

86167_Auto_Edited.docx

Name of Journal: *Artificial Intelligence in Cancer*

Manuscript NO: 86167

Manuscript Type: EVIDENCE REVIEW

Artificial intelligence in the diagnosis of thyroid cancer: Recent advances and future directions

Lakshmi Nagendra, Joseph M Pappachan, Cornelius James Fernandez

Abstract

The diagnosis and management of thyroid cancer is fraught with challenges despite the advent of innovative diagnostic, surgical, and chemotherapeutic modalities. Challenges like inaccuracy in prognostication, uncertainty in cytopathological diagnosis, trouble in differentiating follicular neoplasms, intraobserver and interobserver variability on ultrasound imaging preclude personalised treatment in thyroid cancer. Artificial intelligence (AI) is bringing a paradigm shift to the healthcare, powered by quick advancement of the analytic techniques. Several recent studies have shown remarkable progress in thyroid cancer diagnostics based on AI-assisted algorithms. Application of AI techniques in thyroid ultrasonography and cytopathology have shown remarkable improvement in sensitivity and specificity over the traditional diagnostic modalities. AI has also been explored in the development of treatment algorithms for indeterminate nodules and for prognostication in the patients with thyroid cancer. The benefits of high repeatability and straightforward implementation of AI in the management of thyroid cancer suggest that it is a promise for clinical application. Limited clinical experience and lack of prospective validation studies remain the biggest drawbacks. Developing verified and trustworthy algorithms after extensive testing and validation using prospective, multi-centre trials is crucial for the future use of AI in the pipeline of precision medicine in the management of thyroid cancer.

INTRODUCTION

One or more thyroid nodules are seen in up to 60% of the general population. The main concern is the potential for malignancy, even though only around 5% of cases turn out to be malignant.^[1] Although epidemiological studies have shown a small increase in the incidence of thyroid cancer, probably due to exposure to environmental risk factors, the exponential ¹³ increase in the diagnosis of thyroid cancer is primarily due to the increased adoption of diagnostic imaging technology and medical surveillance together with improved access to general health care.^[2] Overdiagnosis and overtreatment are linked to conceivably exorbitant expenses and patient distress.

In stratifying patients' risk of malignancy, ultrasound assessment is now universally recognised as a critical diagnostic step. However, there is considerable doubt over the diagnostic efficacy of several characteristics assessed during the ultrasonographic examination of thyroid nodules. A meta-analysis involving 31 studies in 13,736 adults concluded that individual ultrasound features are not accurate predictors of thyroid cancer.^[3] This is further complicated by intraobserver and interobserver variability in thyroid ultrasonography.^[4]

Thyroid fine needle aspiration cytology (FNAC) is associated with significant intraobserver and interobserver variability. Its sensitivity ² ranges from 68 to 98% while the specificity ranges from 56 to 100%. Nearly 15 to 30% of thyroid nodules are categorised as indeterminate in thyroid FNAC, of which 25% are later found to be malignant.^[5] The indeterminate thyroid nodules should have received The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) category III, IV, or V in thyroid FNAC is shown in **Table 1**.^[6]

Lobectomy and formal histopathological examination would help in reaching a definitive diagnosis in patients with indeterminate thyroid nodules (ITNs). However, surgery for indeterminate nodules carries a risk of surgical complications and associated costs, both of which can be avoided, if we could confidently identify the likelihood of thyroid cancer in indeterminate nodules.

Molecular testing with next generation sequencing technology that looks for a wide range of genomic alterations in the thyroid FNAC samples in patients with indeterminate thyroid nodules can help to resolve diagnostic uncertainties.^[8] Though this may also help to guide the management, wide applicability of this is limited by the associated costs, the limited availability in only major centers, and the time required to obtain the test results.^[9] Hence, the diagnosis and management of thyroid cancer is fraught with challenges despite the advent of innovative diagnostic, surgical, and chemotherapeutic modalities in recent years.

The ability of machines to approach problem-solving with human-like logic is known as artificial intelligence (AI), in its broadest sense. There are two main computing approaches that make up artificial intelligence: deep learning that uses the neural networks that replicate the structure and function of the brain to analyse the data, and the traditional or classic machine learning that involves computers' learning based on the data supplied by people.^[10] Please see **Figure 1** for further details.

Figure 1. Classical machine learning and deep learning models in artificial intelligence (AI)-based thyroid cancer diagnostic algorithms.

In supervised learning (traditional or classic machine learning) that is more commonly used in medicine, a labelled ⁶dataset of images already classified by human experts is introduced into an algorithm's resources. Based ⁶on these resources, the algorithm tries to identify a function that most accurately classifies the provided dataset with an aim to obtain results as close as possible to the classification made by human experts. Then, the program compares its results with those contained in labelled datasets. The algorithm then repeats the steps to develop a function with the highest possible accuracy. The supervised learning relies on feature engineering (i.e., use of expert knowledge to distil features from data) and it needs tremendous time and human resources.^[11] The support vector machine (SVM), multilayer perceptron (MLP), and random forest are examples of supervised learning.^[12] The classic machine learning algorithms are used in the computer aided diagnosis (CAD) systems to improve the accuracy of diagnosis and to reduce the time required for image interpretation.

On the contrary, unsupervised learning or deep learning (DL) eliminates the need for manual data extraction and hence has improved time and resource efficiency. DL is provided by neural networks made up of neurons. There are many neurons in each layer. The neural network transforms the initial input to the output through connections between layers and nonlinear processing. ²² The output of the upper layer is considered as the input of the subsequent layer in neural networks. This high-level neural network overcomes the main drawback of the traditional or classic machine learning (i.e., need for human data extraction) by automatically learning more abstract and generalised characteristics from the input.^[11] The main limitation of the DL method is the need for large datasets for training the model. Another limitation is the difficulty to understand how the AI arrived at its decision (black-box issue) due to lack of explanations. This creates new challenges for safety, trust and reliability. ⁷ A lack of trust in the AI systems is a significant impediment to its adoption. ⁷ Additionally, adoption of AI systems can be influenced by several factors including properties of the AI system, user experiences, and the design of the human-AI collaboration. The various artificial neural network (ANN) models that facilitate DL are convolutional neural network (CNN), recurrent neural network (RNN) and its variants including gated recurrent unit (GRU), and long short-term memory (LSTM).^[12]

Radiomics is another application of AI which is gaining importance. Radiomics is described as the extraction of high-throughput features from medical images, deeper data mining of the image information using automatic or semi-automatic analytical methods, and association of the extracted features with other clinical data for the purpose of supporting clinical diagnostic and treatment decisions based on the best available scientific evidence.^[13,14] Thus, the radiomics workflow include image acquisition, image segmentation, feature extraction, feature selection (i.e., elimination of unreliable, uninformative or redundant features), and construction of classification and prediction models.

The image segmentation can either be done manually (which is the gold standard) or using semi-automatic methods. Evaluation by multiple clinicians or the use of multiple

algorithms can be used to avoid possible inter-observer bias. The features extracted can be divided into shape features (morphological features), first-order features (distribution of grayscale intensity), second-order features (texture features - heterogeneity within tumour), and higher-order features (repeating patterns, histogram-oriented patterns, or local binary patterns).^[13] Radiomics can either be conventional ML-based where the features to be extracted are predefined (e.g., tumour phenotype) or be DL-based where the features are not predefined but are automatically extracted from the underlying data are used for future prediction (e.g., treatment response). SVM is one of the most common algorithms used for ML-based and CNN is one of the most common algorithms used for DL-based radiomics.

In comparison to conventional manual interpretation, radiomics is more objective, makes greater use of available information, and offer better repeatability of interpretations. Radiomics gives physicians a new tool to diagnose and treat patients in a more precise and individualised way in this era of precision medicine.^[13,14] However, the limitation of radionomics include lack of data images from multiple centres and different equipment for training the AI developed predictive models and for external validation. These are needed to avoid overfitting (i.e., the process in which the predicting model learns exactly the training set but fails to fit new data from the test set), and to improve generalizability and model performance. The 'overfitting' can be mitigated by adopting a cross-validation set-up and by reducing the number of features used by the model through feature selection methods.

AI is bringing a paradigm shift to healthcare, powered by quick advancement of analytic techniques. Several recent studies have shown remarkable progress in thyroid cancer diagnostics based on AI assisted algorithms. In this comprehensive review we provides the evidence favouring AI applications in the diagnosis of thyroid cancer and discuss its future.

Application of AI in thyroid ultrasound (US)

Several risk stratification systems are available for evaluating thyroid nodules on ultrasound. They use varying approaches to classify levels of suspicion for malignancy

leading to variable performance. Some of these include the American Association of Clinical Endocrinologists, American College of Endocrinology and Associazione Medici Endocrinologi (AACE/ACE/AME) system, American College of Radiology Thyroid Imaging Reporting and Data System (ACR TI-RADS), the American Thyroid Association (ATA) risk stratification system, European Thyroid Association Thyroid Imaging Reporting and Data System (EU-TIRADS), the Korean Thyroid Imaging Reporting and Data System (KTA/KSThR-TIRADS) endorsed by the Korean Thyroid Association and the Korean Society of Thyroid Radiology, Chinese Thyroid Imaging Reporting and Data Systems (C-TIRADS), and the Thyroid Imaging Reporting and Data System developed by Kwak *et al* (Kwak TIRADS).^[15]

A systematic review and network meta-analysis by Kim *et al.* compared the diagnostic performance of six of the above risk stratification systems (except C-TIRADS) observed that ACR TI-RADS had the highest diagnostic performance for the detection of thyroid nodules on ultrasound.^[15] A retrospective study by Yang *et al.* compared the diagnostic performance of six of these risk stratification systems (except Kwak TIRADS) vs. an AI system (AI-SONICTM) in differentiating thyroid nodules, especially medullary thyroid carcinoma. The study observed that when assessing medullary thyroid carcinoma, the ACR- TI-RADS had the best sensitivity (90.2%), and negative predictive value (NPV 91.8%), and AI-SONICTM had the best specificity (85.6%) and positive predictive value (PPV 67.5%).^[16]

An artificial intelligence-optimized TI-RADS (AI-TIRADS), developed by Wildman-Tobriner *et al.* and validated against ACR TI-RADS, was found to have minimally improved specificity with comparable sensitivity.^[17] Using this AI-TIRADS, Watkins *et al.* observed that the British Thyroid Association (BTA), ACR-TIRADS and AI-TIRADS have comparable diagnostic performance for predicting thyroid nodule malignancy with high sensitivity of 98.28%, 95.24%, and 93.44% respectively, but relatively low specificity (45.71%, 40.57%, and 45.71% respectively). ACR-TIRADS and AI-TIRADS recommended FNAC at a significantly lower rate in benign thyroid nodules compared to BTA (30.7%, 29.8%, and 46.3% respectively).^[18]

Peng *et al.* developed and trained a DL (CNN)-based AI model known as ThyNet using 18,049 images. Thereafter, they tested the AI model on 4305 images. The area under the receiver operating characteristic curve (AUC) for accurate diagnosis was higher for ThyNet than that of radiologists (0.922 vs 0.839). ThyNet-assisted strategy improved the pooled AUC of the radiologists from 0.837 to 0.875 for reviewing images, and from 0.862 to 0.873 when using images and videos. Furthermore, use of AI-assisted strategy reduced the number of FNACs (from 61.9% to 35.2%) and missed malignancy (from 18.9 to 17.0%).^[19]

Wang *et al.* collected and labelled US images from 2450 benign thyroid nodules and 2557 malignant thyroid nodules as a training data for DL using the YOLOv2 neural network. This was followed by retrospective testing of US images of 276 patients. AI system correctly identified the lesion area, with an AUC of 0.902, which is better than that of the radiologists (0.859). Thus, AI system showed matching sensitivity, PPV, NPV, and accuracy for the diagnosis of malignant nodules (90.5%, 95.22%, 80.99%, and 90.31%, respectively) in comparison to the radiologists, but with superior specificity (89.91% vs 77.98%).^[20]

Park *et al.* developed a DL-based computer-aided diagnosis (CAD) system (dCAD) for the diagnosis of thyroid nodules based on 4919 images of thyroid nodules from 3 institutions. They compared the performance of dCAD with those of a support vector machine (SVM)-based CAD system (sCAD) and that by the radiologists. The study observed that there is no significant difference in overall sensitivity (94.2% vs 91%), specificity (76.9% vs 80%), PPV (83.1% vs 84.5%), NPV (91.7% vs 88.1%) and accuracy (86.4% vs 86%) between radiologists and dCAD ($p > 0.05$). However, radiologists and dCAD showed higher specificity, PPV, and accuracy than sCAD (all $p < 0.001$). In small nodules, experienced radiologists showed higher specificity, PPV and accuracy than sCAD ($p < 0.05$).^[21]

A prospective study by He *et al.* compared the diagnostic performances of two junior and two senior radiologists, an AI system, and an AI-assisted junior radiologist in diagnosing thyroid nodules. The study also looked at the factors affecting its diagnostic

accuracy. The study observed that the diagnostic performance of the senior radiologists, the AI system, and the AI-assisted junior radiologist were better than the junior radiologist ($p < 0.05$). However, the diagnostic performance of the AI system and the AI-assisted junior radiologist were like that of the senior radiologists ($p > 0.05$). Hence, the AI systems could potentially improve the diagnostic efficiency of junior radiologists. The diagnostic error rates of the AI system, junior and senior radiologists were higher for nodules with a maximum diameter of ≤ 1 cm compared to those with a maximum diameter of ≥ 1 cm.^[22]

Rho *et al.* observed that a CNN-based AI model that trained with data from thyroid nodules more than 10 mm in size showed overall better diagnostic performance than radiologists in the diagnosis and categorization of thyroid nodules which are less than 10 mm in size, especially in nodules with size less than 5 mm.^[23] Zheng *et al.* evaluated the diagnostic performance of different ultrasound sections of thyroid nodule using AI-CADS in predicting thyroid malignancy. The study observed that the diagnostic performance was higher in the transverse section compared to longitudinal section.^[24]

Tao *et al.* developed and tested a DL model based on multimodal US images including Grayscale US (GSU), colour Doppler flow imaging (CDFI), strain elastography (SE), and region of interest mask (Mask) images to accurately diagnose TI-RADS 3-5 thyroid nodules. They randomly divided multimodal US images of 1,138 thyroid nodules of TI-RADS 3-5 categories into training set ($n = 728$), validation set ($n = 182$), and test set ($n = 228$). The study observed that the AUCs of DL in the differentiation of thyroid nodules were 0.928 based on GSU, CDFI, SE, and Mask, 0.909 based on GSU and CDFI, 0.906 based on GSU, CDFI, and SE, 0.881 based on GSU and Mask, 0.858 based on GSU and SE, and 0.825 based single GSU alone.^[25]

The study by Tao *et al.* also observed that the specificity was highest (89.5%) based on GSU, CDFI, and SE. The diagnostic accuracy (86.2%) and sensitivity (86.9%) were highest based on GSU, CDFI, and Mask. With assistance from DL model, the AUC of junior radiologists has increased from 0.720 to 0.796, which was slightly higher than that of senior radiologists without assistance from DL model (0.796 vs. 0.794). With the

assistance from DL model, the senior radiologists exhibited higher accuracy and comparable AUC than that of DL based on GSU (83.4% vs. 78.9%). However, the AUC of DL model based on multimodal US images was significantly higher than that based on visual diagnosis by radiologists.^[25]

A systematic review and meta-analysis of twenty-five studies that compared all available AI and radiologists using thyroid US imaging, observed that the pooled sensitivity and specificity of AI were 0.86 and 0.78, respectively, whereas the pooled sensitivity and specificity of radiologists were 0.85 and 0.82, respectively. The AI and radiologists have comparable accuracy with AUC of 0.89 vs 0.91. There was no statistically significant difference in diagnostic odd ratio (23.10 vs 27.12). The DL model of AI had statistically significantly greater sensitivity and specificity compared to the classic ML AI.^[26]

Tong *et al.* studied whether a personalized AI model would help radiologists with varying levels of expertise in decision-making processes. They developed an optimized AI model using 1754 US images. Subsequently, they prospectively studied 300 US images to compare the optimized strategy with the traditional all-AI strategy for diagnostic performance and workload reduction. The study suggested that an optimized AI model in thyroid nodule management may reduce diagnostic time-based costs without sacrificing diagnostic accuracy for senior radiologists. On the other hand, the traditional all-AI model may still be more beneficial for junior radiologists.^[27]

Zhu *et al.* developed and trained a CNN model called the Brief Efficient Thyroid Network (BETNET) using 16,401 US images for localization and automatic diagnosis of thyroid nodules on US images. The BETNET model exhibited better diagnostic performance than three state-of-the-art algorithms, which in turn has a similar diagnostic performance as the experienced radiologists.^[28] Qi *et al.* developed a DL model to look for presence of gross extrathyroidal extension (ETE) in US images for the preoperative evaluation in thyroid cancer patients. The DL model was found to have a similar or even better diagnostic performance than senior radiologists and hence is a helpful tool for preoperative diagnosis of gross ETE.^[29] Jassal *et al.* has recently

15 developed a pilot AI model that can accurately predict malignancy in thyroid nodules using USG features, FNAC, demographics and serum TSH.^[30]

Role of AI in thyroid FNAC and histopathology

The widespread application of ancillary techniques such as liquid-based cytology (LBC), immunocytochemistry, and flow cytometry has increased the diagnostic accuracy of thyroid fine-needle aspiration cytology (FNAC).^[31-33] Additionally, the clinical care of nodules with a FNAC diagnosis of follicular neoplasm (FN) or atypia of uncertain significance (AUS) now includes the possibility of doing a molecular testing in the FNAC sample. Given its strong NPV, the molecular testing is utilised as an exclusion test, and for benign or low-risk FN or AUS nodules, clinical follow-up without resection is advised.^[34] The assessment of thyroid FNAC specimens may be improved further with the AI incorporating machine learning.

12 Interestingly, AI as an adjunct in thyroid FNAC was evaluated by Karakitsos *et al.* as early as 1996 to classify benign and malignant follicular and Hurtle cell lesions. Based on a neural network evaluating the geometric and densimetric nuclear features of approximately 100 nuclei from each of the 51 patients' Giemsa-stained smears, 2 different neural networks were trained and tested. The classifier 3 achieved an overall accuracy of 90.6%, sensitivity of 94.9% and specificity of 98.9%.^[35] CNNs are more often used for analysis in recent studies. A backpropagation algorithm is used in CNN, which has numerous 14 building blocks including convolution layers, pooling layers, and fully connected layers, to enable it to learn spatial hierarchies of features automatically and adaptively.^[36]

10 Sanyal *et al.* created an artificial neural network (ANN) model using 186 microphotographs from Romanowsky/Pap-stained smears of papillary thyroid carcinoma and 184 microphotographs from smears of various thyroid lesions. Two magnifications including x10 and x40 were utilised to train the ANN. Following the training phase, a collection of 174 microphotographs (sixty-six cases of nonpapillary carcinomas and twenty-one cases of papillary carcinomas) shot at x10 and x40 magnifications were used to assess performance. Together, both magnifications

10 revealed good sensitivity (90.48%), moderate specificity (83.33%), very high NPV (96.49%), and 85.06% diagnostic accuracy. However, the vague papillary projections from the benign follicular cells were mistakenly diagnosed as papillary thyroid cancer.^[37]

Dov *et al.* developed and validated a machine learning algorithm (MLA) that can identify regions of interest (ROI) on thyroid FNAC whole slide images (WSI). The study observed almost 100% concordance between the diagnosis of WSI by the cytopathologist and diagnosis using MLA generated ROI image gallery.^[38] Using this MLA, Elliott Range *et al.* developed another MLA aimed at predicting malignancy in thyroid FNAC specimens by utilizing non-pre-processed WSIs. This MLA was able to identify and predict malignancy with a sensitivity of 92% and specificity of 90.5%. The area under the receiver operating curve (AUC) was 0.932, comparable to the AUC of 0.931 achieved by the cytopathologist. When electronic medical record diagnoses are combined into the MLA, the MLA was able to perform even better, with the combined AUC increasing to 0.962. This study showed that MLA would be able to screen WSI to create a gallery of significant ROIs, and to predict TBSRTC and malignancy status for a WSI with excellent sensitivity and specificity equivalent to that of a human cytopathologist.^[39]

Preoperative differentiation of follicular thyroid carcinoma (FC) from follicular adenoma (FA) is challenging. Patients may get unnecessary lobectomies for inevitable histological confirmation.^[40] In order to identify FA from FC, Savala *et al.* attempted to construct an ANN model from the cytological and morphometric aspects of the thyroid FNAC smears. The cytological and morphometric characteristics were analyzed on the FNAC smears from twenty-six histology-proven FA cases and thirty-one histology-proven FC cases. Two independent observers conducted a semi-quantitative analysis of the cytological characteristics. An ANN model was created using this data to distinguish between FA and FC on FNAC material. All cases of FA and FC were effectively discriminated by the ANN model and the AUC was 1.^[41]

Role of AI in risk stratification of indeterminate thyroid nodules

Nodules that have a TBSRTC classification of III, IV, or V following a first evaluation (history, physical examination, ultrasound, and FNAC) are usually referred to as indeterminate thyroid nodules. Up to 40% of all FNACs fall into the uncertain group, which has a malignancy risk between 5% and 75%. We will need to rule out malignancy at the lower end of the range, which is atypia of undetermined significance (AUS) or follicular lesion of undetermined significance (FLUS). Thus, a strategy is needed that has a strong NPV in such lesions and a strategy with a strong PPV towards the higher end of the range.^[42]

Saini *et al.* devised an ANN model to differentiate between malignant and benign cases in FNAC samples of AUS/FLUS in thyroid lesions based on cytological features. Cytological characteristics of twenty-nine cases of histology-proven malignancy, and thirty-two cases that had either been histology-proven to be benign, or for which no progress of malignancy on follow-up had been observed in 2 years were analysed semi-quantitatively by two independent observers and data for ANN was generated. The ANN model successfully distinguished between all benign (5/5) and malignant cases (6/6).^[43]

It may be advantageous to identify the *BRAF* (V600E) mutation in thyroid neoplasia since it is specific for malignancy, portends a poor prognosis, and may be targeted with selective *BRAF* inhibitors. Wang *et al.* explored the use of AI for prediction of *BRAF* mutation in thyroid cytopathology. A total of 118 complete slide pictures were used to assess the proposed DL architecture. The findings demonstrate that the suggested DL-based method achieved an accuracy of 87%, a precision of 94%, a sensitivity of 91%, a specificity of 71%, and a mean of sensitivity and specificity of 81%.^[44]

Role of AI in predicting metastasis in thyroid cancer

Although the differentiated thyroid cancer has a very good prognosis, individuals having metastatic illness on initial presentation often do not fare well.^[45] Although prophylactic lymph node dissection may reduce the likelihood of recurrence and metastasis, extensive surgery may have unfavourable side effects, such as injury to the

laryngeal nerve.^[46] A clinical algorithm that can determine metastasis before surgery is the need of the hour.

Xia *et al.* combined an AI-algorithm and clinico-pathologic data to preoperatively predict the lymph node metastasis in PTC.^[47] In this study, 251 PTC patients with lymph node metastasis and 194 PTC patients without lymph node metastasis were retrospectively studied. The clinico-pathological data including higher ACR TI-RADS score, higher number of tumours, absence of well-defined margin, unclear junction of the lymph node, skin and marrow, presence of rim calcification on US, male sex, young age, large tumour size, and presence of Hashimoto's thyroiditis predict PTCs with central lymph node metastasis than without metastasis. The AI algorithms used in the above study include the support vector machine (SVM) and the probabilistic neural network (PNN) models.^[47] The sensitivity, specificity, and accuracy of the PNN model to predict central lymph node metastasis were found to be 80.4%, 97.7%, and 88.4% respectively. The sensitivity, specificity, and accuracy of the SVM model to predict lateral lymph node metastasis were found to be 100%, 81.8%, and 94.7% respectively.

More recently, Wang *et al.* published an AI based prediction of LN metastasis in patients with PTC using Computed Tomography (CT) images. The AI system based on DL achieved excellent performance with AUCs in the range of 0.84 and 0.8.^[48] Esce *et al.* reported a sensitivity and specificity of 94% and 100%, respectively, when identifying nodal metastases using AI trained on visual data from the primary tumor alone.^[49] Using multivariable logistic regression analysis, Chang *et al.* created an integrated nomogram incorporating DL, clinical traits, and ultrasound data to forecast central lymph node metastasis in PTC patients. 3359 PTC patients who had undergone total thyroidectomy or thyroid lobectomy from two medical centers were divided into three datasets for training, internal validation, and external validation. For the training, internal validation, and external validation cohorts, AUC for the nomogram to predict central lymph node metastasis was 0.812 (95%CI, 0.794-0.830), 0.809 (95%CI, 0.780-0.837), and 0.829 (95%CI, 0.785-0.872) respectively.^[50]

Several studies using various AI models have also reported good accuracy in predicting lymph node metastasis in thyroid cancer.^[51-55] Liu *et al.* established a random forest algorithm model based on machine learning that reliably predicted bone metastasis in patients with newly diagnosed thyroid cancer (with AUC: 0.917 and specificity: 0.905).^[56] Another machine learning model developed by Liu *et al.* could predict lung metastasis from thyroid cancer with an AUC of 0.99.^[57]

Role of AI in prognostication of thyroid cancer

The American Joint Committee on Cancer (AJCC) updated the 8th TNM staging system to better predict the disease-specific survival of PTC patients. However, it has limitations in predicting the likelihood of recurrence and does not consider the biological behaviour of PTC.^[58] Park *et al.* analyzed the prognostic significance of clinico-pathologic factors, including the number of metastatic lymph nodes, lymph node ratio (LNR) and contralateral lymph node metastasis, and attempted to construct a disease recurrence prediction model using machine learning techniques. The study compared the performance of five machine learning models related to prediction of recurrence. The 'Decision Tree' model showed the best accuracy at 95% and the lightGBM and stacking model together showed the accuracy at 93%. All machine learning prediction models showed an accuracy of 90% or more for predicting disease recurrence in PTC.^[59]

Survival prediction is important for both patients and physicians so that they can choose the best management plan. Jajroudi *et al.* used ANN and logistic regression to predict 1-year, 3-year, and 5-year survival in thyroid cancer patients and to compare the accuracy, sensitivity, and specificity. The ANN model was found to be superior to the statistical models in predicting the 1-year survival and found to be equivalent to statistical models in predicting the 3-year, and 5-year survival in thyroid cancer patients. Hence, the ANN model could effectively be used to predict the survival in thyroid cancer patients.^[60]

Future directions

The high repeatability and straightforward implementation of AI in thyroid cancer suggests that it offers significant promise in clinical application, even though there are still many obstacles to overcome before it is extensively employed in clinical practise. Limited clinical experience and lack of prospective validation studies remain the biggest drawbacks with another drawback being the issue regarding the legal responsibility following incorrect diagnosis using AI-based systems. Well conducted prospective validation studies would be able to improve the diagnostic and prognostic performance of these AI-based systems. In ideal world, the trained model should be tested in cross-validation or on an external, independent dataset before being applied on the new dataset. Optimized AI-strategy rather than traditional AI-strategy would improve the performance when used by medical experts.

Supervised learning requires model training using time-intensive and cost-intensive manual labelling of samples by medical experts. In semi-supervised learning both labelled data and unlabelled data are used – labelled data used to train a classifier which is augmented by unlabelled data of the same distribution to derive additional information in order to boost performance.^[61] Supervised learning is advantageous when dealing with a small dataset and medical expert help is available. Semi-supervised learning is advantageous when dealing with a large dataset for which only a limited number of labels can be obtained.

The AI can be useful at various stages in the diagnosis of thyroid nodules including US, cytopathological or histopathological studies and nuclear medicine techniques. The AI-based algorithms has equivalent accuracy to that of experienced radiologists in the US evaluation of thyroid nodules and is particularly helpful for less experienced radiologists. It would significantly reduce the number of unnecessary FNACs. It can be used in the intraoperative diagnosis to decide regarding the degree of radicality of surgery and in the evaluation of thyroid scintigrams as well as SPECT images where it can offer a faster and easier way to evaluate the secretory status of thyroid nodules with least confusion. Finally, it can offer additional assistance for improving the prognostication of thyroid cancer with superior workflow efficiency.

2

The continued development of algorithms capable of selecting regions of interest and performing consistent and accurate diagnoses will enable cytopathologists to focus their attention on the regions of interest, allowing faster and accurate interpretations. Finding a verified and trustworthy algorithm after extensive testing and validation using prospective, multi-centre trials is crucial for the future use of AI in the pipeline of precision medicine in the management of thyroid cancer. A recent study that developed and validated an easily understandable AI report system known as TiNet (easier than the clinical reports) for thyroid cancer prediction is a great development in this aspect.^[62]

CONCLUSION

Artificial intelligence- (AI)- based technologies are revolutionising the diagnosis and management of thyroid cancer in recent years as in other domains of medical practice and clinical care. Development of new scientific advances in this field are expected to further improve the AI technology to help mankind to ensure best care for patients in coming years.

26%

SIMILARITY INDEX

PRIMARY SOURCES

- | | | |
|---|---|----------------|
| 1 | www.researchgate.net
Internet | 256 words — 5% |
| 2 | Charles M. Wong, Brie E. Kezlarian, Oscar Lin. "Current status of machine learning in thyroid cytopathology", Journal of Pathology Informatics, 2023
Crossref | 162 words — 3% |
| 3 | www.ncbi.nlm.nih.gov
Internet | 148 words — 3% |
| 4 | doctorpenguin.com
Internet | 118 words — 2% |
| 5 | 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", American Journal of Roentgenology, 2023
Crossref | 117 words — 2% |
| 6 | www2.mdpi.com
Internet | 93 words — 2% |
| 7 | jamanetwork.com
Internet | 84 words — 2% |
| 8 | www.eurekaselect.com
Internet | |

64 words — 1%

9 www.mdpi.com
Internet

56 words — 1%

10 go.gale.com
Internet

50 words — 1%

11 pubmed.ncbi.nlm.nih.gov
Internet

50 words — 1%

12 www.karger.com
Internet

29 words — 1%

13 www.frontiersin.org
Internet

22 words — < 1%

14 jmeditec.utm.my
Internet

19 words — < 1%

15 link.springer.com
Internet

19 words — < 1%

16 Pimrada Potipimpanon, Natamon Charakorn, Prakobkiat Hirunwiwatkul. "A comparison of artificial intelligence versus radiologists in the diagnosis of thyroid nodules using ultrasonography: a systematic review and meta-analysis", European Archives of Oto-Rhino-Laryngology, 2022

18 words — < 1%

Crossref

17 Luyao Shen. "Editorial Comment: Meta-Analysis Supports ACR TI-RADS for Risk Stratification of Thyroid Nodules Over Other Systems", American Journal of Roentgenology, 2023

16 words — < 1%

Crossref

18 www.scienceworldjournal.org 14 words — < 1%
Internet

19 Antoinette R. Esce, Jordan P. Redemann, Andrew C. Sanchez, Garth T. Olson et al. "Predicting nodal metastases in papillary thyroid carcinoma using artificial intelligence", The American Journal of Surgery, 2021 13 words — < 1%
Crossref

20 Bong-Keun Choe, Su Kang Kim, Hae Jeong Park, Hyun-Kyung Park, Kee Hwan Kwon, Sung Hoon Lim, Sung-Vin Yim. "Polymorphisms of TGFBR2 contribute to the progression of papillary thyroid carcinoma", Molecular & Cellular Toxicology, 2012 13 words — < 1%
Crossref

21 William C. Faquin, Massimo Bongiovanni, Peter M. Sadow. "Update in Thyroid Fine Needle Aspiration", Endocrine Pathology, 2011 13 words — < 1%
Crossref

22 dokumen.pub 12 words — < 1%
Internet

EXCLUDE QUOTES ON
EXCLUDE BIBLIOGRAPHY ON

EXCLUDE SOURCES OFF
EXCLUDE MATCHES < 12 WORDS