MDPI

Article

Ultrasound Image Classification of Thyroid Nodules Using Machine Learning Techniques

Vijay Vyas Vadhiraj [1,2,*], **Andrew Simpkin** [3], **James O'Connell** [1,2] , **Naykky Singh Ospina** [4], **Spyridoula Maraka** [5,6] **and Derek T. O'Keeffe** [1,2,7]

1   School of Medicine, College of Medicine Nursing and Health Sciences, National University of Ireland Galway, H91 TK33 Galway, Ireland; jamesoconnell118@gmail.com (J.O.); derek.okeeffe@nuigalway.ie (D.T.O.)

2   Health Innovation Via Engineering Laboratory, Cúram SFI Research Centre for Medical Devices, Lambe Institute for Translational Research, National University of Ireland Galway, H91 TK33 Galway, Ireland

3   School of Mathematics, Statistics and Applied Maths, National University of Ireland, H91 TK33 Galway, Ireland; andrew.simpkin@nuigalway.ie

4   Division of Endocrinology, Department of Medicine, University of Florida, Gainesville, FL 3210, USA; naykky.singhospina@medicine.ufl.edu

5   Division of Endocrinology and Metabolism, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA; smaraka@uams.edu

6   Medicine Section, Central Arkansas Veterans Healthcare System, Little Rock, AR 72205, USA

7   Lero, SFI Centre for Software Research, National University of Ireland Galway, H91 TK33 Galway, Ireland

*   Correspondence: vijayvyas1@gmail.com; Tel.: +353-(0)91-524411

Abstract: *Background and Objectives*: Thyroid nodules are lumps of solid or liquid-filled tumors that form inside the thyroid gland, which can be malignant or benign. Our aim was to test whether the described features of the Thyroid Imaging Reporting and Data System (TI-RADS) could improve radiologists' decision making when integrated into a computer system. In this study, we developed a computer-aided diagnosis system integrated into multiple-instance learning (MIL) that would focus on benign–malignant classification. Data were available from the Universidad Nacional de Colombia. *Materials and Methods*: There were 99 cases (33 Benign and 66 malignant). In this study, the median filter and image binarization were used for image pre-processing and segmentation. The grey level co-occurrence matrix (GLCM) was used to extract seven ultrasound image features. These data were divided into 87% training and 13% validation sets. We compared the support vector machine (SVM) and artificial neural network (ANN) classification algorithms based on their accuracy score, sensitivity, and specificity. The outcome measure was whether the thyroid nodule was benign or malignant. We also developed a graphic user interface (GUI) to display the image features that would help radiologists with decision making. *Results*: ANN and SVM achieved an accuracy of 75% and 96% respectively. SVM outperformed all the other models on all performance metrics, achieving higher accuracy, sensitivity, and specificity score. *Conclusions*: Our study suggests promising results from MIL in thyroid cancer detection. Further testing with external data is required before our classification model can be employed in practice.

Keywords: computer aided diagnostics; CAD; artificial intelligence; AI; digital health; TI-RADS; big data; ANN; SVM; malignant; benign; cancer

1. Introduction

Evaluating thyroid nodules is clinically challenging. Thyroid nodules are frequently detected incidentally during the diagnostic imaging of the neck [1]. They may be found in 42–76% of people, becoming more prevalent with increasing age [2]. Most thyroid nodules are benign, but 10% may be malignant [3]. Thyroidectomy, radioactive iodine therapy, immunotherapy, or chemoradiotherapy to prevent recurrence and death may be indicated for thyroid cancer treatment depending on the histological subtype, patient preference, and comorbidities [1,3,4].

A fine needle aspirate biopsy (FNAB) of the nodule may be obtained for a cytological examination of the nodule. As well as generating anxiety for patients [5], FNABs may cause localized pain and although generally safe, there is a low risk of hematoma [6]. In 20–30% of FNABs, the outcome of cytological examination is indeterminate, which means they do not always yield clinically useful results [7,8]. Additionally, some older patients with comorbidities may not suffer adverse outcomes within their lifetime from low-risk thyroid cancers detected after the FNAB of a thyroid nodule [9]. In South Korea, thyroid cancer screening led to a sharp increase in the incidence of papillary thyroid cancer without any impact on thyroid cancer mortality [10], raising concerns for the overdiagnosis of thyroid cancer. As an alternative to invasive FNAB, a conservative strategy such as the active ultrasound surveillance of thyroid nodules may be opted for in carefully selected patients, but this strategy risks missing clinically relevant cancers [11].

Evidently, when deciding the best approach to managing a thyroid nodule, clinicians and patients must balance the risks of an FNAB and overdiagnosis against those of misdiagnosis. To risk stratify nodules and aid in decision making with regard to the need for an FNAB, the American College of Radiology as shown in Appendix C, Table A1 and the Korean Society of Thyroid Radiology each have developed their own Thyroid Imaging Reporting and Data Systems (TI-RADS).

Based on nodule ultrasonographic characteristics such as internal composition, echogenicity, calcification, margins, and size, radiologists can report on the risk of thyroid nodules being malignant using a standardized scoring system [5]. The disadvantages of TI-RADS are that by grouping nodules with different risk factors into a small number of categories, they are managed as equivalent despite their risk of malignancy being potentially very different [5]. Inter-observer and intra-observer variability may also be challenging when implementing TI-RADS [12]. This could potentially cause variability in treatment decisions and outcomes in different patients despite having similar tumors [13].

Systems, which can reliably identify patients with benign and malignant thyroid nodules, while avoiding an invasive FNAB and its associated negative consequences for patients as well as avoiding resource utilization by healthcare providers, are needed. Computer aided diagnosis (CAD) may be of practical application here. Described TI-RADS features may be integrated into computer systems, which can be used to aid in the radiological classification of thyroid nodules. With a structured approach for image acquisition, feature extraction, classification, training, and prediction using machine learning (supervised and unsupervised), it is possible to build and train a model that would predict whether thyroid nodules are malignant or benign [13]. Furthermore, with the use of the same model, a graphic user interface (GUI) can display nodule image features to aid in the decision-making process of radiologists. The accuracy score, sensitivity, and specificity of the models would be important determinants of the clinical utility of these systems. The aim of this research was to build a CAD model that would predict whether thyroid nodules are benign or malignant.

Previous Related Work

For this study, a comprehensive literature review, which included 35 research papers on the ultrasound of thyroid nodules using the CAD system, was completed. Numerous CAD systems have been studied for automated thyroid detection in recent years. Most CAD systems are based on the Convolution Neural Network (CNN) and the Radial basis function (RBF) neural network, which use K-NN for high accuracy scores. Moreover, the multitask cascade convolution neural network (MC-CNN) requires a more extensive data set for a training class [14–20].

Given that there was a lack of research on CAD systems based on the Artificial neural network (ANN) (Figure 1), this was chosen as the primary model for this research. Other neural networks featured in the literature were Probabilistic neural networks (PN) and Multilayer Perceptron Neural Network (MLPNN) [1,2,21–27].
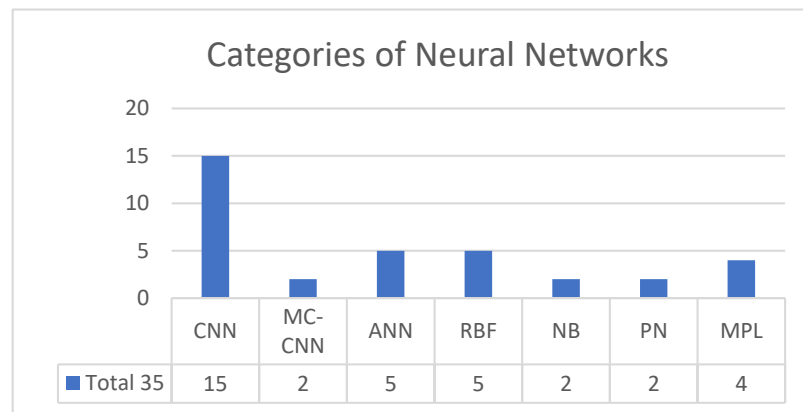
Figure 1. Categories of neural network for automated thyroid nodule detection.

The selection of classifiers mainly depends on how well the model is trained and tested. Figure 2 describes the frequency of the different classifiers used in the literature. SVM is widely used because of its flexible kernel function and threshold. A total of 28% of the research conducted on thyroid nodule detection is based on the SVM algorithm. However, only 2% of the research conducted was based on ANN and SVM. Therefore, ANN and SVM were chosen as the classifiers for this research.
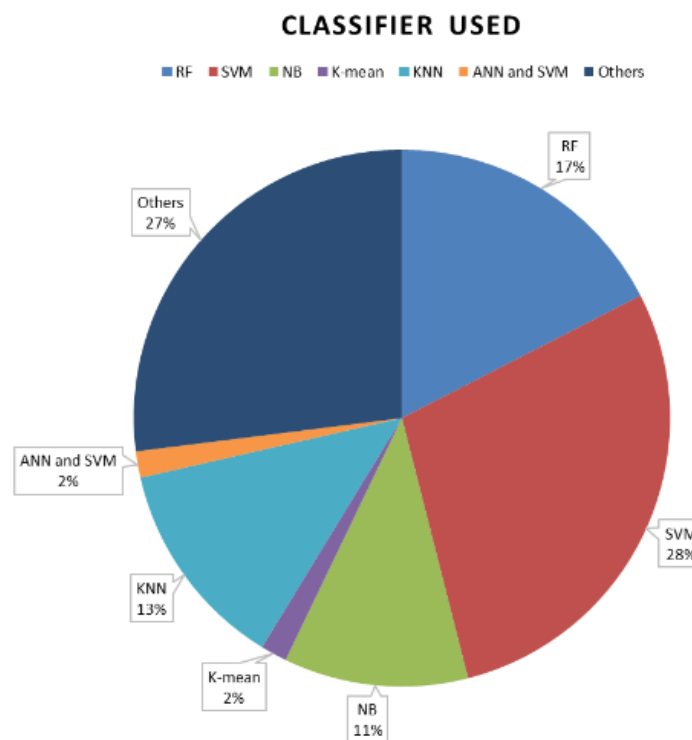

Figure 2. Classifier used.

Regarding image processing and segmentation, median filters are more practical than mean filters or binarization methods when converting greyscale images into black and white. Therefore, these methods were adopted. In computer vision, for any complicated computational task, feature extraction is deployed to extract specific points, compile them, and find solutions. In this research, GLCM was used to extract seven features to determine malignant and benign nodules.

The type of machine learning and the classifiers used were selected for this research based on the literature. A statistical analysis of the categories of neural networks and

their subclassifications that indicated the machine learning technique and classifiers to be used in this study were also based on the literature [5,28–35], as shown in Appendix A, Figure A1.

2. Materials and Methods

Based on literature and expert opinion, a systematic, structured approach was adopted to construct an accurate score-based model for thyroid nodule detection.

2.1. MATLAB Toolbox

There are numerous toolboxes available such as signal processing, fuzzy logic, neural network, control system, and image processing. For this study, machine learning, image processing, the GUI layout toolbox, and the deep-learning toolbox were used. Many toolboxes were readily available in MATLAB 2018a, and the built-in toolbox was readily available as trial software for the student version.

2.1.1. Image Processing Toolbox

This toolbox provided vital tools for image segmentation, image enhancement, and noise reduction. This pre-image processing and segmentation was a significant part of model building in this study.

2.1.2. Statistics and Machine-Learning Toolbox

The machine-learning toolbox was very helpful in building classification and regression models that drew interpretations of data and used them for training and in predictive models. They provided the applications with the capacity to describe, evaluate, analyze, and build a model. This toolbox was vital because both supervised and unsupervised machine-learning algorithms were provided.

2.1.3. GUI Layout Toolbox

The GUI layout toolbox provided options for displaying the interface from the user's perspective. MATLAB 2018a was used to build GUI in our study. They provided the applications with the capacity to describe and analyze results from the user's perspective.

2.2. Data Set

A digital database of thyroid ultrasound images was retrieved from an open-source scientific community on 9 February 2019. The primary aim of the open-source data set was for people to use these images in building CAD models for thyroid nodule analyses [36]. It was uploaded and shared online by the Universidad Nacional de Colombia.

An XML file was presented with each image containing the expert's diagnosis and the relevant patient information. The XML file contained additional information for a better understanding of the characteristics needed for thyroid nodule feature extraction. Patient's age, sex, nodule composition, shape, margin, calcifications, and TI-RADS score were recorded in .mat format and used for validating results.
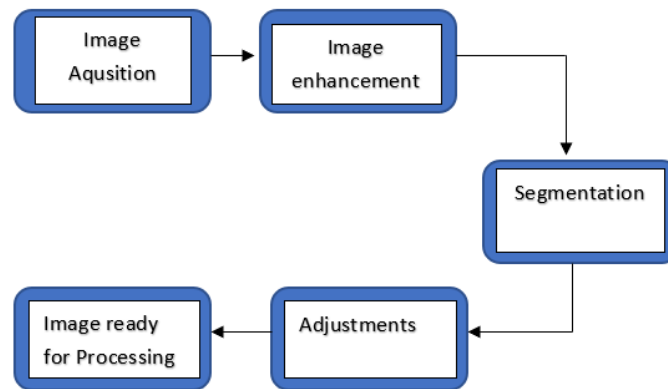
The data set images were divided into two categories—benign and malignant. They were also classified under TI-RADS 2, 3, 4a, 4b, 4c, 5. TI-RADS 2 and 3 were considered benign, and TI-RADS, 4a, 4b, 4c, and 5 were considered malignant. In this research, the American College of Radiology TI-RADS classification was used.

A total of 99 patients and 134 ultrasound examinations, were present in the data set. When multiple images were available for some cases, all images were included in order to optimize nodule feature extraction. Among the 99 patients, there were 33 with nodules classified as benign and 66 with nodules classified as malignant.

2.3. Image Analysis

The images acquired from the data set had to undergo pre-processing and enhancement (Figure 3). All images had to be resized, changed to the same distance scale, and

filtered. Segmentation, which involved the binarization of the image, followed image pre-processing. A threshold was then applied to the image, and further adjustments were made, if required. After this stage, the image was ready for processing.

Figure 3. Image processing flow chart.

2.4. Image Pre-Processing

A median filter was used in this study. A median filter is a non-linear filter, which, when incorporated, removes impulse noise (salt-and-pepper noise). The median level is determined by the kernel size. Note in Figures 4 and 5 that the filtered image is smoothed, and the high-frequency information of the image is reduced. Boundary pixels seem to be distorted because of the zero-padding effect. Even after omission of the distorted signal, the overall quality of the image in the middle portion remained unchanged [37].

Figure 4. Original ultrasound scanning image.

Figure 5. Filtered image.

2.5. Segmentation

Image segmentation is a process of segregating an image into numerous segments. Segmentation is required to change the representation of an image for easy analysis without altering meaningful information. In this study, our objective for using segmentation was to locate boundaries in thyroid nodule images. Figures 6 and 7 exhibit the segmentation of a filtered image.
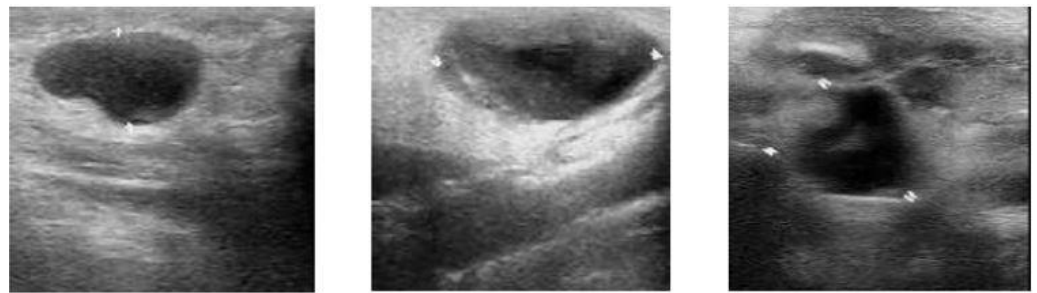
Figure 6. Original Ultrasound scanning Image.

Figure 7. Segmentation process output.

Segmentation can be performed with the use of many different methods. In this study, we used the binarization method to convert greyscale images into black and white. The result of OCR (Optical Character Recognition) is black and white, also considered as 0 or 1. Because of the presence of noise in the original image, a high-quality binarized image was used to enhance the OCR result. Furthermore, we used a fixed thresholding method, which used value to assign 0 s and 1 s, and a global binarization method, which used the single value for images [37].

Feature extraction was used to extract data from the visual content of an image, for retrieval and indexing. First-order and second order are the two types of texture feature measures. In this study, a second-order feature measure was used because this measure considers the relationship between neighboring pixels. With the grey level co-occurrence matrix (GLCM), texture feature can be extracted from a given input.

When GLCM is used, the number of rows and columns are equal to the number of quantized grey levels N. Table 1 shows all seven features [38].

Table 1. Feature Extraction Formulas.

Features	Equations
Energy	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p(i,j)^2$
Correlation	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i \sigma_j}$
Entropy	$-\sum_{i=0}^{N-1} \sum_{j=0}^{n-1} p(i,j) log(p(i,j))$
Homogeneity	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{p(i,j)}{1+|i-j|}$
Cluster Shade	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \left(i+j-\mu_x-\mu_y\right)^3 p(i,j)$
Contrast	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |i-j|^2 p(i,j)$
Inverse Difference Movement	$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \left(1/(1+(i-j))\right)p(i,j)$

2.6. Classification

Image classification consists of two phases, namely Training and Testing. After the feature extraction process, the data of the seven characteristics are calculated and separated into classification categories. Based on these, training class is created. Likewise, in the training phase, the separated features are used to classify image features into benign or malignant. A training class is an integral part of the classification process.

In this study, as with all the models that fall under supervised classification, the objective criteria for building a training class were:

1. Independent: Variation of training class data should not impact other values.
2. Discriminatory: Different image features should indicate different characteristics of the thyroid nodule.
3. Reliable: All image features in the training group should share common definitive characteristics with the training class group [39].

Based on the literature survey, dataset, GLCM feature extraction, and training class, four models under supervised learning were built for testing. Two models under classification and two models under regression. After carefully considering factors such as the nature of the data set, results obtained from feature extraction, and training class, the following models were selected:

- Artificial Neural Network (ANN).
- Support Vector Machine (SVM) [39].

2.7. Performance Metric

In machine learning, evaluating the model by performance measurement is very important. The interpretation of AUC, ROC, accuracy, specificity, and sensitivity is necessary for analysis and in determining if the performance of the model is sufficient or needs further optimization. Each metric or collection of metrics is usually used to check the multi-class classification problem. In this research, the objective was to focus on the accuracy of the model in building computer-aided diagnostics that could classify thyroid nodules as benign or malignant. However, it was also essential to check the performance and efficiency of the model.

Accuracy refers to how well the model can classify nodules correctly. In other words, it evaluates the efficiency and correctness.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

where True Positive (*TP*). False Positive (*FP*). True Negative (*TN*). False Negative (*FN*).

Sensitivity is also called the hit rate or recall. It measures the proportion of the total positive samples correctly identified by the model.

It is usually calculated for the probability of 1, whereas specificity, also called an inverse recall, is the proportion of all negative samples correctly classified as negative. Sensitivity and specificity act as two kinds of accuracy. Sensitivity is for actual positive samples and specificity is for actual negative samples. Therefore, both can be used for evaluating model performance.

$$TPR = \frac{TP}{TP + FN} = \frac{TP}{P}, \ TNR = \frac{TN}{FP + TN} = \frac{TN}{N}$$

where Sensitivity = True Positive Rate (*TPR*), Specificity = True Negative Rate (*TNR*).

The *F1-score* is the harmonic mean of *Precision* and *Recall* and gives a better measure of the incorrectly classified cases than the Accuracy Metric.

$$F1 \ Score = \left(\frac{Recall^{-1} + Precision^{-1}}{2} \right)^{-1} = 2 * \frac{(Precision * Recall)}{(Precison + Recall)}$$

2.8. Artificial Neural Network (ANN)

The neural network model was perceived as a mathematical model, which defines a function, as shown by the following formula:

$$f : X \rightarrow Y$$

According to the literature, a neural network is commonly known as ANN. It is a definition of a class in functions where each node of a class is obtained by changing parameters, connection weights or architecture (number of neurons), or connectivity [40].

In a neural network, the following formula is used:

$$f(x) \rightarrow gi(x)$$

where *f(x)* is the compositions of other functions *gi(x)*; *gi(x)* can be disintegrated into other functions.

Nonlinear weighted sum uses the following formula:

$$f(x) = K\left(\sum i\right)wigi(x))$$

where *K* can be a hyperbolic tangent, a sigmoid function, a softmax function, or a rectifier function.

This is also known as the activation function, which provides a smooth transition for any input variance. Collection of a function is represented as follows:

$$gi \rightarrow Vector\ g$$

$$g = (g1, g2, g3 \ldots gn)$$

Figure 8 shows the disintegration of h1; the arrows specify variables between dependencies. This can be viewed as follows:
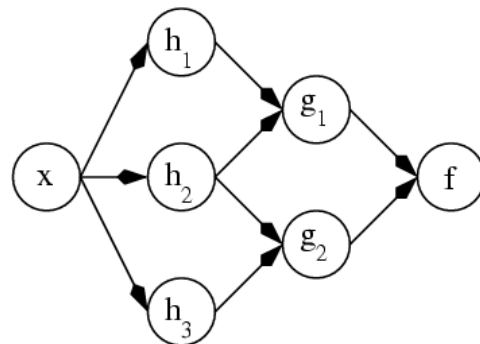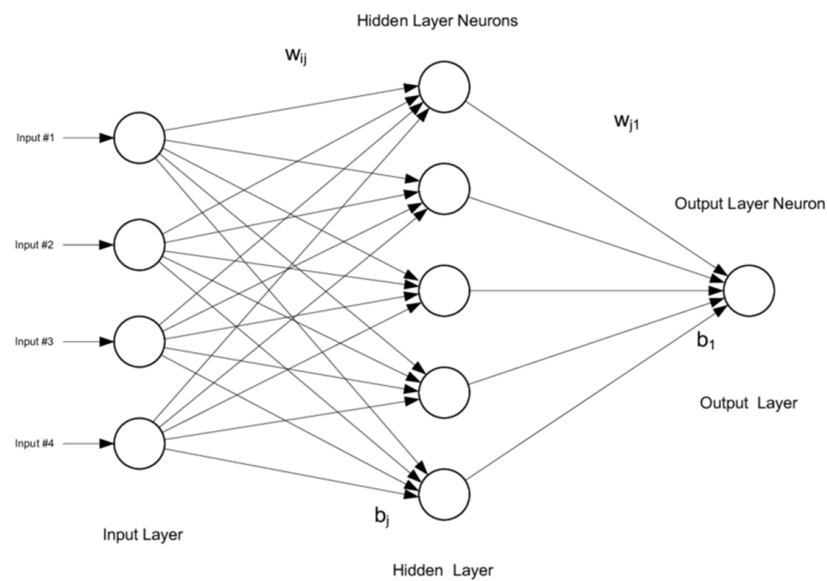

Figure 8. Feedforward artificial neural network [40].

View 1: 3D vector h transformed from input x. 3D vector h further transformed into 2D vector g, and finally transformed into f.

View 2: Random variable F = f(G) depends on G = g(H), and further depends on H = h(X).

The architecture of View 1 and View 2 and its components of layers are independent of one another. Because of the directed acyclic graph, such a network is called feedforward artificial neural network [41]. Figure 9 shows multilayer ANN. Appendix B describes Support Vector Machine-SVM.

Figure 9. Multilayer artificial neural network [42].

3. Results

3.1. Image Analysis and Image Processing

Figure 10a–c shows the image pre-processing and segmentation of the original image of a benign nodule. The original image is resized into 256 × 256 mm. Figure 10b shows a filtered image. The variation between the original image (Figure 10a) and the median filtered image (Figure 10b) is evident. The filtered image is smoothed; the high-frequency information of the image is reduced. Boundary pixels appear to be distorted because of the zero-padding effect. Figure 10c shows a binary segmented image that locates the boundaries of thyroid nodules. In this segmented image, there is a gap in the center pointing out the nodule for feature extraction. This was classified as a benign image.

Figure 11a–c shows the image pre-processing and segmentation of the original image of a malignant nodule. The process of image pre-processing and segmentation is the same as that used for a benign nodule. In the images with multiple nodules (such as Figure 11a,b, which shows two nodules next to one another in the center), the area is pointed out for feature extraction. This image was classified as a malignant nodule.
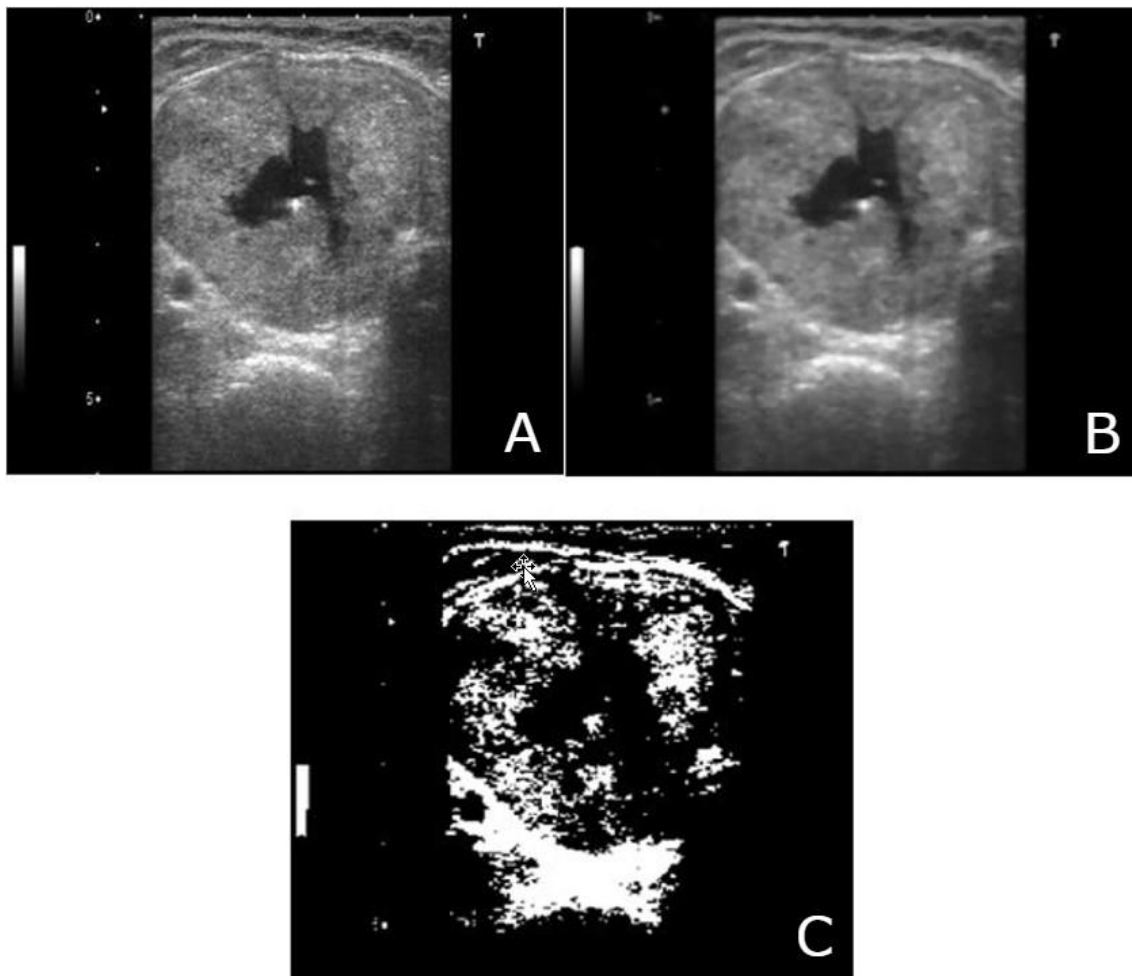
3.2. Accuracy-Based Predictive Model and Optimization

The predictive analysis for the training class was divided into three categories: benign, malignant and test (includes benign, malignant, and additional images). As this study was based on the accuracy score and the model's stability, over-optimization was avoided. Two models, ANN and SVM, achieved accuracy above 70% in the test prediction and are mentioned in this section.

Predictive models for thyroid nodule detection with good accuracy score, sensitivity, and specificity. ANN and SVM produced a decent accuracy score. The accuracy of the SVM model was 96%, and the accuracy of the ANN model was 75%.

Sensitivity describes how often a model correctly predicts a positive result for people with the condition that is tested for. For ANN, the sensitivity is 0.4914, and for SVM, it is 0.7866. Therefore, SVM outperforms ANN in sensitivity. Specificity describes how often a model correctly predicts a negative result for people without the condition tested for. Both ANN and SVM have a high specificity of 1.

Table 2 specifies the performance of both the model. The F1 score represents the harmonic mean of recall, and precision ranges from 0 to 1. The value indicates high-classification performance and the results demonstrated that SVM outperformed ANN. Table 3 shows the Confusion matrix for SVM test phase.
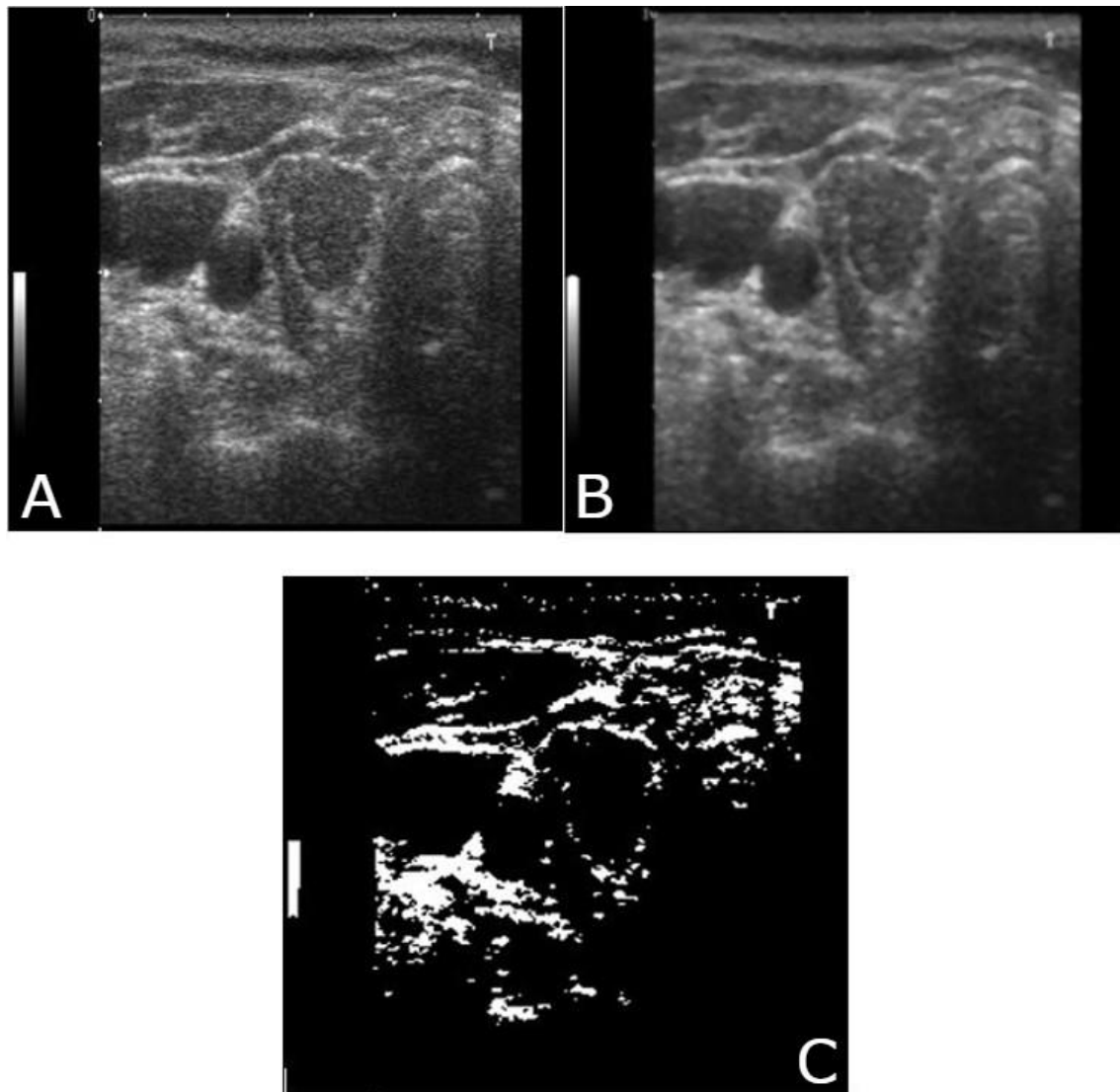
Figure 10. (**A**) Original image (benign)-Top Left; (**B**) median filtered image (benign)-Top Right; (**C**) segmentation-Center Bottom.

Table 2. Performance metric of ANN and SVM test phase.

ROC Curve Parameters	ANN—Test Phase	SVM
Distance	0.5086	0.7416
Threshold	0.0078	16
Sensitivity	0.4914	0.7866
Specificity	1	1
Accuracy	74.5684	96
PPV	1	0.7857
FNR	0.5086	0.2134
FPR	0	0
F1 score	0.659	0.92

Table 3. Confusion Matrix for SVM test phase.

Actual	Predicted		Accuracy = 0.96
	Positive	Negative	
Positive	$TP = 11$	$FN = 3$	P = 14
Negative	$FP = 0$	$TN = 62$	N = 62
	$PP = 11$	$PN = 65$	M = 76

Figure 11. (**A**) Original image (malignant)-Top Left; (**B**) median filtered image (malignant) Top Right; (**C**) binary segmentation-Center Bottom.

3.3. Graphic User Interface (GUI) from the Support Vector Machine (SVM)

After closely examining both ANN and SVM performance, it was decided that a Graphic User Interface would be designed for SVM. As SVM performed better in all parameters with high accuracy, sensitivity, and specificity, feature classification was obtained from the SVM model analysis as it had the best accuracy at 96%.

After running the GUI code, this window opens and prompts image selection from the test images file.

The images are uploaded to the input image window as shown in Figure 12; as soon as the SVM code is completed, the GUI displays the features.

The GUI was designed to display Nodule Composition, Echogenicity, Margin, Classification, and Type as shown in the Figure 13. For a highly suspicious model, a window prompt was used to recommend an FNAB to radiologists and doctors as shown in Figure 14. Four features for the GUI display were selected based on expert opinion. These four features demonstrated a high TPR in the SVM validation phase. Therefore, we chose these features in the GUI in order to give the clinician the most important information.
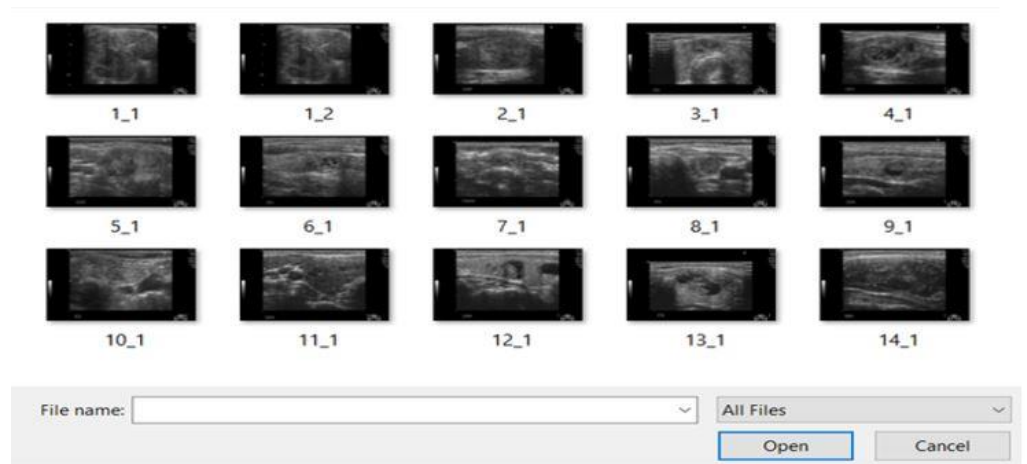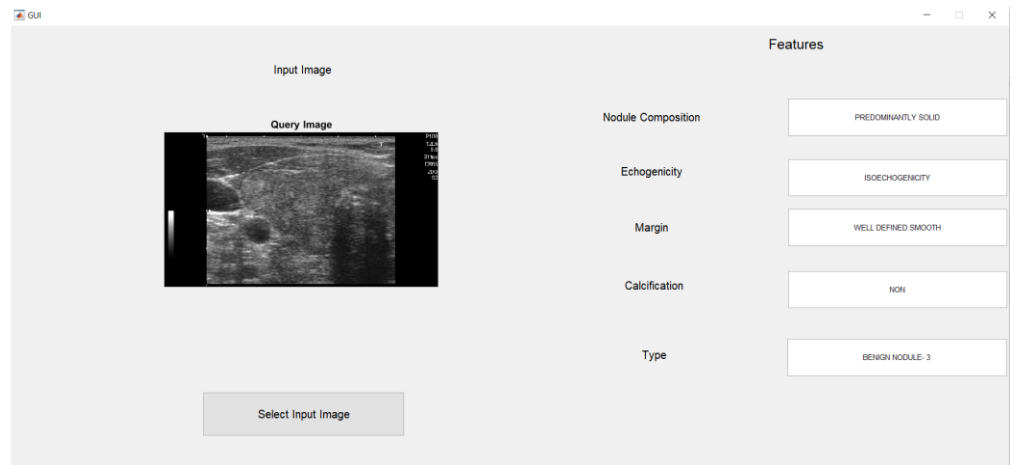
Figure 12. GUI prompt.

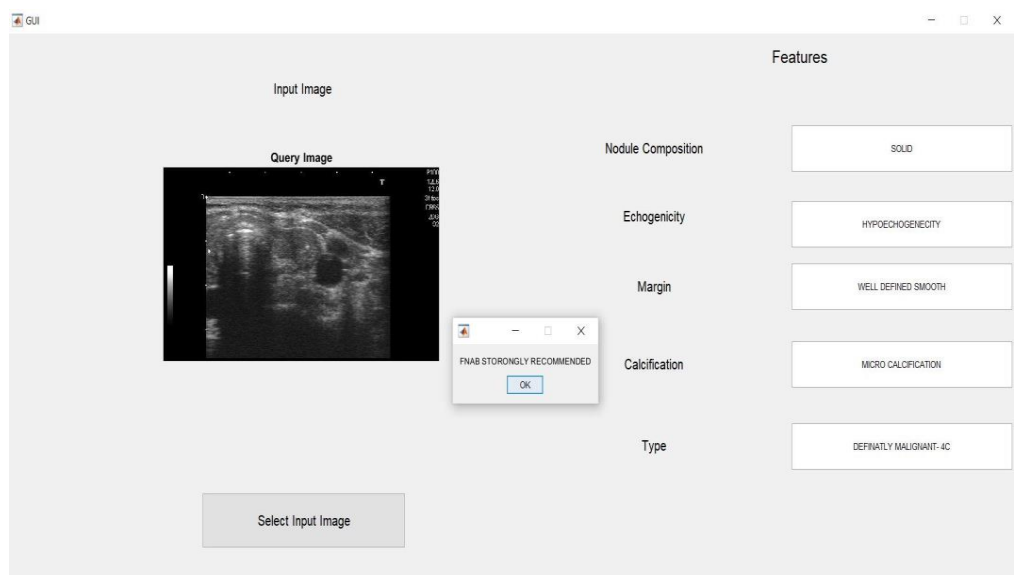**Figure 13.** GUI for benign nodules.

Figure 14. GUI for malignant nodule.

4. Discussion

This research described the ultrasound classification of thyroid nodules using machine learning techniques. An important aspect of this study is the completion of pre-processing with the use of median filters as the median filter is non-linear and can preserve essential details even after omitting noise. In this process, it was noted that median filters were advantageous over mean filters because no neighboring pixel would vary the median value.

Moreover, edge preservation is vital, and median filters did not add new pixels in the image. We found a significant reduction of noise with a change in contrast as compared to the original image. To divide an image into many segments, the segmentation process was used. This process located the boundary of the nodule and indicated the area of interest for feature extraction. Optical Character Recognition (ORC) enhanced the image by reducing the high-frequency noise and assigned values for 0 and 1 (black and white), with a fixed threshold.

In this study, two models were constructed. SVM outperformed ANN on all performance metrics, achieving an accuracy score of nearly 96%.

Regarding the construction of the model, SVMs perform well under a changing threshold with the introduction of the flexible kernel. It is non-linear and non-parametric, supporting different functions from all classifications of data. SVMs can provide out-of-sample generalization. This is due to SVM being robust even when there is bias in the training class data. In a way, SVM outperforms neural networks since the optimality problem is convex. A neural network has multiple solutions associated with local minima.

In this study we took a slightly different approach. According to our literature, we recognized that only 2% of all CAD use ANN with supervised learning. Therefore, we wanted to explore a highly accurate and stable model with minimum optimization.

There is only one published study to compare our results with [29]. However, they used different models such as 10-fold validation and optimization, a larger data set, and different GLCM texture features for image processing. Table 4 shows the comparison ROC curve parameters.

Table 4. Performance metric comparison from literature (ANN) perspective.

ROC Curve Parameters	This Study	Previous Study
Supervised Learning	Classification	Classification
Classification Secondary Model	ANN	Decision tree, random forest model
Sensitivity	0.4914	0.789
Specificity	1	0.785
AROC	0.7475	0.84
Accuracy	74.5684	78.5
PPV	1	0.785
FNR	0.5086	Not Specified
FPR	0	0.215
*F*1 *score*	0.659	0.784

The optimization of SVM and ANN was vital, varying the specific parameters in the training class to obtain better performance from the classifier. To use data effectively, the training class used most of the images as compared to the test class. A ratio of 87:13 was used for the training and testing classes for both SVM and ANN. A ratio of 85:15 was also tried, and ANN' accuracy decreased by a significant margin. For SVM variables, constraints and objective functions used were taken from libraries.

Clinically, our model performance suggests that it could reliably distinguish benign from malignant nodules. It has been suggested that the CAD systems could be used as a first reader or a second reader to the radiologist, or even as an autonomous reader. Further development of CAD systems such as those presented in our study would require careful planning as to how they could be integrated into clinical practice workflows. Validation

studies of integrated CAD systems would be required for real-world application to become evidence based. There is also clinical uncertainty as to the best way to apply CAD models—whether their purpose should be to have a high positive predictive value (i.e., if positive, the nodule is likely to be malignant and an FNAB would be warranted) or to have a high negative predictive value (i.e., if negative, the nodule is not likely to be malignant and an FNAB would not be warranted). If the radiologist is less experienced, a model that has a high negative predictive value may be preferred. Significant consideration should also be given to how integrated CAD models will influence the process of shared decision making between clinicians and patients regarding the management of thyroid nodules.

The successful implementation of CAD systems in real-world clinical practice could potentially reduce time spent by radiologists assessing ultrasounds, reduce the need for FNABs, and reduce interreader variability associated with TI-RADs. Recent studies of deep learning neural algorithms to provide recommendations for thyroid nodule biopsy are promising.

Future research in this area should use much more substantial datasets, and ones that come from two ultrasound sources. Future research should build two models—one under supervised learning and another under unsupervised learning—and try to integrate both models and compare performance metrics. Reinforcement machine learning algorithms should be deployed in future research because they are capable of choosing an action, building on each data point, and reviewing its own decision. These algorithms re-learn from their shortcomings and produce enhanced results each time. Platforms like Python or R should be used instead of MATLAB, as in this research. More information should be extracted to assist radiologists, which may be applied as a performance metric in future research and compared to the performance of the CAD system.

5. Conclusions

This research achieved its aim of building a CAD model that could classify thyroid nodules as benign or malignant, and the aim of evaluating its performance, demonstrating that an SVM model has superior performance as compared to an ANN model. This work demonstrates the significance of using traditional ML approaches with minimum optimization, stability, and successful CAD systems.
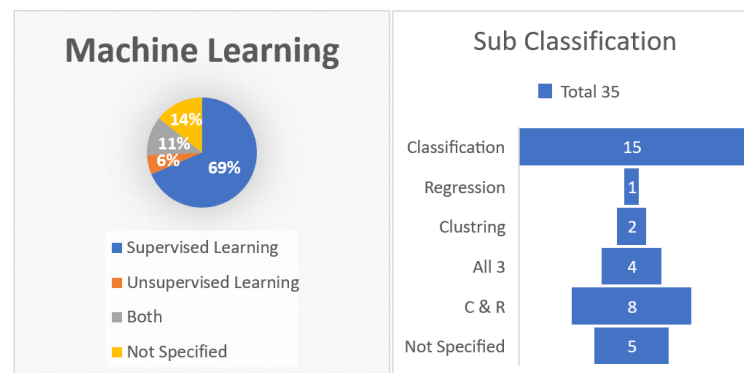
Via Engineering Laboratory, Cúram SFI Research Centre for Medical Devices, Lambe Institute for Translational Research, National University of Ireland Galway.

Conflicts of Interest: The authors declare no conflict of Interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

Most of the papers in the literature review used supervised learning due to the nature of the data. With the feedback loop and the use of both input and output data for the predictive model, high accuracy is guaranteed; however, using supervised learning and further optimizing it by unsupervised learning models produced a high accuracy score. Unsupervised models are specificity-based models which yield low accuracy for the size of our data set. Figure A1 shows that a number of research papers from our literature studies use various approaches.

Figure A1. Literature analysis.

Accuracy, sensitivity, and specificity-based models adopt the classification method, whereas Mean Squared Error and Mean Absolute Error-based models adopt the regression approach. A combined approach yields stable models. Figure A1 shows that a number of research papers from our literature studies use various approaches.

In computer vision, for any complicated computational task, feature extraction is deployed to extract specific points, compile it, and find solutions. In this thesis, GLCM was used to extract seven features to determine malignant and benign nodules.

Appendix B

Appendix B.1. Support Vector Machine (SVM)

Support vector machine (SVM) was the method used to accomplish the classification of benign and malignant by constructing hyperplane in a multidimensional space. They later separate different class labels. In our study, SVM was used to support classification tasks under supervised learning. SVM creates a dummy variable 0 or 1. Therefore, the dependent variable consists of three variables. Dummy variables are represented as X, Y, and Z.

$$X: \{1\ 0\ 0\}, Y: \{0\ 1\ 0\}, Z: \{0\ 0\ 1\}$$

To negate the error function, SVM uses an iterative training algorithm to build an optimal hyperplane.

Because we were using a classification model in our study, SVM models are categorized into two. They are as follows:

- Type 1 SVM Classification(C-SVM);
- Type 2 SVM Classification (nu-SVM);

To negate the error function, we used the following formula:

$$\frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi i$$

Constraints are as follows:

$$y_i\left(w^T\phi(xi)+b\right)\geq 1-\xi_i \text{ and } \xi_i\geq 0, 1\ldots.N$$

where the following are represented:

C—Capacity constant; w—vector of coefficients; b—constant; ei nonseparable data (inputs).

The index I tags the N, which is the training cases. These represent the class labels, while xi represents independent variables. I was the kernel for the transformation of data from input to feature [40].

Type 2 SVM Classification (nu—SVM).

To negate the error function, we used the following formula:

$$\frac{1}{2}w^T w - vp + \frac{1}{N}\sum_{i=1}^{N}\xi_i$$

Constraint is as follows:

$$yi\left(w^T\phi(x_i)+b\right)\geq p-\xi_i, \xi_i\geq 0, i=1\ldots N \text{ and } p\geq 0$$

In type 2 SVM, there was a need to estimate the functional dependence of variable y on the independent variable x. The relationship between x and y is represented by deterministic function f plus.

SVM is supported by many kernels such as linear, polynomial, radial basis function (RBF), and sigmoid.

$$K(X_i,X_j)=\begin{cases} X_i.X_j \text{ Linear}\\ (\gamma X_i.X_j+C)^d \text{ Polynomial}\\ exp\left(-\gamma|X_i-X_j|^2\right) RBF\\ tanh(\gamma X_i.X_j+C) \text{ Sigmoid} \end{cases}$$

where,

$$K(X_i,X_j)=\phi(X_i).\phi(X_j)$$

A dot product of input data points is mapped into a higher dimensional feature by ϕ.

Appendix C

Table A1. ACR TI-RADS.

No.	Scoring	Classification	Risk of Malignancy	Recommendations
TR1	0 Points	Benign	0.3%	No FNA required
TR2	2 Points	Not Suspicious	1.5%	No FNA required
TR3	3 Points	Mildly Suspicious	4.8%	Radiologist Decision (FNA or No FNA)
TR4	4–6 Points	Moderately Suspicious	9.1%	FNA required
TR5	Above 7 Points	Highly Suspicious	35%	FNA required

References

1. Ospina, N.S.; Iñiguez-Ariza, N.M.; Castro, M.R. Thyroid nodules: Diagnostic evaluation based on thyroid cancer risk assessment. *BMJ* **2020**, *368*. [CrossRef]
2. Guth, S.; Theune, U.; Aberle, J.; Galach, A.; Bamberger, C.M. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. *Eur. J. Clin. Investig.* **2009**, *39*, 699–706. [CrossRef] [PubMed]
3. Bomeli, S.R.; LeBeau, S.O.; Ferris, R.L. *Evaluation of a Thyroid Nodule. Otolaryngologic Clinics of North America*; Elsevier: Amsterdam, The Netherlands, 2010; Volume 43, pp. 229–238.
4. Burman, K.D.; Wartofsky, L. Thyroid nodules. *N. Engl. J. Med. Mass Med. Soc.* **2015**, *373*, 2347–2356. [CrossRef]
5. Ha, E.J.; Baek, J.H. Applications of machine learning and deep learning to thyroid imaging: Where do we stand? *Ultrasonoraphy* **2021**, *40*, 23. [CrossRef] [PubMed]
6. Cappelli, C.; Pirola, I.; Agosti, B.; Tironi, A.; Gandossi, E.; Incardona, P.; Marini, F.; Guerini, A.; Castellano, M. Complications after fine-needle aspiration cytology: A retrospective study of 7449 consecutive thyroid nodules. *Br. J. Oral Maxillofac. Surg.* **2017**, *55*, 266–269. [CrossRef]
7. Sebo, T.J. What are the keys to successful thyroid FNA interpretation? *Clin. Endocrinol.* **2012**, *77*, 13–17. [CrossRef]
8. Altavilla, G.; Pascale, M.; Nenci, I. Fine needle aspiration cytology of thyroid gland diseases. *Acta Cytol.* **1990**, *34*, 251–256.
9. Wang, Z.; Vyas, C.M.; van Benschoten, O.; Nehs, M.A.; Moore, F.D., Jr.; Marqusee, E.; Krane, J.F.; Kim, M.I.; Heller, H.T.; Gawande, A.A.; et al. Quantitative analysis of the benefits and risk of thyroid nodule evaluation in patients ≥70 years old. *Thyroid* **2018**, *28*, 465–471. [CrossRef]
10. Ahn, H.S.; Kim, H.J.; Kim, K.H.; Lee, Y.S.; Han, S.J.; Kim, Y.; Ko, M.J.; Brito, J.P. Thyroid cancer screening in South Korea increases detection of papillary cancers with no impact on other subtypes or thyroid cancer mortality. *Thyroid* **2016**, *26*, 1535–1540. [CrossRef] [PubMed]
11. Nilubol, N.; Kebebew, E. Should small papillary thyroid cancer be observed? A population-based study. *Cancer* **2015**, *121*, 1017–1024. [CrossRef]
12. Tappouni, R.R.; Itri, J.N.; McQueen, T.S.; Lalwani, N.; Ou, J.J. ACR TI-RADS: Pitfalls, solutions, and future directions. *Radiographics* **2019**, *39*, 2040–2052. [CrossRef]
13. Thomas, J.; Haertling, T. AIBx, artificial intelligence model to risk stratify thyroid nodules. *Thyroid* **2020**, *30*, 878–884. [CrossRef]
14. Ko, S.Y.; Lee, J.H.; Yoon, J.H.; Na, H.; Hong, E. *Deep Convolutional Neural Network for the Diagnosis of Thyroid Nodules on Ultrasound*; PubMed: Seoul, Korea, 2011.
15. Wang, L.; Yang, S.; Yang, S.; Zhao, C.; Tian, G.; Gao, Y.; Chen, Y.; Lu, Y. Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the YOLOv2 neural network. *World J. Surg. Oncol.* **2019**, *17*, 12. [CrossRef] [PubMed]
16. Seo, J.K.; Kim, Y.J.; Kim, K.G.; Shin, I.; Shin, J.H.; Kwak, J.Y. Differentiation of the follicular neoplasm on the grayscale us by image selection subsampling along with the marginal outline using convolutional neural network. *BioMed Res. Int.* **2017**. [CrossRef] [PubMed]
17. Chi, J.; Walia, E.; Babyn, P.; Wang, J.; Groot, G.; Eramian, M. Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. *J. Digit. Imaging* **2017**, *30*, 477–486. [CrossRef]
18. Gao, L.; Liu, R.; Jiang, Y.; Song, W.; Wang, Y.; Liu, J.; Wang, J.; Wu, D.; Li, S.; Hao, A.; et al. Computer-aided system for diagnosing thyroid nodules on ultrasound: A comparison with radiologist-based clinical assessments. *Head Neck* **2018**, *40*, 778–783. [CrossRef]
19. Yu, Q.; Jiang, T.; Zhou, A.; Zhang, L.; Zhang, C.; Xu, P. Computer-aided diagnosis of malignant or benign thyroid nodes based on ultrasound images. *Eur. Arch. Otorhinolaryngol.* **2017**, *274*, 2891–2897. [CrossRef]
20. Ma, J.; Wu, F.; Zhu, J.; Xu, D.; Kong, D. A pre-trained convolutional neural network-based method for thyroid nodule diagnosis. *Ultrasonics* **2017**, *73*, 221–230. [CrossRef] [PubMed]
21. Xia, J.; Chen, H.; Li, Q.; Zhou, M.; Chen, L.; Cai, Z.; Fang, Y.; Zhou, H. Ultrasound-based differentiation of malignant and benign thyroid Nodules: An extreme learning machine approach. *Comput. Methods Programs Biomed.* **2017**, *147*, 37–49. [CrossRef]
22. Wu, H.; Deng, Z.; Zhang, B.; Liu, Q.; Chen, J. Classifier model based on machine learning algorithms: Application to differential diagnosis of suspicious thyroid nodules via sonography. *Am. J. Roentgenol.* **2016**, *207*, 859–864. [CrossRef]
23. Ardakani, A.A.; Gharbali, A.; Mohammadi, A. *Application of Texture Analysis Method for Classification of Benign and Malignant Thyroid Nodules in Ultrasound Images*; Pubmed: Urmia, Iran, 2015.
24. Liu, Y.I.; Kamaya, A.; Desser, T.S.; Rubin, D.L. A bayesian network for differentiating benign from malignant thyroid nodules using sonographic and demographic features. *Am. J. Roentgenol.* **2011**, *196*, W598–W605. [CrossRef]
25. Tsantis, S.; Dimitropoulos, N.; Cavouras, D.; Nikiforidis, G. Morphological and wavelet features towards sonographic thyroid nodules evaluation. *Comput. Med. Imaging Graph.* **2009**, *33*, 91–99. [CrossRef]
26. Zhu, L.C.; Ye, Y.L.; Luo, W.H.; Su, M.; Wei, H.P.; Zhang, X.B.; Wei, J.; Zou, C.L. A model to discriminate malignant from benign thyroid nodules using artificial neural network. *PLoS ONE* **2013**, *8*, e82211. [CrossRef]
27. Lim, K.J.; Choi, C.S.; Yoon, D.Y.; Chang, S.K.; Kim, K.K.; Han, H.; Kim, S.S.; Lee, J.; Jeon, Y.H. *Computer-Aided Diagnosis for the Differentiation of Malignant from Benign Thyroid Nodules on Ultrasonography*; Pubmed: Seoul, Korea, 2008.
28. Raghavendra, U.; Acharya, U.R.; Gudigar, A.; Tan, J.H.; Fujita, H.; Hagiwara, Y.; Molinari, F.; Kongmebhol, P.; Ng, K.H. Fusion of spatial gray level dependency and fractal texture features for the characterization of thyroid lesions. *Ultrasonics* **2017**, *77*, 110–120. [CrossRef]

29. Song, G.; Xue, F.; Zhang, C. A model using texture features to differentiate the nature of thyroid nodules on sonography. *J. Ultrasound Med.* **2015**, *34*, 1753–1760. [CrossRef] [PubMed]
30. Schenke, S.A.; Wuestemann, J.; Zimny, M.; Kreissl, M.C. Ultrasound Assessment of Autonomous Thyroid Nodules before and after Radioiodine Therapy Using Thyroid Imaging Reporting and Data System (TIRADS). *Diagnostics* **2020**, *10*, 1038. [CrossRef] [PubMed]
31. Gomes Ataide, E.J.; Ponugoti, N.; Illanes, A.; Schenke, S.; Kreissl, M.; Friebe, M. Thyroid Nodule Classification for Physician Decision Support Using Machine Learning-Evaluated Geometric and Morphological Features. *Sensors* **2020**, *20*, 6110. [CrossRef] [PubMed]
32. Kwon, M.R.; Shin, J.H.; Park, H.; Cho, H.; Kim, E.; Hahn, S.Y. Radiomics Based on Thyroid Ultrasound Can Predict Distant Metastasis of Follicular Thyroid Carcinoma. *J. Clin. Med.* **2020**, *9*, 2156. [CrossRef]
33. Nguyen, D.T.; Kang, J.K.; Pham, T.D.; Batchuluun, G.; Park, K.R. Ultrasound image-based diagnosis of malignant thyroid nodule using artificial intelligence. *Sensors* **2020**, *20*, 1822. [CrossRef] [PubMed]
34. Daniels, K.; Gummadi, S.; Zhu, Z.; Wang, S.; Patel, J.; Swendseid, B.; Lyshchik, A.; Curry, J.; Cottrill, E.; Eisenbrey, J. Machine learning by ultrasonography for genetic risk stratification of thyroid nodules. *JAMA Otolaryngol. Head Neck Surg.* **2020**, *146*, 36–41. [CrossRef]
35. Colombia, Universidad Nacional de Colombia. Digital Database of Thyroid Ultrasound Images. Universidad Nacional de Colombia. Available online: http://cimalab.intec.co/?lang=en=project=31 (accessed on 26 July 2019).
36. Tan, L.; Jiang, J. Chapter 13—Image Processing Basics. s.l.: Science Direct. 2019. Available online: https://www.globalspec.com/reference/81808/203279/chapter-13-image-processing-basics (accessed on 26 July 2019).
37. Puneet, P.; Garg, N.K. Binarization Techniques used for Grey Scale Images. *Bathinida Int. J. Comput. Appl.* **2013**, *71*, 8–11. [CrossRef]
38. Anuradha, K. Statistical feature extraction to classify oral cancers. *J. Glob. Res. Comput. Sci.* **2013**, *4*, 8–12.
39. Rai, H. Support Vector Machine-Part 2. Neural Networks the Way Information Moves. Available online: https://neuralnetset.blogspot.com/2016/02/support-vector-machine-part-2.html (accessed on 6 February 2019).
40. Wikipedia. Artificial Neural Network. Available online: https://en.wikipedia.org/wiki/Artificial_neural_network (accessed on 29 July 2019).
41. Parveen, K. *Artificial Neural Networks—A Study*; IJEERT: Ambala, India, 2014; Volume 2, pp. 143–148.
42. Alam, M. Codes in MATLAB for training artificial neural network using particle swarm optimization. *Res. Gate* **2016**, 1–16. [CrossRef]