From Pixels to Prognosis: A Review of Image Analysis Models for Thyroid Nodule Risk Stratification

Parul Chauhan, Anudeep Reddy Beeravolu

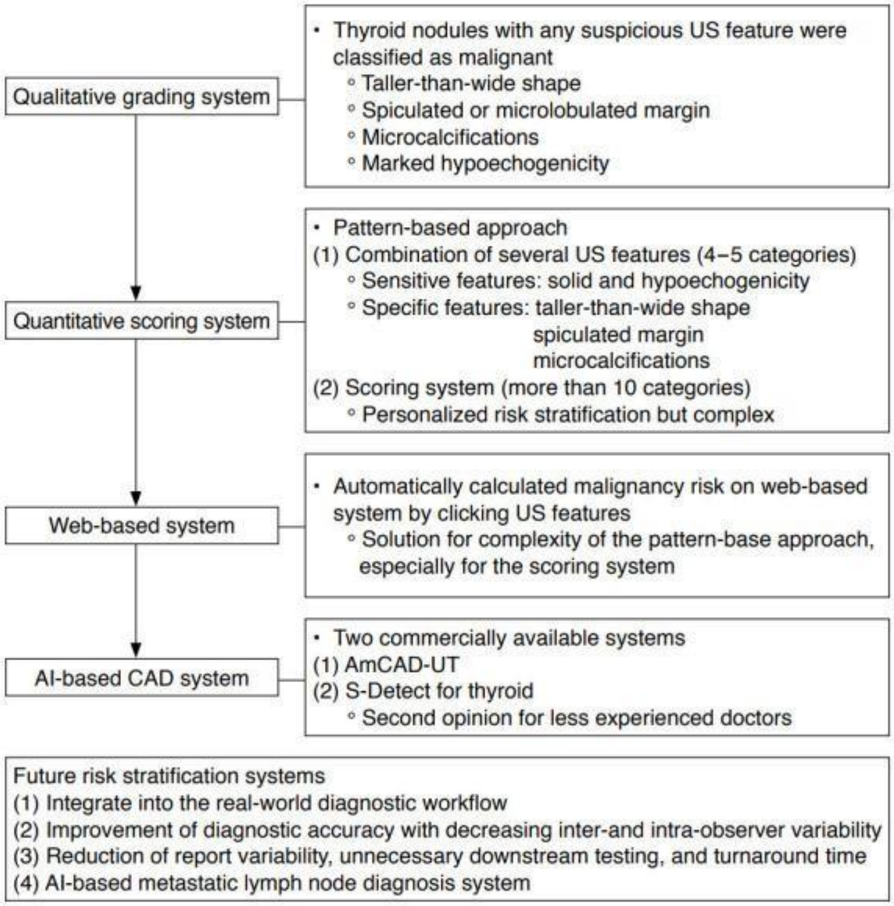
1. Abstract

In the past few years, there have been many advancements in image analysis of thyroid nodules in diagnosing benign and malignant thyroid nodules. Accurate diagnosis in the right time frame is the key to efficient treatment of the patients and proper recovery. Artificial intelligence has become an integral part of diagnosis. This paper explores the machine learning and deep learning models currently employed in Image Analysis for Thyroid Nodule Risk Stratification. We will delve into the efficiency of these models, identify areas for improvement, their use in clinical setup, and discuss potential future developments in the domain. Reviewing the state-of-the-art in thyroid nodule evaluation goes beyond simply explaining it; it also emphasizes the importance of these developments and explores their broader implications in diagnosis accuracy and, eventually, improving patient outcomes.

2. Introduction

Several factors go into determining if a thyroid nodule is a high risk. During image analysis, many features for risk stratification are taken into evaluation. The most benevolent quality is determining if the size is ≥10mm. Even with a nodule belonging to <10mm in diameter, certain factors such as the presence of microcalcifications, irregular margins, anteroposterior to transverse diameter ratio (AP/T ratio), hypo echogenicity, increased vascularity (especially chaotic vascularity) are considered. A taller-than-wide shape is defined as an AP/T ratio greater than one associated with a higher risk of malignancy. Microcalcifications are found in benign and malignant nodules, but their presence in more significant amounts is considered high risk. Among rims, punctuate, and coarse calcifications, punctuate is associated with the most suspicion.

The American College of Radiology developed the Thyroid Imaging Reporting and Data System (TI-RADS) [5]. It helps to report and stratify nodules into risk categories by interpreting ultrasound Images, which gives a clear picture of the malignancy of the nodule in the patient. TI-RADS score is essential in stratifying thyroid nodules and determining the risk factor of the nodule that might be detected. It is considered an effective system to reduce the use of FNP biopsies and helps improve the accurate diagnosis of thyroid cancer. TI-RADS is deemed very reliable in the medical field and has enhanced the quality of diagnosis and patient care.



1.Evolution of risk stratification systems.
(n.d.).https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7758100/.

In several cases, FNA is inevitable as it is the clearest indicator of a high-risk nodule. CAD systems improve the diagnostic capability with image analysis, bypassing several limitations and leading to fewer FNA procedures. To eliminate avoidable FNA procedures, researchers deployed Machine learning models into the realm of image analysis. Feature extraction is a standard practice with ML algorithms. Several algorithms, including random forest, support vector machines, and regression, are being used to differentiate between malignant and benign nodules. With the rise of deep learning methodology in technical fields, convolutional neural networks are emerging in thyroid nodule risk stratification. CNN's ability to capture patterns and relationships within data promises accurate risk stratification within ultrasound image models. We also briefly introduce transfer learning and ensemble learning methodologies. The diverse Machine-learning and deep learning techniques are further discussed and highlighted in the paper.

3. Methods

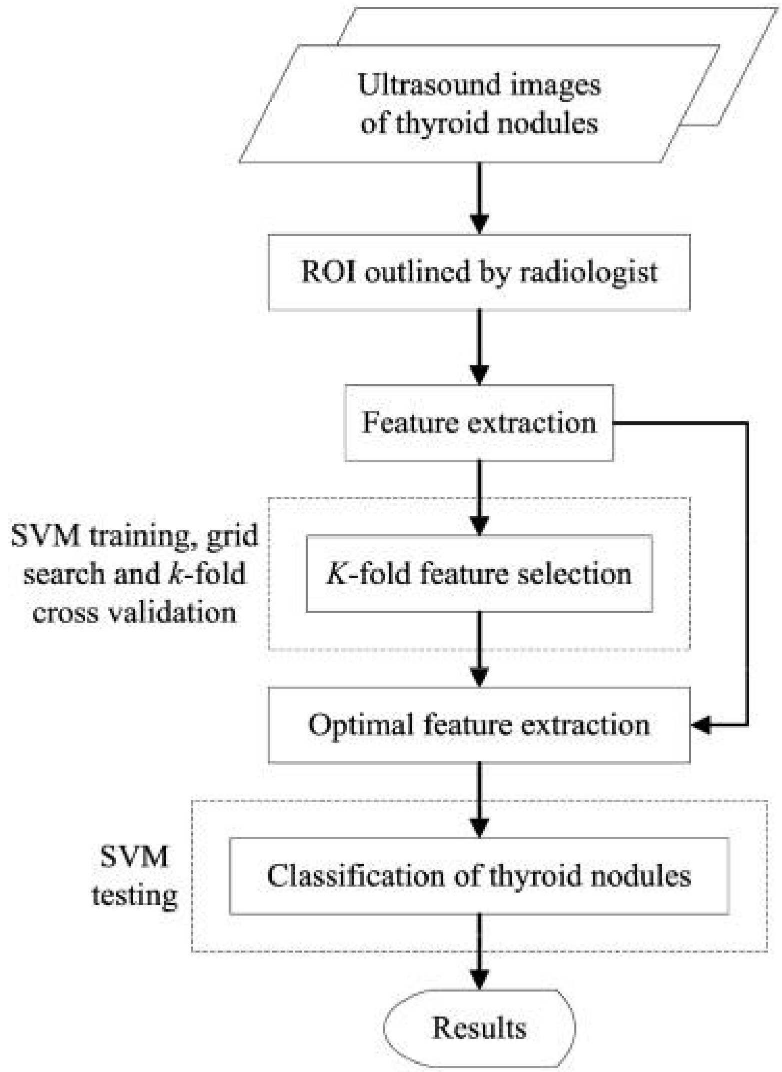
Using Fine Needle Procedure (FNP) to diagnose thyroid nodules is expensive and more complicated, so ultrasound imaging is preferable. Thyroid ultrasonography is a non-evasive, widely used procedure for detecting malignant nodules. Ultrasonography is applicable but also has drawbacks. These limitations include visual image artifacts, making it harder to assess the nodule on the image. A low signal-to-noise ratio is a contributing factor in switching methods for diagnosis. Diagnostic accuracy using ultrasound images highly depends on the level of expertise as they are only sometimes reproducible. Computer-aided systems are introduced for thyroid risk stratification to mitigate complete dependency on medical practitioners. CAD systems help in minimizing human oversight in image analysis. Additionally, CAD systems are promising in the efficient characterization of thyroid nodules.

Before we proceed and delve into various algorithms deployed in clinical practice, a comprehensive understanding of how the algorithms differentiate in thyroid nodules is essential. Different algorithms have their unique set of operations. There are specific common steps involved in the deployment by every algorithm, which gives a broader view of the factors for the efficient working of the model.

First, the objective is defined, and the required data is collected. Then, based on the data set, we choose an appropriate AI model, train it on a data set, and then go for evaluation to determine the proficiency. The model's viability for clinical workflow is decided based on performance. Once deployed, the AI model must undergo constant training to continuously learn to stay updated with everyday changes in thyroid risk stratification. Every model is reliable, and reliability is subjective to various factors, but constant monitoring and timely updation are imperative in addressing any ethical considerations.

3.1 Support Vector Machines (SVM):

SVM-based CAD systems have been one of the pioneers in thyroid nodule risk stratification in the medical field. The first commercialized thyroid using the US system utilized a support vector machine model. One of the most dominant systems in the market, TI-RADS (Thyroid Imaging and Reporting Data System) [5], is a standardized reporting system used in thyroid risk stratification and employs SVM. A study published in JCEM (Journal of Clinical Endocrinology & Metabolism) 2022 found using SVM in TI-RADS reduces the number of unnecessary biopsies by 20% [6]. SVM's feature extraction enables TI-RADS to output risk scores for malignancy. In addition to developing scores for TI-RADS, SVM is the algorithm that classifies the nodules as benign or malignant.



2.SVM risk stratification model.
(n.d.).https://www.sciencedirect.com/science/article/pii/S0031320310001998.

A 2023 review stated SVM is the algorithm used for 28% of research studies [4]. The ability of SVM models to work with high-dimensional data in thyroid nodule classification makes it a standard choice. Support vector machines are popular due to their flexible kernel function and threshold. It is non-linear and non-parametric, not making strong assumptions about the underlying distribution of data.SVM models are robust in handling the imbalance in data distribution across thyroid risk stratification levels. SVM models are also considered practical based on their functionality to work well with unseen data.

3.2 Deep Learning (DL):

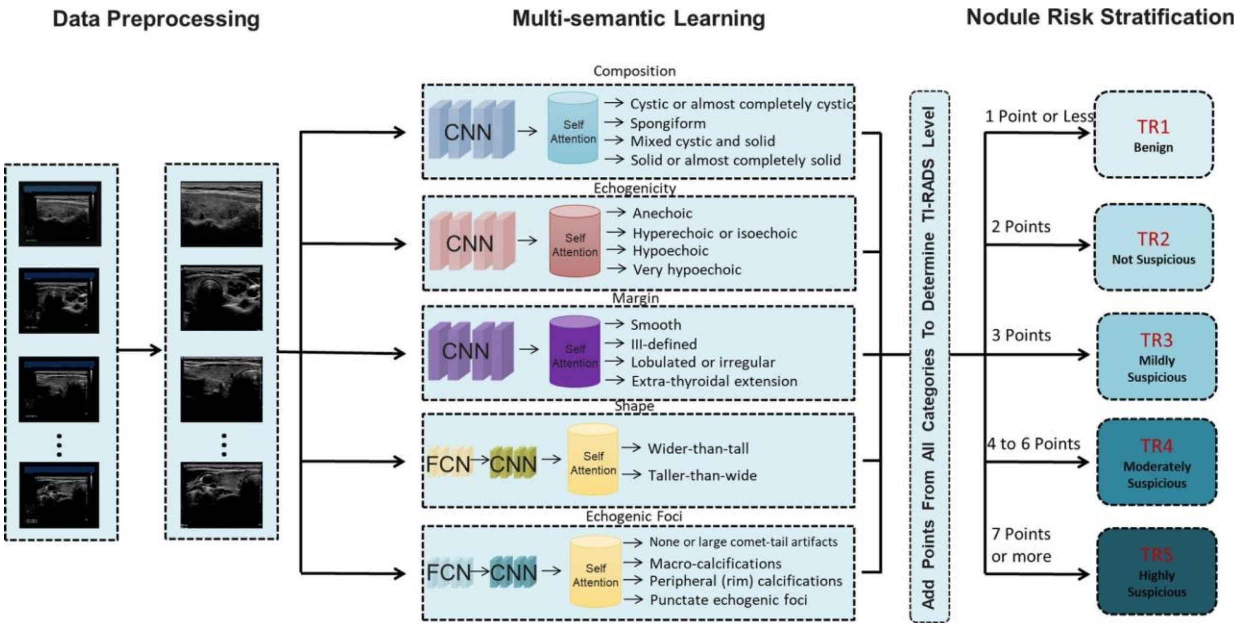
Several developments in image analysis using Deep learning methodology led to its significant deployment in thyroid risk stratification. Deep Learning algorithms are practical in pattern recognition, image classification, and detection. They also have a considerable advantage over Machine Learning algorithms as they train well on unseen medical data. DL algorithms automate feature extraction for the classification of ultrasound images. Human expertise is optional while training the models. However, DL-based CAD systems come with a particular workload. The deeper the network goes, the more complex it gets. Sub-optimal solutions arise, leading to a decrease in performance. The implementation of DL algorithms in real-world models is relatively new. The ability to train varies proportionally in learning the intricacies of its functioning, which is an evolving field.

The DCNN (Deep Conventional Neural network) is an artificial neural network that mimics the network present in the human visual cortex [12]. DCNN models exist in other medical domains but not evidently in thyroid nodule risk imaging. Transfer learning is an alternate solution to this drawback. Deep learning networks are proficient with small data sets and provide accurate results. Since it struggles with large datasets and is prone to overfitting, it is difficult to use in certain medical domains. The reliance on such systems by medical experts remains limited. Similarly, we are studying ResNet (Residual network) [7] models to minimize the drawbacks of deeper networks.ResNet model is employed in S-Detect and uses skip connections to prevent vanishing gradient. Other ways to achieve accuracy include regularization and min-max pooling.

3.3 Hybrid Models:

Recent developments in risk stratification of medical image models include hybrid models. Fine-tuning DL algorithms for feature extraction and then classifying using ML models is being surveyed. The experiments aim to achieve a "best-of-both-worlds" scenario. Pre-processed images were fed into the pre-existing Google Net Model in an investigation, leading to a cost-sensitive RF classifier resulting in benign or malignant categories[14]. The model achieves close to 90 percent specificity, sensitivity, and accuracy.

Feature extraction is an essential step for any thyroid risk stratification model. The HAIbrid framework consists of a CNN that trains with the inclusion of TIRADS classification features [15]. The final malignancy score is a result of pathological and model-obtained features. The aim is to minimize overseen secondary features of malignancy. The model claims to achieve better accuracy than the methods separately. As a pushback, the score can debilitate with either overfitting, lack of human expertise, or both.



3. RS-Net. (n.d.). https://aapm.onlinelibrary.wiley.com/doi/full/10.1002/mp.14543.

A CNN named RS-Net incorporates the clinical expertise of ACR TI-RADS within its framework [18]. The network inputs ultrasound images and then classifies them into one of the five risk categories according to the clinical guidelines. Risk Stratification Network boasts the capabilities of deep learning model's feature extraction following the physician's guidelines. It shows the possibility of improving diagnostics accuracy while integrating human touch. As the model does not classify nodules into benign or malignant but depicts a risk factor score, it bridges the gap of clinical expertise. The doctors can read the score on the automated system and decide regardless.

4. Discussion

The field of thyroid risk stratification with ultrasonography is constantly evolving. Systems like ACR-TI-RADS have remained prevalent for a long time and are non-automated. Classification systems like these usually heavily rely on the skill of practitioners, and interpreting results is dependent on them. With advancements, ML algorithms come into the picture. SVM, RF, and other supervised classifiers require input data. It is more complicated to have standardized, clear-cut medical images. Ultrasound results rely on machinery and often are non-replicable across systems. Even a standardized approach cannot guarantee 100 percent medical accuracy in malignancy.

On the other hand, deep learning models do not encompass the issue.DL models cannot inherently mitigate bias as they depend on training data. If the training data skews towards a specific population, results get misinformative. Fully automated CNN models also lack interpretability in the real world. We can only integrate them into clinical practice with apprehension.

Many real-time AI thyroid risk stratification systems, such as S-Detect, employ a mixture of both. S-Detect 1 [3] employs the SVM method, and S-Detect 2 [3] uses convolutional networks. These AI systems do not replace human expertise but aim to guide them. Doctors currently rely on such AI models as a secondary opinion.

The complete automation of risk stratification is a long way ahead. Systems at the moment are helping doctors reduce unnecessary FNA procedures along with reducing workload. Hybrid systems appear to be the future of the field, promising better accuracy with evolving compromise between technicians and machines.

5. Conclusion

Incorporating artificial intelligence for thyroid risk stratification and accurate categorization of nodules has revolutionized the medical field, reducing the number of FNP biopsy tests. The AI models, like Support Vector Machines and DL models, efficiently identify different malignancy scales by utilizing machine learning and deep learning algorithms. It leads to faster and cost-effective diagnosis and is helping doctors diagnose and treat patients better. It has also facilitated the customization of treatment plans, which examine various factors in providing the best possible results for medical professionals to cater to the needs of unique patient situations.

While these models are great, they have the limitations of needing a wide variety of data sets to train. Constant monitoring and supervision are required to ensure proper functionality. The most challenging problem would be the medical professional accurately interpreting the results provided by the models, which can sometimes be complicated.

We feel the future scope of these models has excellent prospects for effective treatment and timely detection of thyroid cancer. We hope the AI models evolve with time to provide accurate time thyroid nodule monitoring or supporting dynamic risk assessment. In all this, we should make sure we use AI effectively in changing the field of cancer detection and treatment; it is imperative to solve ethical concerns and data protection issues and should be able to guarantee effective communication between technologists and healthcare practitioners.

6. References

1. M. Javanmardi, M. H. Anisi, M. J. Hosseini, A. Mehrabian, M. R. Akhlaghi, M. R. Sadeghi, and R. Safdari, "A Deep Learning Framework for the Characterization of Thyroid Nodules from Ultrasound Images Using Improved Inception Network and Multi-Level Transfer Learning," 2023, PubMed Central
2. Cao, C.-L., Li, Q.-L., Tong, J., Shi, L.-N., Li, W.-X., Xu, Y., Cheng, J., Du, T.-T., Li, J., & Cui, X.-W.Artificial intelligence in thyroid ultrasound. Frontiers in Oncology, 13,2023, Frontiers in Oncology
3. M. Ludwig, B. Ludwig, A. Mikuła, S. Biernat, J. Rudnicki, and K. Kaliszewski, "The Use of Artificial Intelligence in the Diagnosis and Classification of Thyroid Nodules: An Update," IEEE Trans. Med. Imaging, vol. 15, no. 3, pp. 708–713, Feb. 2023., PubMed Central
4. Do Hyun Kim, Sung Won Kim, Mohammed Abdullah Basurrah, Jueun Lee, Se Hwan Hwang, Diagnostic Performance of Six Ultrasound Risk Stratification Systems for Thyroid Nodules: A Systematic Review and Network Meta-Analysis, 2023, PubMed
5. American College of Radiology. Thyroid Imaging Reporting and Data System (TI-RADS): A User's Guide. 2023, American College of Radiology
6. N. Burgos, N. S. Ospina, & J. A. Sipos, The future of thyroid nodule risk stratification ,2022, PubMed Central
7. Lei Zhang, Jia-Qian He, Jie-Ping Fan, Xiao-Ming Li, Yan-Ling He, Yu-Feng Zheng, Wei-Wei Guo, Wei-Hong Zhang, Thyroid Nodules Risk Stratification Using Deep Learning,2021,PubMed
8. Eun Ju Ha, Jeong Hoon Lee, Da Hyun Lee, Dong Gyu Na, Ji-Hoon Kim, Development of a machine learning-based fine-grained risk stratification system for thyroid nodules using predefined clinicoradiological features, 2023, PubMed
9. ScienceDirect
10. Sajad Khodabandelu, Naser Ghaemian, Soraya Khafri, Mehdi Ezoji, Sara Khaleghi, Development of a Machine Learning-Based Screening Method for Thyroid Nodules Classification by Solving the Imbalance Challenge in Thyroid Nodules Data, 2022, PubMed Central
11. Vijay Vyas Vadhiraj, Andrew Simpkin, James O'Connell, Naykky Singh Ospina, Spyridoula Maraka, Derek T. O'Keeffe. Ultrasound Image Classification of Thyroid Nodules Using Machine Learning Techniques. 2021, PubMed Central
12. Kwon, Soon Woo, et al. "Ultrasonographic Thyroid Nodule Classification Using a Deep Convolutional Neural Network with Surgical Pathology." Journal of Digital Imaging 33.5 (2020): 1202-1208. PMC. Web. 24 Oct. 2020, PubMed Central
13. S Pavithra, G Yamuna, R Arunkumar. Deep Learning Method for Classifying Thyroid Nodules Using Ultrasound Images,2022, IEEE Xplore
14. Jianning Chi, Ekta Walia, Paul Babyn, Jimmy Wang, Gary Groot, Mark Eramian, Thyroid Nodule Classification in Ultrasound Images by Fine-Tuning Deep Convolutional Neural Network, 2017, PubMed
15. Jia, X., Ma, Z., Kong, D., Li, Y., Hu, H., Guan, L., Yan, J., Zhang, R., Gu, Y., Chen, X., et al. (2022). Novel Human Artificial Intelligence Hybrid Framework Pinpoints Thyroid

Nodule Malignancy and Identifies Overlooked Second-Order Ultrasonographic Features, MDPI

16. Ha, Eun Ju, and Jung Hwan Baek. "Applications of machine learning and deep learning to thyroid imaging: where do we stand?" Ultrasonography (US) v.40(1); 2021 Jan,PubMed Central
17. Vijay Vyas Vadhiraj, Andrew Simpkin, James O'Connell, Naykky Singh Ospina, Spyridoula Maraka, Derek T. O'Keeffe. Ultrasound Image Classification of Thyroid Nodules Using Machine Learning Techniques. 2021, PubMed Central
18. Ziyu Bai, Luchen Chang, Ruiguo Yu, Xuewei Li, Xi Wei, Mei Yu, Zhiqiang Liu, Jie Gao, Jialin Zhu, Yulin Zhang, Shuaijie Wang, Zhuo Zhang, Thyroid nodules risk stratification through deep learning based on ultrasound images,2020, AAPM
19. Giorgio Grani, Livia Lamartina, Valeria Ascoli, Daniela Bosco, Marco Biffoni, Laura Giacomelli, Marianna Maranghi, Rosa Falcone, Valeria Ramundo, Vito Cantisani, Sebastiano Filetti, Cosimo Durante,Reducing the Number of Unnecessary Thyroid Biopsies While Improving Diagnostic Accuracy: Toward the “Right” TIRADS,2019, Oxford Academic