International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Comprehensive Analysis of Machine Learning Techniques for Thyroid Cancer Detection

Monika D. Kate [1], Dr. Vijay K. Kale[2]

[1] Research Student, Department of Computer, [2] Associate Professor, Department of Computer Science,
[1,2] Dr.G.Y.Pathrikar College of Computer Science and Information Technology MGM University
Chhatrapti Sambajinagar, Maharashtra, India
[1]monikaborkar1405@gmail.com, [2]vkale@mgmu.ac.in

Abstract

Thyroid nodules are commonly identified in clinical practice, and although the majority are benign, a sizable portion may be malignant, requiring prompt and precise diagnosis. It is essential to distinguish between benign and malignant nodules in order to direct suitable clinical interventions and steer clear of needless procedures. This study proposes a comprehensive MATLAB-based system for ultrasound imaging-based thyroid nodule detection and classification. This study offers a multi-step process to enhance the quality of US thyroid images and highlight relevant structures, beginning with image preprocessing techniques like edge identification, contrast enhancement, and noise reduction. In order to find any patterns that stand out, the morphological and textural characteristics of the nodules are examined in the feature extraction step that follows. Finally, using machine learning techniques, the nodules are categorized as either benign or malignant. The objective of this classification is to improve clinical decision-making by offering a dependable, automated tool that facilitates risk assessment and early detection in the treatment of thyroid illness.

Keywords: Thyroid Nodules, Ultrasound Imaging, Machine Learning

1. Introduction

Thyroid nodules are among the most frequently encountered abnormalities in endocrine clinical practice, with epidemiological studies suggesting their presence in a considerable portion of the general population—particularly in older adults and individuals with iodine deficiency. Although the majority of these nodules are benign and asymptomatic, approximately 5–15% carry a potential risk of malignancy. This underscores the critical importance of accurate and timely diagnosis to ensure appropriate clinical intervention, minimize unnecessary biopsies or surgeries, and optimize patient outcomes.

Ultrasonography imaging has long been established as the first-line diagnostic method for thyroid nodule evaluation due to its non-invasive nature, cost-effectiveness, and high sensitivity in detecting structural abnormalities. However, a significant challenge with ultrasound-based diagnosis lies in its inherent subjectivity. The correctness of image interpretation is heavily based on the expertise of the operator and radiologist, often resulting in inter-observer variability and inconsistent diagnostic conclusions.

To address these limitations, the development of computer-aided diagnosis (CAD) systems has emerged as a potentially effective method to support radiologists by offering objective, reproducible, and standardized assessments. These systems make use of advanced image analysis techniques to assist in the detection, characterization, and risk stratification of thyroid nodules, thereby reducing diagnostic uncertainty and enhancing clinical confidence.

Recent progress in artificial intelligence (AI), particularly in the domain of deep learning and machine learning, has further accelerated the capabilities of CAD systems. Various deep learning models, such as CNNs, have shown great success in recognizing images, even in medical imaging. Recently, Zhang et al.(2022) built the Multistep Automated Data Labeling Procedure (MADLaP), a specialized deep learning software for clarifying ultrasound pictures of thyroid nodules. With their mechanism, the team created labeled data from pathology findings automatically and reached an accuracy of over 80% in assigning the right labels. This current study introduces a MATLAB-supported computer system that can automatically detect and sort thyroid nodules seen in ultrasound images. My approach consists of using a set process, beginning with image preprocessing, gathering the needed features, and then classifying them with machine learning methods. Unlike other techniques that can only tell if nodules are benign or malignant, our approach sorts them into three groups—mild, moderate, and severe—which helps give more accurate risk analysis. Its purpose is to help clinicians choose the best approaches for every patient. Our system seeks to help clinicians by pairing familiar image processing techniques with modern machine learning in MATLAB. This allows for simple, reliable support in early evaluation of thyroid nodules and higher consistency in diagnostic outcomes.

2. Literature Review

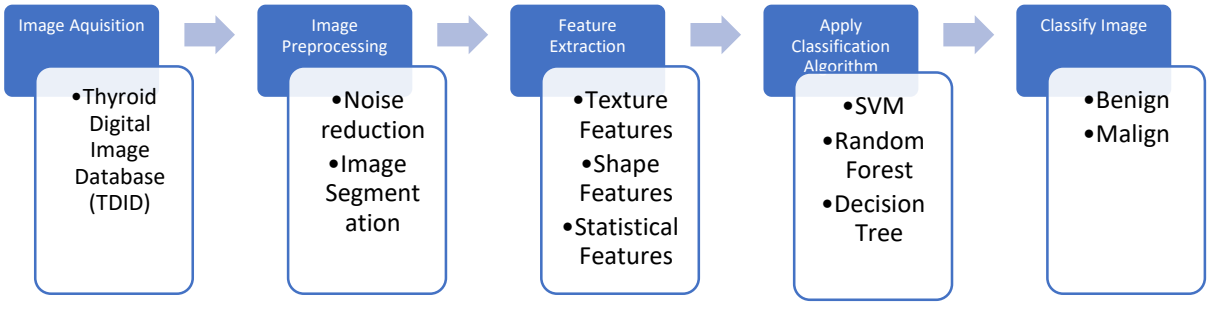
Recent developments in computer-aided diagnosis (CAD) systems have greatly improved the assessment of thyroid nodules by applying machine learning and deep learning techniques. For instance, Zhang et al. (2022) developed the Multistep Automated Data Labelling Procedure (MADLaP), a deep-learning-based tool designed to automate the data annotation process for ultrasound thyroid nodules images. Their study demonstrated that MADLaP could successfully create curated datasets of labeled ultrasound images, achieving an accuracy of 83% in correctly assigning pathology labels. Similarly, Weng et al. (2022) applied a previously validated deep learning algorithm to a new US thyroid nodule image dataset and compared its performance with that of radiologists. The study found that the algorithm achieved an Area Under the Curve (AUC) of 0.69, comparable to the AUCs of four experienced radiologists, which ranged from 0.63 to 0.66. This recommends that deep learning algorithms can perform on par with human experts in classifying thyroid nodules. Furthermore, Manh et al. (2022) proposed a Multi-Attribute Attention Network (MAA-Net) designed to mimic the clinical diagnosis process. The model predicts nodular attributes and infers their malignancy based on clinically relevant features, providing interpretable predictions that may better suit clinical needs. Validation experiments on a large dataset demonstrated that MAA-Net outperformed other state-of-the-art methods in diagnosing thyroid nodules. These studies underscore the potential of integrating artificial intelligence into thyroid nodule evaluation to improve diagnostic consistency and efficiency. Building upon these developments, our study proposes a MATLAB-based CAD system aimed at detecting and classifying thyroid nodules in ultrasound images, thereby assisting medical professionals in making informed decisions regarding patient management. Kamta Nath Mishra et al. (2024). Recent research highlights the growing role of ML in

cancer detection and prevention. Studies show that ML algorithms can accurately classify cancerous cells using features like cell size, shape, nucleus structure, and gene expression. These techniques improve early diagnosis and treatment planning by differentiating between cancerous and non-cancerous tumors. Automated feature extraction and diverse data usage enhance detection accuracy. However, researchers emphasize the necessity of high-quality data, model validation, and continued refinement of algorithms. Future work includes expanding ML applications to various cancer types and integrating deep learning for even greater accuracy. Ya Guo et al. (2024) highlight the growing use of machine learning (ML) and deep learning (DL) in breast cancer detection, enhancing accuracy in classifying benign and malignant tumors through medical imaging. While promising, ML encounters challenges like data privacy, the need for large datasets, and limited ability to diagnose unfamiliar cases. Further study is needed to improve clinical integration and standardization. Yogesh Kumar et al. (2024) explore deep learning models to cancer detection, demonstrating high accuracy across multiple cancer types. DenseNet121, in particular, achieved a validation accuracy of 99.94%, outperforming other CNN architectures. However, limited cancer type coverage and lack of cost-effectiveness analysis highlight the need for broader datasets and more practical, resource-efficient models for clinical use.

3. Methodology

The proposed Computer-Aided Diagnostic (CAD) model follows a structured pipeline of several key stages. It begins with image acquisition, where ultrasound (US) images of thyroid nodules are collected from an online public repository. In the preprocessing phase, these images undergo enhancement and segmentation to isolate the region of interest (ROI), ensuring that only relevant areas are analyzed. Following this, various features are extracted from the ROI, including texture features (such as contrast and homogeneity), shape features (like area and circularity), and statistical features (such as mean and standard deviation). These extracted features serve as input to a set of classification methods, including (SVM) Support Vector Machine, Random Forest, and Decision Tree, which are used to classify the nodules as either benign or malignant accurately.

Figure 1: Preposed Workflow



3.1. Data Acquisition

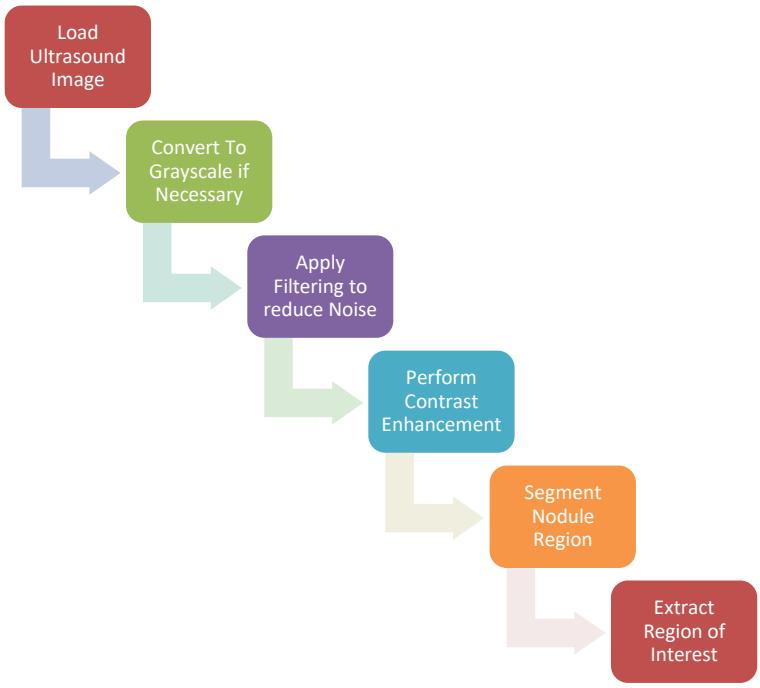
The dataset is obtained from the Thyroid Digital Image Database (TDID), hosted by the University of Columbia, and is publicly accessible. It consists of B-mode ultrasound images of thyroid nodules, which are two-dimensional grayscale images where each bright dot represents the strength of an ultrasound echo. These images have been annotated and diagnosed by experienced radiologists, making the dataset

highly reliable for research and analysis. The dataset includes information from 389 patients, with an additional 99 patient records, 33 classified as benign and 66 as malignant. Thyroid nodules are categorized according to the TIRADS (Thyroid Imaging Reporting and Data System), which scores nodules based on suspicious ultrasound features such as irregular margins, hypoechogenicity, microcalcifications, and shape. In this system, TIRADS 1 indicates normal thyroid, while higher levels indicate increasing risk of malignancy. Among the benign images, 22 are TIRADS 2 and 11 are TIRADS 3. The malignant cases include 20 images with TIRADS 4A, 16 with TIRADS 4B, 19 with TIRADS 4C, and 11 with TIRADS 5. Cancer risk increases significantly with higher TIRADS levels, from 0.9% in TIRADS 2 to 86% in TIRADS 5. For this research, we have used 30 benign and 38 malignant images selected from the dataset. The ultrasound image annotations are stored in .xml file format, enabling easy integration with machine learning and image processing tools.

3.2. Image Preprocessing

To improve image quality, median filtering was employed to reduce speckle noise commonly present in ultrasound images. Histogram equalization was then applied to enhance image contrast, making important features more distinguishable. For the segmentation process, adaptive thresholding was utilized to effectively isolate the thyroid nodules from the surrounding tissue, enabling precise region of interest (ROI) extraction.

Figure 2: Image Preprocessing Steps



3.3 Feature Extraction

Feature extraction is a critical step in the image analysis pipeline, as it involves quantifying meaningful characteristics from the segmented thyroid nodules that can be used for classification. In this study, three main categories of features were extracted from the region of interest (ROI) corresponding to each nodule: texture features, shape features, and statistical features. These features serve as the foundation for distinguishing between different classes of nodules based on their visual and structural properties.

3.3.1 Texture Features

Texture analysis captures the spatial distribution and relationship of pixel intensities within the nodule, offering insights into its internal structure. For this purpose, we employed the **Gray-Level Co-occurrence Matrix (GLCM)**, a widely used method for analyzing texture in grayscale images. From the GLCM, we extracted the following features:

Contrast: Measures the intensity contrast between a pixel and its neighbors over the whole image. High contrast often indicates heterogeneity, which is common in malignant nodules.

- **Correlation**: Indicates the degree of linear dependency of gray levels between neighboring pixels. A higher correlation implies a more regular or structured texture pattern.
- **Energy**: Also known as angular second moment, it reflects the textural uniformity of the image. Homogeneous areas (like benign nodules) tend to have higher energy values.
- **Homogeneity**: Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. Higher homogeneity suggests a smoother texture, typical of benign lesions.

These texture metrics help in differentiating nodules with irregular and complex internal textures (often seen in malignant cases) from those with smoother, more uniform patterns (commonly benign).

3.3.2 Shape Features

Shape-based features provide geometric information about the nodules, which can be critical for malignancy prediction, as cancerous nodules often display irregular or asymmetric shapes. The extracted shape-related parameters include:

- **Area**: Represents the total number of pixels within the nodule boundary. Larger or smaller areas can be indicative of different stages or types of lesions.
- **Perimeter**: The length of the boundary enclosing the nodule, useful in detecting irregular outlines.
- **Circularity**: Defined as 4π×(Area)/(Perimeter2)4\pi \times (\text{Area}) / (\text{Perimeter}^2)4π×(Area)/(Perimeter2), this metric evaluates how close the nodule shape is to a perfect circle. Benign nodules often exhibit higher circularity.
- **Aspect Ratio**: The ratio of the major axis to the minor axis of the bounding ellipse around the nodule. Malignant nodules often have a taller-than-wide shape, resulting in a higher aspect ratio.

These shape features help in capturing structural irregularities that may signal malignancy.

3.3.3 Statistical Features

Statistical features summarize the intensity distribution of pixel values within the segmented nodule and are useful for capturing brightness and contrast variations. The extracted statistical descriptors include:

- **Mean**: The average pixel intensity, giving an overall measure of the brightness level within the nodule.
- **Standard Deviation**: Measures the variability in pixel intensities. A higher standard deviation may indicate heterogeneous regions, often associated with malignant growths.

- **Skewness**: Describes the asymmetry of the intensity histogram. It helps in understanding whether the pixel distribution is skewed toward darker or brighter intensities.
- **Kurtosis**: Indicates the "peakedness" or flatness of the intensity distribution, which can reflect the existence of extreme intensity variations or uniform regions.

Collectively, these statistical metrics deliver valuable information about the nodule's internal composition, aiding in the distinction between different pathological types.

3.4. Classification Algorithm

In this study, we developed and evaluated three widely used classification models—Support Vector Machine (SVM), Random Forest, and Decision Tree—with the objective of accurately distinguishing between benign and malignant thyroid nodules. To assess the robustness and generalizability of these models, we conducted experiments using three different train-test split ratios: 80:20, 70:30, and 60:40. Adopting this approach, we could judge how each model behaved when there were different amounts of training and testing data, representing situations that occur in real clinical and other data.

For each classifier and split configuration, we computed and analyzed various evaluation metrics like accuracy, precision, recall, and F1-score results. These measurements offer a clear view of a model's performance in different ways: accuracy describes how accurate the predictions are overall, precision shows whether the model is correct in marking positive instances, recall denotes how easily the model catches actual positive examples, and the F1-score combines precision and recall into one summary. With this framework in place, we were able to compare classifiers and ways to split the data, which led us to pick the most appropriate model for accurate and easy thyroid nodule classification.

4. Experiment and Results

This study examined the application of machine learning for the classification of thyroid ultrasound images into benign and malignant categories to aid in cancer detection. Three distinct algorithms for classification were explored: Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel, a single Decision Tree, and the ensemble learning method of Random Forest. Each Method relied on a common practice: the initial loading and preprocessing of image data from designated 'benign' and 'malignant' directories, which included resizing images to a uniform dimension and dividing the dataset into training and testing subsets. Subsequently, feature extraction was performed using Histogram of Oriented Gradients (HOG) to capture shape and textural characteristics, the Gray-Level Co-occurrence Matrix (GLCM) to quantify statistical texture properties. To make sure the optimal model performance, the features obtained from the images were standardized before starting the training process. Initially, each of the trained classifiers worked with the prepared training set, and then their accuracy was checked against the independent evaluation data using F1-score, precision, recall, accuracy, and a confusion matrix. The implementations also included optional visualization of the classification results and the capability to save the trained models and extracted features for potential future use or comparison.

5. Comparison of Results

In the current study, we applied three machine learning classifiers—SVM, Random Forest, and Decision Tree—to the given dataset. Figure 3 presents sample test images along with predictions made by the SVM model, which achieved a test accuracy of 70%. Figure 4 displays the predictions of the Decision Tree classifier, with a test accuracy of 60%, while Figure 5 shows the predictions from the Random Forest model, which attained a test accuracy of 65%.

Figure 3:Sample Test Images with Prediction (SVM)

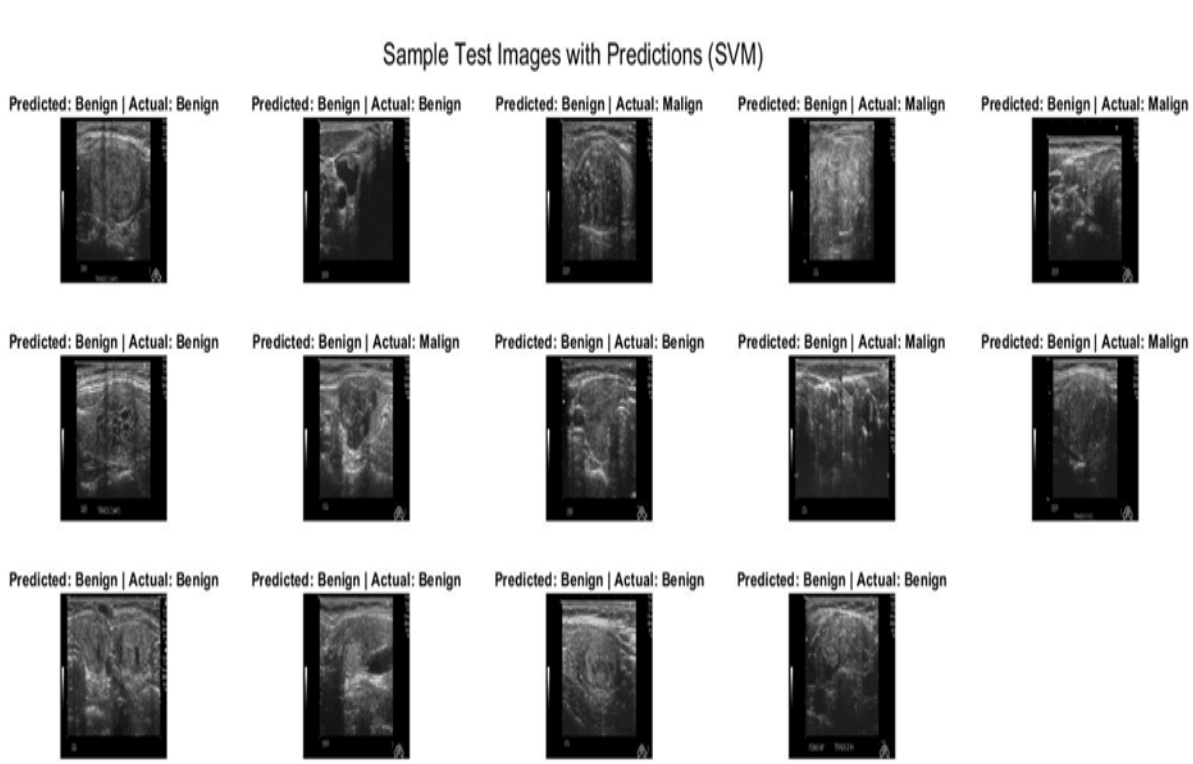


Figure 4:Sample Test Images with Prediction (Decision Tree)

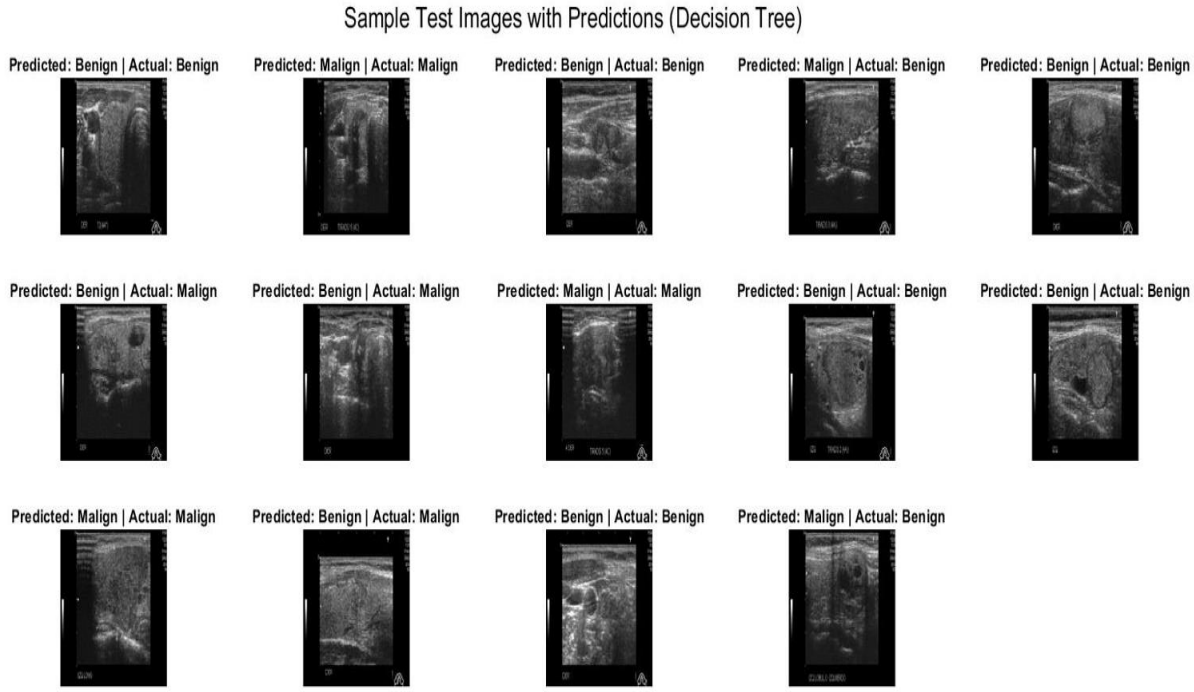


Figure 5:Sample Test Images with Prediction (Random Forest)

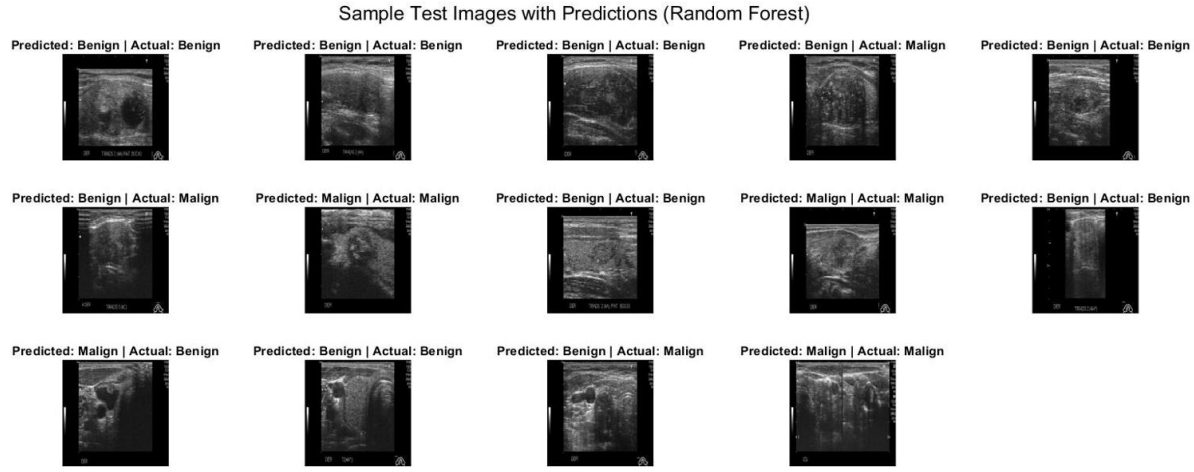
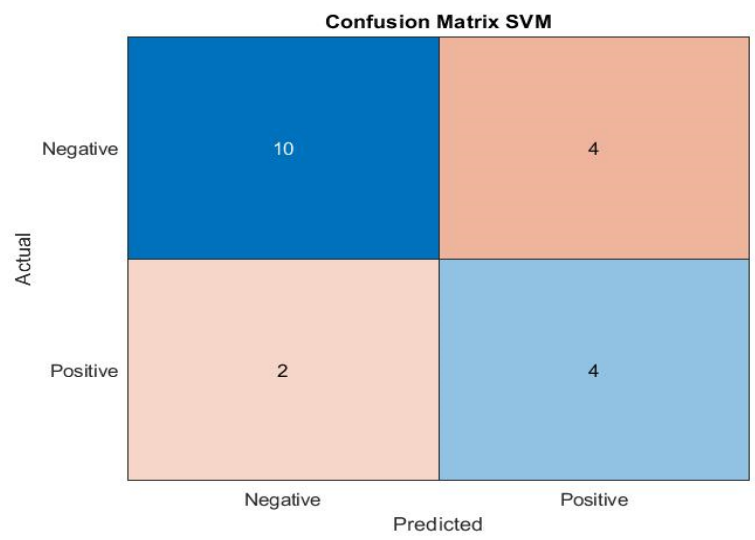


Table 1 :Performance Evaluation Parameters

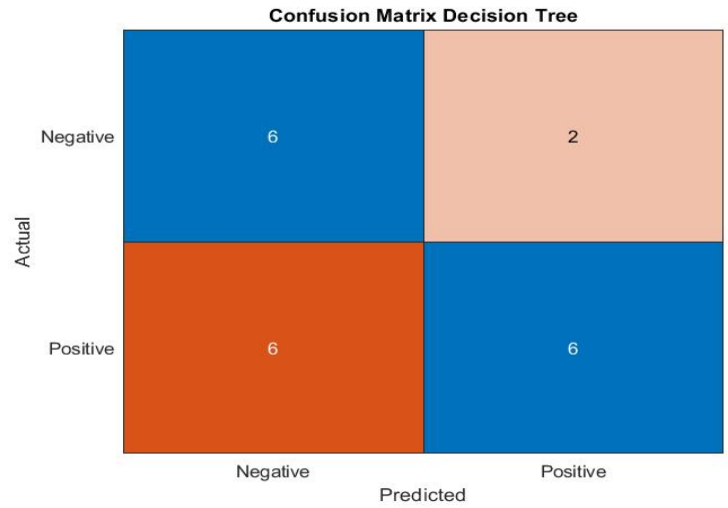
Model	Precision	Recall	F1-score	Test Accuracy	5-Fold Cross-validated Accuracy
SVM	0.67	0.50	0.57	70.00%	62.96%
Random Forest	0.63	0.56	0.59	65.00%	64.71%
Decision Tress	0.50	0.75	0.60	60.00%	45.59%

The performance parameters of the three evaluated models—Support Vector Machine (SVM), Random Forest, and Decision Tree—are shown in Table 1 above.

Among the three models tested—Support Vector Machine (SVM), Random Forest, and Decision Tree, SVM achieved the highest test accuracy at 70%, indicating strong performance on unseen data. It also maintained a good balance between precision (0.67) and F1-score (0.57), along with a cross-validation accuracy of 62.96%, showing reasonable consistency. Random Forest offered a well-rounded performance with slightly lower test accuracy (65%) but the best cross-validation accuracy (64.71%), suggesting strong stability across different data splits. The Decision Tree model, while achieving the highest recall (0.75), had the lowest test accuracy (60%) and cross-validation accuracy (45.59%), implying that although it is good at detecting positive cases, it may be overfitting the training data. Overall, SVM is the best choice for general accuracy, Random Forest provides the most consistent performance, and Decision Tree is useful when recall is the priority.



Confusion Matrix Random Forest



Confusion Matrix Decision Tree

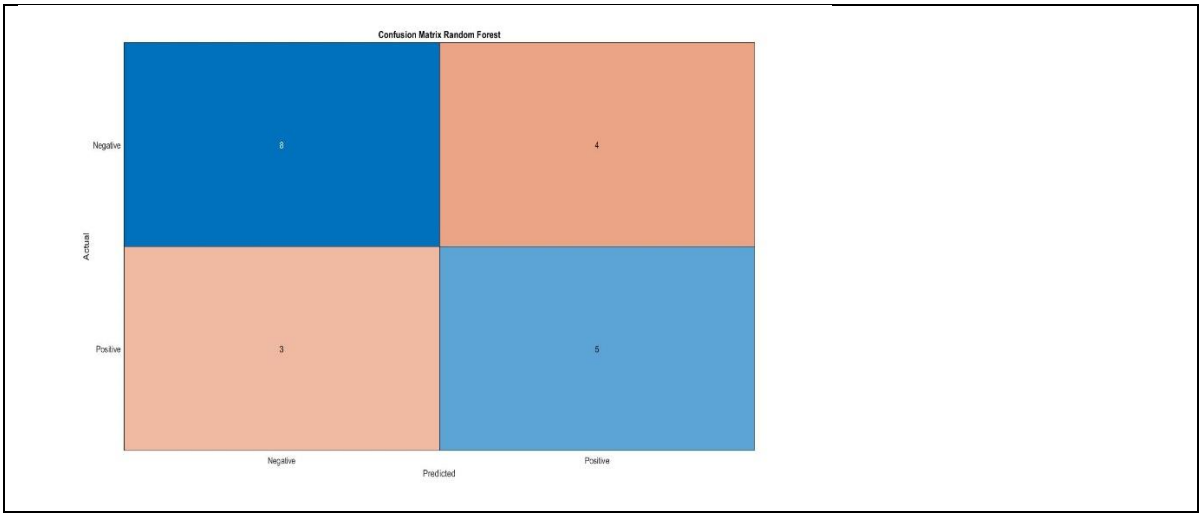


Figure 6 :Confusion Matrix

The confusion matrix images compare the classification performance of three models: SVM, Decision Tree, and Random Forest.:

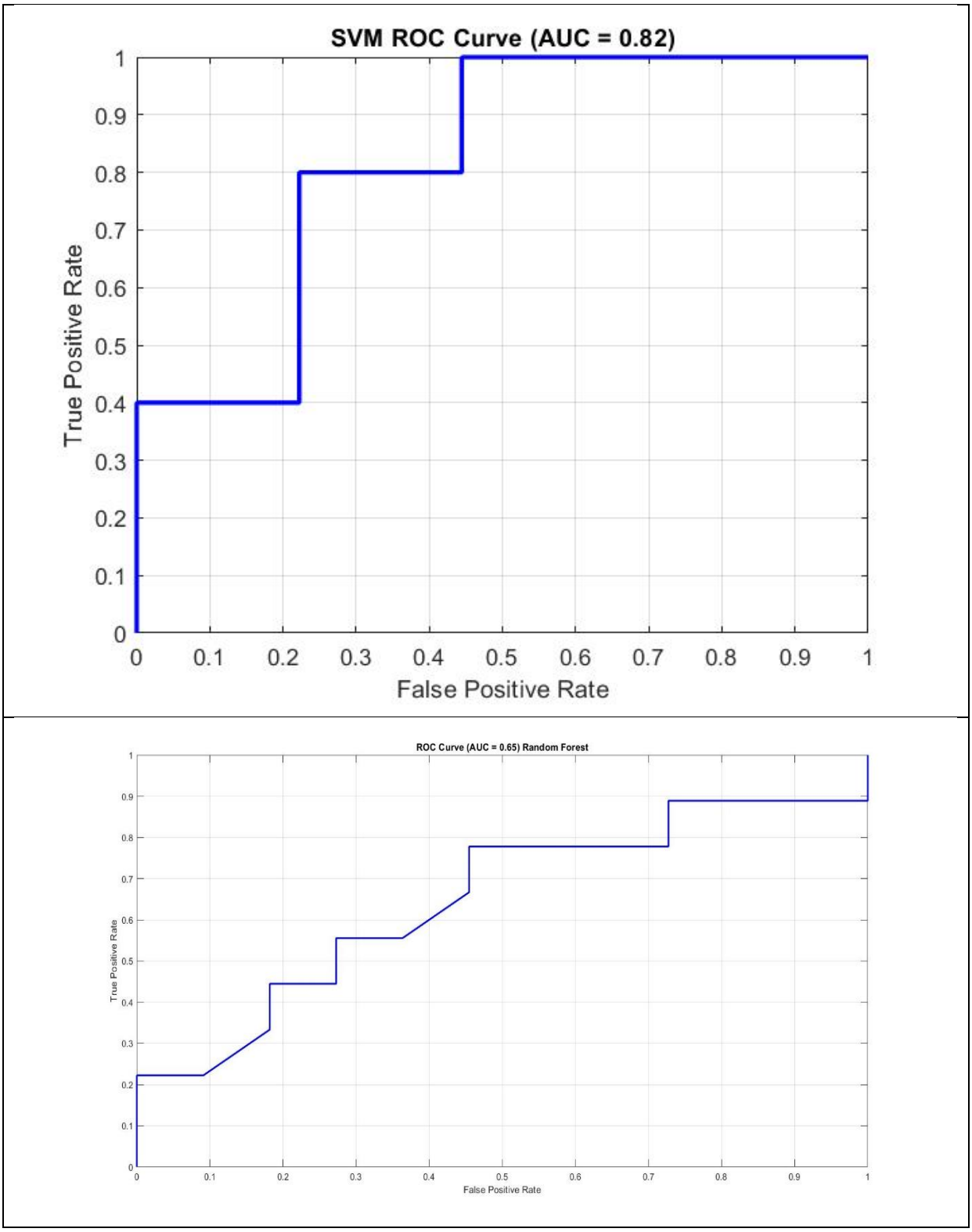
The **SVM model** shows a good balance between true positives (4) and true negatives (10), with only 2 false negatives and 4 false positives. This indicates decent precision and moderate recall, making it reliable at identifying both classes.

The **Decision Tree model** has an equal number of correct positive predictions (6) and false negatives (6), but fewer false positives (2), which leads to higher recall but lower precision due to the higher misclassification of actual positives.

The **Random Forest model** performs moderately with 5 true positives and 8 true negatives, but also shows 4 false positives and 3 false negatives. This results in a fairly balanced trade-off between precision and recall.

Overall, SVM demonstrates stronger true negative identification, Decision Tree captures more positives correctly (high recall), and Random Forest offers a middle ground between them, balancing both classes with reasonable accuracy.

Figure 7 : ROC Curve



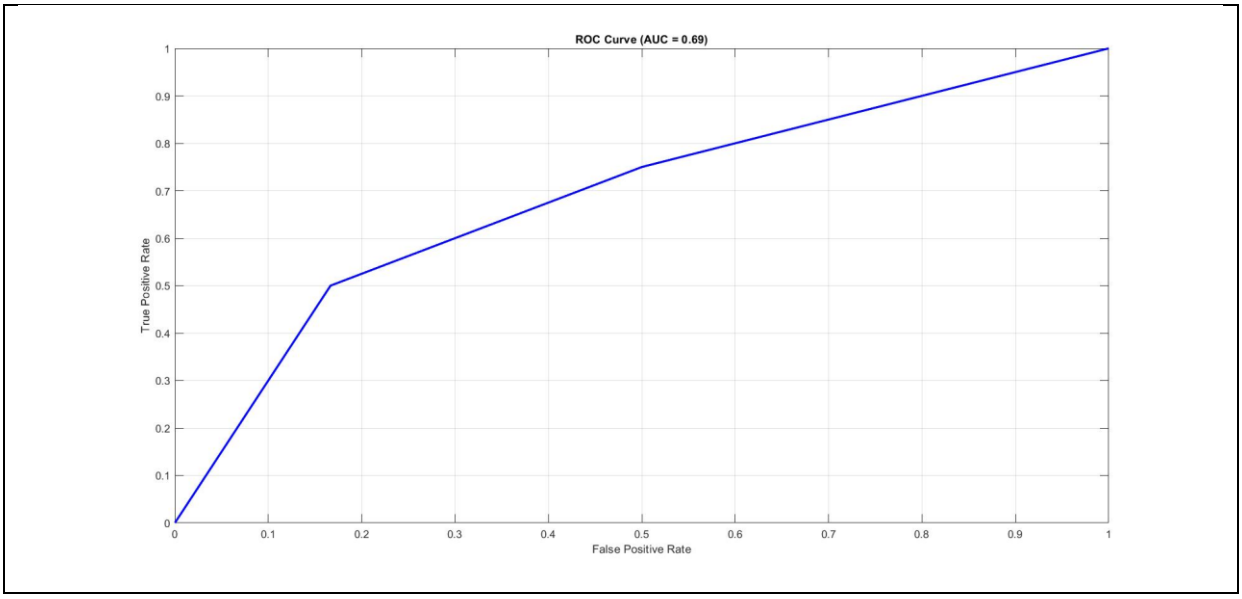


Table 2 : ROC Curve Analysis Summary

Model	AUC Value	Interpretation	Performance
SVM	**0.82**	Indicates strong classification ability with high separability between classes. Most predictions are correct.	Excellent
Decision Tree	**0.69**	Moderate performance. Better than random guessing but with some misclassifications.	Moderate
Random Forest	**0.65**	Also, moderate but slightly worse than Decision Tree. Could be improved with parameter tuning or more data.	Fair

6. Conclusion

This study evaluated the performance of three classification models—Support Vector Machine (SVM), Random Forest, and Decision Tree—using multiple metrics, including accuracy, cross-validation scores, confusion matrices, and ROC-AUC curves. Among these, the Support Vector Machine (SVM) model emerges as the best overall performer. It achieved the highest test accuracy of 70% and demonstrated strong class separability with an AUC of 0.82, indicating excellent classification ability. The SVM also maintained a good balance between precision and recall, making it a reliable model for general-purpose classification. The Random Forest model, while slightly lower in test accuracy at 65%, exhibited the highest cross-validation accuracy of 64.71%, suggesting it is the most consistent and stable across different data splits. It strikes a balance between identifying positives and negatives, making it a dependable second choice. On the other hand, the Decision Tree model recorded the highest recall of 0.75, meaning it effectively identified most positive cases, but it had the lowest test and cross-validation accuracy, pointing to a risk of overfitting. Its moderate AUC of 0.69 supports this interpretation.

Therefore, while SVM is ideal for overall performance, Random Forest is suitable when model consistency is essential, and Decision Tree is preferred in scenarios where capturing all positive cases is a priority.

References

1. Vadhiraj, S., et al. (2021**).** "Ultrasound Image Classification of Thyroid Nodules Using Machine Learning Techniques." *Medicina*, 57(6), 527. https://doi.org/10.3390/medicina57060527
2. Ma, J., et al. (2022**).** "Ultrasound Image Classification of Thyroid Nodules Based on Deep Learning." *Frontiers in Oncology*, 12, 933594. https://doi.org/10.3389/fonc.2022.933594
3. Zhang, J., et al. (2022**).** "Multistep Automated Data Labelling Procedure (MADLaP) for Thyroid Nodules on Ultrasound: An Artificial Intelligence Approach for Automating Image Annotation." *arXiv preprint arXiv:2206.14305*. https://arxiv.org/abs/2206.14305
4. Manh, V. T., et al. (2022**).** "Multi-Attribute Attention Network for Interpretable Diagnosis of Thyroid Nodules in Ultrasound Images." *arXiv preprint arXiv:2207.04219*. https://arxiv.org/abs/2207.04219
5. Sharifi, M., et al. (2022). "Thyroid Ultrasound Image Database." Data in Brief, 42, 108086. https://doi.org/10.1016/j.dib.2022.108086
6. Pedraza, L., et al. (2015). "Open Access Thyroid Ultrasound Image Database." Data in Brief, 5, 106-109. https://doi.org/10.1016/j.dib.2015.08.033
7. Sathesh and Rasitha K "A Nonlinear Adaptive Median Filtering Based Noise Removal Algorithm" Proceedings of First International Conference on Modeling, Control, Automation and Communication (ICMCAC–2010)
8. Jae-Chern Yoo and Chang Wook Ahn "Image restoration by blind-Wiener filter" Published in IET Image Processing. doi: 10.1049/iet-ipr.2013.0693
9. Arin H. Hamad, Hozheen O. Muhamad and Sardar P. Yaba "De-noising of medical images by using some filters" International Journal of Biotechnology Research Vol. 2(2), Available online http://academeresearchjournals.org/journal/ijbr,ISSN 2328-3505
10. Prashant Dwivedy , Anjali Potnis and Madhuram Mishra "Performance Assessment of Several Filters for Removing Salt and Pepper Noise, Gaussian Noise, Rayleigh Noise and Uniform Noise" International Journal of Engineering and Applied Computer Science (IJEACS) Volume: 02, Issue: 06, ISBN: 978-0-9957075-7-3.
11. Suman Shrestha "Image Denoising Using New Adaptive Based Median Filter" Signal & Image Processing: An International Journal (SIPIJ) Vol.5, No.4.
12. Kamta Nath Mishra, et al. "Enhancing cancer detection and prevention mechanisms using advanced machine learning approaches" information medicine unlocked 50 (2024).
13. Ya Guo, et al. "Machine learning and new insights for breast cancer diagnosis" in Journal of International Medical Research 2024, Vol. 52(4) 1–29
14. Yogesh Kumar, et al. "Automating cancer diagnosis using advanced deep learning techniques for multi-cancer image classification" in Scientific Reports | (2024) 14:25006 | https://doi.org/10.1038/s41598-024-75876-2 4