ISSN No:-2456-2165

Enabling Non-Invasive Diagnosis Of Thyroid Nodules With High Specificity And Sensitivity

Sai Maniveer Adapa[1]
B.Tech CSE-DS(AI)
Dr. M.G.R Educational and Research Institute, Tamilnadu, India

Sai Guptha Perla[2]
B.Tech CSE-DS(AI)
Dr. M.G.R Educational and Research Institute, Tamilnadu, India

Adithya Reddy. P[3]
B.Tech CSE-DS(AI)
Dr. M.G.R Educational and Research Institute, Tamilnadu, India

Abstract:- Thyroid nodules can often be diagnosed with ultrasound imaging, although differentiating between benign and malignant nodules can be challenging for medical professionals. This work suggests a novel approach to increase the precision of thyroid nodule identification by combining machine learning and deep learning. The new approach first extracts information from the ultrasound pictures using a deep learning method known as a convolutional autoencoder. A support vector machine, a type of machine learning model, is then trained using these features. With an accuracy of 92.52%, the support vector machine can differentiate between benign and malignant nodules. This innovative technique may decrease the need for pointless biopsies and increase the accuracy of thyroid nodule detection.

Keywords:- *Thyroid Tumor Diagnosis, Ultrasound Images, Deep Learning, Machine Learning, Convolutional Auto-Encoder, Support Vector Machine.*

I. INTRODUCTION

Computer Vision started with ambitious expectations in the 1960s, but early attempts at AI-powered object recognition proved challenging due to the complexity of visual perception. Researchers then turned to studying biological visual systems, leading to the influential Neocognitron model in the 1980s. However, its lack of efficient training methods left it neglected.

A breakthrough came in the late 1980s with the back-propagation algorithm and LeCun's successful implementation of CNNs. Despite its reduced complexity and efficient training compared to Neocognitron, CNNs remained limited by computing power and the vanishing gradient problem. As a result, CNNs remained silent for a decade and kernel machines with hand-crafted visual descriptors became the dominant approach for object recognition.

The rise of GPUs and deep learning in the 2010s revolutionized the field. In 2012, Krizhevsky et al. shocked the world by winning the ImageNet challenge with a deep CNN, reigniting interest in this approach. Since then, deep learning has made remarkable progress in image recognition, surpassing human performance on tasks like object recognition.

Recent years have witnessed significant advancements in medical image recognition using deep learning, revolutionizing areas like anomaly detection, diagnosis, segmentation, and even brain decoding. Within this domain, diagnosing tumors based on medical images stands as a crucial task with immense potential for clinical impact. This paper delves into the application of deep learning for diagnosing thyroid tumors using ultrasound images, a common challenge faced by radiologists.

Thyroid nodules, often presenting as benign growths, can occasionally be cancerous. Differentiating between these types holds critical importance for patient management. Traditionally, radiologists rely on visual assessment of ultrasound images, employing the Thyroid Imaging Reporting and Data Systems (TIRADS) for characterization. But this procedure is sensitive to interpretation and prone to human mistake.

This paper explores the potential of deep learning to overcome these limitations and pave the way for more accurate and objective thyroid tumor diagnosis. We delve into the challenges faced in this specific domain and present our proposed deep learning-based approach, outlining the methodology, datasets, and experimental results. Ultimately, we aim to demonstrate how deep learning can contribute to improved clinical outcomes by aiding in earlier detection and more precise diagnosis of potentially malignant thyroid tumors.

ISSN No:-2456-2165

II. CHALLENGES IN MEDICAL AI

Medical image recognition presents distinct challenges for deep learning, demanding careful consideration of risk and complexity. Diagnoses based on probabilities, unlike "Yes/No" object recognition, necessitate weighing the costs and consequences of both false positives (unnecessary biopsies) and false negatives (missed malignancies). This study tackled this dilemma by adjusting the model's focus towards identifying true positives without overemphasizing them, and by collaborating with radiologists to define a scoring function that balances clinical and economic factors.

Another hurdle relates to the uncertainties surrounding performance limits in medical diagnoses. Unlike tasks like ImageNet where human experts excel, diagnosing malignancy from images is subject to limitations due to the intricate relationship between tumor genotype and phenotype. This study addressed this by prioritizing human-level accuracy as a benchmark, which the final model successfully surpassed, and by employing complex models capable of capturing the inherent complexities of medical diagnosis.

However, such complexity comes with a data bottleneck. Medical datasets are often limited by cost and privacy concerns, typically containing only hundreds of samples. The study countered this challenge by carefully designing the model architecture for efficient data utilization and by leveraging data augmentation techniques to artificially increase the data size while preserving its integrity.

These inherent challenges in medical image recognition necessitate tailored methodologies to overcome them effectively. This study's approach demonstrates how deep learning can be adapted to address these concerns and contribute to more accurate and nuanced medical diagnoses.

While Convolutional Neural Networks (CNNs) have powered incredible advances in image recognition, their application in medical image diagnosis faces unique hurdles. This is because medical images, unlike pictures of faces or cats, lack the consistent spatial organization of features that CNNs rely on. Tumors can appear in diverse shapes, sizes, and arrangements, making it difficult for a CNN to generalize its understanding of "benign" or "malignant" from training data to new cases.

Further complicating matters, tumors with similar appearances can have vastly different clinical implications. This means simply "remembering" what a benign tumor looks like won't guarantee accurate diagnosis. Additionally, medical data often suffers from class imbalance, with far fewer examples of the disease class (e.g., malignant tumors). This can bias model training, further jeopardizing accuracy.

This study tackles these challenges by approaching CNNs as feature extractors instead of end-to-end predictors. The model essentially "learns" what constitutes ultrasound tumor images by analyzing local patterns. These learned features are then fed into other classifiers designed specifically for the complexities of medical diagnosis. This two-step approach allows the model to capture the diverse and nuanced nature of medical images without being limited by rigid spatial assumptions.

Furthermore, the study addresses data imbalance through algorithm manipulation, such as placing additional weight on the minority class (malignant tumors) during training and developing a customized scoring function for model selection. This ensures the model prioritizes accurate identification of the less frequent yet critical cases.

III. DATASET

This study analyzes a dataset of 3,183 ultrasound images of thyroid nodules collected over five years. Trained radiologists evaluated each image and extracted the relevant areas for further analysis. A key element is the Bethesda Grade system, which categorizes the malignancy risk of these nodules from 1 (outlier) to 6 (highest risk).

For this study, outlier images (grade 1) were excluded. The remaining images were grouped into two categories: benign (grades 2 and excluded 1) and suspicious (grades 3 to 6). This resulted in most benign nodules (77%) and a smaller, but crucial, group of suspicious nodules requiring further investigation (23%). This distribution reflects the real-world prevalence of thyroid nodules, where most are benign but a significant portion warrants additional attention due to potential malignancy risk.

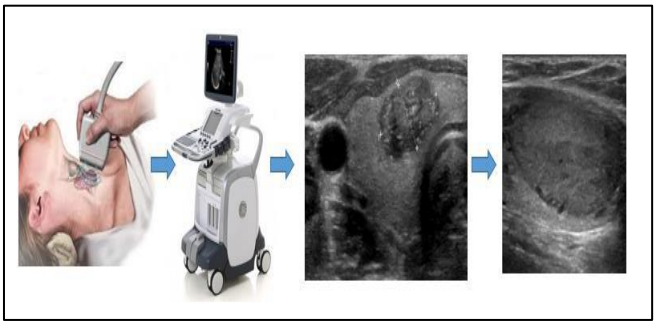

Fig 1 Thyroid Nodule Image Data Acquisition

This study incorporates not only the ultrasound images themselves but also detailed characterizations of 1434 of those images by comptent radiologists. Using the Thyroid Imaging Reporting and Data Systems (TIRADS) framework, radiologists evaluated features like composition, echogenicity, shape, and margin, assigning scores based on a pre-defined system. These scores indicate the likelihood of malignancy and guide recommendations for follow-up or biopsy.

ISSN No:-2456-2165

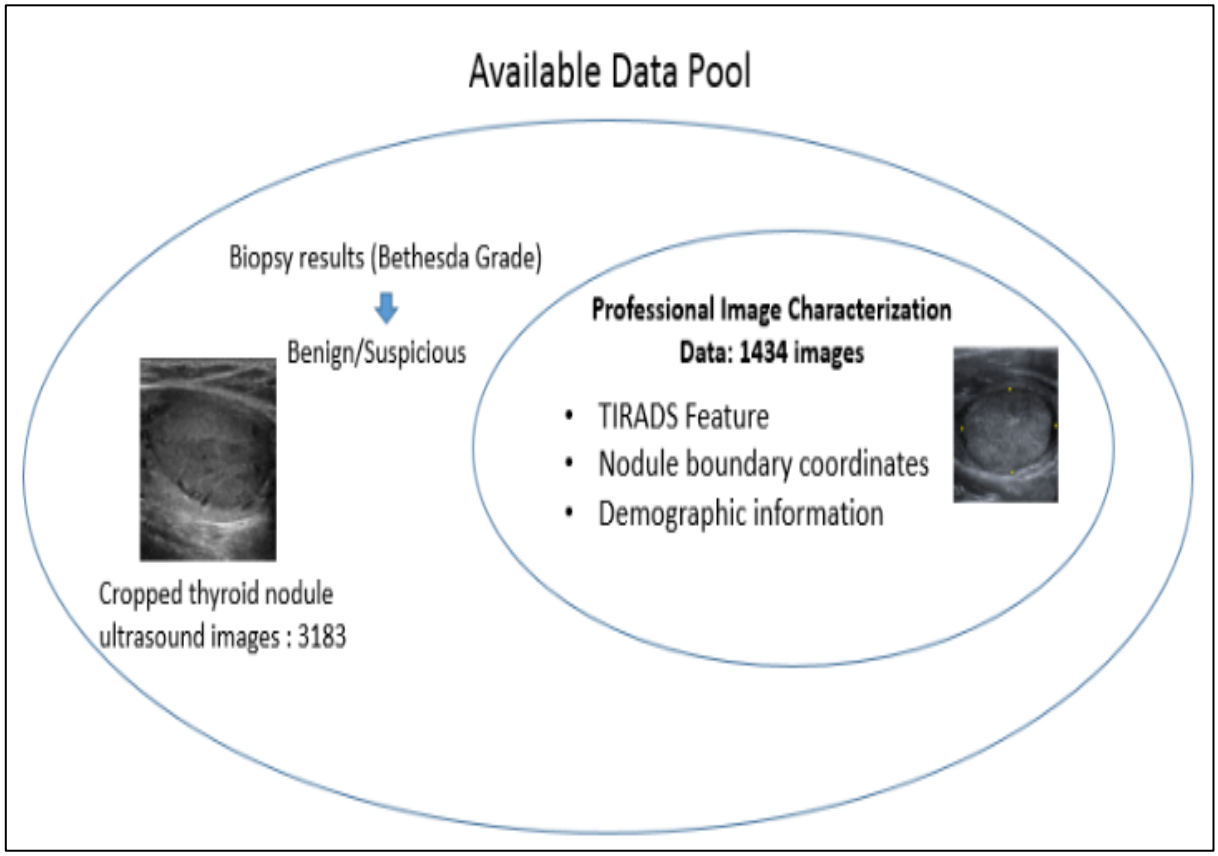

Fig 2 Summary of Available Data for the Study

Interestingly, the characterized nodules in this study were predominantly benign (78.17%), reflecting the real-world prevalence. Additionally, the patients were mostly middle-aged to elderly females (77.6%). Figure provides a concise overview of the entire dataset using a Venn diagram.

IV. METHODOLOGY

This study takes a multi-pronged approach to analyzing thyroid ultrasound images and diagnosing potential malignancy. Starting with a cropped image of the nodule, the study extracts information using both computer vision techniques and radiologists' assessments. Local Binary Pattern (LBP) and Histogram of Oriented Gradients (HOG) capture texture and shape details, while a Convolutional Auto-Encoder (CAE) uncovers hidden patterns. These features are then combined with TIRADS scores (based on radiologists' evaluation) and patient demographics to create a comprehensive profile of each nodule. Finally, sophisticated machine learning techniques identify the optimal way to analyze and interpret these features, ultimately leading to accurate diagnoses of benign or suspicious nodules. This multi-layered approach leverages the strengths of both computer vision and human expertise to provide a robust and reliable basis for thyroid nodule diagnosis.

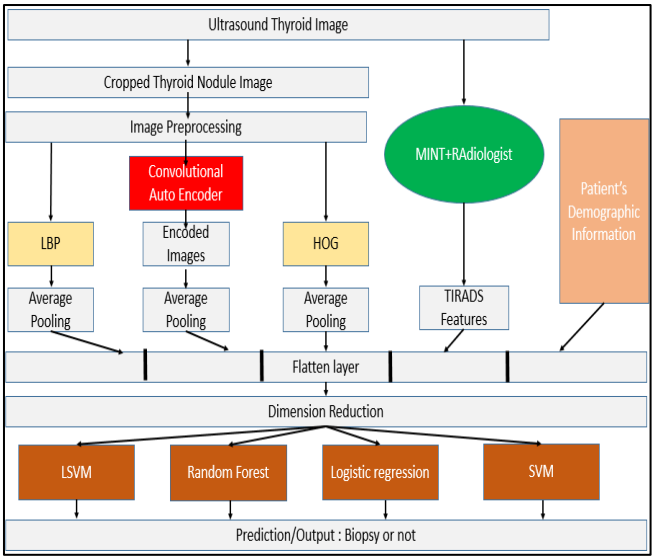

Fig 3 Methodology Workflow in this Study

➢ *Image Preprocessing*

After the nodule image has been cropped, preprocessing is needed to align the nodule center in each image and normalize the sizes of the photos. Initially, the border coordinates are used to calculate each nodule's center. Subsequently, the centers of every nodule picture are positioned at the same location.

ISSN No:-2456-2165

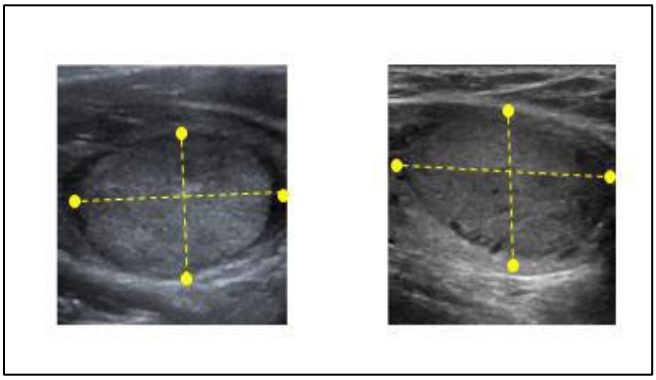

Fig 4 The Geographic Center is the Intersection of Two Boundary Lines

Following alignment, all picture sizes are standardized to 224 x 224 by stretching relative to each center. The stretch ratios, both horizontal and vertical, are noted during the stretch to reconstruct the image afterwards. To make CAE teaching easier, the grey value pixel is finally rescaled between 0 and 1.

- *Convolutional Auto-Encoder*

Convolutional Neural Networks (CNN) and Auto-encoders (AE) are the two concepts that combine to form the notion of Convolutional Auto-encoders (CAE). Similar to an AE, a CAE's main concept is to "compress" two-dimensional images over time into a representation that is smaller in size. We refer to this as the encoding procedure. The next step, known as the decoding procedure, is for CAE to extract the encoded image back to its original form. Making sure the rebuilt image resembles the original as much as possible is the aim of this technique.

The encoding and decoding processes used by CAE can accomplish two objectives. Firstly, it may successfully remove high frequency disturbances from the input by acting as a dimensionality reduction technique.

Second, and maybe more significantly, CAE can extract highly helpful patterns from the image by discovering an effective method of compression and extraction. As a result, CAE is a useful technique for extracting picture features in addition to dimension reduction.

- *Local Binary Patterns & Histogram Of Oriented Gradients*

In computer vision, Local Binary Patterns (LBP) are a straightforward, well-liked, and effective visual descriptor. It is mostly employed to characterize an image's texture. An LBP descriptor's input and output often have the same dimension. The LBP code of the original image is represented by the output of every pixel.

The purpose of the Histogram of Oriented Gradients (HOG) is to depict the form and appearance of a nearby object. The histogram of local intensity gradients can be used to illustrate the image's directional pattern. A centered filtered mask calculates the gradients' orientation and amplitude given the input image. Certain texture and contour information can be captured by the magnitude and gradients. The gradients are then quantized into a local 1-D histogram of gradients over all the pixels in each of the tiny spatial cells that make up the image window. The gradient angle range is divided into a predetermined number of bins by the histogram, and the orientation histogram is determined by the gradient's magnitude.

In addition to directional patterns outside of the borders, it is capturing the form of the nodule. Later machine learning can make use of the histogram count for each direction in a pooling layer as a feature.

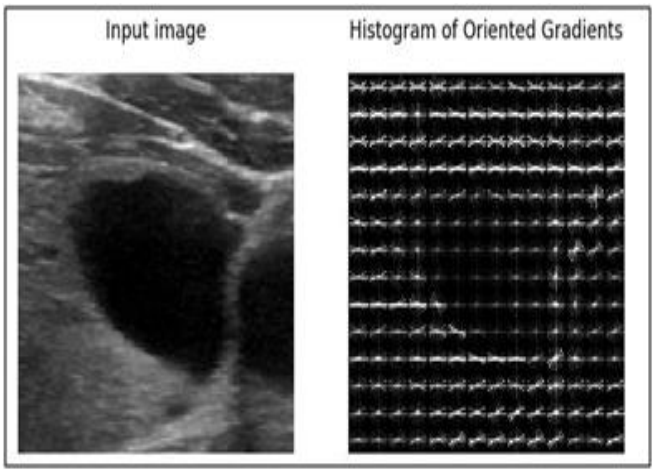

Fig 5 Histogram of Oriented Gradients of Nodule Image

V. RESULTS AND DISCUSSION

➢ *The Series of Visuals Presented Below Showcase the Sequential Development of our Project.*

- *The Main Artifact:*

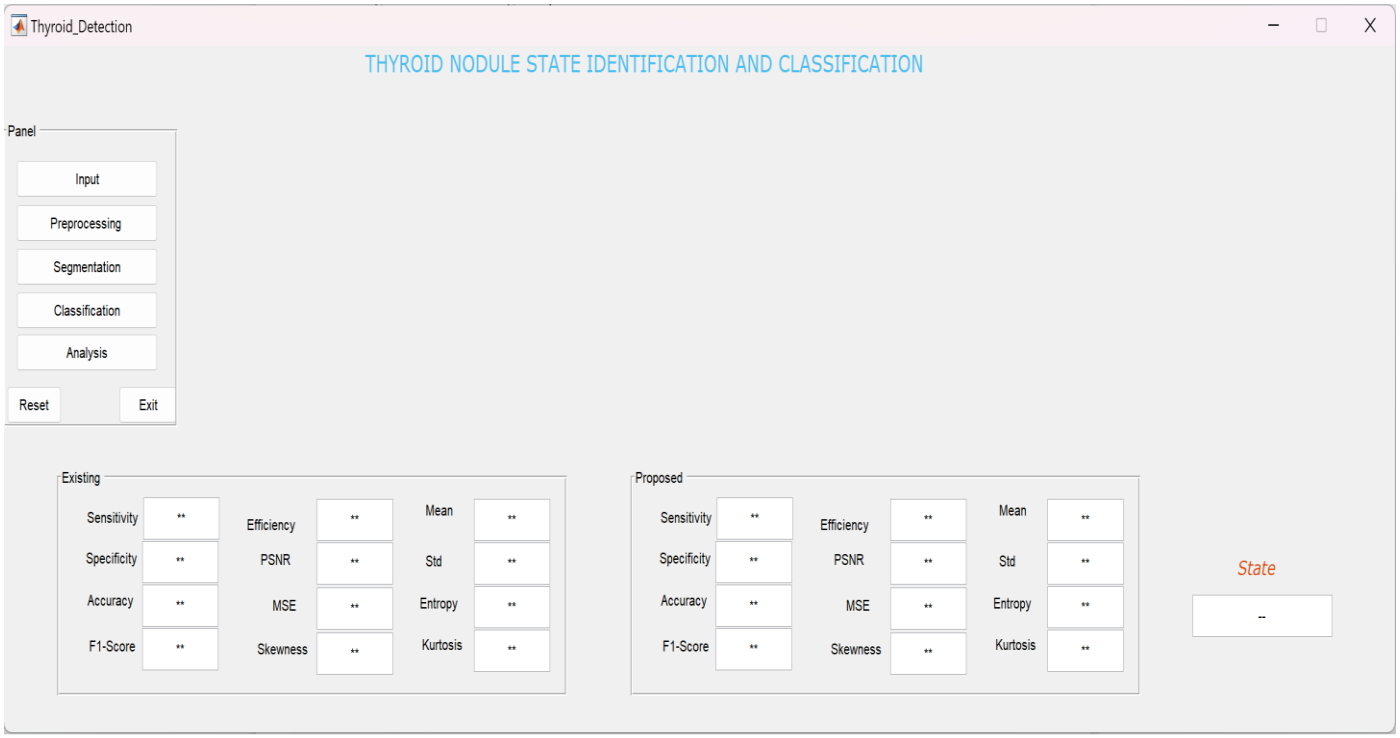

Fig 6 The Main Artifact

- *Input the Data:*

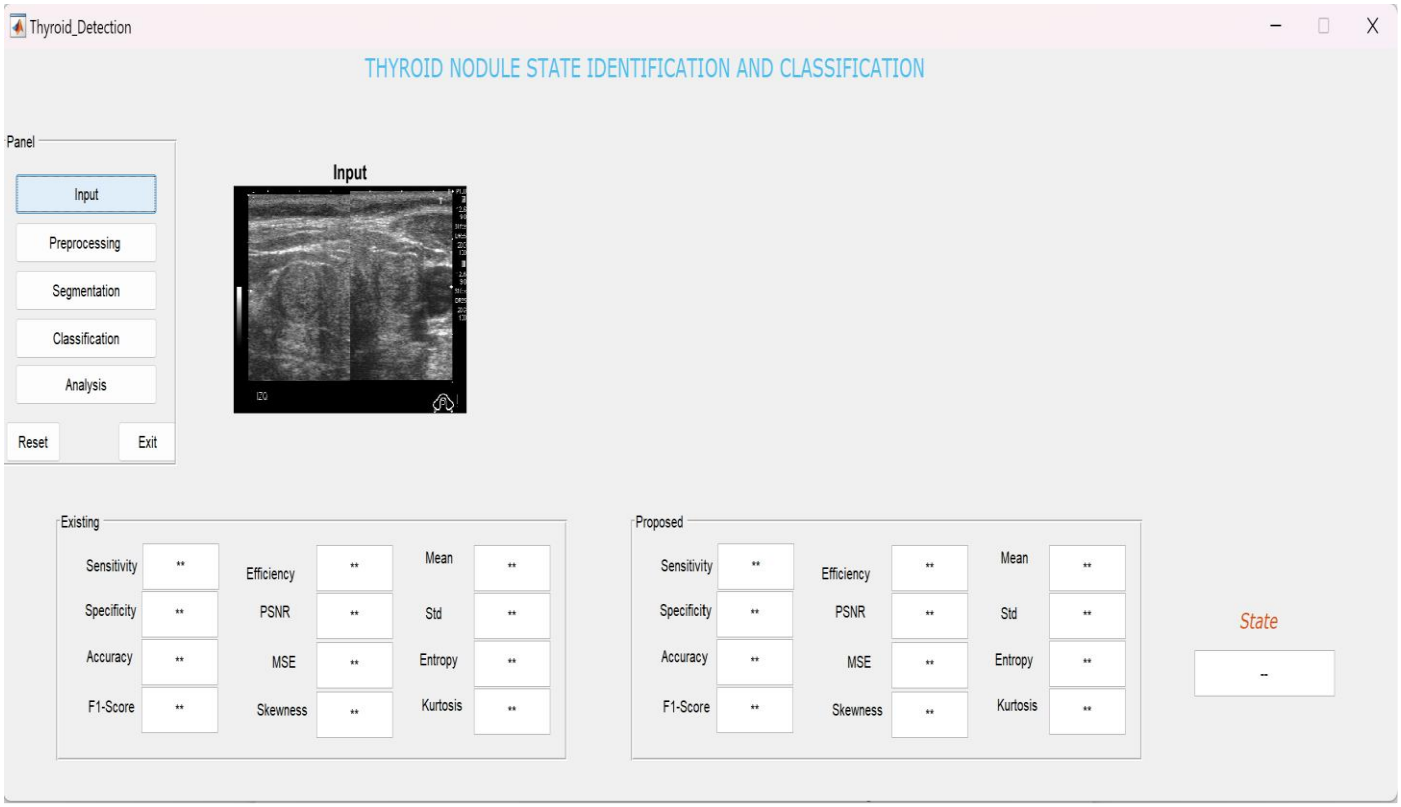

Fig 7 Input the Data

- *Preprocessing:*

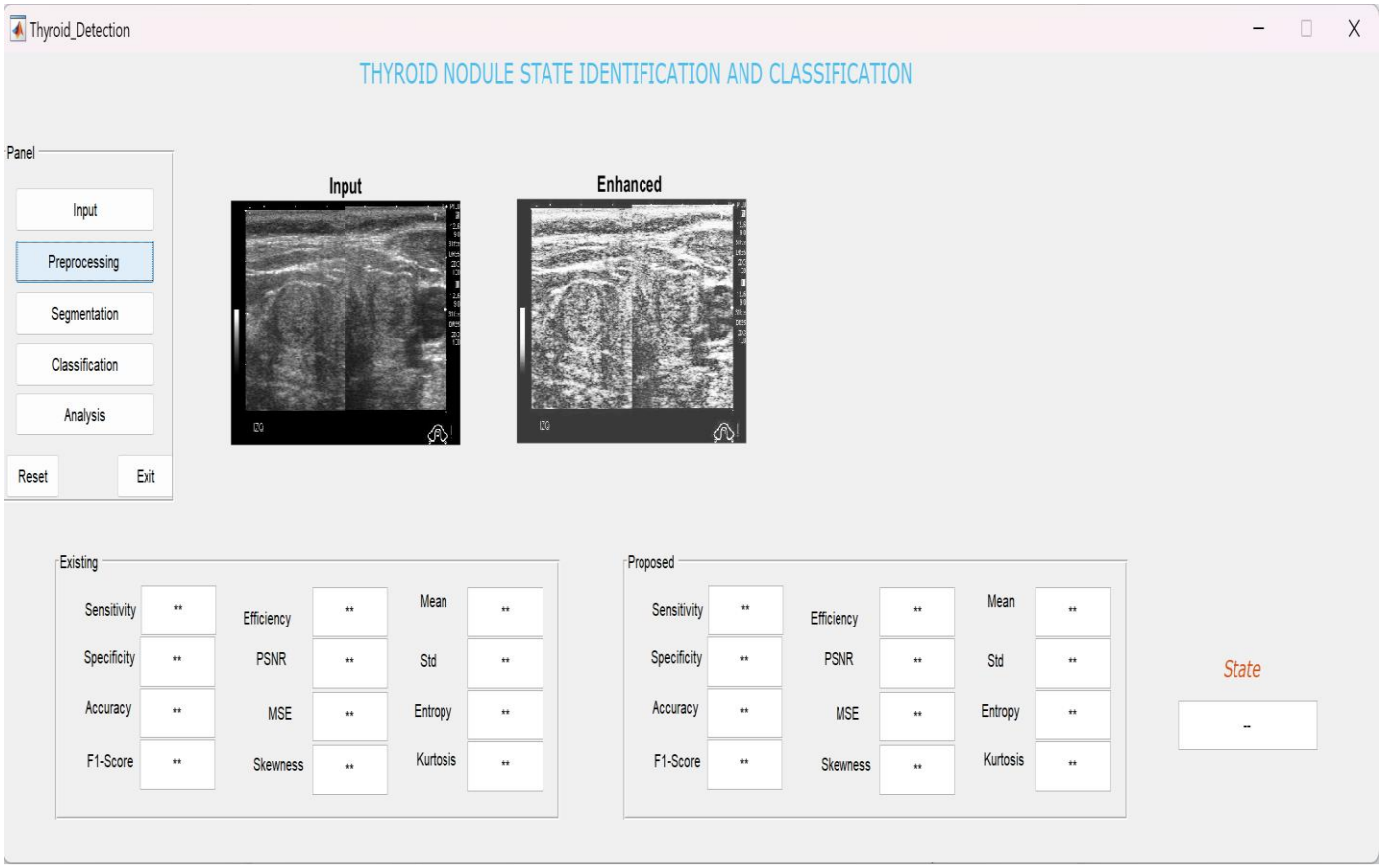

Fig 8 Preprocessing

- *Segmentation:*

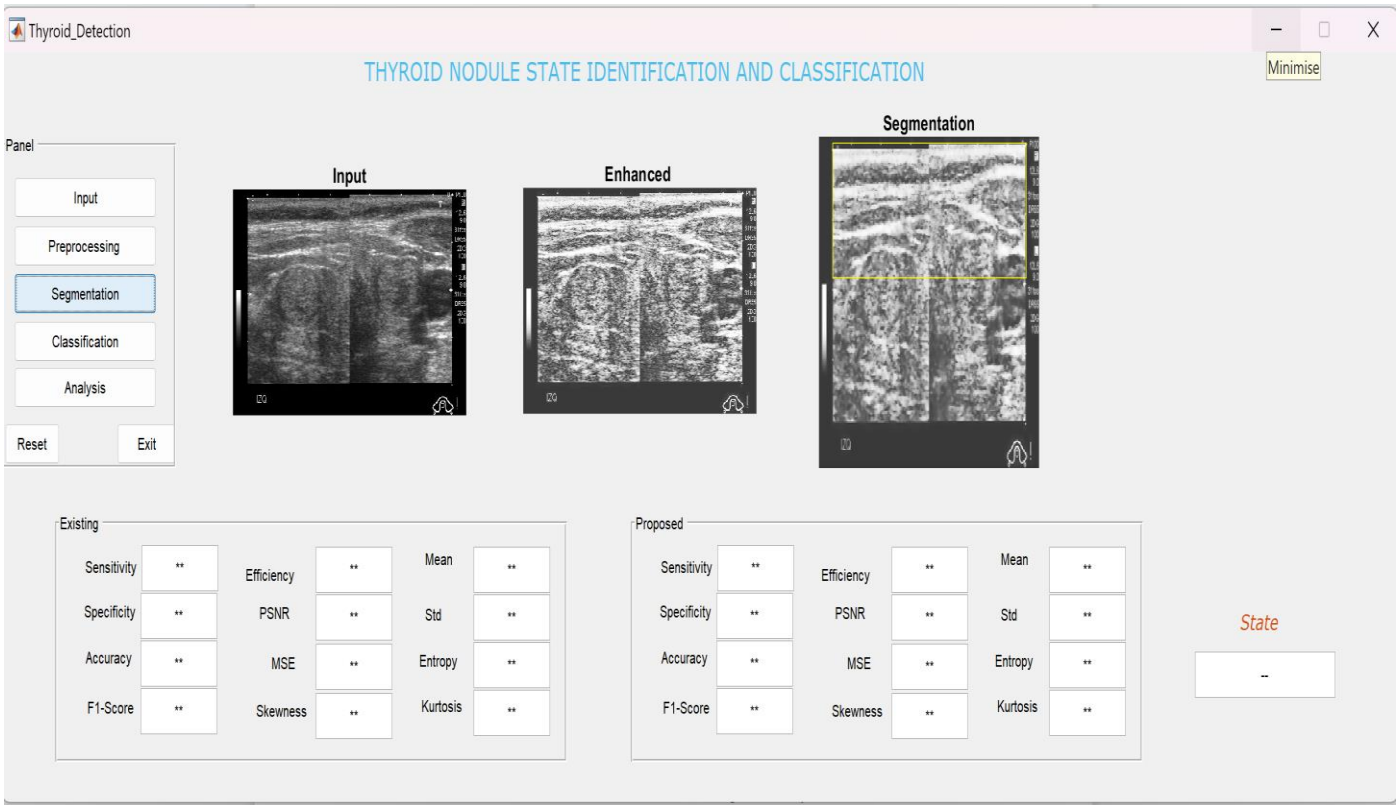

Fig 9 Segmentation

- *Classification:*

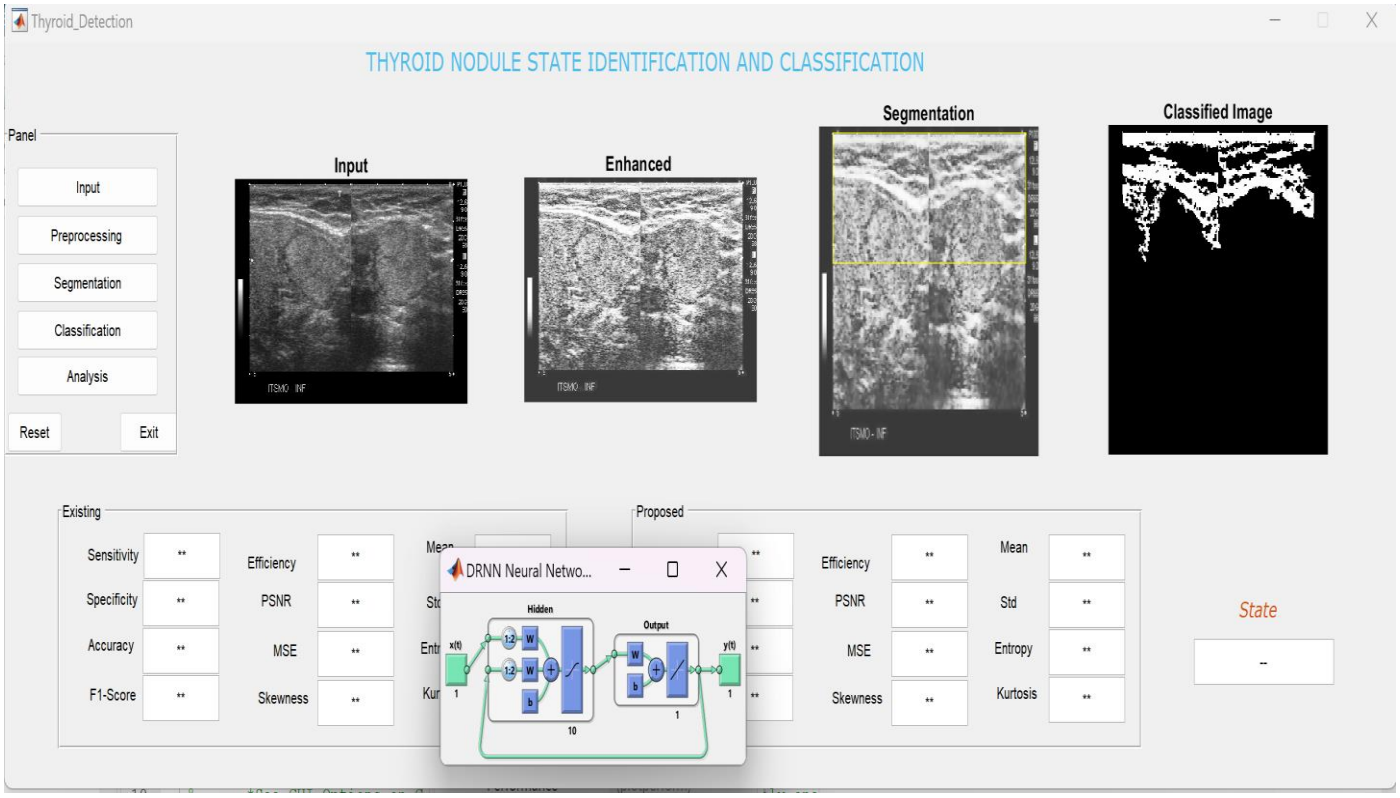

Fig 10 Classification

- *Final Output & Result:*

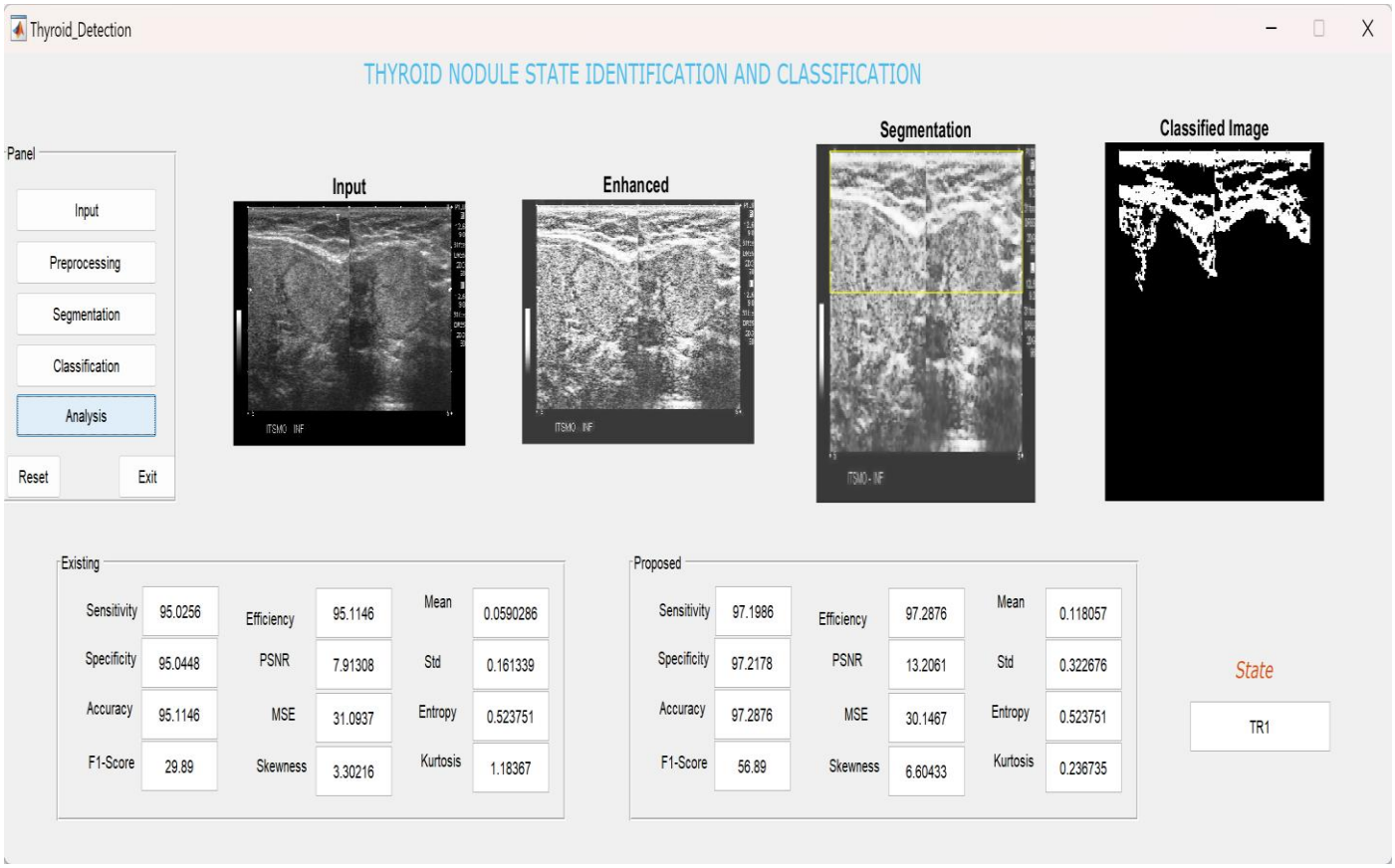

Fig 11 Final Output & Result

- *Comparison Table*

Table.1 Test Performance for Each Algorithm

Classifier	Performance Metrics			Final Score
	Accuracy	Precision	Sensitivity	
Random Forest	0.666	0.323	0.187	2.15
Logistic Regression	0.575	0.316	0.458	3.33
Linear SVM	0.544	0.305	0.505	3.50
RBF SVM	0.417	0.308	0.888	**5.19**
Human Performance	0.262	0.254	0.860	4.80

The highest accuracy is produced by the Random Forest. However, due to the imbalance in the dataset, accuracy may be deceiving as it favors the dominant class over the minor class in most cases. This results in extremely low sensitivity and a large false negative rate. Therefore, according to our scoring formula, the random forest produces the lowest total score.

Despite the fact that Linear SVM has superior recall and Logistic Regression has somewhat more accuracy, their performances are extremely equal. Although both of these models are linear, their performance suggests that, even with the optimal preprocessing plan, the true decision boundary is substantially non-linear.

Surprisingly, the highest score is produced by Support Vector Machines (SVM) using radial basis function kernels (RBF)—even higher than human performance. Its accuracy may not be the best, but it can decrease false negatives the most, leading to the most sensitivity and, hence, the highest score.

It is not surprising that human performance has the highest sensitivity and the lowest accuracy. For the majority of thyroid nodules, radiologists typically advise biopsy in order to minimize false negatives. When comparing the SVM model with human performance, we find that the SVM model with the RBF kernel has superior precision and sensitivity. This indicates that the SVM model can decrease the number of malignancy cases that are missed as well as the number of unnecessary biopsy cases.

VI. CONCLUSION

This study explored how machine learning can assist radiologists in diagnosing thyroid cancer through ultrasound images. It presented a new approach that overcomes data limitations and unlocks hidden patterns within the images to improve diagnosis accuracy. While traditional CNNs falter due to scarce data and lack of consistent spatial features in thyroid nodules, this study proposes several key tactics: medical image augmentation to artificially expand the data, convolutional auto-encoders to extract hidden patterns, and a multi-classifier approach to leverage the strengths of different algorithms. Remarkably, the best-performing model even surpassed human expert accuracy on the test set, showcasing the immense potential of machine learning to revolutionize cancer diagnosis and lead to better patient care.

REFERENCES

[1]. Richard S (30 September 2010). Computer Vision: Algorithms and Applications. Springer Science & Business Media. pp. 10–16. ISBN 978-1-84882-935-0.
[2]. Papert, S (1966-07-01). "The Summer Vision Project". MIT AI Memos (1959 - 2004). Retrieved 2 August 2016.
[3]. Margaret AB (2006). Mind as Machine: A History of Cognitive Science. Clarendon Press. p. 781. ISBN 978-0-19-954316-8.
[4]. Fukushima,K. (1980), Neocognitron: A self-organizing neural network model for a mechanism pattern recognition unaffected by shift in position. Biological Cybernetics, 36(4),193-202
[5]. Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning representations by back-propagating errors. Nature, 323, 533--536.
[6]. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2323. DOI: 10.1109/5.726791
[7]. Cortes, C., Vapnik, V. (1995). "Support-vector networks". Machine Learning. 20 (3): 273–297
[8]. T. Ojala, M. Pietikainen, and T. Maenpa (2002). "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." IEEE Transactions on pattern analysis and machine intelligence, vol 24, pp: 971-987.
[9]. N. Dalal, and B. Triggs (2005). “Histograms of oriented gradients for human detection." Computer Vision and Pattern Recognition. CVPR 2005. IEEE Computer Society Conference on. vol. 1. IEEE, 2005
[10]. Robert M Haralick; K Shanmugam; Its'hak Dinstein (1973). "Textural Features for Image Classification". IEEE Transactions on Systems, Man, and Cybernetics. SMC-3 (6): 610–621.
[11]. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M.S., Berg, A.C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision, 115, 211-252.
[12]. Alex K & Ilya S & E. Hinton G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25.. 10.1145/3065386.
[13]. Hu, J & Shen L & Sun G. (2017). Squeeze-and-Excitation Networks. arXiv preprint arXiv:1709.01507

[14]. Ferrucci D, Brown E, Chu-Carroll J et al (2010). Building Watson: An overview of the DeepQA project. AI magazine;31(3):59-79
[15]. IBM Watson for Oncology. IBM. http://www.ibm.com/smarterplanet/us/en/ibmwatson/watson-oncology.html, 2016. Last Accessed on May 2018
[16]. DeepMind Health. Google DeepMind. https://www.deepmind.com/health, 2016. Last Accessed on May 2018
[17]. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA.;316(22):2402–2410. doi:10.1001/jama.2016.17216