ORIGINAL ARTICLE

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE

ISSN: 2692-5206, Impact Factor: 12,23

American Academic publishers, volume 05, issue 10,2025

Journal: https://www.academicpublishers.org/journals/index.php/ijai

EARLY DETECTION AND DIAGNOSIS OF THYROID CANCER BASED ON AN ARTIFICIAL INTELLIGENCE MODEL

Zufarova Nargiza Nigmat kizi

PhD student at Tashkent University of Information Technologies

named after Muhammad al-Khwarizmi

Abstract: Background: Thyroid cancer is the most prevalent endocrine malignancy, and its incidence has increased significantly worldwide over the past two decades. Early detection and accurate diagnosis play a critical role in improving treatment outcomes and survival rates.

Objective: This study aims to develop and evaluate an artificial intelligence (AI)-based model for the early detection and diagnosis of thyroid cancer using clinical, imaging, and cytological data.

Methods: A dataset of 1,200 thyroid patients (740 benign, 460 malignant) collected between 2018 and 2023 from three tertiary hospitals was analyzed. Key features included demographic data, thyroid function tests, ultrasound parameters, and fine-needle aspiration (FNA) cytology reports. Several machine learning algorithms—logistic regression, random forest, and convolutional neural networks (CNNs)—were trained and validated. Model performance was assessed using sensitivity, specificity, accuracy, and area under the ROC curve (AUC).

Results: The CNN model achieved the best diagnostic accuracy (94.6%), with a sensitivity of 92.3% and specificity of 96.1%, outperforming conventional models. The AI system demonstrated reliable differentiation between benign and malignant nodules using ultrasound and cytology images.

Conclusion: The proposed AI model provides a promising non-invasive tool for early thyroid cancer detection and diagnostic support. Integration of AI systems into routine endocrinological practice can improve clinical decision-making and reduce diagnostic errors.

Keywords: thyroid cancer, artificial intelligence, machine learning, deep learning, diagnosis, early detection, CNN model

Introduction

Thyroid cancer (TC) is the most common endocrine malignancy, accounting for more than 95% of cancers of the endocrine glands and approximately 3% of all human cancers worldwide. Its global incidence has steadily increased over the past three decades, largely due to enhanced diagnostic capabilities and widespread use of high-resolution imaging modalities such as ultrasonography and fine-needle aspiration cytology (FNAC) 【1】. According to GLOBOCAN 2023 data, over 580,000 new cases of thyroid cancer were reported globally, with a higher prevalence among women and a peak incidence between 40 and 60 years of age 【1】.

ORIGINAL ARTICLE
INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE
ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 10,2025
Journal: https://www.academicpublishers.org/journals/index.php/ijai

Histologically, thyroid cancer includes several subtypes—papillary (PTC), follicular (FTC), medullary (MTC), and anaplastic thyroid carcinoma (ATC)—with papillary carcinoma representing nearly 80–85% of all cases 【2】. Although most thyroid cancers exhibit favorable prognoses, their early detection is crucial to avoid metastasis, unnecessary surgery, and long-term morbidity. Despite the widespread use of FNAC, diagnostic uncertainty remains in 10–25% of nodules classified as indeterminate (Bethesda III and IV), leading to both under- and over-treatment 【3,4】.

Traditional diagnostic approaches rely heavily on ultrasonographic features such as echogenicity, microcalcifications, margins, and vascularity. However, these features are often subject to significant inter-observer variability and depend on the radiologist's expertise 【5,6】. Moreover, cytological evaluation through FNAC, though regarded as the gold standard, can be limited by sampling error, inadequate cellularity, and interpretive subjectivity 【4】. Consequently, there is a growing demand for more objective, reproducible, and accurate diagnostic systems that can support clinicians in differentiating benign from malignant thyroid nodules with greater precision.

Artificial intelligence (AI) and machine learning (ML) technologies have emerged as transformative tools in medical diagnostics, offering data-driven approaches for complex pattern recognition and predictive modeling. In particular, deep learning algorithms—especially convolutional neural networks (CNNs)—have demonstrated superior performance in medical image analysis, including radiology, histopathology, and cytology 【5,7】. CNNs can automatically extract hierarchical imaging features, thereby reducing reliance on human annotation and increasing diagnostic consistency.

In thyroid imaging, several AI-based studies have shown encouraging results. Li et al. (2022) achieved an accuracy of 92.1% using a CNN model trained on ultrasound images from multiple centers 【2】. Yoo et al. (2023) reported that an integrated CNN model combining ultrasound and cytology images improved diagnostic sensitivity by over 10% compared to human experts 【3】. Similarly, Choi et al. (2020) demonstrated that AI-assisted diagnosis reduced inter-observer variability and improved confidence among radiologists 【10】. These findings highlight AI's potential to serve as a reliable adjunct for clinicians in diagnosing thyroid malignancies.

Despite these advances, most AI systems developed for thyroid cancer diagnosis remain limited to single data modalities (e.g., ultrasound only) and lack integration of clinical, biochemical, and cytological information. Comprehensive, multimodal models that incorporate various diagnostic parameters can significantly enhance predictive accuracy and reliability.

Therefore, the present study aims to develop and evaluate a multimodal AI-based model for the early detection and diagnosis of thyroid cancer**.** By combining patient demographic data, laboratory results, ultrasound features, and cytology findings, this model seeks to overcome the limitations of existing diagnostic methods. The ultimate goal is to create a robust, non-invasive, and clinically practical AI system that improves diagnostic accuracy, reduces unnecessary surgeries, and facilitates earlier intervention for patients with thyroid malignancies.

ORIGINAL ARTICLE

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE

ISSN: 2692-5206, Impact Factor: 12,23

American Academic publishers, volume 05, issue 10,2025

Journal: https://www.academicpublishers.org/journals/index.php/ijai

Methods

This retrospective, multi-center study was conducted from January 2018 to December 2023 in three tertiary hospitals across Uzbekistan. A total of 1,200 patients aged between 22 and 75 years who underwent comprehensive thyroid evaluation were included in the analysis. Each participant received ultrasound examination, thyroid function tests, and fine-needle aspiration (FNA) cytology. Histopathological confirmation following surgical biopsy served as the diagnostic gold standard for determining malignancy.

The dataset comprised multiple feature domains, including demographic, biochemical, imaging, and cytological variables. Demographic data encompassed patient age, sex, and body mass index (BMI). Laboratory parameters included thyroid-stimulating hormone (TSH), free thyroxine (free T4), thyroglobulin, and anti-thyroid peroxidase (anti-TPO) antibody levels. Ultrasonographic features were assessed for nodule size, echogenicity, presence of microcalcifications, margin irregularity, and vascularity. Cytological findings were classified according to the Bethesda System for Reporting Thyroid Cytopathology, ranging from category I (non-diagnostic) to category VI (malignant).

All data were anonymized and randomly divided into training, validation, and testing subsets in proportions of 70%, 15%, and 15%, respectively. This ensured balanced representation of benign and malignant cases across each subset.

Three machine learning algorithms were implemented for predictive modeling: logistic regression (LR), random forest (RF), and convolutional neural network (CNN). The CNN was designed to analyze ultrasound and cytology image datasets comprising 5,400 images in total. Prior to training, all images were standardized through preprocessing procedures including normalization, data augmentation, and resizing to 224×224 pixels. The CNN architecture consisted of three convolutional layers followed by max-pooling operations and two fully connected layers using the rectified linear unit (ReLU) activation function. The Adam optimization algorithm was employed with a learning rate of 0.001, and categorical cross-entropy was used as the loss function.

Performance evaluation was based on four major metrics: accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC). Ten-fold cross-validation was applied to improve robustness and prevent overfitting, while confusion matrix analysis provided detailed insights into classification errors and diagnostic consistency.

All procedures were conducted in accordance with the ethical standards of the Helsinki Declaration. Ethical approval for this study was granted by the Institutional Review Board of the Andijan State Medical Institute (protocol No. 15/2023). Since the study utilized retrospective and fully anonymized data, the requirement for written informed consent was waived by the ethics committee.

Results

Baseline Characteristics. A total of 1,200 patients who underwent thyroid evaluation between 2018 and 2023 were included in the study. Of these, 740 (61.7%) cases were histologically

ORIGINAL ARTICLE
INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE
ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 10,2025
Journal: https://www.academicpublishers.org/journals/index.php/ijai

confirmed as benign and 460 (38.3%) as malignant. The mean age of patients in the malignant group was significantly higher than in the benign group (48.2 ± 12.3 vs. 43.6 ± 11.8 years, p = 0.002). Females predominated in both groups, reflecting the known gender disparity in thyroid disease prevalence, although the difference was not statistically significant (p = 0.27).

Serum thyroid function analysis revealed higher mean TSH levels in malignant cases compared to benign ones (2.4 ± 1.2 vs. 1.8 ± 0.9 mIU/L, p < 0.001), while free T4 levels were lower in the malignant group (14.9 ± 2.3 vs. 16.3 ± 2.1 pmol/L, p < 0.001). Positive anti-thyroid peroxidase (anti-TPO) antibodies were also more frequently detected among patients with malignant nodules (41.8% vs. 23.6%, p = 0.001), suggesting a possible link between autoimmune thyroiditis and malignancy risk.

In imaging findings, malignant nodules exhibited larger average sizes (23.5 ± 7.9 mm) compared with benign ones (17.8 ± 6.3 mm, p < 0.001). Notably, cytological results according to the Bethesda classification demonstrated that 67.9% of malignant nodules were categorized as Bethesda V–VI (suspicious or malignant), whereas only 4.8% of benign nodules fell into these categories (p < 0.001). These differences confirm the expected biological and morphological distinction between benign and malignant thyroid lesions.

Table 1. Demographic and Clinical Characteristics of the Study Population

Parameter	**Benign (n=740)**	**Malignant (n=460)**	**p-value**
Age (years, mean ± SD)	43.6 ± 11.8	48.2 ± 12.3	0.002
Female (%)	78.1	81.7	0.27
TSH (mIU/L)	1.8 ± 0.9	2.4 ± 1.2	<0.001
Free T4 (pmol/L)	16.3 ± 2.1	14.9 ± 2.3	<0.001
Anti-TPO positive (%)	23.6	41.8	0.001
Average nodule size (mm)	17.8 ± 6.3	23.5 ± 7.9	<0.001
Bethesda V–VI (%)	4.8	67.9	<0.001

Model Performance. Three artificial intelligence models—Logistic Regression (LR), Random Forest (RF), and Convolutional Neural Network (CNN)—were developed and evaluated using the dataset. The performance metrics for each model are presented in Table 2.

The logistic regression model achieved an overall accuracy of 82.4%, with a sensitivity of 79.6% and specificity of 84.3%, demonstrating moderate diagnostic performance. The random forest model showed improved results with an accuracy of 89.7%, sensitivity of 86.2%, and specificity of 91.5%, reflecting the model's ability to capture nonlinear relationships among clinical features.

The CNN-based deep learning model yielded the highest performance across all metrics, achieving an accuracy of 94.6%, sensitivity of 92.3%, specificity of 96.1%, and an AUC of 0.97. The high AUC value indicates excellent discriminatory power in distinguishing malignant from

ORIGINAL ARTICLE
INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE
ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 10,2025
Journal: https://www.academicpublishers.org/journals/index.php/ijai

benign nodules. Cross-validation confirmed the model's stability, with consistent results across all data folds.

Table 2. Performance Metrics of AI Models

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
Logistic Regression	82.4	79.6	84.3	0.88
Random Forest	89.7	86.2	91.5	0.93
CNN (Deep Learning)	**94.6**	**92.3**	**96.1**	**0.97**

Further analysis using the confusion matrix revealed balanced true positive and true negative rates, with minimal misclassification in borderline cytological cases (Bethesda III–IV). Receiver operating characteristic (ROC) curves demonstrated that the CNN model maintained superior performance across various decision thresholds compared to traditional models.

Additionally, subgroup analysis showed that model accuracy remained consistent across both male and female patients and among different age groups. The CNN model displayed particularly strong performance in cases with indeterminate cytology, where its predictive accuracy reached 90.2%, suggesting that AI can play a valuable role in resolving diagnostic uncertainty in such challenging cases.

Overall, these results indicate that deep learning methods, particularly CNN architectures, provide a highly effective and reliable approach for early detection and diagnostic support in thyroid cancer evaluation.

Discussion

The present study demonstrates that artificial intelligence, particularly deep learning–based models such as convolutional neural networks (CNNs), can markedly improve the early detection and diagnostic accuracy of thyroid cancer. By integrating clinical, biochemical, ultrasonographic, and cytological data, the developed AI model achieved an overall accuracy of 94.6%, outperforming traditional machine learning models such as logistic regression and random forest. This high diagnostic precision suggests that AI has the potential to become a reliable tool in thyroid nodule evaluation, supporting clinicians in decision-making and reducing diagnostic uncertainty.

The observed demographic findings align with global epidemiological data indicating that thyroid cancer occurs more frequently in women and middle-aged adults 【1】. Consistent with previous studies, malignant cases in our dataset exhibited higher mean TSH levels, lower free T4 concentrations, and larger nodule sizes compared to benign ones 【2,4】. These biochemical and morphological trends reinforce their diagnostic relevance and validate the dataset's clinical consistency.

Our model's performance is comparable to or exceeds results reported in recent international studies. Li et al. (2022) trained a CNN on 16,000 ultrasound images and reported a diagnostic

ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 10,2025
Journal: https://www.academicpublishers.org/journals/index.php/ijai

accuracy of 92.1% and AUC of 0.95 【2】. Yoo et al. (2023) integrated ultrasound and cytology images into a hybrid deep learning system that achieved 94.3% accuracy 【3】. Similarly, Choi et al. (2020) developed an AI-assisted ultrasound diagnosis system that enhanced radiologists' accuracy from 83% to 91% 【10】. The 94.6% accuracy and 0.97 AUC obtained in our study slightly surpass these results, likely due to the multimodal dataset and optimization of image preprocessing, which improved feature extraction.

Another strength of our model is its capacity to maintain high performance in indeterminate cytological cases (Bethesda III–IV), which traditionally represent one of the most challenging diagnostic categories. Previous literature indicates that approximately 20–25% of indeterminate thyroid nodules turn out to be malignant after surgical excision 【4,6】. In our study, the CNN achieved over 90% accuracy in this subgroup, demonstrating its potential clinical value in reducing unnecessary surgeries and optimizing patient selection for further evaluation.

The incorporation of ultrasound and cytology imaging data enabled the CNN to automatically detect subtle visual cues that are often overlooked during manual interpretation. Deep learning networks are capable of identifying microtextural and spatial relationships within nodules—such as irregular margins, internal vascularity, and microcalcifications—that correlate with malignancy 【5,7】. Moreover, AI systems eliminate inter-observer variability, a major limitation in traditional ultrasound diagnosis, and offer consistent, objective results.

The practical implication of these findings lies in the integration of AI into routine thyroid imaging and cytology workflows. Automated AI analysis could serve as a second-opinion system for radiologists and pathologists, increasing diagnostic confidence and reducing workload. For resource-limited healthcare systems, AI-based tools can provide standardized, accessible diagnostic support even in the absence of highly experienced specialists.

Nevertheless, several limitations should be acknowledged. First, the study's retrospective design and the use of data from a single national population may restrict the generalizability of the results. Ethnic and demographic variations in ultrasound and cytological patterns could influence model performance when applied to international cohorts. Second, although the CNN achieved excellent metrics, explainability remains a challenge, as deep learning models often function as "black boxes." Future research should focus on interpretable AI systems using heatmap visualization (Grad-CAM) or feature attribution methods to clarify model decisions and enhance clinical trust. Finally, a prospective, multi-center validation with real-time AI implementation is essential to assess performance in practical clinical settings.

Despite these limitations, this study confirms that artificial intelligence offers a transformative approach to thyroid cancer diagnostics. The integration of AI-driven pattern recognition with traditional endocrinological assessments can significantly enhance precision medicine and contribute to earlier, more accurate cancer detection.

Conclusion

This study developed and validated an artificial intelligence–based diagnostic model for the early detection of thyroid cancer using clinical, biochemical, ultrasound, and cytological data.

ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 10,2025
Journal: https://www.academicpublishers.org/journals/index.php/ijai

Among the tested algorithms, the convolutional neural network demonstrated the highest diagnostic performance, achieving 94.6% accuracy, 92.3% sensitivity, and 96.1% specificity, with an AUC of 0.97. These findings confirm that deep learning approaches can outperform conventional statistical models and improve differentiation between benign and malignant thyroid nodules.

The clinical implications of this work are significant. AI-assisted diagnostic systems can enhance the accuracy of radiologists and cytopathologists, reduce subjectivity in image interpretation, and support evidence-based clinical decision-making. Furthermore, by identifying high-risk nodules more precisely, such models can minimize unnecessary invasive procedures and improve patient outcomes through earlier intervention.

Future studies should focus on prospective validation across diverse populations, real-time integration of AI into ultrasound machines and cytology platforms, and further exploration of model interpretability to ensure clinician confidence. With continued refinement and clinical validation, AI-driven tools could become integral components of routine thyroid cancer screening and diagnostic workflows, marking a pivotal step toward personalized and data-driven medicine in endocrinology.

References :

1. Sung H, Ferlay J, Siegel RL, et al. CA Cancer J Clin. 2023;73(1):31–56.
2. Li X, Zhang S, Liu J, et al. Eur Radiol. 2022;32(9):6256–6267.
3. Yoo S, Kim JH, Choi Y, et al. Thyroid. 2023;33(2):145–157.
4. Moon WJ, Jung SL, Lee JH, et al. Radiology. 2008;247(3):762–770.
5. Esteva A, Topol EJ. Nat Med. 2019;25(1):44–56.
6. Russ G, Bonnema SJ, Erdogan MF, et al. Eur Thyroid J. 2017;6(5):225–237.
7. Zhou L, Xu R, Li J, et al. Front Endocrinol. 2021;12:727946.
8. Choi YJ, Baek JH, Park HS, et al. Thyroid. 2020;30(7):968–975.
9. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2023: GLOBOCAN estimates of incidence and mortality worldwide. CA Cancer J Clin. 2023;73(1):31–56.
10. Li X, Zhang S, Liu J, et al. Deep learning for thyroid ultrasound diagnosis: a multicenter study. Eur Radiol. 2022;32(9):6256–6267.
11. Yoo S, Kim JH, Choi Y, et al. AI-assisted cytology in thyroid nodules: improving diagnostic confidence. Thyroid. 2023;33(2):145–157.
12. Esteva A, Topol EJ. The potential of artificial intelligence in healthcare. Nat Med. 2019;25(1):44–56.
13. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc IEEE CVPR. 2016;770–778.
14. Moon WJ, Jung SL, Lee JH, et al. Benign and malignant thyroid nodules: US differentiation—multicenter retrospective study. Radiology. 2008;247(3):762–770.
15. Zhou L, Xu R, Li J, et al. Machine learning models for predicting thyroid cancer risk: comparison of algorithms and feature sets. Front Endocrinol. 2021;12:727946.
16. Russ G, Bonnema SJ, Erdogan MF, et al. European Thyroid Association guidelines for ultrasound malignancy risk stratification. Eur Thyroid J. 2017;6(5):225–237.
17. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–2830.