communities or sub-nodes depending upon the similarities between features (highly correlated). During initialization, each node (feature) is treated as an individual community. For each iteration, two neighbors (say  $a$  and  $b$ ) of a node is chosen. A modular gain factor is evaluated by eliminating the node  $a$  from its own community and inserting it to one of the communities of node  $b$ . This procedure is continued until all the neighbors of node  $b$  are visited. The node  $a$  is then added into the community that results in higher positive modular gain factor. If the modular gain of all the neighboring communities is negative, then  $a$  remain in its own community. This procedure is repeated until no change is found in the modular gain value and a new network is drawn based upon the final communities. The modularity gain factor that is obtained after inserting the isolated node  $a$  to any of one of the  $b$  communities is determined as:

$$\Delta G = \left[ \frac{\sum_{\text{input}} W_{ab}^c - Y_{a,ac}}{2k} - \left( \frac{\sum_{\text{total}} W_{ab}^c + Y_a}{2k} \right)^2 \right] - \left[ \frac{\sum_{\text{input}} W_{ab}^c}{2k} - \left( \frac{\sum_{\text{total}} W_{ab}^c}{2k} \right)^2 \right] - \left( \frac{Y_a}{2k} \right)^2$$

Where  $\sum_{\text{input}} W_{ab}^c$  is the total weight of interior edges of community  $C$ ,  $\sum_{\text{total}} W_{ab}^c$  is the total weight of edges that are incident to nodes in community  $C$ ,  $Y_a$  is the total weight of edges that are connected to node  $a$ ,  $Y_{a,ac}$  is the total weight of edges from node  $a$  to every node in community  $C$  and  $k = \sum_{\text{input}} W_{ab}^c$ . To enhance the outcomes of feature clustering, the weights less than  $\theta$ , a preset threshold value used to control the amount of clusters are excluded, where  $\theta$  ranges between 0.3 and 0.8.

**ii.Ant Colony Optimization.** The working of Ant Colony Optimization algorithmis based on the movement of ants through the path travelled by other ants identified by pheromone chemical imprintsleft by them.In GCACO, an ant is randomly allocated to one of the clusters produced by feature clustering approach. On each iteration, the algorithm considers two random values( $N_1$  and  $N_2$ )along with a parameter ( $\lambda_0$ ) and threshold value ( $\mathcal{E}$ ). The value of  $N_1$ ,  $N_2$ ,  $\lambda_0$  and  $\mathcal{E}$  lies between therange 0 and 1. If  $N_1 > \lambda_0$ , the ant selects a feature from the cluster on the basis of roulette wheel concept. If  $N_2 < \mathcal{E}$ , the ant remains in the same cluster and choose another feature. When  $N_2 > \mathcal{E}$ , the ant leaves the current cluster and goes to another cluster. The parameter  $\lambda_0$  is used to switch among exploitation and exploration phase and the threshold value  $\mathcal{E}$  is utilized to control the number of features to be selected within a cluster. The above process is continued until selecting features from all the available clusters. After going through all the clusters, the features selected by first ant are stored and the next ant enters into feature selection process. The same procedure is repeated in a cyclic order for the required amount of iterations. The pheromone values of the features are maintained by Multiple

Discriminant Analysis (MDA) and the final subset of features is sorted according to their pheromone value. The implementation process of GCACO algorithm is explained below:

**Step1:**The algorithm is initialized by setting up the following parameters: total iterations( $i$ ), number of ants( $A$ ), evaporation coefficient of pheromone( $\gamma$ ), initial pheromone quantity( $\eta_i, i = 1,2, \dots, n$ ) and other constants  $\lambda_0, \mathcal{E}, \alpha, \beta$ .

**Step2:** In GCACO, the relevant feature analysis isdone by the Fisher score value.Based on the F score value, the features are sorted according to their significance. The Fisher score for the  $q$ th feature in the feature set  $f = \{f_1, f_2, \dots, f_n\}$ is defined as:

$$Fscore (F_s) = \frac{\sum_{x=1}^X u_x \cdot (\rho_x^s - \rho^s)^2}{\sum_{x=1}^X u_x \cdot (\sigma_x^s)^2}$$

Where  $\rho_x^s$  and  $\sigma_x^s$  are mean and standard deviation of the  $x$ th class with  $u_x$  samples respectively, and  $\rho^s$  is the mean of the samples in the  $s$ th feature vector. The F score value is normalized within range 0 and 1 by the softmax scaling method. The features with largest F-score value are considered to be better discriminant features.

**Step 3:** In GCACO, the redundancy analysis is achieved by calculating the absolute value of Pearson's correlation. To obtain this, the cross-correlation mean( $m$ )values within  $x$ th feature and all other features ( $VF_b$ ) visited by the  $b$ th ant from all previous clusters is first evaluated. Finally, the following function is estimated to know about the redundancy:

$$\Psi(F_s, VF_b) = Fscore (F_s) - \frac{1}{S_{VF_b}} \sum_{t=1}^{S_{VF_b}} W(F_s, F_t)$$

This function is utilized to know about exploitation/exploration of feature during feature selection task. In this function,  $S_{VF_b}$ represents the size of  $VF_b$ .

(i) If  $N_1 > \lambda_0$ , then the  $b$ th ant choose the succeeding feature as follows:

$$f_{next} = \arg \max_{F_s \in UF_t^b} \{(\tau^i(F_s))^{\alpha} \times \Psi(F_s, VF_b)^{\beta}\}$$

Where  $UF_t^b$ represents the features that are still not visited by the  $b$ th ant from the present cluster ( $t$ th cluster),  $\tau^i(F_s)$  is the total pheromone of the  $s$ th feature, and  $\alpha$  and  $\beta$  are the relative importance of the pheromone value and heuristic information, respectively.

(ii) If  $N_1 < \lambda_0$ , a probability function is estimated for the remaining features in the present cluster. It is described as follows:

$$Pr(F_s) = \frac{(\tau^i(F_s))^{\alpha} \times \Psi(F_s, VF_b)^{\beta}}{\sum_{F_s \in UF_t^b} (\tau^i(F_s))^{\alpha} \times \Psi(F_s, VF_b)^{\beta}} \text{ for } F_s \in UF_t^b$$

Then, the next suitable feature is chosen on the basis of roulette wheel rule.

**Step 4:**On every iteration $i$ , the pheromone amount ( $\tau^i(F_s)$ )of the  $s$ th feature is updated based on MDA as follows:



$$\tau^{i+1}(F_s) = (1 - \gamma)\tau^i(F_s) + \sum_{b=1}^A \Delta_b^i(F_s)$$

$$\Delta_b^i(F_s) = \begin{cases} \eta_b^i & F_s \in SF_b^i \\ 0 & F_s \notin SF_b^i \end{cases}$$

Where  $SF_b^i$  represents the feature selected on  $i$ th iteration by the  $b$ th ant,  $A$  denotes the number of ants and  $\eta_b^i$  is defined as the separability index of  $b$ th selected subset in the  $i$ th iteration, i.e.,  $M_B$  and  $M_W$  are the between and within scatter matrices respectively,  $M_T$  is the transform matrix from the  $Q$ -dimensional space to the  $q$ -dimensional space, where  $Q$  is the number of the features selected by the  $b$ th ant in the  $i$ th iteration and  $q$  is an integer value between 1 to  $K - 1$  with  $K$  as the total number of classes.

**Step 5:** After completing the overall iterations( $I$ ), the value of pheromone is utilized to select the needed optimal feature set. In each cluster, the features are sorted according to the amount of pheromone content and the first  $\omega$ set of features from each cluster is chosen for further processing. Therefore, for  $C_N$  clusters,  $\omega \times C_N$  features are selected.

### B. Extreme Learning Machine based classification

A brief description about classification using Extreme Learning Machine is explained in this section. Consider a training set  $\mathcal{N} = \{(\mathbf{x}_i, \mathbf{y}_i) | \mathbf{x}_i \in \mathbb{R}^l, \mathbf{y}_i \in \mathbb{R}^I, i = 1, 2, \dots, n\}$ , where  $\mathbf{x}_i$  and  $\mathbf{y}_i$  denotes the input feature vector of size  $n \times l$  and target vector of size  $n \times I$  respectively. The conventional SLFNs hold an activation function  $\sigma(\mathbf{x})$ , and the amount of hidden neurons  $\tilde{N}$  can be mathematically framed as follows:

$$\sum_{i=1}^{\tilde{N}} \mathbf{w}_i \sigma(\mathbf{x}_i \cdot \mathbf{w}_i + b_i) = \mathbf{y}_i, i = 1, 2, \dots, n$$

Where  $\mathbf{w}_i$  and  $\mathbf{b}_i$  represent the weight vectors among the input layer and output layer of  $i$ th neuron in the hidden layer respectively.  $b_i$  is the bias of the  $i$ th neuron in the hidden layer and  $\mathbf{y}_i$  is the target vector of the  $i$ th input data. The inner product of  $\mathbf{w}_i$  and  $\mathbf{x}_i$  be  $x_i \cdot w_i$ . If it is possible for the SLFNs to approximate the  $n$  samples with zero errors, there will be  $\sum_{i=1}^n \|\mathbf{y}_i - \mathbf{y}_i\| = 0$  i.e., and there exists  $\mathbf{w}_i, \mathbf{b}_i, i$  such that  $\sum_{i=1}^{\tilde{N}} \mathbf{w}_i \sigma(\mathbf{x}_i \cdot \mathbf{w}_i + b_i) = \mathbf{y}_i, i = 1, 2, \dots, n$ . The above Equation can be reformulated as follows:

$$\mathbf{W} \mathbf{H} = \mathbf{Y}$$

Where,

$$= \begin{pmatrix} \sigma(\mathbf{x}_1 \cdot \mathbf{w}_1 + b_1) & \cdots & \sigma(\mathbf{x}_{\tilde{N}} \cdot \mathbf{w}_1 + b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ \sigma(\mathbf{x}_1 \cdot \mathbf{w}_{\tilde{N}} + b_1) & \cdots & \sigma(\mathbf{x}_{\tilde{N}} \cdot \mathbf{w}_{\tilde{N}} + b_{\tilde{N}}) \end{pmatrix}_{\tilde{N} \times \tilde{N}}$$

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_{\tilde{N}} \end{bmatrix}_{\tilde{N} \times l} \quad \text{and} \quad \mathbf{H} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_{\tilde{N}} \end{bmatrix}_{\tilde{N} \times l}$$

As explained by Huang et al. [50], the output matrix of the hidden layer of neural network with  $\square$ th column of  $\square$  being the  $\square$ th output of hidden neuron with respect to the input variables,  $\square_1, \square_2, \dots, \square_n$ . Further, they showed that the hidden layer bias and input weights of SLFNS are not expected to be modified and are provided in a random manner. With this assumption, the output weights are analytically estimated through least square solution  $\square$  of the linear system,  $\square \square = \square$ :

$$\begin{aligned} \|\square(\square_1, \dots, \square_{\tilde{N}}, \square_1, \dots, \square_{\tilde{N}})\|^2 \\ - \|\square\|^2 = \min_{\square} \|\square(\square_1, \dots, \square_{\tilde{N}}, \square_1, \dots, \square_{\tilde{N}})\|^2 \\ - \|\square\|^2 \end{aligned}$$

The above Equation can be easily determined by a generalized linear approach like Moor-Penrose (MP) by finding the inverse of  $\square$ , as is shown in the Equation given below.

$$\square \square = \square \Rightarrow \square = \square^T \square$$

Where  $\square^T$  is the generalized inverse matrix  $\square$  obtained from MP approach. Utilizing this generalized inverse may result in minimized solution for the resulting least square norms. It yields the unique as well as smallest least square norms compared to existing least square solutions. After performing effective analysis, Huang et al. [49] explained that the generalized inverse of MP obtains better ELM performance with dramatically improved learning speed. The learning process of ELM is proceeded as follows: Initially, consider a training set,  $\mathcal{N} = \{(\mathbf{x}_i, \mathbf{y}_i) | \mathbf{x}_i \in \mathbb{R}^l, \mathbf{y}_i \in \mathbb{R}^I, i = 1, 2, \dots, n\}$ , an activation element  $\sigma(\mathbf{x})$ , and total hidden neurons  $\tilde{N}$ . (a) Randomly allocate the input weights  $\mathbf{w}_i$  and bias  $b_i$ ,  $i = 1, 2, 3, \dots, \tilde{N}$ . (b) Evaluate the output of Hidden layer matrix  $\mathbf{H}$ . (c) Estimate the resultant weight  $\mathbf{W} = \mathbf{W}^T \mathbf{H}$ ,  $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{\tilde{N}}]^T$ .

## V. RESULTS AND DISCUSSION

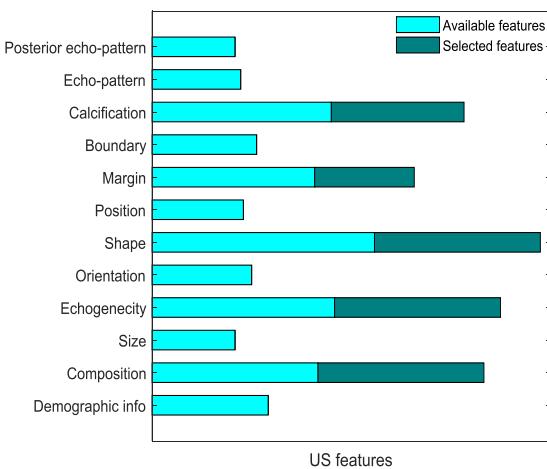
The implementation and performance analysis of the proposed work is performed on MATLAB R2018a software running on windows operating system with 1.7 GHz CPU and 4.00 GB of RAM. This analysis is performed to know about the functioning of feature selection as well as classification approaches. The extreme learning machine is built on the basis of 10-fold cross validation process on the thyroid disease dataset.

### Experimental Design

#### A. Feature selection

The retrospective analysis of thyroid disease is performed on pathologically verified thyroid nodules with the help of different characteristics of US images. The thyroid dataset taken for evaluation consists of 1427 nodules with 1180 benign nodules and 247 malignant nodules. The benign nodules are considerably lengthier than the malignant nodules. For evaluating the performance of feature selection algorithm, different characteristics of US images were considered. It includes different features like demographic information, boundary, echo pattern, posterior acoustic pattern, margin, orientation, position, thyroid shape, tumor size and calcification. These features are extracted from the US images.

After applying GCACO algorithm, the important features are selected from the overall available features as shown in Figure 2 and Table 1. The definitions of selected features are provided below.



**Fig 2. Features selected from the thyroid dataset by using GCACO approach**

Table 1. Feature subset selected by GCACO algorithm	
Size of subset	Features selected on each iteration
1	{Calcification}
2	{Calcification, Composition}
3	{Calcification, Composition, Echogenicity}
4	{Calcification, Composition, Echogenicity, Shape}
5	{Calcification, Composition, Echogenicity, Shape, Margin}

- Calcification.** Calcification is categorized into three types like microcalcification, macrocalcification or no calcification. Microcalcification and macrocalcification are calcification with diameter less than 1 mm and larger than 1 mm respectively. If a nodule consists of both types of these calcifications, then it is remarked as microcalcification.
- Composition.** The proportion of fluid or soft tissue in a nodule is termed as composition. It may be solid or liquid or cystic. Solid is comprised of soft tissues with liquid lesser than 10%. Predominantly solid substances are consisted of >10% liquid on <50% volume of the nodule. Incase of cystic composition, the nodule is fully or almost fully filled with liquid. One special appearance of composition is spongiform appearance that resembles like minute cystic spaces detached by thin pieces of septa.
- Echogenicity.** In solid portions, the echogenicity is classified as iso/hyper-echogenicity, hypo-echogenicity and marked hypo-echogenicity. If the echogenicity in the nodule looks similar to the thyroid parenchyma present in their surroundings, it is termed as iso-echogenicity. If the echogenicity is low as compared to that found in strap muscles, it is called as marked hypo-echogenicity.
- Shape.** The shape of the thyroid gland may be oval or round and it is either taller than wide or taller than long. The shape of the nodules is identified from the diameter of an anteroposterior nodule. If the anteroposterior diameter is smaller than the diameter of longitudinal and transverse planes, then the shape is said to be oval shape. Otherwise, if the anteroposterior diameter is equal to the diameter of longitudinal and transverse planes, then it is called as round shape. If the ratios of anteroposterior to

transverse and longitudinal diameters are greater than one, then the structure of nodule is taller than wide and taller than long respectively.

- Margin.** The outline of thyroid nodule is called as margin of the nodule. The margin of the nodule takes different structures like smooth margin, ill-defined margin, irregular margin and microlobulated.

**Table 2. Ultrasound features of thyroid nodules**

Features	Number of benign nodules (n=1180)	Number of malignant nodules (n=247)	p-value
<b>Calcification</b>			<0.001
Macrocalcification (n=303)	257	46	
Microcalcification (n=127)	50	77	
No calcification (n=997)	873	124	
<b>Composition</b>			<0.001
Solid (n=983)	763	220	
Mixed (n=444)	417	27	
<b>Echogenicity</b>			<0.001
Hyper-echogenicity (n=854)	650	204	
Hypo-echogenicity (n=412)	370	42	
Marked hypo-echogenicity (n=161)	160	1	
<b>Margins</b>			<0.001
Smooth (n=1010)	985	25	
Microlobulated (n=243)	178	65	
Irregular (n=174)	17	157	
<b>Shape</b>			<0.001
Wider than tall (n=1253)	1117	136	
Taller than wide (n=174)	63	111	

The information about the extracted features is given in Table 2. These features are significant to identify the risk of malignancy associated with thyroid nodules.

# Extreme Learning Machine for Thyroid Nodule Classification with Graph Cluster Ant Colony Optimization Based Feature Selection

From Chi-square analysis, the nodules with the following criteria are considered as malignant nodules with p-value<0.001. It includes nodules larger than its width, solid type of composition, irregular margins, microcalcification and hypo-echogenicity. Therefore, the following characteristics well describes the benign nodules: nodules that are wider than larger shape, non-solid type of composition, smooth margins, macro-calcifications and iso-echogenicity.

## **B. Confusion matrix for classifier analysis**

The performance of the classifier is evaluated from the confusion matrix shown in Table 3. The quality of ELM classifier is determined through different criteria like correctly predicted nodules (true positives), correctly predicted but that does not belong to the malignant category (true negatives), the nodules that belongs to malignancy but predicted as benign (false negatives) and the nodules that are benign but predicted as malignant (false positives). Thus the performance of prediction through classification is estimated by analyzing four different parameters like accuracy, area under curve, sensitivity and specificity. Off these, the parameters such as accuracy, sensitivity and specificity can be directly determined from the confusion matrix of the thyroid dataset.

**Table 3. General structure of the confusion matrix**

Type of classifier		Predicted	
		Malignant nodules	Benign nodules
Actual	Malignant nodules	True positive	False Negative
	Benign nodules	False Positive	True Negative

$$\frac{\square\square + \square\square}{\square\square + \square\square + \square\square + \square\square}$$

$$\frac{\square\square}{\square\square + \square\square}$$

$$\frac{\square\square}{\square\square + \square\square}$$

The equations for calculating accuracy, sensitivity and specificity from the confusion matrix are described above. The region under the receiver operating characteristic curve is termed as the area under curve which draws the true positives rather than the false positive rates. Finally, the classifier with higher AUC is remarked as the best classifier than the classifier that produces smaller AUCs. A classifier with AUC equal to one is concluded as a perfect classifier.

## **C. Performance Analysis**

The performance of the proposed thyroid disease nodule classification method is evaluated by analyzing the functions of both GCACO feature selection and extreme machine learning classification algorithms. The effectiveness of classification based on the selected features is compared with existing methods such as L-Score, F-Score, ReliefF and UFSACO. Table 4-6 shows the comparison table for classification accuracy, sensitivity and specificity obtained with different feature selection algorithms. Different feature criteria that decides the efficiency of classification includes composition, calcification, margin, shape, solid part of echogenicity and size of nodules that are equal to or larger than 5 mm.

**Table4. Comparison of classification accuracy with different feature selection algorithms**

No. of features in feature subset	L-score	F-score	ReliefF	UFS ACO	Proposed
1	82.37±5.37	81.85 ±1.56	94.37± 5.38	88.27 ±3.34	<b>98.00 ±1.67</b>
2	81.24±4.69	82.78 ±2.56	<b>96.27± 1.38</b>	86.45 ±2.95	93.19 ±0.40
3	82.25±2.68	84.89 ±1.56	<b>95.32± 2.46</b>	84.35 ±4.27	93.28 ±0.37
4	82.55±1.52	86.58 ±2.56	93.23± 3.24	84.56 ±3.48	<b>95.21 ±0.59</b>
5	89.78±0.78	76.45 ±5.64	91.42± 3.28	89.58 ±4.58	<b>97.00 ±1.29</b>

The comparison of classification accuracy of GCACO with existing feature extraction methods such as L-score, F-score, ReliefF and UFSACO are shown in Table 4. From this analysis, it is clear that GCACO produce better accuracy in classification than the existing methods. The maximum classification accuracy obtained for GCACO is 98%, 95% and 97% with feature subsets 1, 4 and 5 respectively. The featuresubsets extracted from relief methods also produced better results with 96% and 95% accuracy for number of features 2 and 3 respectively. On the other hand, L-score, F-score and UFSACO produced worst results on classification accuracy. The classification sensitivity for GCACO outperformed the existing methods with 98%, 96% and 94% sensitivity with subsets 1, 3 and 4 respectively. It is listed in Table 5. The ReliefF method utilized 2 and 5 number of featuresfor performing the classification. The sensitivity of these featuresubsets are found to be 97% and 95% respectively. As shown in Table 6, the classification specificity of GCACO for the subsets 1, 2, 3 and 4 is found to be 99%, 95%, 98% and 95%

**Table5. Comparison of classification sensitivity with different feature selection algorithms**

No. of features in feature subset	L-score	F-score	ReliefF	UFSACO	Proposed
1	86.11±6.56	89.3 5±1. 52	94.67±2.7 7	95.35±3.4 5	<b>98.5 6±1. 24</b>
2	89.46±4.38	95.6 7±4. 45	<b>97.46±1.5 6</b>	86.56±2.3 5	93.6 7±4. 16
3	84.46±2.45	90.7 8±7. 54	90.67±3.8 6	88.67±4.2 6	<b>96.6 7±4. 24</b>
4	94.45±6.35	94.2 4±7. 87	91.35±7.4 5	92.57±4.2 6	<b>94.4 6±1. 23</b>
5	95.47±3.88	95.2 4±1. 48	<b>95.67±2.5 6</b>	94.67±3.5 4	94.5 7±2. 65

respectively. But, L-score produced 91% better classification specificity with 4 features in the feature subset.

Thus, compared to univariate feature selection approaches like L-score, F-score and Relief F, the proposed

multivariate feature selection approach functions significantly better.

**Table6. Comparison of classification specificity with different feature selection algorithms**

No. of features	L-score	F-score	ReliefF	UFSA CO	Proposed
1	97.65±2.32	98.45±2.12	97.55±1.21	92.54±4.61	<b>99.54±5.64</b>
2	94.54±2.56	93.45±3.31	93.24±2.65	91.56±2.32	<b>95.65±2.61</b>
3	97.01±1.35	97.54±0.65	97.54±5.65	94.64±3.61	<b>98.68±1.31</b>
4	<b>91.26±5.32</b>	84.52±2.64	81.65±5.56	81.25±5.45	85.65±6.26
5	80.32±2.26	94.48±2.79	93.54±1.52	90.15±5.06	<b>95.54±3.31</b>

The performance of proposed ELM classifier is evaluated by comparing it with other classification techniques such as SVM and KNN. The average accuracy, sensitivity and specificity obtained for all these techniques for the features of thyroid datasets are presented in Table 5. From there, it is observed that the ELM classifier outperforms the existing approaches used for classifying the thyroid dataset with selected feature subsets. Further, the improved accuracy, sensitivity and specificity resemble the higher grade of relevancy and redundancy between the selected features from the input dataset. Thus the significant performance obtained from the ELM classifier is due to high relevancy as well as low redundancy among the features considered for classification.

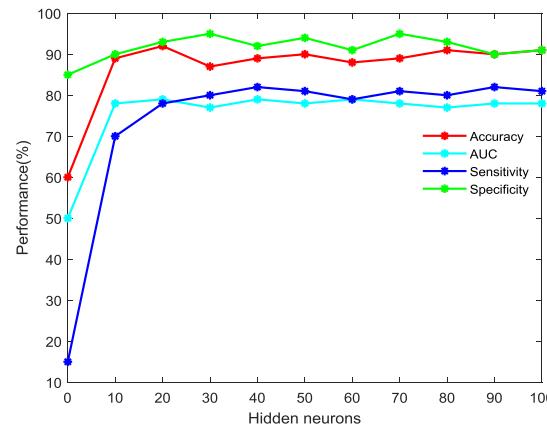
The confusion matrix outcomes of three classifiers such as ELM, ANN and SVM are presented in Table 7. It is identified that ELM has properly differentiated 57 malignant nodules and 87 benign nodules. Further, it misidentified 16 malignant nodules as benign and 10 benign nodules as malignant. On the other hand, ANN perfectly classified 49 malignant nodules and 23 benign nodules. In addition, it misjudges 23 malignant nodules as benign and 17 benign nodules as malignant. Finally, the SVM classifier predicted 51 malignant nodules and 83 benign nodules. However, it misjudges 20 malignant nodules as benign and 14 benign nodules as malignant. This visualizes that ELM outperforms the other two methods in differentiating the type of nodules.

**Table 7. Confusion matrix of ELM**

	Type of classifier	Predicted nodules	
		Malignant	Benign
Actual nodules	ELM classifier		
	Malignant	57	16
	Benign	10	87
	ANN classifier		
	Malignant	49	23
	Benign	17	80
	SVM classifier		

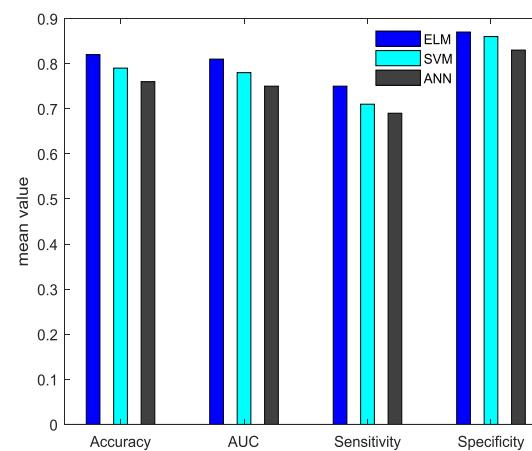
	Malignant	51	20
	Benign	14	83

The performance of ELM based on different hidden neurons is shown in Figure 3. As shown, the accuracy and specificity fluctuates through certain limit and remains stable as the total amount of neurons is increased. The sensitivity and AUC is found to be almost stable throughout the hidden neurons. At 20 neuron case, the accuracy of the extreme learning machine classifier is found to be higher and so, it is taken as the optimal number of the neurons.



**Figure 3. Performance of ELM based on the hidden neurons**

The comparative results for accuracy, AUC, sensitivity and specificity of ELM and existing techniques like SVM and ANN classifiers are shown in Figure 4. It is provided in terms of the mean value of different parameters taken for analysis. The figure depicts that the performance of ELM is better than other machine learning classification techniques such as SVM and ANN. Further, the performance of SVM is slightly higher than that of the artificial neural network classifier. This describes that ELM is the best method to be used for analyzing the thyroid disease using US characteristics.



**Figure 4. Comparative analysis of classifier algorithms in terms of: accuracy, AUC, sensitivity and specificity**

## VI. CONCLUSION

The ultrasound is defined as a non-invasive tool that is utilized for the diagnosis of thyroid lesions because of its affordable price and ease of availability. During thyroid treatment, differentiation of malignancy nodules remain a critical task due to the challenges faced by texture analysis and employing machine learning algorithms in modern diagnostic procedures. Thus, it is necessary to modernize and authorize these processes for the purpose of making it to be desire in the field of thyroid diagnostics. In this paper, Graph-Clustering Ant Colony Optimization based Extreme Machine Learning approach is introduced for the detection of malignancy risk associated with thyroid nodules. Using GCACO, the features are partitioned into group of clusters and the clusters are assembled as an undirected graph with community detection algorithm. Afterwards, ACO is employed to select the optimal features from the group of clusters. In this work, five of the ultra-sonographic features are taken as discriminant features from the US thyroid dataset. The GCACO feature selection method is multivariate and it is compared with existing univariate methodologies like L-score, F-score, ReliefF and UFSACO. It is identified that GCACO significantly identifies the suitable features and functions better than that of the univariate algorithms introduced for the same purpose. Further, the simulation results demonstrate that ELM classifier accurately differentiates the malignant nodules from benign nodules. Thus, GCACO based ELM classifier can be efficiently applied for clinical diagnosis of thyroid disorders and produce effective result in thyroid treatment.

## REFERENCES

1. Acharya, U.R., Chowriappa, P., Fujita, H., Bhat, S., Dua, S., Koh, J.E., Eugene, L.W.J., Kongmebhol, P. and Ng, K.H., 2016. Thyroid lesion classification in 242 patient population using Gabor transform features from high resolution ultrasound images. *Knowledge-Based Systems*, 107, pp.235-245.
2. Ardashani, A.A., Gharbali, A. and Mohammadi, A., 2015. Application of texture analysis method for classification of benign and malignant thyroid nodules in ultrasound images. *Iranian journal of cancer prevention*, 8(2), p.116.
3. Bakshi, N.A., Mansoor, I. and Jones, B.A., 2003. Analysis of inconclusive fine-needle aspiration of thyroid follicular lesions. *Endocrine pathology*, 14(2), pp.167-175.
4. Chang, C.Y., Chen, S.J. and Tsai, M.F., 2010. Application of support-vector-machine-based method for feature selection and classification of thyroid nodules in ultrasound images. *Pattern recognition*, 43(10), pp.3494-3506.
5. Chen, F.L. and Ou, T.Y., 2011. Sales forecasting system based on Gray extreme learning machine with Taguchi method in retail industry. *Expert Systems with Applications*, 38(3), pp.1336-1345.
6. Choi, W.J., Park, J.S., Kim, K.G., Kim, S.Y., Koo, H.R. and Lee, Y.J., 2015. Computerized analysis of calcification of thyroid nodules as visualized by ultrasonography. *European journal of radiology*, 84(10), pp.1949-1953.
7. Cooper, D.S., Doherty, G.M., Haugen, B.R., Kloos, R.T., Lee, S.L., Mandel, S.J., Mazzaferri, E.L., McIver, B., Sherman, S.I. and Tuttle, R.M., 2006. Management guidelines for patients with thyroid nodules and differentiated thyroid cancer: The American Thyroid Association Guidelines Taskforce. *Thyroid*, 16(2), pp.109-142.
8. Erdem, G., Erdem, T., Muammer, H., Mutlu, D.Y., Firat, A.K., Sahin, I. and Alkan, A., 2010. Diffusion-weighted images differentiate benign from malignant thyroid nodules. *Journal of Magnetic Resonance Imaging*, 31(1), pp.94-100.
9. Frates, M.C., Benson, C.B., Doubilet, P.M., Kunreuther, E., Contreras, M., Cibas, E.S., Orcutt, J., Moore Jr, F.D., Larsen, P.R., Marqusee, E. and Alexander, E.K., 2006. Prevalence and distribution of carcinoma in patients with solitary and multiple thyroid nodules on sonography. *The Journal of Clinical Endocrinology & Metabolism*, 91(9), pp.3411-3417.
10. Gomathi, M. and Thangaraj, P., 2010. A computer aided diagnosis system for lung cancer detection using support vector machine. *American Journal of Applied Sciences*, 7(12), p.1532.
11. Han, F., Huang, D.S., Zhu, Z.H. and Rong, T.H., 2006. The forecast of the postoperative survival time of patients suffered from non-small cell lung cancer based on PCA and extreme learning machine. *International journal of neural systems*, 16(01), pp.39-46.
12. Helmy, T. and Rasheed, Z., 2009, May. Multi-category bioinformatics dataset classification using extreme learning machine. In *2009 IEEE Congress on Evolutionary Computation* (pp. 3234-3240). IEEE.
13. Huang, G.B., Zhu, Q.Y. and Siew, C.K., 2004. Extreme learning machine: a new learning scheme of feedforward neural networks. *Neural networks*, 2, pp.985-990.
14. Kabir, M.M., Shahjahan, M. and Murase, K., 2011. A new local search based hybrid genetic algorithm for feature selection. *Neurocomputing*, 74(17), pp.2914-2928.
15. Li, L.N., Ouyang, J.H., Chen, H.L. and Liu, D.Y., 2012. A computer aided diagnosis system for thyroid disease using extreme learning machine. *Journal of medical systems*, 36(5), pp.3327-3337.
16. Li, X., Zhang, S., Zhang, Q., Wei, X., Pan, Y., Zhao, J., Xin, X., Qin, C., Wang, X., Li, J. and Yang, F., 2019. Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. *The Lancet Oncology*, 20(2), pp.193-201.
17. Li, X., Zhang, S., Zhang, Q., Wei, X., Gao, M., Zhang, W. and Chen, K., 2019. Deep convolutional neural network models for the diagnosis of thyroid cancer—Authors' reply. *The Lancet Oncology*, 20(3), p.e131.
18. Liu, N., Lin, Z., Koh, Z., Huang, G.B., Ser, W. and Ong, M.E.H., 2011. Patient outcome prediction with heart rate variability and vital signs. *Journal of Signal Processing Systems*, 64(2), pp.265-278.
19. Koundal, D., Gupta, S. and Singh, S., 2018. Computer aided thyroid nodule detection system using medical ultrasound images. *Biomedical Signal Processing and Control*, 40, pp.117-130.
20. Ma, J., Wu, F., Zhu, J., Xu, D. and Kong, D., 2017. A pre-trained convolutional neural network based method for thyroid nodule diagnosis. *Ultrasonics*, 73, pp.221-230.
21. Moradi, P. and Rostami, M., 2015. Integration of graph clustering with ant colony optimization for feature selection. *Knowledge-Based Systems*, 84, pp.144-161.
22. Ouyang, F.S., Guo, B.L., Ouyang, L.Z., Liu, Z.W., Lin, S.J., Meng, W., Huang, X.Y., Chen, H.X., Qiu-gen, H. and Yang, S.M., 2019. Comparison between linear and nonlinear machine-learning algorithms for the classification of thyroid nodules. *European Journal of Radiology*, 113, pp.251-257.
23. Shankar, K., Lakshmanaprabu, S.K., Gupta, D., Maseleno, A. and de Albuquerque, V.H.C., 2018. Optimal feature-based multi-kernel SVM approach for thyroid disease classification. *The Journal of Supercomputing*, pp.1-16.
24. Sivagaminathan, R.K. and Ramakrishnan, S., 2007. A hybrid approach for feature subset selection using neural networks and ant colony optimization. *Expert systems with applications*, 33(1), pp.49-60.
25. Sollini, M., Cozzi, L., Chiti, A. and Kirienko, M., 2018. Texture analysis and machine learning to characterize suspected thyroid nodules and differentiated thyroid cancer: Where do we stand?. *European journal of radiology*, 99, pp.1-8.
26. Tabakh, S., Moradi, P. and Akhlaghian, F., 2014. An unsupervised feature selection algorithm based on ant colony optimization. *Engineering Applications of Artificial Intelligence*, 32, pp.112-123.
27. Tsantis, S., Dimitropoulos, N., Cavouras, D. and Nikiforidis, G., 2009. Morphological and wavelet features towards sonographic thyroid nodules evaluation. *Computerized Medical Imaging and Graphics*, 33(2), pp.91-99.
28. Wu, Y., Yue, X., Shen, W., Du, Y., Yuan, Y., Tao, X. and Tang, C.Y., 2013. Diagnostic value of diffusion-weighted MR imaging in thyroid disease: application in differentiating benign from malignant disease. *BMC medical imaging*, 13(1), p.23.
29. Yong, Z., Dun-wei, G. and Wan-qiu, Z., 2016. Feature selection of unreliable data using an improved multi-objective PSO algorithm. *Neurocomputing*, 171, pp.1281-1290.
30. Zhang, R., Huang, G.B., Sundararajan, N. and Saratchandran, P., 2007. Multicategory classification using an extreme learning machine for microarray gene expression cancer diagnosis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 4(3), pp.485-495.

## AUTHORS PROFILE



**Rasheeduddin Sayyad** Department of Computer Science & Engineering , Nagarjuna University, GUNTUR Received the M.Tech degree from Nagarjuna University . He is currently working in Malla Reddy College of Engineering as a Asst. Professor in Computer Science & Engineering . His research interest include Data Mining & Machine Learning.



**Dr.Kurra Rajasekhar Rao** is a Professor of Computer Science & Engineering, having more than **33+ years of Teaching and Research experience**. Dr.KKR garu is actively engaged in the active research related to Embedded Systems, Software Engineering and Knowledge Management. He published more than **200** papers in various International/National Journals and Conferences.