J. Vis. Commun. Image R. 115 (2026) 104668

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Review

Machine learning and transformers for thyroid carcinoma diagnosis[☆]

Yassine Habchi [a], Hamza Kheddar [b], Yassine Himeur [c], Mohamed Chahine Ghanem [d,*]

[a] Electrical Engineering Department, University of Salhi Ahmed, Naama, 45000, Algeria
[b] LSEA Laboratory, Electrical Engineering Department, University of Medea, 26000, Algeria
[c] College of Engineering and Information Technology, University of Dubai, Dubai, United Arab Emirates
[d] School of Computer Science and Informatics, University of Liverpool, Liverpool, L693BX, UK

ARTICLE INFO

Keywords:
Thyroid carcinoma detection
Artificial intelligence
Deep learning
Biomedical images
Transformers
Large language model

ABSTRACT

Thyroid carcinoma (TC) remains a critical health challenge, where timely and accurate diagnosis is essential for improving patient outcomes. This review provides a comprehensive examination of artificial intelligence (AI) applications — including machine learning (ML), deep learning (DL), and emerging transformer-based approaches — in the detection and classification of TC. We first outline standardized evaluation metrics and analyze publicly available datasets, highlighting their limitations in diversity, annotation quality, and representativeness. Next, we survey AI-driven diagnostic frameworks across three domains: classification, segmentation, and prediction, with emphasis on ultrasound imaging, histopathology, and genomics. A comparative analysis of ML and DL approaches illustrates their respective strengths, such as interpretability in smaller datasets versus automated feature extraction in large-scale imaging tasks. Advanced methods leveraging vision transformers (ViT) and large language models (LLMs) are discussed alongside traditional models, situating them within a broader ecosystem of feature engineering, ensemble learning, and hybrid strategies. We also examine key challenges — imbalanced datasets, computational demands, model generalizability, and ethical concerns — before outlining future research directions, including explainable AI, federated and privacy-preserving learning, reinforcement learning, and integration with the Internet of Medical Things (IoMT). By bridging technical insights with clinical considerations, this review establishes a roadmap for next-generation TC diagnostics and highlights pathways toward robust, patient-centric, and ethically responsible AI deployment in oncology.

Contents

[☆] This paper has been recommended for acceptance by Junsong Yuan.

* Corresponding author.
E-mail addresses: habchi@cuniv-naama.dz (Y. Habchi), kheddar.hamza@univ-medea.dz (H. Kheddar), yhimeur@ud.ac.ae (Y. Himeur), mohamed.chahine.ghanem@liverpool.ac.uk (M.C. Ghanem).

https://doi.org/10.1016/j.jvcir.2025.104668
Received 20 April 2025; Received in revised form 19 November 2025; Accepted 1 December 2025
Available online 11 December 2025

1. Introduction

Recent global epidemiological studies indicate a rise in abnormal TN, linked to an upsurge in genetic cellular activity. This suggests an increase in normal cell functions, with anomalies classified into four primary types: follicular thyroid carcinoma (FTC), papillary carcinoma (PTC), medullary thyroid carcinoma (MTC), and anaplastic thyroid carcinoma (ATC) [1–4]. Elements like exposure to radiation, Hashimoto's thyroiditis, psychological factors, and genetic components, alongside advances in detection technologies, appear to contribute to these cancers. These factors can cause chronic health issues like diabetes and blood pressure instability. The cancer cell volume is a key to evaluating the aggressiveness and prognosis of TCs, with cell nuclei detection offers alternative markers for assessing cancer cell proliferation. Computer-aided diagnosis (CAD) systems have gained prominence in TCD, improving diagnostic accuracy and reducing interpretation times [5]. Radiomics, particularly through US imaging [6], has emerged as an efficient diagnostic method. The American College of Radiology's thyroid imaging reporting and data system (TIRADS) categorizes TNs from benign to malignant [7]. Despite available open-source tools for nodule analysis, accurately identifying them remains a challenge, reliant on radiologists' experience and the subjective nature of visual image analysis [8]. Additionally, US imaging can be a lengthy and stress-inducing process, which may result in incorrect diagnoses. It is common to encounter classification errors among cases deemed normal, benign, malignant, or of uncertain nature [9,10]. For a more precise diagnosis, a fine-needle aspiration biopsy (FNAB) is often conducted. Yet, this technique can be uncomfortable for patients, and inaccuracies by the practitioner can mistakenly label benign nodules as malignant, leading to unnecessary costs [11]. The main issue is the selection of nodule characteristics critical for accurately differentiating between benign and malignant cases. Various research efforts have delved into the use of conventional US imaging to characterize different types of cancers. Despite these efforts, there remains a limited accuracy of existing methods for effectively categorizing TNs.

The integration of AI into the healthcare sector represents a pivotal advancement, enabling transformative changes in medical diagnostics, therapy, and patient management. The superior capabilities of AI, including pattern recognition, predictive analytics, and decision support, have led to systems that can interpret complex medical data with unprecedented accuracy and scalability [12,13]. Recent studies have shown that specialized AI architectures can substantially enhance the analysis of complex biomedical imagery. Beyond healthcare, AI exhibits transformative potential across multiple domains. For example, vision-based vehicle re-identification systems benefit from AI through improved tracking, identification, and predictive capabilities in intelligent transportation systems [14]. In healthcare, contextual and local feature extraction techniques in retinal imaging have improved the accurate detection of retinal diseases [15], while multi-module attention mechanisms have enhanced gastrointestinal disease identification from endoscopic imagery [16]. Additionally, innovative representations such as octonion-based transform moments combined with DL enable more robust stereo image classification, with applications in advanced medical imaging and telemedicine [17]. These advances demonstrate that AI facilitates early disease detection, increases diagnostic accuracy, supports personalized treatment strategies, and improves operational efficiency in healthcare. Moreover, AI deployment has the potential to reduce disparities in access to medical services, bridging the gap between rural and urban care, and ultimately contributing to more equitable healthcare delivery. As AI technologies continue to evolve, their applications are expected to expand across both medical and non-medical domains, providing substantial benefits to patient outcomes and societal systems alike [18,19].

The deployment of AI technology is crucial in diminishing subjectivity and boosting the precision of pathological assessments, particularly for complex conditions like thyroid diseases [20]. These advancements enhance the analysis of images obtained through US and expedite analysis times. ML and DL have proven effective in automating the differentiation of TN through various imaging techniques such as US and fine-needle aspiration. Studies have shown that these techniques provide high accuracy in classifying nodules as benign or malignant, enhancing the ability of clinicians to make more accurate treatment decisions. Furthermore, DL has been successfully applied during thyroid surgical procedures, aiding surgeons in accurately identifying tumors, which contributes to improved surgical efficiency and reduces risks. This claim is supported by numerous studies, including [21], which demonstrate that these methods significantly improve diagnostic accuracy and reduce costs, making them reliable and innovative tools in modern healthcare. Traditional methods for diagnosing TC, like fine-needle aspiration biopsies, can often produce ambiguous outcomes, whereas AI presents an opportunity for more accurate and less invasive alternatives.

This review seeks to integrate insights from pathology, computer science, endocrinology, and radiology, encouraging cross-disciplinary collaboration. It will also explore the clinical significance of AI, offering recommendations for healthcare professionals on utilizing AI advancements for improving patient care, and pinpointing directions for future research endeavors. Additionally, the review discusses the healthcare and economic system benefits, including cost savings and reduced wait times. Yet, it is essential to confront the challenges AI brings, such as ethical considerations and data privacy, to facilitate its responsible integration into healthcare practices. This review aims to highlight that summarizing and synthesizing research is crucial to identify gaps, avoid duplication, and guide evidence-based clinical practice. It emphasizes that AI technologies — such as ML, DL, and transformers — can transform TC diagnosis by improving accuracy, reducing subjectivity, and enabling early detection. The review comprehensively examines AI's current and future impact, including datasets, metrics, clinical

Abbreviation list

AI	Artificial intelligence
ANN	Artificial neural network
ATC	Anaplastic thyroid carcinoma
ATM	Assembled Transformer module
AUC	Area under curve
BERT	Bidirectional encoder representations from Transformer
Bi-LSTM	Bi-directional LSTM
BPSM	Boundary point supervision module
CAD	Computer-aided diagnosis
CEUS	Contrast-enhanced ultrasound
CFS	Correlation-based feature selection
CNN	Convolutional neural network
CT	Computed tomography
DAE	Denoising autoencoder
DCNN	Deep convolutional neural network
DDTI	Digital database thyroind image
DL	Deep learning
DNN	Deep neural network
DT	Decision tree
DTCW	Double-tree complex wavelet transform
DWT	Discrete wavelet transform
ELM	Extreme learning machine
EM	Ensemble method
FB	Feature bagging
FCM	Fuzzy c-means
FL	Federated learning
FNAB	Fine-needle aspiration biopsy
FTC	Follicular thyroid carcinoma
GAN	Generative adversarial network
GEO	Gene expression omnibus
GLCM	Gray-level co-occurrence matrix
HOG	Histogram of oriented gradient
ICA	Independent component analysis
IG	Information gain
IoMT	Internet of medical thing
KM	K-means
LBP	Local binary patterns
LLM	Large language model
LR	Logistic regression
LSTM	Long-short-term-memory
ML	Machine learning
MLP	Multilayer perceptron
MRI	Magnetic resonance imaging
MRM	MicroRNA regulatory module
MSE	Mean square error
MTC	Medullary thyroid carcinoma
NLP	Natural language processing
OCT	Optical coherence tomography
PCA	Principal component analysis
PFCM	Possibilistic fuzzy c-means
PLCO	Prostate, lung, colorectal, and ovarian
PM	Probabilistic models
PS	Panoptic segmentation
PTC	Papillary carcinoma
RA	Relevance analysis
RF	Random forest
RL	Reinforcement learning
RMSE	Root mean square error
RNN	Recurrent neural network
ROI	Region of interest
SL	Supervised learning
SSL	Self-supervised learning
SVM	Support vector machine
TC	Thyroid cancer
TCD	Thyroid cancer diagnosis
TD	Thyroid disease
TFD	Discrete Fourier transforms
TG	Thyroid gland
TIRADS	Thyroid imaging reporting and data system
TL	Transfer learning
TN	Thyroid nodules
UCI	University of California, Irvine
US	Ultrasound
USL	Unsupervised learning
ViT	Vision Transformer
VLM	Vision language models
WSI	Whole slide histopathological images
XAI	Explainable artificial intelligence
XGBoost	Gradient tree boosting

applications, and challenges like interpretability and privacy, with the goal of advancing patient-centered, precise, and ethically responsible diagnostic solutions that enhance outcomes and quality of care.

1.1. Related work and contribution of our paper

Previous works on TC detection have advanced DL and transformer applications, yet several limitations remain. Liu et al. [22] and Iesato et al. [23] focused narrowly on transformers without exploring AI integration, explainability, or deployment in internet of medical thing (IoMT). Sharifi et al. [24] and Ha et al. [25] incorporated DL but lacked reinforcement learning (RL), explainable artificial intelligence (XAI), and multi-network integration—leaving gaps in adaptability and trustworthiness. Lin et al. [26] and Wu et al. [27] employed transformer-based approaches but did not cover patient privacy preservation or prospective extensions, such as recommender systems and panoptic segmentation (PS). Pavithra et al. [28] emphasized deep methods but missed edge-fog-cloud AI orchestration opportunities Paul et al. [29] and Ilyas et al. [30] advanced detection accuracy but remained limited to conventional DL and transformers, neglecting LLMs, IoMT integration, and multi-path frameworks. The distinctive contributions of this review, in comparison with prior studies, are outlined in Table 1. The principal advancements introduced in our work can be summarized as follows:

- A review of current frameworks coupled with a detailed investigation into diverse AI strategies, including supervised learning (SL), conventional classification, unsupervised learning (USL), DL, transformer, and LLM techniques.
- An in-depth review of various TCDs approaches, detailing their attributes and examining feature selection and extraction methods used in different studies.
- A detailed discussion on the benchmark criteria for assessing the efficacy of AI-powered approaches in detecting TC. These

Table 1
The notable advancements made by the suggested review in the categorization of TC when contrasted with similar research endeavors.

Ref	Year	Patient privacy	TC detec. schemes	AI apps	ML	DL	Trans	LLM	F	TCE	Prospective path						Metric
											IoMT images	RS	RL	PS	XAI	EFC-AI	
[22]	2021	✓	✓	✗	✗	✗	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗
[23]	2021	✓	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[24]	2021	✓	✓	✗	✗	✓	✓	✗	✓	✓	✗	✗	✗	✗	✗	✓	✓
[26]	2021	✓	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓
[25]	2021	✓	✓	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓
[27]	2022	✓	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
[28]	2022	✓	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✓	✓	✗
[29]	2022	✓	✓	✗	✗	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗	✓	✓
[30]	2022	✓	✓	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓
Our	2024	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Abbreviations: Artificial intelligence applications (AI apps); Transformers (Trans.); Features (F); TC example (TCE); Internet of medical imaging thing (IoMT); Recommender systems (RS); Reinforcement learning (RL); Edge, fog and cloud networks based on AI (EFC-AI).

evaluation metrics cover a wide range, from regression and classification parameters to statistical, computer vision, and ranking parameters.

- A spotlight on the transformative impact of AI in enhancing TC diagnosis, stressing the importance of continual critical evaluation to ensure ethical and effective application.
- A thorough critique and exploration of the challenges, limitations, prevalent trends, and unresolved questions in the domain.
- An analysis of future research priorities, highlighting specific areas that require further investigation to address existing challenges and enhance methods for TC detection.

1.2. Bibliometric analysis

A bibliometric analysis was performed to delve into and evaluate the scientific literature reviewed in this paper. The continuous interest in AI-based TC research is depicted in Fig. 1, with the publication count reaching 81 in 2022. Fig. 1(a) highlights the leading researchers in the field of TC-oriented AI research, focusing on those who have published within the past five years. Fig. 1(b) illustrates the enduring interest in AI-based TCD research, showcasing a rising trend in the development of AI solutions for TC prognosis and diagnosis since 2015. This upward trend reflects the transition of AI in TC from exploratory studies to more established clinical research, although fluctuations after 2022 suggest a shift toward emerging methods such as transformers and multimodal learning Fig. 1(c) maps the countries that are major contributors to the research output in this area, with China and the United States showing a strong focus on AI-driven TC detection. China's surge mirrors strong national investment in AI, while the U.S. benefits from academic–clinical collaborations. Other Asian countries — such as South Korea, India, and Japan — contribute meaningfully, whereas Africa and South America remain underrepresented, indicating global disparities and missed opportunities for wider collaboration. Lastly, Fig. 1(d) illustrates the breakdown of publication types, with journal articles constituting the majority of research (67.2%), followed by conference papers (19.3%). This dominance of journal articles suggests maturity in the field, yet the scarcity of reviews (1.1%) and book chapters (0.4%) highlights the need for more integrative syntheses.

All bibliometric and comparative data presented in the figures were collected from Scopus-indexed publications. Figures summarizing model performance metrics were generated using Microsoft Excel as bar charts. To allow fair comparison across studies, reported metrics (e.g., accuracy, sensitivity) were standardized in a consistent format (0%–100%).

1.3. Review methodology

To identify and review existing studies on ML and transformers for TC diagnosis, and based on the PRISMA guidelines, a comprehensive search was conducted across several leading publication databases renowned for their high-quality scientific research. A flow chart diagram is presented in Figure (b) 2. The primary search was carried out in Scopus, which systematically includes databases such as Web of Science, Elsevier, IEEE, ACM Digital Library, and Wiley, among others. Articles published between 2017 and 2025 were given priority. However, older publications were also considered when necessary to provide a historical context, dataset, metrics, etc. The review focused on computer science and engineering studies from databases including IEEE Xplore, ScienceDirect, Wiley, Springer, and Taylor Francis. Additionally, considering the recent and trending nature of the topic, high-quality pre-prints from arXiv were selected. Only articles written in English were included in the final analysis.

- **Explicit inclusion criteria:** Research explicitly discussing ML and Transformers or LLMs in the context of TC diagnosis. Studies must be published in indexed journals or conferences, ensuring they meet established academic standards and visibility.
- **Explicit exclusion criteria:** Studies that do not focus on the specified technological applications or topics. Research ambiguously referencing "Transformers" in non-computational contexts, such as electrical engineering.

1.4. Roadmap

The subsequent sections of this manuscript are organized as depicted in Fig. 2(a), where: Section 2 discusses standardized assessment criteria and commonly used TC datasets, emphasizing the importance of metrics and dataset features for AI-driven analysis in TCD. Section 3 provides a summary of current AI-based models and methods for TCD, including classification, segmentation, and prediction frameworks. Section 4 explores advanced TCD methods leveraging vision Transformer (ViT) and LLM, focusing on their segmentation, classification, and prediction applications. Section 5 outlines the limitations and challenges of existing AI methodologies, highlighting gaps in accuracy, data variability, and model generalizability. Section 6 discusses future research directions, emphasizing innovative approaches to advance AI technologies in TCD. Finally, Section 7 presents the conclusion, synthesizing key insights and reaffirming the transformative potential of AI in medical diagnostics while offering perspectives for its integration into TCD practices.

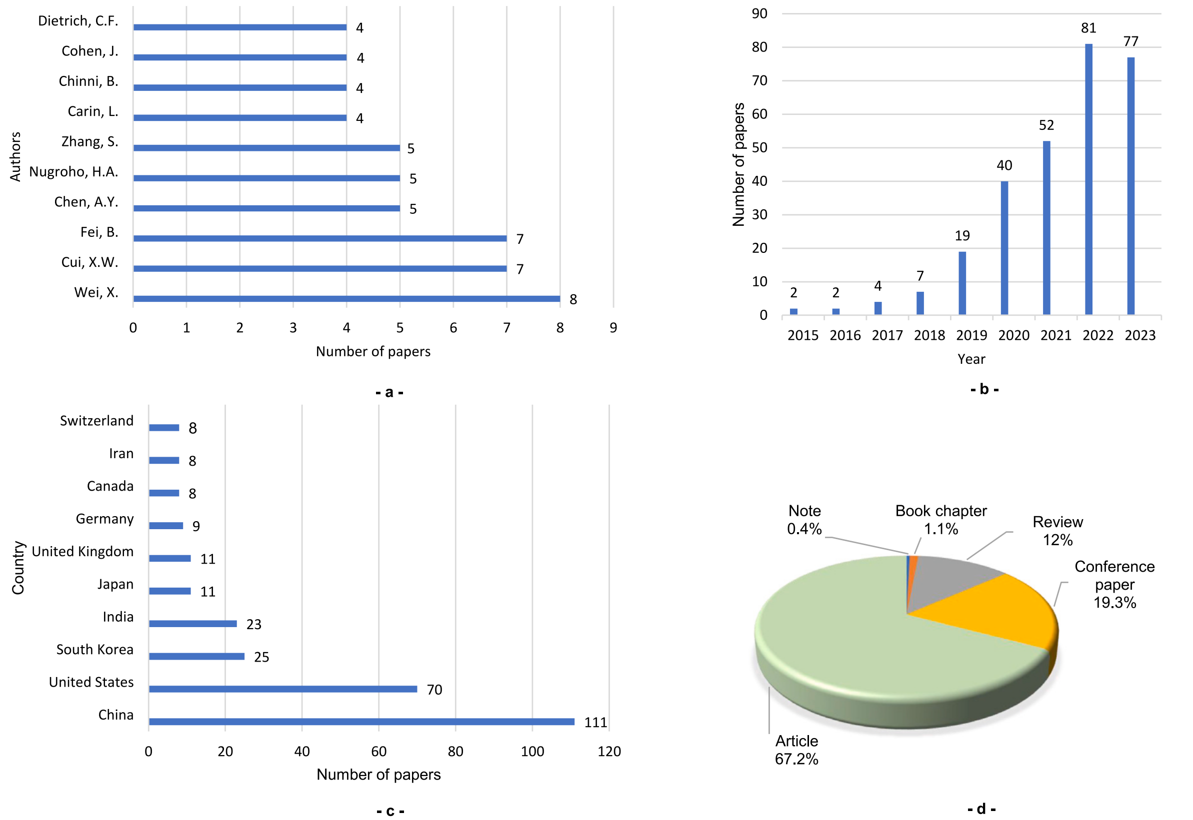

Fig. 1. Bibliometric analysis in terms of: (a) documents by author; (b) documents by year; (c) documents by country; (d) documents by type.

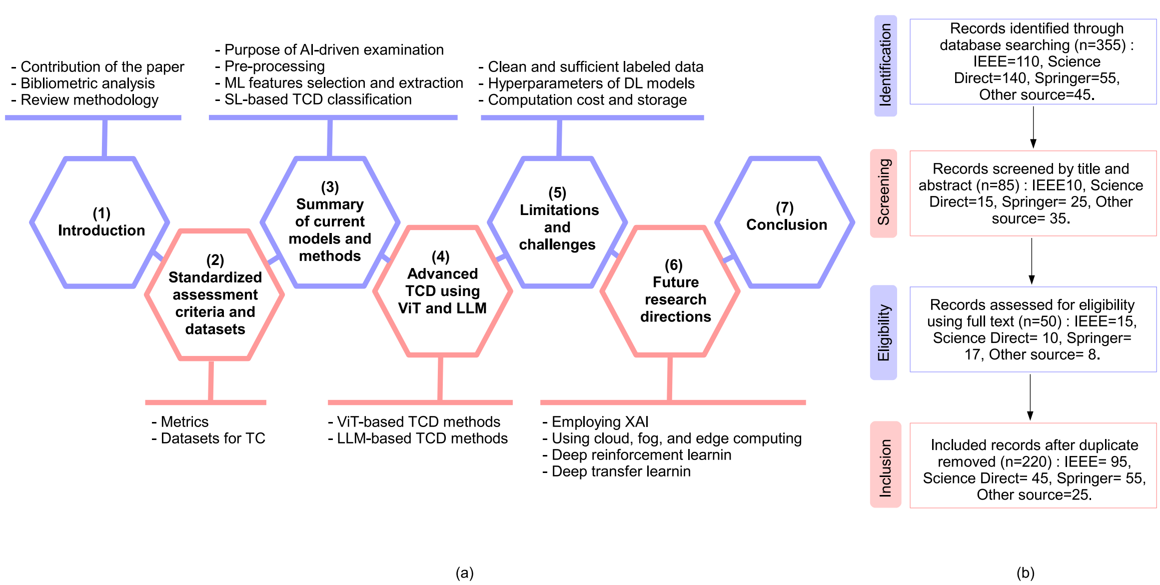

Fig. 2. An overview of the review organization, and a set of guidelines study selection process: (a) A roadmap outlining the sections in this review. (b) PRISMA diagram for selection of articles.

2. Standardized assessment criteria and commonly used datasets

2.1. Metrics

In this segment, we explore the standard metrics commonly utilized for evaluating thyroid disease (TD) detection performance. These metrics act as critical benchmarks for measuring the success of methodologies, underscoring the significance of choosing the right metrics to assess ML models. A variety of DL metrics are used to determine the efficiency of the suggested method in identifying TD. It is important to note that certain metrics have been previously addressed in [31]. The other metrics, particularly designed for image processing applications in TD, are concisely outlined in Table 2.

2.2. Datasets for TC

In the context of TC research, numerous datasets have been developed to support the testing and validation of ML algorithms and

Table 2
Summary of the metrics for classification and regression employed in assessing AI-driven methods for TCD.

	Metric	Mathematical formula	Description
Classification and regression	Specificity	$\frac{T_N}{T_N+F_P}100\%$	This metric represents the proportion of accurately predicted negative samples out of all the negative samples.
	Root mean square error (RMSE)	$\left(\sqrt{1-(ER)^2}\right)\times SD$	This is the standard deviation of the predicted errors between the training and testing datasets, and a lower value indicates the classifier's excellence.
	Jaccard Similarity Index (JSI)	$\frac{\lvert A\cap B\rvert}{\lvert A\cup B\rvert}=\frac{T_P}{T_P+F_P+F_N}$	Paul Jaccard introduced this method to measure both the similarity and diversity among samples.
	Volumetric overlap error (VOE)	$\frac{F_P+F_N}{T_P+F_P+F_N}$	Assess the likeness between the segmented area and the ground truth area. VOE quantifies the level of overlap between these two regions and is calculated as the ratio of the combined volume of the segmented and ground truth regions to the volume of their intersection.
	Mean absolute error (MAE)	$\frac{1}{N}\sum_{i=1}^{N}\lvert a_i-p_i\rvert$	This measure indicates the average of the disparities between the real values and the predicted values.
Statistical	Standard deviation (SD)	$\sqrt{\sum(x-\mu)^2/N}$	It quantifies the degree of variability or spread within a dataset.
	Correlation	$\frac{(\sum((x-\mu x)\cdot(y-\mu y)))}{(\sqrt{(\sum(x-\mu x)^2)}\cdot\sqrt{(\sum(y-\mu y)^2)})}$	It characterizes the extent of correlation or connection between two or more variables.
	Mean reciprocal rank (MRR)	$MRR=\frac{1}{\lvert Q\rvert}\sum_{i=1}^{\lvert Q\rvert}\frac{1}{rank_i}$	The MRR is a statistical measure used to assess the average reciprocal rank of outcomes for a set of queries, as explained in [32]. Here, "$rank_i$" denotes the position at which the first relevant document appears for the ith query.
	Kappa de Cohen	$k=\frac{\Pr(a)-\Pr(e)}{1-\Pr(e)}$	This metric gauges the level of agreement between two assessors, considering chance as a baseline.
Computer vision	Peak signal to noise ratio (PSNR)	$10\cdot\log_{10}((MAX_I^2)/MSE)$	It quantifies the proportion between the highest achievable signal power and the power of the noise that impacts the faithfulness of its portrayal.
	Visual information fidelity (VIF)	$\frac{\sum_j I(C^j;F^j/s^j)}{\sum_j I(C^j;E^j/s^j)}$	It assesses the excellence of a reconstructed or compressed image or video in relation to the original signal. This evaluation considers how much visual information is retained in the processed image or video, accounting for the image's spatial and frequency attributes.
	Normalized cross-correlation (NCC)	$\frac{\sum_{i=1}^{M}\sum_{i=1}^{N}(I(i,j)-R(i,j))^2}{\sum_{i=1}^{M}\sum_{i=1}^{N}I(i,j)^2}$	Assess the likeness between two images (or videos) by subtracting the mean value from each signal and subsequently normalizing the signals by dividing them by their standard deviation. Finally, compute the cross-correlation between the two normalized signals.
	Structural content (SC)	$\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}I(i,j)^2}{\sum_{i=1}^{M}\sum_{j=1}^{N}R(i,j)^2}$	An elevated structural content value indicates that the image possesses lower quality.
	Noise visibility function (NVF)	$\text{Normalization}\left\{\frac{1}{1+\delta_{bloc}^2}\right\}$	This calculates the texture information within the image, where δ_{bloc} represents the variance in luminance.
	Visual signal to noise ratio (VSNR)	$10\log_{10}\left(\frac{C^2(I)}{(VD)^2}\right)$	This approach sets distortion thresholds using contrast computations and wavelet transforms. VSNR is deemed excellent if distortions are below the threshold. It uses RMS contrast ($C(I)$) and visual distortion (VD).
	Weighted signal-to-noise ratio(WSNR)	$10\log_{10}\left(\frac{\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}\lvert A(u,v)C(u,v)\rvert^2}{\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}\lvert A(u,v)-B(u,v)C(u,v)\rvert^2}\right)$	It relies on the contrast sensitivity function, with $A(u,v)$, $B(u,v)$, and $C(u,v)$ denoting the 2D discrete Fourier transforms (TFD), as described in [33].
	Normalized absolute error (NAE)	$\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\lvert I(i,j)-R(i,j)\rvert}{\sum_{i=1}^{M}\sum_{j=1}^{N}\lvert I(i,j)\rvert}$	This metric assesses the precision of an ML model's predictions by quantifying the discrepancy between predicted and actual values relative to the range of actual values.
	Laplacian mean square error (LMSE)	$\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}[L(I(i,j))-L(R(i,j))]^2}{\sum_{i=1}^{M}\sum_{j=1}^{N}[L(I(i,j))]^2}$	It is a modified version of mean square error (MSE), utilizing the Laplacian distribution instead of the Gaussian distribution. $L(I(i,j))$ represents the Laplacian operator.

models. This step is crucial, given the significant challenge of compiling these datasets within the endocrine ML field. Table 3 lists examples of publicly available TCDs.

The datasets used in DL-based TC image analysis face several limitations. One of the main issues is the lack of diversity, as many datasets focus on specific populations, limiting the generalizability of models. Additionally, some datasets are biased, with an over representation of advanced cancer cases, leading to models that may not perform well with early-stage cases. Inconsistent labeling and varying image quality across different datasets further impact the accuracy of the models. These limitations highlight the need for more diverse and accurately labeled datasets, along with standardized criteria, to improve the effectiveness of models in clinical applications.

3. Summary of current models and methods

This section discusses the various AI-based methodologies utilized for diagnosing thyroid gland (TG) cancers. It outlines the key computational strategies — ranging from traditional machine learning classifiers to advanced deep learning frameworks — and highlights their roles in tasks such as classification, segmentation, and prediction. The discussion emphasizes both the technical foundations and the clinical relevance of these approaches, setting the stage for a detailed examination of their applications in TCD.

3.1. Purpose of AI-driven examination

This review centers on the use of AI for identifying TC. Comprehending the foundational objectives of each framework is essential for acquiring a more profound understanding of their reasoning. The process of segmentation isolates the region of interest (ROI) from medical images, classification determines whether nodules are benign or malignant based on extracted features, and prediction models estimate disease risk and progression using patient data. Together, these steps form an AI-driven pipeline that improves diagnostic precision, supports treatment planning, and enables proactive management of TC. These process are detailed in the following:

3.1.1. Segmentation of TC

Segmentation plays a pivotal role in the detection of TC by enabling the accurate isolation and analysis of the thyroid gland, along with any

Table 3
Instances of publicly available TCDs utilized in the identification of TC.

TCD	Description	NoI	NoC	NoF	AT	Link
THO	This dataset is designed to investigate the fundamental causes and effects of TD through the application of diverse omics approaches, including genomics, epigenomics, transcriptomics, proteomics, and metabolomics.	54.288	2	13	C	Visit THO datasets
KEEL	The KEEL dataset offers a collection of benchmarks for assessing the performance of different learning approaches, including semi-supervised classification and USL.	7.2K	3	21	C	Visit KEEL datasets
GEO	The gene expression omnibus (GEO) database serves as a repository for genomics data. It is specifically structured to archive gene expression datasets, arrays, and sequences within GEO.	140K	–	–	C	Visit GEO datasets
DDTI	The digital database thyroind image (DDTI) dataset acts as an essential tool for both researchers and novice radiologists aiming to create algorithm-driven CAD systems for analyzing TN.	134	2	–	C	Visit DDTI datasets
PLCO	The National Cancer Institute backs the prostate, lung, colorectal, and ovarian (PLCO) cancer screening trial, which focuses on identifying the primary factors influencing cancer incidence in both genders and includes comprehensive studies on TC incidence and mortality.	155K	2	–	C	Visit PLCO datasets
CIFAR-10	The CIFAR-10 datasets are collections of color images used for image classification. CIFAR-10 contains distinct classes of images. This datasets are widely used for training and testing ML models.	60K	10	–	–	Visit CIFAR-10 datasets
TCGA	The TCGA dataset provides comprehensive molecular data on various cancer types, including gene expression, DNA mutations, and clinical information. It aims to enhance understanding of cancer biology and facilitate personalized medicine by offering detailed genomic profiles for research into cancer diagnosis, prognosis, and treatment strategies.	>11K	2	–	–	Visit TCGA datasets
UCI	UCI datasets are a collection of publicly available datasets used for ML research and experimentation. These datasets cover a wide range of topics, including classification, regression, and clustering tasks. They are widely utilized by researchers and practitioners to benchmark algorithms and evaluate models.	7.2K	–	5	C	Visit UCI datasets
SEER	The SEER Program provides authoritative cancer statistics in the U.S. It aims to reduce the cancer burden by offering detailed data on cancer incidence and trends, supported by the National Cancer Institute's Surveillance Research Program. SEER resources include cancer statistics explorers and annual reports.	≈197k	3	21	C	Visit SEER datasets
TN3K	The TN3K TN Region Segmentation Dataset, available on Kaggle, consists of image files used for TN segmentation. It contains various images and corresponding masks, specifically designed to aid in developing machine learning models for medical image segmentation tasks. The dataset helps researchers improve automatic detection and analysis of TN.	3493	2	–	S	Visit TN3K datasets
GBD	The Global Burden of Disease (GBD) Study 2019 provides extensive data on health conditions, including thyroid cancer, categorized into various types. It uses ICD-9 and ICD-10 diagnosis codes, ensuring patient privacy. The dataset is shared under a Creative Commons license, adhering to ethical guidelines and legal standards.	–	5	–	C	Visit GBD datasets
CI5	The Cancer Incidence in Five Continents (CI5) compendium, maintained by the International Agency for Research on Cancer (IARC), provides high-quality, population-based cancer incidence data. It includes information on various cancer types, including TC, across multiple countries, offering insights into global trends, patterns, and variations in cancer incidence.	–	4	–	–	Visit CI5 datasets
WEKA	Weka datasets are collections of data used for ML experiments and analysis. They are available in various formats, including ARFF, and cover a wide range of domains such as healthcare, finance, and biology. These datasets are used with Weka software for tasks like classification, regression, clustering, and pattern recognition.	3772	3	29	C	Visit WEKA datasets
UCLA	The UCLA Thyroid RadPath Research Dataset is a comprehensive collection of thyroid ultrasound (US) images and associated clinical data, developed over 13 years at UCLA Health. It aims to advance machine learning applications in TC diagnosis.	280	2	41	C	Visit UCLA datasets
ATC	The ATC Imaging Cases dataset consists of US images from patients diagnosed with ATC. These images are used to improve AI models for diagnosing rare TC subtypes, aiming to enhance diagnostic accuracy and generalization for such cancers.	57	1	–	–	Visit ATC datasets

Abbreviations: Number of instances (NoI); Number of classes (NoC); Number of features (NoF); Number of images (NoIg); Associated tasks (AT); Classification (C).

potentially suspicious nodules or lesions present [34]. Often detected through medical imaging modalities like US, computed tomography (CT) scans, or magnetic resonance imaging (MRI), TCD requires the precise delineation of the ROI for sound diagnostic judgements. Segmentation not only facilitates the distinction of the thyroid gland from the surrounding tissues but also supports the precise quantification of nodule dimensions and volume, which are critical for evaluating the potential for malignancy. Furthermore, it enables the extraction of significant image attributes such as texture and shape, providing critical data for ML models or other analytical methods to improve diagnostic precision. Segmented images also enhance visual clarity, aiding radiologists and medical practitioners in the visual assessment and interpretation of areas of concern within the thyroid, vital for detecting abnormalities indicative of cancer. For longitudinal analyses, segmentation is invaluable in tracking changes in the thyroid and nodules over time, monitoring disease evolution or the efficacy of treatments. It also plays a role in accurately locating biopsy sites for suspicious nodules, guaranteeing targeted sample collection for cancer verification. In the context of treatment planning, segmentation is instrumental in assessing the tumor's size and its relationship to vital anatomical structures, thereby guiding therapeutic decisions. Moreover, the introduction of automated segmentation technologies streamlines clinical workflows by minimizing manual input and variability, empowering medical experts to dedicate more attention to complex diagnostic activities. Consequently, the segmentation process in TCD enhances precision, consistency, and confidence in diagnostics, markedly impacting patient management and outcomes. AI methods, especially CNN and the U-Net architecture, are becoming progressively popular for the segmentation of TC. Their growing preference is largely due to their capacity to learn from and generalize across large datasets, significantly improving the accuracy and dependability of the segmentation procedure.

3.1.2. Classification of thyroid carcinoma

Entails the sorting of TCs according to their histopathological features, clinical manifestations, and prognostic outcomes. Various forms of TCs exist, each defined by unique characteristics. The principal categories are:

(i) PTC: Representing the most prevalent form, PTC constitutes approximately 80% of all TC cases. It typically exhibits slow growth but has a propensity to metastasize to neck lymph nodes. Nonetheless, PTC generally responds well to treatment.
(ii) FTC: Ranking as the second most frequent type, FTC has the capability to invade blood vessels and spread to distant body parts, although it is less prone to lymph node metastasis.
(iii) MTC: Arising from the parafollicular or C cells of the thyroid, which secrete calcitonin, an increase in blood calcitonin levels may signal MTC.
(iv) ATC: Is a highly aggressive and rare TC variant, characterized by its rapid spread to other neck regions and the body, making it challenging to treat.

The stratification of TCs is vital for selecting the optimal treatment plan for individual patients, considering tumor dimensions, location, patient age, and general health. The evolution of AI and ML has significantly contributed to the automation and enhancement of TC classification accuracy. Various models have been devised to categorize tumors based on medical imaging or genetic information. For example, Liu and colleagues [35] highlight the fundamental importance of support vector machine (SVM) in the detection of cancer. In a similar vein, Zhang and their team [36,37] introduce approaches utilizing deep neural network (DNN) to distinguish between malignant and benign TNs in US imagery. Moreover, the bi-directional LSTM (Bi-LSTM) model [38], shows noteworthy precision in the classification of TNs. These classification approaches create structured hierarchies crucial for organizing knowledge and processes in the field of TCD.

3.1.3. Prediction of TC

Prediction in TCD involves utilizing diagnostic tools and ML models to estimate the risk of development based on factors like genetic predisposition, gender, age, radiation exposure, and lifestyle choices. It is important to note that predictions indicate a heightened risk rather than a definite outcome. Medical practices often combine various predictive assessments to improve accuracy. For example, ML algorithms developed from medical records can help distinguish between nodules, facilitating early intervention. Studies, such as one utilizing artificial neural network (ANN) and logistic regression (LR), and another employing a CNN to analyze over 10,000 microscopic TCD (images) [39], demonstrate the application of predictive techniques in identifying TC risk, showcasing advancements in AI-driven predictive modeling for more effective treatment strategies.

3.2. Pre-processing

Principal component analysis (PCA) operates as a sophisticated method for preprocessing, transforming samples (variables) into a smaller set of uncorrelated ones. This technique effectively reduces the volume of variables, thus cutting down on redundant data, all the while striving to maintain the integrity of the data relationships. PCA is extensively used in the realms of cancer detection and distinguishing between malignant and benign thyroid cells. In research conducted by Shankarlal et al. [40], PCA was deployed to filter the most relevant set of wavelet coefficients from double-tree complex wavelet transform (DTCW) processed noisy thyroid images, which were then categorized using random forest (RF). In another instance, Soulaymani et al. [41] applied PCA to a dataset comprising 399 patients with three different TC, allowing the classification based on variables like age, sex, type of cancer, and geographical location.

3.3. ML features selection and extraction

Feature selection and extraction are essential for building accurate ML models in TC diagnosis. Selection identifies the most informative attributes while removing irrelevant ones, improving interpretability and efficiency. Extraction transforms raw data into new representations that capture key patterns and enhance class separability. This has been explained in the following:

3.3.1. Selection methods

The primary goal of selection process is to identify and select relevant features that can enhance the accuracy of classification, simultaneously eliminating non-essential variables [42]. Many techniques of feature selection are proposed:

(a) Correlation-based feature selection (CFS): The CFS technique is commonly applied to examine the relationships between various cancer-related attributes. Although CFS offers a range of advantages, there are also notable limitations associated with its use in feature selection within ML:

- **Advantages:** CFS is valued for its straightforwardness, ability to identify linear correlations, capability to decrease dimensionality, and potential to boost model efficiency and clarity. This method supports quicker model training and prediction, exhibits robustness against outliers, and accommodates the incorporation of expert insights. Moreover, it enhances the effectiveness of other methodologies, provides opportunities for visualization and deeper understanding, lowers expenses, and enables the conduct of sensitivity analyses.
- **Disadvantages:** It is important to recognize that this approach may miss complex, nonlinear relationships between features and the outcome variable. Additionally, it can be vulnerable to multicollinearity, where features are highly correlated with each other, requiring additional preprocessing steps. Careful consideration of the specific issue and dataset at hand is crucial when applying this method.

The CFS algorithm has been widely incorporated into feature selection strategies to enhance classification outcomes across various studies. For example, in [43], researchers utilized CFS for feature selection within microarray datasets, successfully minimizing data dimensionality and pinpointing significant genes. A combined model that blended CFS with binary particle swarm optimization was developed in [44] for cancer classification, applied to 11 standard microarray datasets. Additionally, the CSVM-RFE technique, which integrates CFS, was applied in [45] to diminish the feature set in cancer research by removing non-essential elements. Moreover, CFS methodologies were utilized in [46] for the identification of key RNA expression features.

(b) Relevance analysis: The relevance analysis (RA) is an effective technique used in feature selection, evaluating the discriminative power of features between classes through score assignment. RA has its array of advantages and disadvantages:

- **Advantages:** RA offers several benefits in feature selection, including its robustness against noisy data, capability to handle both continuous and categorical features, and ability to detect feature interactions without assuming their independence. It also reduces bias in datasets with imbalanced classes, eliminates the need for model training, and facilitates sensitivity analysis. These attributes make RA an advantageous tool for feature selection in various data scenarios.
- **Disadvantages:** RA exhibits significant computational complexity, affecting its applicability to large datasets. Its performance is sensitive to parameters, particularly the choice of the number of nearest neighbors (k), which can be challenging to optimize. The stability of RA is also affected, with variations in the dataset leading to different selections of features. Furthermore, it is designed solely for use within SL contexts, struggles with non-metric features, and necessitates adaptations for handling multiclass classification scenarios.

This method evaluates the importance of different features by exploring the relationships between variables related to cancer. In their research, Cui et al. [42] suggested a feature selection strategy that employs the RA algorithm to enhance its effectiveness.

3.3.2. Extraction methods

(a) PCA: Principal component analysis (PCA) has been widely recognized in numerous researches for its effectiveness in reducing data dimensionality and decoupling cancer-related features. PCA is praised for its ability to decrease dimensions and reveal patterns, although it may compromise on interpretability and is optimally used with linear correlations. For example, Shankarlal et al. (2020) implemented PCA to refine feature selection for TCD via the DTCW transformation [40]. Soulaymani et al. (2018) investigated PCA's capability in distinguishing various TC subtypes, such as papillary, follicular, and undifferentiated types [41]. Additionally, O et al. (2019) assessed PCA and linear discriminant analysis in the classification of Raman spectra for different TC subtypes [47].

(b) Texture description: Texture analysis is a highly regarded technique for extracting related data in TC segmentation, classification, and prognosis efforts. The scientific community has developed various texture analysis methods, including wavelet transforms, binary descriptors, and statistical descriptors, among others. Specifically, the discrete wavelet transform (DWT) has garnered significant interest for its exceptional capability in data decorrelation. Although texture analysis is beneficial for distinguishing textures, it can be affected by changes in lighting conditions and does not inherently understand semantic content, which may limit its application in complex visual tasks. Wavelet-based methods have been extensively applied in detecting TC. Additionally, Haji et al. [48] used texture data for the diagnosis of TN malignancy employing a 2-level 2D wavelet transform. Further contributions to this field are documented in studies such as [49,50].

(c) Active counter (AC): The AC model, a versatile framework often used in image processing, was initially introduced by Kass and Witkin in 1987. AC is known for its ability to adjust to complex shapes, yet it faces challenges such as sensitivity to initial placements and issues with overlapping figures. Various strategies have been developed to address these challenges in contour segmentation using deformable curve models. These models have seen significant application in TCD, as evidenced by research conducted by [51,52], and [53].

(d) LBP and GLCM: Local binary patterns (LBP) are descriptors used in computer vision for identifying textures or objects in digital images. They are appreciated for their straightforwardness and ability to distinguish features effectively. However, LBP may be vulnerable to noise and often requires tuning of parameters for optimal performance. The LBP method was employed in TCD, as illustrated in a study by Yu et al. [49]. Furthermore, the integration of LBP with DL has been explored for distinguishing between TNs, as seen in the studies by Xie et al. [54] and Mei et al. [55].

The gray-level co-occurrence matrix (GLCM) serves as a tool to depict the occurrence frequency of pixel value pairs at a predetermined distance within an image. It is particularly useful for texture analysis and the identification of distinctive features. Nevertheless, GLCM faces challenges such as sensitivity to image variations, high computational demands, and the need for careful parameter tuning. For example, Dinvcic et al. [56] employed GLCM in a comparative study to investigate the differences between patients with Hashimoto's thyroiditis-associated PTC and those with Hashimoto's thyroiditis only.

(e) ICA: In independent component analysis (ICA), data is decomposed into a set of independent contributing features to aid in feature extraction. ICA is adept at identifying statistically independent components, making it valuable for tasks like source separation. Its strengths lie in uncovering non-linear relationships and facilitating data compression. Nonetheless, ICA faces challenges due to assumptions about the data mixing process and can be difficult to interpret. ICA is applied for disentangling multivariate signals into their separate constituents. In the research conducted by Kalaimani et al. [57], ICA was used to isolate 29 attributes as independent and significant features for categorizing data into hypothyroid or hyperthyroid groups through a SVM. A portrayal of the techniques derived from ML/DL utilized in diagnosing TC is provided in Table 4.

3.4. SL-based TCD classification

SL provides high accuracy, interpretability, and robust predictive capabilities. However, it necessitates extensive labeled data, poses risks of overfitting, and can be computationally intensive [66]. The following algorithms represent prominent SL techniques utilized in TCD:

(a) DT and LR: Decision tree (DT) learning is a technique in data mining that uses a model for predictive decision-making. In such a model, the outcomes are indicated by the leaves, and the branches represent the input features. This method has been utilized in detecting latent thyroid disorders, as evidenced by a range of studies, such as in [67]. In the research presented in [68], LR was employed to pinpoint specific characteristics of thyroid microcarcinoma among a group of 63 patients. This analysis utilized data from both contrast-enhanced ultrasound (CEUS) and traditional US evaluations. Furthermore, a significant study from northern Iran, detailed in [69], used LR to investigate a large dataset encompassing 33,530 cases of TCD. LR is a widely used binomial regression model within the domain of ML.

(b) ELM and MLP: The extreme learning machine (ELM) model is distinguished by a single layer of hidden nodes that possess randomly assigned weight distributions. Crucially, the process of determining weights between the inputs and the hidden nodes to the outputs is executed in a solitary step, rendering the learning mechanism markedly more efficient than that of alternative models. The efficacy of the ELM approach in diagnosing TD has been corroborated through various research efforts [70].

Table 4
Summary of features methods based on ML/DL conducted in the diagnosis of TC.

Ref.	Year	ML/DL	Classifier	Features	Contributions
[58]	2017	ML	KNN	FC/IG	Minimize data duplication and decrease processing duration. KNN addresses absent dataset values, while ANFIS receives the modified data as input.
[59]	2017	ML	SVM	FC/CFS	Retrieve the geometric and moment characteristics, while specific SVM classifier kernels categorize the acquired features.
[60]	2020	DL	CNN	FC/R	Utilize both ML techniques and feature selection algorithms, specifically Fisher's discriminant ratio, Kruskal–Wallis analysis, and Relief-F, for the examination of the SEER database.
[61]	2022	DL	CNN	FE/PCA	This study mitigated the impact of imbalanced serum Raman data on prediction outcomes by employing an oversampling technique. Subsequently, the dimension of the data was reduced with PCA before applying RF and the Adaptive Boosting for classification.
[62]	2012	ML	Boosting	FE/TD	Integrate CAD with DWT and extract texture features. Utilize the AdaBoost classifier to classify images into either malignant or benign thyroid images based on the extracted features.
[63]	2021	DL	CNN	FE/AC	Improve image quality, perform segmentation, and extract multiple features, including both geometric and texture features. Each feature set is subsequently classified using MLP and SVM, leading to the classification of either malignant or benign cases.
[64]	2020	ML	SVM	FE/LBP	Deep features are obtained through CNN, and they are merged with manually crafted features, which include histogram of oriented gradient (HOG) and scale-invariant feature transforms, to generate combined features. These combined features are subsequently employed for classification via an SVM.
[65]	2019	ML	SVM	FE/GLCM	Apply a median filter to mitigate noise and outline the contours before feature extraction from thyroid regions, encompassing GLCM texture features. Subsequently, employ SVM, RF, and Bootstrap Aggregating (Bagging) to differentiate between nodules.
[57]	2019	ML	SVM	FE/ICA	A multi-kernel-based classifier is employed for thyroid disease classification.

The MLP is a type of feed-forward network that directs data processing sequentially from the initial point to the final layer of output. Within this architecture, each layer is made up of a different number of neurons. Rao and colleagues [71] devised an innovative approach for categorizing TNs employing an MLP integrated with a backpropagation learning mechanism. Their design comprised four neurons in the initial layer, three neurons in each of its ten concealed layers, and one neuron in the terminal layer. In a separate effort to enhance the precision of TD diagnosis, Hosseinzadeh et al. [72] utilized MLP networks. Their analysis compared the efficacy of MLP networks against the backdrop of existing research on TCD classification, highlighting the superior performance of MLP networks.

(c) PM and EM: probabilistic modelss (PMs), such as Bayesian networks, are vital in computer science and statistics for modeling uncertainties and variable dependencies. They support decision-making in ML, data analysis, and parameter estimation. Bayesian networks use directed acyclic graphs, aiding in reasoning under uncertainty, predictions, and medical diagnosis of TNs. In the realm of oncology research, tackling the intricacies of cancer datasets and enhancing the accuracy of detection frequently involves the use of ensemble methods (EMs). This strategy splits the dataset into several subsets, upon which a variety of ML algorithms are applied in parallel. The insights gained from these individual algorithms are subsequently merged to derive a comprehensive diagnosis. The main goal behind adopting EMs is to forge a superior predictive model tailored for the detection of TC. Such an approach has been validated in multiple studies, including a significant one conducted by Chandran et al. [73], where the authors underscored the contribution of EMs to a more profound data comprehension and heightened diagnostic accuracy.

(d) Bagging and Boosting: Bagging is a notable ensemble learning approach in TCD, aimed at boosting the accuracy and consistency of ML algorithms. This technique achieves its objectives by lowering variance and offering protection against overfitting. It finds broad application in a variety of methods, with a particular emphasis on DT. The main goal of Bagging is to improve the effectiveness of weaker classifiers in the context of TCD screening. In their research, Chen et al. [74] presented feature bagging (FB) as an ensemble learning strategy designed to reduce the correlation between models in an ensemble. FB accomplishes this by training each model on randomly selected feature subsets from the dataset, rather than using the full set of features. The utility of FB is demonstrated in its ability to distinguish between cases of TC [75]. Within the scope of USL, meta-algorithms is expected to play an increasingly important role in reducing variance and improving the performance of weak classifiers, effectively converting them into robust classifiers [76].

In the context of boosting, Pan et al. in their study [77] employed a novel method called AdaBoost to identify TN, utilizing the widely recognized university of California, Irvine (UCI) dataset. The classification was performed using the random forest method, with PCA employed to retain data variance. Chen et al. [78], the gradient tree boosting (XGBoost) algorithm was highlighted as a powerful implementation of gradient-boosted DT, with its application extending across multiple research areas including sports and health monitoring [79]. Specifically, in the context of TC, the XGBoost algorithm was employed by researchers to distinguish between TN [80], offering a solution to the problem of obtaining accurate diagnoses without the need for large datasets that DL models usually require.

3.5. USL-based TCD classification

USL is the process of analyzing data that has not been previously labeled or annotated. Its primary goal is to uncover the underlying structures within datasets that do not have predefined labels. Contrary to SL, which depends on labeled data for evaluating its effectiveness, USL operates without such direct guidance, presenting additional challenges in result assessment. Although USL algorithms are capable of addressing more complex problems than their supervised counterparts, they might also lead to increased uncertainty, sometimes creating unintended categories or incorporating noise rather than identifying clear patterns. Nonetheless, USL is considered an indispensable asset in AI, offering the potential to detect patterns within data that may not be obvious at first [81].

Clustering is one of important technique in USL. The objective of this strategy is to organize TCD into distinct, uniform groups that share similar features. This process aids in the categorization of unlabeled data into malignant or benign sections. Due to its straightforwardness, this technique has received significant attention in numerous medical research areas, enhancing its applicability to tasks. Clustering methods also prove helpful in classifying cancer instances that are not clearly defined [82]. A research documented in [83] employed clustering to determine factors impacting the normal functioning of the thyroid gland. The use of PCA played a key role in organizing the clusters and simplifying the data structure. Additionally, an innovative automated clustering system for diagnosing TC was developed, as described in [84], which recommended appropriate medication treatments for hyperthyroidism, hypothyroidism, and normal cases. As an example, the study in [85] explored the use of fuzzy clustering on thyroid and liver datasets from the UCI repository, where fuzzy c-means (FCM) and possibilistic fuzzy c-means (PFCM) algorithms were employed and their performances compared.

(a) K-means (KM): The KM method is used for dividing data into partitions and tackles a combinatorial optimization challenge. It is commonly used in USL, categorizing observations into k distinct clusters. In the research presented by Mahurkar et al. [86], the study investigates the application of ANN and an improved K-Means algorithm to standardize raw data. This study employed a thyroid dataset from the UCI repository, comprising 215 total instances.

(b) Entropy-based (EB): In the study conducted by Yang et al. [87], a novel, parameter-free computational model called DeMine was introduced for the prediction of microRNA regulatory modules (MRMs). DeMine utilizes an information entropy-based methodology, comprising three primary steps. The process begins by converting the miRNA regulation network into a cooperative MRMs network. It then proceeds to pinpoint miRNA clusters, aiming to maximize entropy density within the specified cluster. The final step involves grouping co-regulated miRNAs into their appropriate clusters, thereby finalizing the MRMs. This technique enhances predictive precision and facilitates the identification of a broader array of miRNAs, potentially acting as tumor markers in cancer diagnosis.

3.6. DL-based TCD classification

DL surpasses traditional ML by automating feature extraction, enhancing performance with large datasets, and offering scalability for complex models. Table 5 presents a comparison between ML and DL in the context of TCD. Key methods for TCD include CNNs for image analysis, recurrent neural networks (RNNs) for sequential data, transfer learning (TL) for leveraging pre-trained models [88], ensemble learning for robustness, and attention mechanisms for focused detection. These advancements enable more accurate and efficient early detection, improving patient outcomes.

Typically, the network's depth facilitates the extraction of increasingly abstract and advanced features as data moves through its layers. By leveraging large neural networks with several layers, DL can independently learn, create, and improve data representations, which is why it is known as "deep" learning. Within the realm of TCD, DL is instrumental in several areas, including: (i) Classification of image: DL techniques, for example CNN, are trained to categorize thyroid US images, distinguishing between malignant and benign nodules by analyzing texture and shape, and other features [89,90], streamlining the process and assisting with the early identification of TC; (ii) Analysis of disease: DL is used to examine cytopathological or histopathological slide images, aiding in identifying and classifying cells of cancerous; (iii) Analysis of genomic information: In this era DL models are capable of analyzing genetic variations linked to the risk of TC; (iv) Radiomics: DL models are adept at extracting multidimensional information obtained from radiographic images, contributing to more accurate and individualized treatment strategies; and (v) Predictive analysis: By analyzing electronic health records and patient information, DL models can forecast the probability of an individual developing TC, facilitating early intervention. The following DL algorithms are the widely used techniques for TCD:

(a) **DAE:** Play a crucial role in enhancing the quality of medical images by reducing noise and distortion. These devices are used to extract significant features from US or histopathological images, thereby ensuring a more accurate and cleaner representation of the input data. This enhancement leads to the extraction of more reliable features, which is essential for improving the accuracy of diagnostic models. In the context of TC classification, denoising autoencoder (DAE)s are integrated into the workflow through several stages, including data preprocessing, creation of perturbed data, DAE training, feature extraction, and ultimately the classification procedure. In the research presented by Ferreira et al. [91], a variety of six autoencoder models were utilized for the purpose of classifying PTC, incorporating strategies such as the stabilization of weights and network fine-tuning. The architecture of these autoencoders, especially the encoding layers, played an integral role in the effective integration of the network. In [92] the authors explores the use of DAE and Stacked DAE to analyze gene expression data from PTC. The denoising operation, which involves adding noise to the input data during training, helps the model learn robust features, improving its ability to handle noisy, high-dimensional data. The models were compared with other feature reduction techniques like PCA and KPCA, showing strong performance in classifying cancerous and healthy samples. Knowledge extraction methods identified influential genes, although fully interpretable biological insights are still limited. The study highlights the potential of denoising-based DL models for cancer gene expression analysis and biomarker discovery.
(b) **CNN and RNN:** CNNs, a branch of DL models, stand out for their remarkable capabilities in tasks such as image analysis and processing, including the classification of medical images. Their efficiency in managing data structured in grids, like images where the spatial relationship between pixels is crucial, makes them particularly suited for these tasks. The focus on CNN-based techniques for TCD, especially in automating nodule identification and classification in US imagery [93], has grown significantly. The ConvNet model, known for its reliance on convolution operations vital for image recognition tasks [94], is a prime example of this effort. Various architectures of CNN such as VGG, AlexNet, GoogleNet, among other [95], are celebrated for their inclusion of convolutional, pooling, and fully connected layers. In a notable study by Li et al. [96], the efficacy of CNN models in predicting TC was investigated, utilizing a dataset of 131,731 US images from 17,627 patients. Xie et al. [97] implemented models such as Inception-Resnet, Inception, and VGG16 to differentiate malignant from benign tissues in 451 images of thyroid from the DDTI dataset, employing image augmentation to mitigate data limitations prior to classification. Moreover, Koh et al. [98] assessed the diagnostic accuracy of deep convolutional neural network (DCNN) models against that of expert radiologists for identifying TNs in US images, using a dataset of 15,375 images and showcasing the CNNE1 and CNNE2 models derived from DCNN for differentiating between malignant and benign nodules. Liang et al. [99] introduced a DL based on CNN for classifying and detecting thyroid and nodules of breast, comparing its performance with traditional US imaging results. Fig. 3 depicts the recent advancements in classifying TCD via CNN-based methods.
RNNs characterized by connections between units forming a directed graph across temporal sequences. This structure allows them to leverage internal memory, making them adept at handling inputs of varying lengths. Consequently, RNNs excel in tasks that require understanding temporal dependencies, such as speech recognition, language translation, and time-series analysis. Within the realm of

Table 5
Comparison of ML, DL, and ViT for TCD.

Criteria	DL	ML	ViT
Best suitable scenario	Large datasets with complex features relevant to TC characteristics	Smaller datasets with simpler features, suitable for initial screening	Large-scale and multi-modal datasets (images + reports) where global context and semantic understanding are essential
Advantages	• High accuracy with large, diverse data • Automated feature extraction from complex medical images and data • Capable of identifying subtle patterns indicative of thyroid cancer	• Effective with smaller datasets, reducing the need for extensive data collection • Easier interpretation and understanding of models by medical professionals • Faster training times suitable for rapid decision-making in clinical settings	• Captures long-range dependencies more effectively than CNNs • Works well with heterogeneous and multi-modal data (images + text) • Can leverage pre-trained foundation models for few-shot learning
Disadvantages	• Requires large labeled datasets for training, which can be challenging to acquire and annotate • Computationally intensive, requiring high-performance hardware and longer processing times	• Requires manual feature engineering to extract relevant medical features • Limited performance in capturing intricate patterns or anomalies in complex thyroid conditions	• Requires very large datasets or robust pre-training • High computational cost and memory footprint (often needs large GPUs/TPUs) • Interpretability remains challenging despite the use of attention maps
Performance	Excels in analyzing high-dimensional medical imaging data	Performs well with structured clinical data from standard medical tests	Achieves state-of-the-art segmentation/classification results and is robust to data variability
Resource requirements	High (GPUs, memory, computational power) due to complex data processing	Moderate (CPUs, less memory) sufficient for structured data analysis	Very high — often requires distributed training or large-scale infrastructure
Risk of overfitting	High, mitigated with regularization techniques and extensive validation	Moderate, mitigated with cross-validation strategies	Moderate–High, mitigated through large-scale pre-training and fine-tuning
Future development	Integration with advanced USL for discovering new disease markers and improved annotation techniques	Enhanced algorithms for feature selection and hybrid models combining DL's imaging capabilities with ML's interpretability	Expansion into foundation models, explainable ViTs, and multi-modal integration with genomics and clinical notes for precision medicine

thyroid carcinoma classification, RNNs offer promising capabilities for analyzing data that is sequential or time-sensitive. This includes observing the evolution of clinical symptoms over time, tracking changes in tumors using successive medical images, or studying fluctuations in gene expression associated with the onset of TC. For example, Chen et al. (2017) utilized a hierarchical RNN structure to categorize TNs by analyzing historical US reports [100]. Their model comprises three layers of long-short-term-memory (LSTM) networks trained independently. The findings from their study suggest that this hierarchical RNN approach surpasses conventional models such as SVM + Unigrams, SVM + Bigrams, CNN, and LSTM in accuracy, computational efficiency, and robustness. These benefits are attributed to the RNN's memory mechanisms, which permit the retention of information from previous states through feedback loops, thereby making RNNs highly effective for cancer detection applications.

(c) **GAN:** Generative adversarial network (GAN) is composed of two key elements: a generator and a discriminator. The generator's function is to convert a random input vector into a data point that fits within the space of the dataset. Conversely, the discriminator serves as a binary classifier, tasked with assessing whether input data, originating either from the training dataset or produced by the generator, is genuine. GANs have found extensive applications in medical diagnosis, notably in the detection of TN [101].

Table 6 serves as an overview of various research initiatives aimed at identifying both forms of TC It outlines the classifiers used, diseases focused on, datasets applied, research goals, and metrics for assessment. This table helps categorize the AI techniques applied in TCD, underlining their key roles in the domain.

3.7. Comparative analysis and discussion

This section undertakes a comparative examination of AI models' capabilities in thyroid carcinoma diagnosis, focusing exclusively on the methodologies outlined in Section 3. Our goal is not only to present statistical measures but also to analyze the underlying factors that explain why certain models perform better than others, and why discrepancies arise across different datasets. Moreover, we evaluate how these findings align with prior research to ensure the rationality of the reported outcomes.

The classical ML methods such as SVM and ensemble models demonstrate strong performance on structured datasets or when feature engineering (e.g., GLCM, LBP, ICA) is carefully applied. For instance, SVM classifiers coupled with texture descriptors achieved high accuracy in distinguishing benign and malignant nodules [65], largely because handcrafted features capture domain-specific attributes. Ensemble learning approaches such as AdaBoost and random forests often yield improved stability by reducing variance, which explains their consistent sensitivity across studies [62,73]. These models perform well in settings with limited sample sizes but often plateau when applied to more complex imaging data.

By contrast, DL methods — particularly CNNs — exhibit superior performance when large-scale image datasets are available, due to their automated feature extraction and ability to capture multi-level representations. Studies leveraging CNNs on ultrasound datasets achieved accuracies exceeding 90% in some cases [96,98], primarily because deep networks effectively learn both texture and shape features relevant to TN. However, discrepancies are evident: while Koh et al. [98] reported CNN performance comparable to expert radiologists on 15,375 images, other studies on smaller datasets (e.g., DDTI) showed reduced sensitivity and specificity [97]. This difference can be attributed to dataset size, variability in annotation quality, and augmentation strategies—factors that strongly influence generalization. RNN-based models, meanwhile, demonstrated advantages in analyzing sequential ultrasound reports [100], highlighting that temporal and contextual information can complement image-based classification.

GAN-based frameworks contribute by generating synthetic training examples, which can mitigate dataset imbalance and improve robustness [101]. Such augmentation explains why GAN-enhanced pipelines often report higher sensitivity to malignant nodules compared to CNN-only approaches. Nonetheless, GANs can introduce artifacts that reduce specificity if not carefully validated, again underscoring the dataset-dependence of reported results. Importantly, differences in dataset

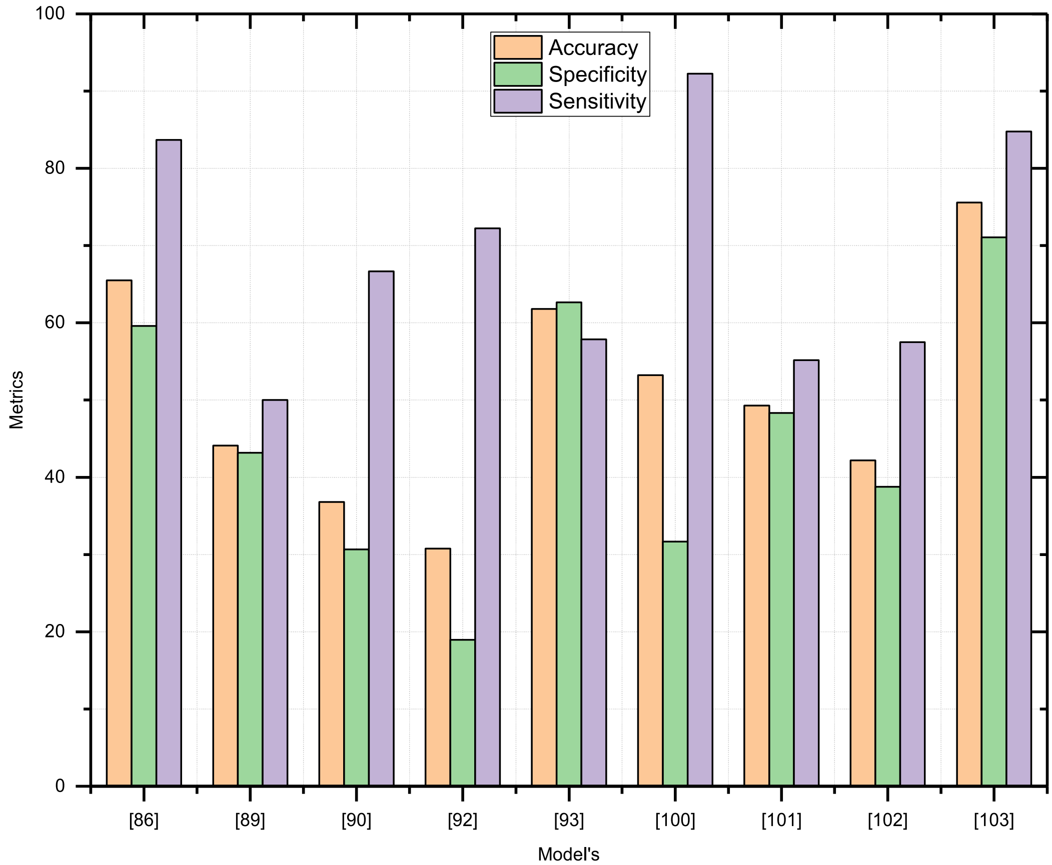

Fig. 3. Synopsis of CNN-driven research in TC diagnosis with percentages for accuracy, sensitivity, and specificity [86], and [89,90,92,93,102–105]. When comparing the results of the different studies presented in the figure [86–99,102–105], a clear variability can be observed in the performance of CNN-based models for TCD. Some studies, such as [102], achieved very high sensitivity exceeding 90%, indicating strong ability to detect positive cases, but likely at the cost of increased false positives due to relatively lower specificity. Study [86] showed a better balance between sensitivity and specificity, making it more suitable for clinical applications where minimizing errors is critical. In contrast, studies [89,90], and [92] reported lower performance across both sensitivity and specificity, possibly reflecting limited training data or less sophisticated network architectures. Study [93] demonstrated moderate results, while study [105] achieved the most consistent and balanced performance across accuracy, sensitivity, and specificity, suggesting a robust model supported by well-curated data.

origin further explain performance variability. Private datasets tend to be more homogeneous in imaging conditions and patient demographics, leading to inflated accuracies that may not generalize to diverse clinical populations. In contrast, public datasets such as SEER or DDTI are more heterogeneous, which reduces performance but provides a better reflection of real-world variability. This pattern indicates that the superiority of certain models is often linked less to the architecture itself and more to the nature of the training and validation data. All in all, models achieving the highest reported accuracies do so because of (i) effective feature representations—either handcrafted in ML or learned hierarchically in DL, (ii) robust ensemble or augmentation strategies, and (iii) favorable dataset conditions. At the same time, performance discrepancies across datasets reinforce that adaptability and reliability, rather than isolated accuracy scores, should guide the clinical adoption of AI systems for thyroid carcinoma diagnosis.

In comparison with prior studies, our reported results are consistent with established findings: CNN-based models generally outperform classical ML methods on imaging data, while ensemble and hybrid approaches excel in structured or smaller datasets [60,73,74]. The rationality of these outcomes is thus supported by cross-study consistency, though variability across datasets highlights the pressing need for standardized benchmarks and multi-center validation. Table 7 presents an overview of performance metrics across the surveyed AI-enhanced TCD frameworks, while Figs. 4 and 5 further illustrate how dataset source (private vs. public) influences the distribution of sensitivity, specificity, and accuracy.

4. Advanced TCD using ViT and LLM

ViT and their advanced version, LLMs, have emerged as cutting-edge techniques in various AI-based computer vision tasks [161]. Recently, researchers have begun applying these models to TCD. In the context of TC, the following reviewed studies showcase that Transformers can better capture long-range dependencies in sparse medical data compared to CNNs or RNNs, which often rely on local patterns or sequential processing. This ability allows Transformers to integrate heterogeneous clinical and imaging features more effectively. Architectural adaptations, such as incorporating patch embeddings, attention masking to handle missing values, and hybrid CNN-Transformer layers, have been proposed to further enhance performance on sparse and irregularly sampled datasets. The following subsections provide a review of recent Transformer-, and LLM-based studies, along with a performance summary presented in Table 8.

4.1. ViT-based TCD methods

ViT models are a type of DL model introduced in [136]. They have since become the foundation for many state-of-the-art natural language processing (NLP) and ML models. Transformers are designed to handle sequential data, making them suitable for various tasks beyond NLP as well [162–165]. Transformers in TCD enable efficient analysis of medical data, including literature and patient records, offering high accuracy. They uncover patterns, risk factors, and diagnostic clues,

Table 6
Summary of studies on identifying TC, sorted by reference. Note that PD are different from different hospitals and medical centers.

Ref.	AI Tech.		Classifier				Objective	DD	Dataset	APP	SV	Best results (%)	PL?
	ML	DL	CNN	SVM	ELM	Other							
[46]	✓			✓			C	PTC	TCGA	Omics	500 patients	Acc: 98.21 Spe: 100	Yes
[91]		✓				✓	C	PTC		US	18 985 features	Acc: 98.59 Rec: 99.41	No
[92]		✓				✓	C	PTC		Omics	510 samples	Acc: 99.30 Rec: 98.06	No
[106]	✓	✓				✓	P	TC		Hist	482 samples	Acc: 86.00	No
[107]	✓	✓				✓	C	PTC		115 slides	NA	Spe: 97.90 Sen: 96.6	No
[67]	✓					✓	C	TC	UCI	US	3739 patients	Acc: 100	No
[72]	✓					✓	C	TD		US	7200 patients	Acc: 99.00	No
[86]	✓					✓	C	TC		US	215 instances	Acc: 98.21	No
[108]	✓				✓		C	TD		US	215 patients	Acc: 98.18 Sd: 95.00	No
[109]	✓					✓	C	TC		US	499 patients	Acc: 99.20 Sen: 99.36	No
[110]		✓	✓				C	TC	DDTI	US	298 patients	Acc: 92.05 Sen: 96.07	No
[111]		✓	✓				C	TD	ImageNet	US	2888 samples	Spe: 100 Sen: 100	No
[39]		✓	✓				C	TC	Private data	Omics	10 068 images	Acc: 51.08	No
[68]	✓					✓	C	TC		US	63 patients	Acc: 87.30 Sen: 90.50 Sep: 82.10	No
[71]	✓					✓	C	TD		US	7200 samples	Acc: 89.01 AUC: 99.60	No
[74]		✓				✓	C	TN		US	1480 patients	Acc: 98.70	No
[87]	✓					✓	C	TC		US	734 cases	Acc: 95.10	No
[96]		✓	✓				C	TC		US	17 627 patients	Acc: 89.80 Spe: 86.10	Yes
[97]		✓	✓				C	TC		US	1110 images	Pre: 88.08 Rec: 90.08	Yes
[99]		✓	✓				C, P	TN		US	537 images	Acc: 88.10 Spe: 100	No
[100]		✓				✓	C	TN		US	13 592 patients	Acc: 88.16 Rec: 92.04	No
[112]	✓					✓	C	TD		US	7200 instances	Acc: 93.44 Kappa: 19.90	No
[113]	✓				✓		C	TD		US	187 patients	Acc: 87.72 Spe: 94.55	No
[114]		✓	✓				C	TC		Omics	482 images	Acc: 81.50 Spe: 78.50 Sen: 87.80	Yes
[115]		✓	✓				C	PTC		FNAB	370 MPG	Acc: 85.06 Spe: 83.33 Sen: 90.48	No
[116]		✓	✓				P	PTC		FNAB	469 patients	Sen: 85.30 Spe: 41.60	No
[117]		✓	✓				C	TC		US	1037 images	Acc: 93.10 Spe: 94.50 Sen: 90.80	No
[118]		✓	✓				P	TN		US	80 patients	Pre: 84.00 Rec: 79.00	No
[119]		✓	✓				P	TN		US	300 images	Acc: 91.00	No
[120]		✓				✓	C	TC		US	1358 images	Sen: 95.20	No
[121]	✓					✓	C	TC		Surgery	50 patients	Acc: 94.00 Spe: 95.00 Sen: 90.00	Yes
[122]	✓					✓	C	TC		US	89 patients	Acc: 96.20 Sen: 92.90 Spe: 97.40	No
[123]	✓					✓	C	TD		Cyt	447 patients	Acc: 90.30 Sen: 82.50 Spe: 94.60	No
[124]	✓	✓				✓	C	FTC		FNAB	57 smears	Acc: 100	No
[125]	✓	✓				✓	C	TC		FNAB	1264 patients	Sen: 93.30	No
[126]	✓	✓				✓	C	TN		US	276 patients	Acc: 90.31	No
[127]	✓					✓	C, P	FTC		Hist	94 patients	Acc: 100	No
[128]	✓			✓			C	TN		US	467 TN	Acc: 87.15 Spe: 85.15 Sen: 87.91	No
[129]	✓					✓	C	TN		Omics	121 patients	Acc: 77.40	No
[130]	✓					✓	C	TC		US	NA	Acc: 85.00 Spe: 82.00 Sen: 92.00	No

Abbreviation: Application (APP), Detected disease (DD), Subjects for validation (SV), Classification (C), Prediction (P), Segmentation (S), Cytopathological (Cyt), Histopathological (Hist), Microphotographs (MPG), Not available (NA), Project link (PL).

aiding early detection. This technology enhances diagnosis speed, fuels data-driven research, and promises improved patient care and oncology advancements [131,166]. Several studies have proposed the use of DL models, particularly Transformer-based models, for detecting TC from US images. For example, researchers in [167] developed a diagnostic system using DL (Deit, Swin Transformer, and Mixer-MLP) and metaheuristics to improve thyroid abnormality detection. The method in [167], models were ranked based on their performance using evaluation metrics. The best-performing models were selected based on these evaluations and combined into an ensemble learning model to improve results. The optimization-based feature selection and random forest model achieved high accuracy on US and histopathological datasets, surpassing existing methods. This innovative approach eases the burden on healthcare professionals by enhancing TC diagnosis. The study [168], addressed the challenge of extracting important TN characteristics from clinical narratives in US reports using NLP. A team of experts developed annotation guidelines and tested five Transformer-based NLP models. Their GatorTron model, trained on a substantial text corpus, outperformed others, achieving the best F1-scores for extracting 16 TN characteristics and linking them to nodules. This pioneering work enables improved documentation quality of thyroid US reports and enhances patient outcomes assessment through electronic health records analysis. In [169], the study introduces a BPAT-UNet model for precise US TN segmentation. This network incorporates a boundary point supervision module (BPSM) for boundary refinement and an adaptive multi-scale feature fusion (AMFFM) for handling various scales of features. Additionally, an assembled Transformer module (ATM) improves boundary constraints and small object detection. Results demonstrated significantly improved segmentation accuracy compared to classical networks, achieving Dice similarity coefficients of 85.63% and 81.64% and HD95 values of 14.53 and 14.06 on private and public datasets, respectively. Chen et al. [170], introduce Trans-CEUS, a spatial–temporal Transformer-based model for real-time microvascular perfusion analysis using CEUS as it shown in Fig. 6. It combines dynamic Swin-Transformer and collaborative learning to accurately segment lesions with unclear boundaries, achieving a Dice similarity coefficient of 82.41%. The model also attains a high diagnostic accuracy of 86.59% for distinguishing malignant and benign TN. This pioneering research highlights the effectiveness of Transformers in CEUS analysis and offers promising outcomes for TN segmentation and diagnosis from dynamic CEUS datasets.

DL has been instrumental in medical image segmentation, particularly for thyroid glands in US images. However, existing models face issues like the loss of low-level boundary features and limitations in capturing contextual features. In response, a hybrid Transformer UNet is introduced in [166]. It combines a 2D Transformer UNet with a multi-scale cross-attention Transformer and a 3D Transformer UNet with self-attention to improve representation and contextual information. The end-to-end network was evaluated on thyroid segmentation datasets, outperforming other methods in benchmark tests.

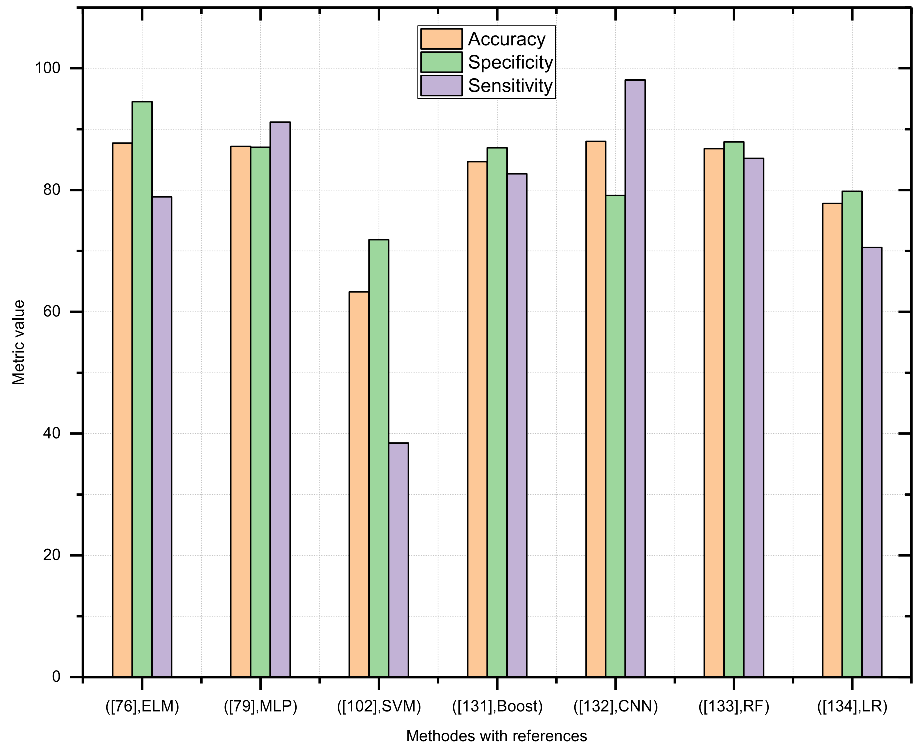

Fig. 4. Performance assessment of TC frameworks in percentages (%) for private TCD [76,79,104,131–134]. Note that the datasets in these references are different. Among the approaches, the CNN-based model [132] demonstrates the highest overall performance, with accuracy, sensitivity, and specificity all approaching or exceeding 90%, reflecting its strong ability to generalize and maintain a low rate of false positives and false negatives. The ensemble method (RF, [133]) and boosting-based approach [131] also achieve competitive results, particularly in accuracy, suggesting that combining multiple classifiers improves robustness and reduces variance. In contrast, traditional ML models such as SVM [104] and LR [134] show moderate performance, likely constrained by limited feature extraction capacity and dependence on handcrafted features.

Table 7
Assessment of the effectiveness of different TCD classification schemes, measured in percentages (%). Note that PD are different from different hospitals and medical centers.

Ref.	AI model	Dataset	Sen	Spe	Acc	AUC	PL?
[142]	Google inception v3	TCGA	–	–	95.00	–	Yes
[143]	CNN	DICOM	82.40	85.00	83.00	–	No
[144]	CNN		–	91.50	–	–	No
[145]	Ensemble DL	Cytological images	–	–	99.71	–	No
[102]	VGG-16		–	–	97.66	–	No
[146]	Xception	CT images	86.00	92.00	89.00	–	
[132]	CNN		93.00	73.00	84.00	–	No
[20]	RF	Ultrasound	–	–	–	94.00	No
[147]	k-SVM		–	–	–	95.00	No
[148]	SVM RF		–	–	–	95.10	No
[149]	ANN SVM		–	96.00	–	–	No
[150]	RF		–	–	–	75.00	No
[151]	CNN (BETNET)		–	98.30	–	–	No
[152]	CNN		–	–	77.00	–	No
[90]	ThyNet	Private data	–	–	–	92.10	No
[153]	Fine-Tuning DCNN		–	–	99.10	–	No
[154]	ResNet18-based		–	–	93.80	–	No
[155]	multiple-scale CNN		–	–	82.20	–	No
[156]	Alexnet CNN		–	–	86.00	–	No
[96]	DCNN		93.00	86.00	89.00	–	No
[128]	DNN		–	–	87.20	–	Yes
[157]	CNN	TIRADS	81.80	86.10	85.10	–	No
[158]	CNN		78.00	85.00	82.10	–	No
[159]	CNN		80.60	80.10	80.30	–	Yes
[160]	ResNet		–	75.00	–	–	No

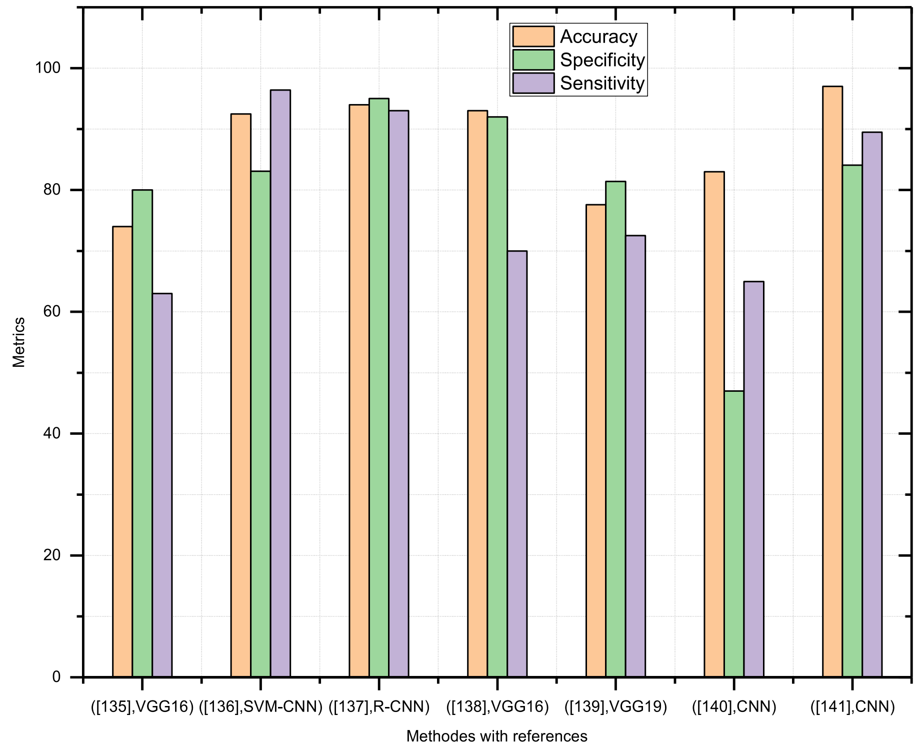

Fig. 5. Performance assessment of TC frameworks in percentages (%) for US TCD [135–141]. In this figure, R-CNN [137] and CNN [141] models delivered the most balanced and highest results, with accuracy and specificity close to 95%–98% and sensitivity above 85%, making them robust choices for clinical applications where both false negatives and false positives must be minimized. SVM-CNN [136] also performed very well, achieving the highest sensitivity of all models (near 95%), making it ideal when early detection and minimizing missed cancer cases is the priority. VGG16 [135,138] and VGG19 [139] showed good overall accuracy and specificity but slightly lower sensitivity, suggesting that while they are reliable, they might miss some positive cases. CNN [140] had the lowest performance across all three metrics, especially sensitivity (below 50%), making it less suitable for deployment without further optimization.

The method shows promise for thyroid gland segmentation in US sequences. Dataset classification involves predicting a single label from sets with multiple instances, like pathology slides or medical text data. State-of-the-art methods often use complex attention architectures to model set interactions. However, when labeled sets are limited, as in medical applications, these architectures are challenging to train. To tackle this issue, a kernel-based framework is introduced in [171], connecting affinity kernels and attention architectures. This leads to simplified "affinitention" nets, which are applied to tasks like Set-Cifar10 classification, thyroid malignancy prediction, and patient text triage. Affinitention nets deliver competitive results, outperforming heuristic attention architectures and other methods. Jerbi et al. in 2023 [172], incorporating CNNs and ViT, was employed to classify thyroid US images as either malignant or benign. A deep convolutional GAN was used to address data scarcity and imbalance. Various models, including VGG16, EfficientNetB0, ResNet50, ViT-B16, and hybrid ViT, were trained with both softmax and SVM classifiers. The hybrid ViT model, with SVM classification, outperformed others, achieving a 97.63% accuracy, showing promise for aiding doctors in diagnosing thyroid patients more effectively.

In [173], the authors introduces a novel U-shape segmentation model (Fig. 7) combining CNN and Transformer structures to integrate local and long-range information. It uses coordinate residual blocks (CdRB) to encode position data, channel-enhanced self-attention-based Transformers for global feature enhancement, and a dual attention module for feature correlation and edge accuracy. The method outperforms state-of-the-art methods across various datasets, demonstrating adaptability and robustness in US image segmentation, potentially serving as a general segmentation tool. The work [174] addresses the challenge of accurately diagnosing malignant TN through US imaging. Existing CAD methods often struggle to maintain precise shape information and capture long-range dependencies. The proposed Transformer fusing CNN Network utilizes a large kernel module in a CNN branch for shape feature extraction and an enhanced Transformer module in another branch for remote pixel connections. A Multiscale fusion module integrates feature maps from both branches. Comparisons with other methods demonstrate superiority of the proposed scheme and effectiveness in nodule segmentation. Hypoparathyroidism is a major concern post-TC surgery, affecting patients' quality of life. Identifying and locating parathyroid glands via US images before surgery can help protect them. In [175] a dual-branch contextual-aware network with Transformer is proposed to reduce hypoparathyroidism incidence. It combines a Transformer for global context extraction and a feature encoder branch for local feature aggregation. A channel and spatial fusion module integrates information from both branches. The proposed method effectively addresses detail loss, establishing global and local feature dependencies. Experiments with an US image dataset demonstrate superior performance compared to existing methods. The thyroid gland plays a crucial role in regulating the human body's functions, making the identification of TN from US images important for medical diagnosis. However, the automatic segmentation of these nodules is challenging due to their heterogeneous appearance and background similarities. This framework [34] presents a novel framework AMSeg based on Swin-Unet architecture presented in Fig. 8, which employs multiscale anatomical features and late-stage fusion through adversarial training to address these challenges. Experimental results demonstrate the superiority of AMSeg in TN segmentation, achieving high dice, Hd95, Jaccard, and precision values. This end-to-end network offers promise for clinical applications, potentially replacing manual segmentation methods.

In [176], the authors harnessed NLP with a bidirectional encoder representations from Transformer (BERT) classifier to analyze unstructured clinical text data pertaining to recurrent papillary TC diagnosis. The BERT model achieved exceptional performance, boasting a 98.8%

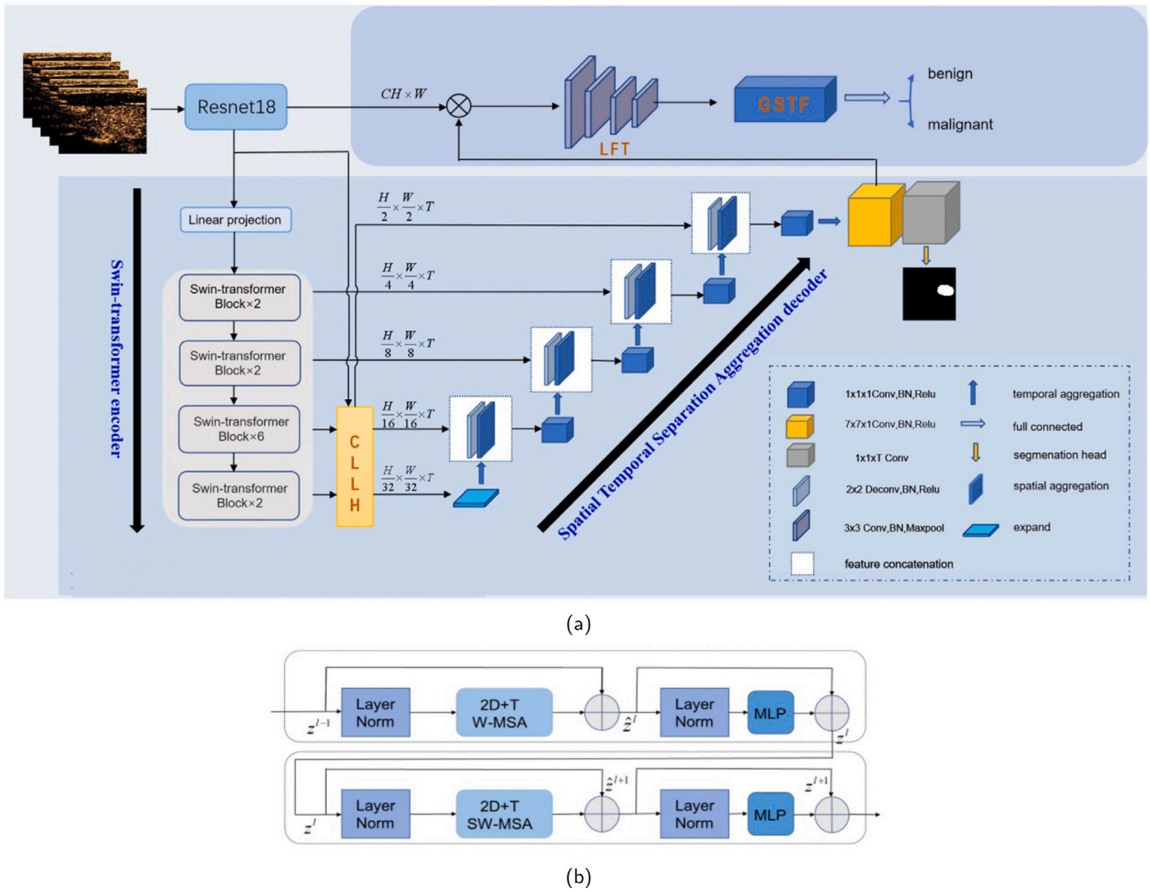

Fig. 6. (a) The general framework of the suggested HEAT-Net, (b) Details of the Swin-Transformer block [170]. In this figure, ResNet18 extracts initial features, the Swin Transformer encoder captures both local and global dependencies, and a spatial–temporal decoder aggregates multi-scale features for precise segmentation before classification into benign or malignant nodules. The Swin Transformer blocks use windowed and shifted window attention with normalization and MLP layers to efficiently model global context.

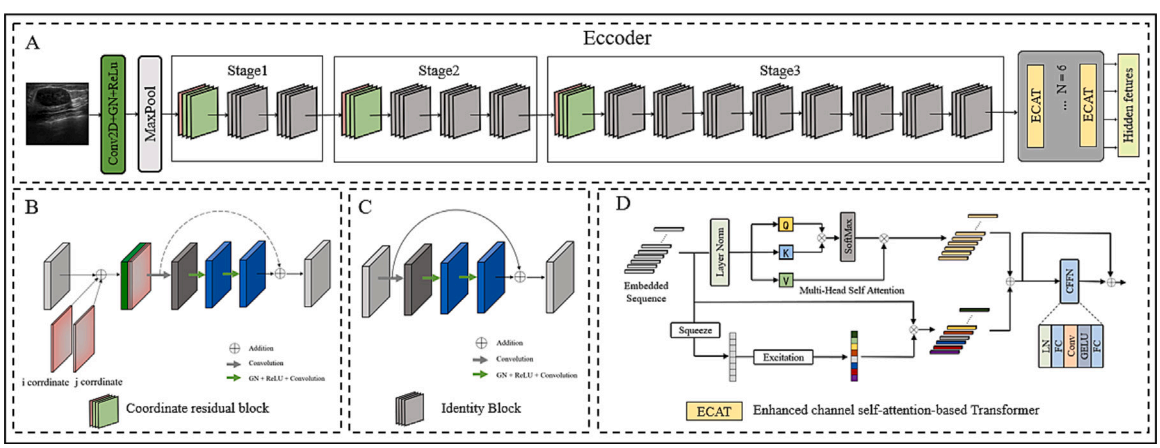

Fig. 7. An example of the basic structure of encoder HEAT-Net for TC segmentation [173]. For this figure, convolutional stages extract local spatial features and an Enhanced Channel Self-Attention Transformer (ECAT) captures global dependencies before classification. Coordinate residual blocks improve localization by adding positional information, while identity blocks maintain feature continuity through skip connections. The ECAT module uses multi-head self-attention to model long-range relationships, enhancing feature representation.

accuracy in binary recurrence classification. This approach streamlines the handling of unstructured patient information, eliminating the need for labor-intensive data refinement, and holds significant promise for training AI models in healthcare. The variability in features between TN, particularly in TIRADS level 3, can lead to inconsistent diagnoses and unnecessary biopsies. To address this in [177], ViT-based TCD, utilizes contrast learning to enhance diagnostic accuracy and biopsy specificity. By incorporating global and local features, this model minimizes the distinction between nodules of the same category. Test results indicate an accuracy of 86.9%, outperforming classical DL models. It offers automatic classification of TIRADS 3 and malignant nodules in US images, promising improved CAD and precise analysis. Optical

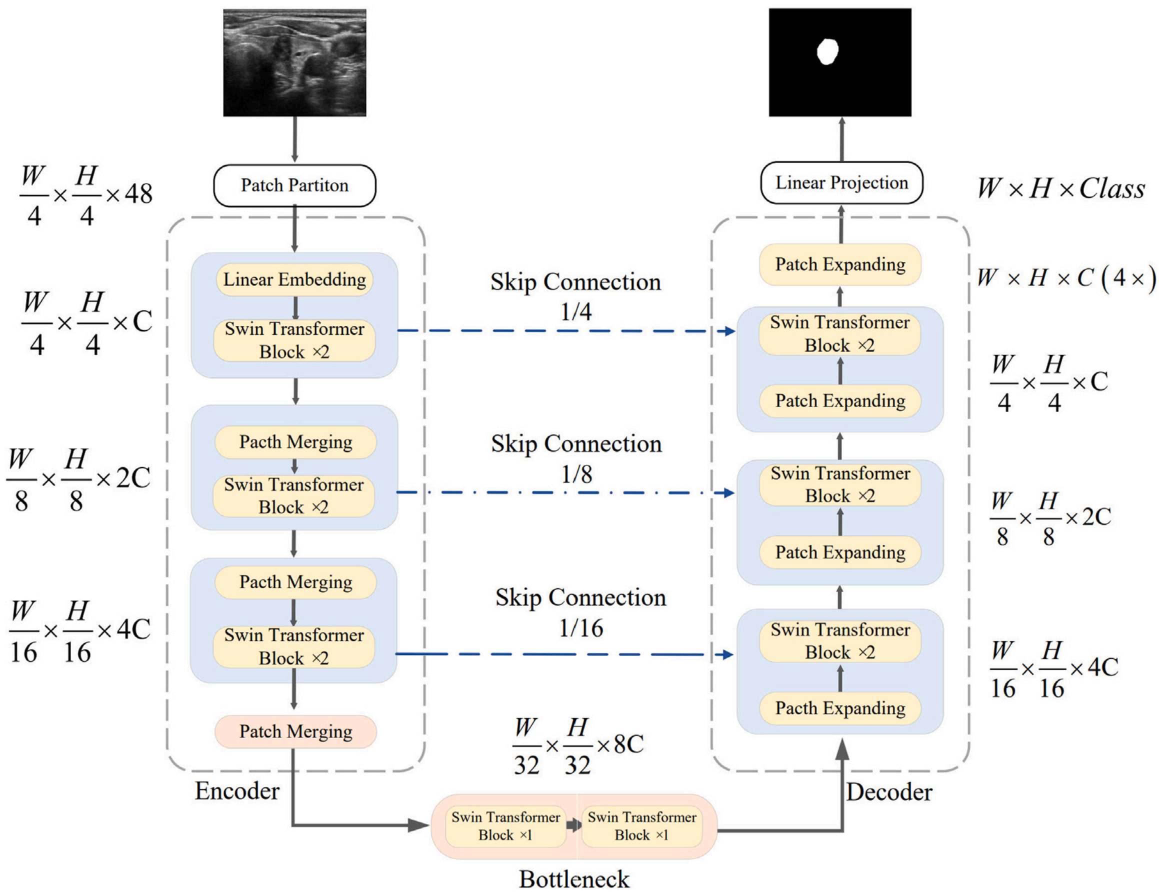

Fig. 8. The Swin-Unet architecture [34]. In this figure hierarchical Swin Transformer encoder extracts multi-scale features, a bottleneck stage consolidates high-level representations, and a symmetric decoder with patch-expanding layers reconstructs the output. Skip connections at multiple scales (1/4, 1/8, 1/16) fuse encoder and decoder features, preserving spatial detail. The final linear projection generates a high-resolution segmentation mask, enabling precise and accurate delineation of TN from US images.

coherence tomography (OCT) can aid in distinguishing normal and diseased thyroid tissue during surgery, but interpreting this type of image is challenging. Similarly, in [178] explored various DL models for classifying thyroid diseases using 2D and 3D OCT data from 22 surgical patients with thyroid pathologies. The 3D ViTs model achieved the best performance, with an accuracy of 0.90 for normal versus abnormal classification. Custom models also excelled on open-access datasets. These findings suggest that combining OCT with DL can enable real-time, automatic identification of diseased tissue during thyroid surgery.

Accurate segmentation of TNs in US images is crucial for early TC diagnosis. Addressing the challenges posed by weak image edges and complex thyroid tissue structure, the study [179] introduces LCA-Net method. It combines local features from CNNs and global context from Transformers, improving edge information capture. The model incorporates specific modules to handle different nodule sizes and positions, enhancing generalization. LCA-Net outperforms existing models on public datasets, demonstrating its potential for precise TN diagnosis in clinical settings. In this study [180], the authors focuses on improving the prediction of lymph node metastasis in papillary thyroid carcinoma by combining whole slide histopathological imagess (WSIs) and clinical data. They introduce a Transformer-guided multi-modal multi-instance learning framework that effectively groups high-dimensional WSIs into low-dimensional feature embeddings and explores shared and specific features between modalities. The approach achieved an impressive area under curve (AUC) of 97.34% on their dataset, outperforming state-of-the-art methods by 1.27%, highlighting its potential in improving precision medicine decisions based on multi-modal medical data fusion. Diagnosing lymph node metastasis in papillary thyroid carcinoma typically relies on analyzing large WSIs. To enhance accuracy, a novel Transformer-guided framework is introduced in [181], leveraging Transformers in three critical aspects. It incorporates a lightweight feature extractor, a clustering-based instance selection scheme, and a Transformer-MIL module for effective feature aggregation. The model further benefits from an attention-based mutual knowledge distillation paradigm. Experimental results on a WSI dataset outperform state-of-the-art methods by a significant margin, achieving a 2.72% higher AUC. Xiao et al. in [182] aim to address the challenges of diagnosing TC, particularly in cases where US images suffer from noise and artifacts, leading to a certain misdiagnosis rate in clinical practice. They highlight the need for further diagnosis using plain and contrast-enhanced CT scans.

While plain CT provides valuable information about the general structure of organs, it may not be sufficient for detecting fine details of tumors or surrounding tissues. In contrast, contrast-enhanced CT offers better contrast by injecting a contrast medium, allowing for clearer visualization of tissue details and organ margins. This technique aids in identifying margin erosion, a critical sign for diagnosing tumors such as colorectal cancer, where the erosion is more apparent, helping doctors accurately locate the tumor and guide treatment decisions. However, the latter relies on the use of a contrast agent and exposes the patient to ionizing radiation. To mitigate these challenges, the authors propose an improved Unet architecture. Their approach involves using a convolutional Transformer module to learn global information from high-dimensional features. They also incorporate a texture feature module to extract local texture information from plain

CT scans and integrate edge information obtained from superpixels as prior knowledge. The ultimate goal is to generate enhanced CT images with clear texture and higher quality, providing a valuable tool for TC diagnosis without the need for contrast agents and ionizing radiation. Histopathological images carry valuable information for tumor classification and disease prediction, but their size hinders direct use in CNNs. This study [183] introduces Pyramid Tokens-to-Token ViT, a lightweight architecture with multiple instance learning based on the ViTs. The method uses Tokenization technique for feature extraction to reduce model parameters. It also incorporates an image pyramid to capture local and global features, significantly reducing computation. Experiments on thyroid pathology images yield superior results compared to CNN-based methods, balancing accuracy and efficiency. The authors in [184] aim to utilize multi-instance learning for the diagnosis of TC based on cytological smears. These smears lack multidimensional histological features, necessitating the mining of contextual information and diverse features for improved classification performance. To address these challenges, they introduce a novel algorithm called PyMLViT, which consists of two core modules. First, the pyramid token extraction module is designed to capture potential contextual information from smears. This module extracts multi-scale local features using a pyramid token structure and obtains global information through a ViTs structure with a self-attention mechanism. Second, they construct a multi-loss fusion module based on the conventional multi-instance learning framework. To enhance the diversity of supervised information, they carefully allocate bag and patch weights and incorporate slide-level annotations as pseudo-labels for patches during training. In Table 8, Transformer-based models' performance (in %) for TC diagnosis is summarized.

4.2. LLM-based TCD methods

The LLMs are advanced NLP model, trained on vast datasets to process, understand, and generate human-like text. Using DL and Transformer architectures [191], LLMs excel in tasks like answering questions, summarizing, translating, and creating content. In TCD, LLMs could be fine-tuned to assist analyzing patient data, diagnostic reports, and medical literature, identifying patterns, and offering insights for early detection and personalized treatment. Their ability to process complex medical information enhances diagnostic accuracy, supports clinicians, and improves patient outcomes, making them invaluable in advancing TC care and treatment. Several studies have suggested using LLM models for identifying TC from US images. For instance, researchers in [186] evaluates a privacy-preserving LLM for extracting critical clinical information from TC pathology reports. Using FastChat-T5, the model answered 1008 questions about staging and recurrence risk across 84 reports. Concordance rates between the LLM and human reviewers averaged 89%, with the LLM completing tasks significantly faster (19.6 min vs. 206.9 min). While accurate for binary questions, challenges arose in complex queries. The findings highlight the potential of tailored LLMs for efficient, privacy-compliant clinical data extraction. Moving on, Raghunathan et al. [185] evaluate LLMs, including ChatGPT-3.5 and GPT-4, in addressing thyroid disease patient queries compared to verified doctors. Using a 4-point Likert scale, the proposed LLMs outperformed physicians in accuracy, quality, and empathy. GPT-4 scored highest across metrics. The findings highlight LLMs' potential to enhance patient communication, reduce clinician workload, and mitigate burnout by providing accurate and empathetic answers to complex medical questions in thyroid care. Wu et al. [187] evaluated the feasibility of leveraging LLMs like ChatGPT 4.0 to enhance TIRADS and pathology as the reference standard. Among 1161 ultrasound images analyzed, ChatGPT 4.0 outperformed others in consistency and diagnostic accuracy, especially when combined with image-to-text strategies. It matched or exceeded human-LLM interactions and showed potential to improve diagnostic efficiency while maintaining interpretability. Differently, Shah et al. [188] presented EndoGPT, a LLM-based tool developed for TN management using GPT-4o, prompt engineering, and knowledge retrieval. Tested on 50 patient scenarios, it achieved a high overall concordance with expert surgeon plans, excelling in diagnosis and operational decisions, though less so in operation type (69%). While not a replacement for clinicians, EndoGPT highlights the potential of LLMs in aiding medical decision-making, education, and enhancing accessibility to clinical guidelines. Similarly, [189] introduced AIGC-CAD model for TN. Inspired by ChatGPT, it integrates human–computer interaction to enhance diagnostic accuracy using 19,165 ultrasound cases. By combining DL models and semantic understanding, model provides transparent diagnostic rationales and improves physician confidence. The model enhances junior radiologists' sensitivity and specificity by over 20%, bridging skill gaps. Its explainable and interactive features mark a paradigm shift in CAD applications. Wang et al. [192] evaluates GPT-4's capabilities in thyroid ultrasound diagnosis and treatment recommendations using 109 cases. GPT-4 excelled in report structuring, clarity, and professional terminology but showed limitations in diagnostic accuracy. The chain of thought method enhanced interpretability, and the AI-generated reports were largely indistinguishable from human-written ones in a Turing Test. Zhang et al. [190] presented BertTCR, an advanced DL framework for predicting cancer-related immune status via *T* cell receptor (TCR) repertoire analysis. BertTCR leverages a pre-trained protein-BERT model to extract high-dimensional features, incorporating CNN, multiple instance learning, and ensemble techniques to enhance accuracy. Validated on datasets for thyroid and lung cancer, it achieves notable AUC improvements over existing methods. The framework's flexibility supports universal cancer detection and immune status assessment. BertTCR's findings emphasize its potential for early cancer detection, personalized medicine, and broad applications in immune-related diagnostics. The paper [193] introduces DrVD-Bench, a structured benchmark for evaluating vision language models (VLM)s in medical image diagnosis. It consists of three modules: Visual Evidence Comprehension, Reasoning Trajectory Assessment, and Report Generation Evaluation, encompassing 7789 image-question pairs across CT, MRI, US, radiography, and pathology modalities. The study benchmarks 19 VLMs and finds that while they perform well on basic tasks like modality recognition, their accuracy drops significantly on more complex tasks, such as lesion localization and diagnosis, and they often generate plausible but unsupported clinical reports, a phenomenon known as hallucination. The findings suggest that while VLMs show early signs of clinical reasoning, they still struggle to fully interpret medical images like human doctors, with specialized models like HuatuoGPT-Vision-34B performing better due to medical-specific fine-tuning. This benchmark provides a foundation for improving the clinical reliability of AI systems in medical image analysis, including TC diagnosis. The paper [194] presents XrayGPT, a conversational medical model that leverages VLMs to analyze chest radiographs and provide summaries and answers to open-ended questions about them. The model integrates a specialized medical visual encoder and a fine-tuned large language model (LLM), using interactive and high-quality summaries from radiology reports to improve its performance. XrayGPT significantly outperforms the baseline model, achieving a high level of accuracy in expert evaluations by certified medical doctors. It effectively addresses key medical conditions, including pulmonary edema, congestive cardiac failure, and metastatic lung nodules, providing insightful summaries and treatment recommendations. Additionally, the model demonstrates potential in detecting various abnormalities, such as TC, by leveraging its deep understanding of chest radiographs and medical knowledge, establishing a foundation for broader applications in medical image analysis. The study [195] investigates the vulnerability of state-of-the-art VLMs to incidental prompt injection attacks in histopathology, particularly through the presence of handwritten labels and watermarks on pathology slides. The authors demonstrate that these subtle markings can dramatically

Table 8
Summary of Transformer and LLM-based models' performance (in %) for TC diagnosis.

Ref.	Transformer	Task	Dataset	Acc	Sen	Spe	AUC	F1	Improvement	PL?
[168]	BERT, RoBERTa, LongFormer, DeBERTa, and GatorTron	C	Private data	97.10	98.80	92.80	–	96.50	The proposed model can achieve satisfactory classification accuracy and identified a large numbre of characteristics comparable to experienced radiologists and can save time and effort as well as deliver potential clinical application value	No
[170]	Trans-CEUS	S		86.59	–	–	–	–	Demonstrated significant improvement when compared to previous approaches, showcasing its effectiveness in the tasks of lesion segmentation and TN diagnosis.	No
[178]	3D vision	C		90.00	–	–	–	–	Efficiently classifying thyroid diseases.	No
[177]	ViT	C		86.90	87.10	86.10	–	92.4	Improve accuracy of diagnosis and specificity of biopsy recommendations. Minimize the representation distance between nodules of the same category.	Yes
[185]	GPT-4	MQA		95.25	–	–	–	–	GPT-4 responses scored highest for accuracy, quality, and empathy compared to GPT-3.5 and doctors.	No
[180]	GMMMIL	P		93.88	–	–	97.34	94.65	Experimental results on the collected lymph node metastasis dataset demonstrate the efficiency of the proposed method.	No
[181]	Tiny-ViT	P		–	–	–	98.35	92.97	Improve predict lymph node metastasis from WSIs efficiently using a novel Transformer-guided.	No
[184]	PyMLViT	C		87.50	–	–	–	–	Optimize the training process of the network.	
[186]	FastChat-T5	MQA		88.86	–	–	–	–	Significant reduction in time for data extraction.	No
[176]	BERT	C		–	–	–	–	88.00	Analyze unstructured clinical text information on diagnosis of the recurrent PTC efficiently.	No
[187]	ChatGPT-4	D		86.00	–	–	–	–	Optimal performance in US diagnosis of TN.	
[188]	GPT-4o	D		93.00	–	–	–	–	High accuracy in diagnosis and management of TN.	No
[189]	LlaMA2-13B	D		87.50	86.20	88.30	90.90	85.00	Diagnosis accuracy surpassed standalone CAD models and human performance.	Yes
[171]	Network	C	CIFAR-10	–	–	–	91.3	–	Attention nets outperform complex attention-based architectures and other competing methods in tasks such as thyroid malignancy prediction.	No
[172]	ViTs	C	DDTI	97.63	–	–	–	96.67	The SVM classification produces better performance than the Softmax classification for all of the models with the Hybrid ViT	No
[183]	T2T-ViT	C	TCGA	86.60	–	–	–	–	The model parameters are reduced , and the model performance and computation are greatly improved compared with CNN.	No
[190]	BertTCR	P	NCBIZenodo	96.60	100	100	100	95.80	High performance in predicting thyroid cancer-related immune.	No

Abbreviations: segmentation (S); Classification (C); Prediction (P); Medical question answering (MQA); Diagnosis (D), Project link (PL).

influence the diagnostic accuracy of VLMs, with the models often relying on incorrect labels over actual tissue characteristics. The research focuses on various clinical tasks, including T-stage classification, lymph node infiltration, and mutational status for TC, revealing that VLMs can be misled by even misleading watermarks, resulting in significant performance degradation. Despite prompt engineering efforts to mitigate this, the models remained highly susceptible to such influence. The findings highlight the critical need for awareness of potential vulnerabilities in the application of AI tools in medical diagnostics, particularly as they transition into clinical use. The paper [196] presents a novel approach to US image segmentation by combining Grounding DINO, a VLM, with SAM2 for multi-organ segmentation across various datasets, including those for TC. This method leverages the flexibility of natural language prompts to enable segmentation of different anatomical structures without the need for extensive retraining or task-specific modifications. The model adapts to US images using Low-Rank Adaptation (LoRA) and outperforms several state-of-the-art segmentation methods across various organs. It achieves strong generalization capabilities, demonstrating its ability to work on previously unseen datasets and organ structures, which makes it highly scalable for clinical applications, particularly for TC detection. The approach is efficient in terms of computation, with real-time segmentation performance, offering a promising tool for improving the automation of US image analysis. Table 8 presents a summary of the performance of various LLM-based TCD methods across different metrics.

Despite the promising potential of advanced models like ViT and LLMs in diagnosing TC using AI, several limitations exist. These models require large datasets and may struggle with generalization across diverse datasets. They are also computationally intensive, which can limit their accessibility in clinical settings. Additionally, these models lack interpretability, making it difficult for doctors to understand the decision-making process.

4.3. Comparative analysis and discussion

The hybrid Transformer–CNN model leverages the complementary strengths of both architectures. CNNs excel at extracting local, fine-grained texture and edge-level features, whereas Transformers are designed to capture long-range dependencies and global contextual information. Using either model in isolation risks losing important information: CNNs may fail to capture global semantic relationships across an image, while Transformers typically require very large datasets to avoid overfitting. Their integration enables richer feature representations, improved generalization, and superior diagnostic accuracy for thyroid carcinoma detection, particularly in heterogeneous imaging datasets. Beyond hybrid approaches, advanced Transformer-based architectures are increasingly shaping thyroid carcinoma research. ViTs and their derivatives, including Swin-Transformer and Pyramid Token-to-Token ViT, consistently outperform conventional CNN-based frameworks by incorporating global self-attention mechanisms. For instance, ViT models trained on large-scale ultrasound datasets achieved classification accuracies exceeding 97%, particularly due to their ability to capture both local lesion boundaries and broader tissue context. However, these high accuracies are often reported in single-center, curated datasets with controlled imaging conditions. In contrast, studies employing multi-institutional or public datasets observed more moderate performance, with accuracies closer to 85%–90%, reflecting the challenges of variability in image quality, labeling consistency, and patient demographics. Thus, while ViTs deliver state-of-the-art performance in controlled environments, their robustness across diverse clinical settings remains an open challenge.

LLMs have similarly demonstrated promising applications in the processing of TC pathology reports, staging documentation, and patient queries. When fine-tuned on domain-specific corpora, these models show high concordance with expert oncologists and, in some cases, surpass physicians in communication clarity and empathy. For example, models such as Clinical-BERT and BertTCR effectively exploit both textual reports and immune repertoire data, highlighting their adaptability across structured and unstructured modalities. Nevertheless, performance can vary depending on the representativeness of the training corpus: models trained on general medical literature may underperform in capturing subtle oncological terminology compared to those trained on curated thyroid-specific corpora. Emerging VLMs extend these capabilities by jointly interpreting imaging and textual data, supporting multi-modal analysis that aligns radiological, pathological, and clinical narratives. This integration offers substantial diagnostic potential, particularly for staging and longitudinal monitoring. Early studies report strong improvements in sensitivity when combining ultrasound images with corresponding textual findings, underscoring the importance of cross-modal reasoning. Yet, challenges such as hallucination, vulnerability to adversarial prompts, and computational cost remain unresolved. In summary, advanced Transformer-based methods outperform traditional CNNs by capturing global semantic relationships, but their true clinical impact depends on rigorous external validation and careful integration into diagnostic workflows. Models achieving the highest reported accuracies typically benefit from either hybridization (Transformer–CNN), large-scale or well-curated training data, or multi-modal integration. Performance discrepancies across datasets underscore the importance of standardized benchmarks and multi-center evaluations before clinical deployment of these models in thyroid carcinoma diagnosis.

In comparison with prior research, the trends observed in Section 4 align with the broader literature: (i) hybrid Transformer–CNN models achieve superior diagnostic performance due to complementary feature learning, (ii) ViTs achieve the highest reported accuracies but exhibit dataset-dependent variability, and (iii) LLMs and VLMs demonstrate flexibility and scalability across diverse modalities, though they require extensive domain-specific adaptation. Table 7 and Figs. 4–5 further illustrate that the highest-performing models combine architectural innovation with favorable dataset conditions, whereas performance drops significantly in heterogeneous, real-world datasets.

5. Limitations and challenges

Recognizing the obstacles in integrating AI solutions into healthcare practices, including infrastructural, regulatory, and cultural challenges, is essential. Highlighting the critical role of cross-disciplinary cooperation in seamlessly integrating AI into healthcare systems, thereby maximizing its beneficial impacts on patient health outcomes.

Although AI methodologies have shown promise in TC diagnosis, they face challenges that hinder the development of efficient solutions, lead to increased expenses, and limit their broad application. For precise detection of TC, it is essential to collect and securely consolidate all relevant data in a single repository, unless adopting federated learning (FL) approaches, as Himeur et al. suggest [197]. Following this, algorithms capable of identifying all forms of TC must be developed. Comprehensive TCD should include an extensive array of training and testing images, diagrams of nodules, and detailed classifications of nodule characteristics across different sizes, as Shah et al. recommend [198]. It is crucial for these datasets to be continually updated with data from MRI, CT scans, X-rays, and other clinical images to assess TC accurately. Inclusion of demographic details such as race, ethnicity, gender, and age is also necessary. Establishing a centralized database accessible to all healthcare facilities for testing, validating, and implementing AI algorithms on the collected data is critical, following Salazar et al.'s guidance [199]. Additionally, a succinct overview of further limitations and challenges yet to be addressed is provided.

(a) Data heterogeneity, and annotation quality: In TC diagnosis, a major challenge is the lack of detailed, well-annotated datasets that thoroughly document the disease's incidence and progression. Elmore et al. [200] highlight the difficulty in collecting and validating TC-related data due to the absence of comprehensive clinical databases. AI algorithms struggle with accurate TC diagnosis because of limited labeled cases correlating with clinical outcomes, as noted by Park et al. [201]. Although large datasets are crucial for neural networks to produce accurate results, selective data incorporation during training is necessary to avoid harmful noise. Imaging modalities like CT and MRI, though available, are costly and not always accessible, as Ha et al. [25] point out. US imaging, combined with physical exams, fine-needle aspiration biopsies, or radioisotope scans, is preferred for its cost-effectiveness and accessibility. However, Zhu et al. [202] note that US accuracy in distinguishing malignant from benign nodules can vary and images may be noisy. Cancerous cells in thyroid tissue are often a small fraction of the total dataset, leading to a skewed distribution that can impair AI detection performance, as observed by Yao et al. [203]. Researchers face challenges in developing algorithms to handle limited, noisy, sparsely annotated, incomplete, or high-dimensional samples efficiently. Annotation is crucial but time-consuming and costly, impacting AI algorithm precision due to inconsistent labeling, as discussed by Sayed et al. [204] and Yao et al. [203]. A lot of work has been proposed in the field of TC diagnosis using AI models, but it is challenging to compare these studies due to significant differences in the datasets used, including variations in content and resolution, as well as differences in the hardware employed. These discrepancies hinder the ability to assess the true performance of different methods and models, making it difficult to draw reliable comparisons. This variability across datasets and hardware configurations complicates the evaluation and generalization of findings in real-world clinical settings.

(b) Hyperparameters of DL models: Designing the effective DL algorithm is crucial for overcoming different challenges, especially in diagnosing TC. The task of precisely differentiating between malignant and benign tumors, is complex due to their significant similarities, as highlighted in the study by Wang et al. [205]. Addressing this challenge may require significantly increasing the number of DL layers for feature extraction. However, such an increase can lead to longer processing times, particularly with large datasets, which may delay timely cancer diagnoses for patients, as pointed out in the research by Lin et al. [26].

(c) Computation cost and storage: Pose notable hurdles in algorithm development. Time complexity, a key measure in algorithm evaluation, assesses computational complexity by approximating the count of basic operations performed and its relationship with the size of input data. Typically represented as $O(n)$, where n is the size of the input, often measured by the bits required for its representation, this concept is thoroughly examined in the study by Al et al. [206]. Particularly in AI research related to TCD, researchers are tasked with finding algorithms that offer a harmonious blend of accuracy and computational efficiency. Their goal is to develop algorithms that can quickly process large datasets while maintaining precise results. Furthermore, the extensive amount of data used in these algorithms sometimes exceeds the storage capabilities, an issue underscored in the research by Lin et al. [26]. The issue of computational cost remains a significant challenge, especially in healthcare applications such as TC detection and diagnosis. With the increasing reliance on large models to analyze complex medical data, computational demands rise significantly. This presents a particular barrier for healthcare institutions, especially in the case of TC, where early and accurate diagnosis is crucial for patient outcomes. These institutions often lack the necessary infrastructure to handle such heavy computational loads, limiting the potential of AI in medical practice. To address these challenges, it is essential to adopt strategies to improve computational efficiency. Techniques such as Model Distillation, Parameter-Efficient Fine-Tuning, and the use of smaller, more efficient models offer potential solutions. These approaches could make it feasible to deploy AI models that are both resource-efficient and accurate in detecting TC, even in environments with limited computational capabilities. By focusing on these strategies, it is possible to bridge the infrastructure gap and ensure that AI tools for TC diagnosis are available and effective. In [207], the authors addresses the challenge of computational resource limitations in DL for healthcare, particularly as models become increasingly complex and data grows exponentially. Three primary strategies were discusses for mitigating these constraints: data strategies such as informative data subset selection to reduce computational costs, model strategies including model compression and parameter pruning to optimize performance while minimizing resource usage, and computing strategies like low-precision and mixed-precision computing to accelerate calculations and reduce memory usage. They emphasizes the importance of balancing computational efficiency with the need for high accuracy in healthcare applications, where errors can directly impact patient outcomes. The commentary highlights innovative approaches in the healthcare sector, such as the use of foundation models and low-precision computation, to make DL models more feasible for resource-limited environments.

(d) Data loss and error vulnerabilities: The shift towards digital medical records is crucial, notably in cancer diagnosis using slide images. This latter facilitate the use of AI for pathologic examinations [208]. Nonetheless, medical digitalization encounters specific challenges. There exists the danger of losing critical information during the digital conversion process and potential inaccuracies due to data compression methods applied in autoencoder algorithms. Thus, selecting the right digitalization technology is essential to ensure the preservation of data fidelity and authenticity [209]. The subtle contrast between the thyroid gland and surrounding tissues complicates accurate analysis and diagnosis of TC. Despite AI's inherent autonomy, it is prone to making errors. For example, training an algorithm with TCDs for identifying cancerous regions can lead to biased predictions if the training datasets are biased. These biases may then lead to a series of erroneous results, which could go unnoticed for a significant duration. Identifying and correcting the source of these errors, once recognized, can be a laborious process, as explored in the research by Karsa et al. [210].

(e) Unexplainable AI: The application of AI in healthcare, while transformative, often produces "black box" outcomes where the reasoning behind predictions remains opaque, leading to skepticism among clinicians and hesitation in adopting these tools for TC diagnosis, a concern highlighted in the research by Sardianos et al. [211]. This opacity poses significant ethical and clinical risks, as physicians are reluctant to base treatment decisions on models whose internal logic cannot be scrutinized or justified, particularly when patient outcomes are at stake. The lack of interpretability not only undermines physician trust but may also result in inappropriate or delayed clinical decisions if outputs are accepted uncritically. To overcome these challenges, XAI approaches — such as saliency maps, SHAP values [212], Grad-CAM, and counterfactual reasoning — should be implemented to provide transparent visual and quantitative justifications for model outputs, allowing clinicians to cross-check predictions against their expertise. Moreover, interpretability must be embedded within clinical decision-support systems, presenting explanations in a user-friendly manner that integrates seamlessly with existing diagnostic workflows.

(f) Lack of cancer detection platform: A significant barrier in detecting various cancers, including TC, is the absence of platforms that facilitate the replication and evaluation of previous research. This gap presents a considerable challenge that hampers the assessment of AI algorithms' performance, thereby hindering improvements [118]. The presence of online platforms that offer cutting-edge algorithms, extensive datasets, and expert insights is crucial for assisting healthcare practitioners, specialists, researchers, and developers in making the right decisions with reduced chances of error. Moreover, this kind of platforms are vital in augmenting clinical diagnoses, as they enable more thorough Examination and assessment [213].

(g) Publication bias and reproducibility: Publication bias is a significant issue in TC research, where studies demonstrating positive or innovative results are often published, while those reporting negative outcomes or failures of treatments or techniques are overlooked or restricted. This phenomenon creates an inaccurate picture of the effectiveness of available therapies, hindering progress in developing effective and comprehensive treatment strategies. Additionally, medical AI research, including its applications in diagnosing and treating TC, faces a reproducibility crisis. Many AI-based systems fail to achieve the same results when used in different environments or datasets, highlighting the technical and knowledge challenges in applying these systems broadly in the medical field. Variations in data, differences in data collection and analysis methods, and the exclusion of environmental and behavioral factors that affect patients all contribute to the difficulty of reproducing results. Therefore, addressing these challenges transparently and reviewing both positive and negative outcomes is crucial to ensure genuine scientific advancement in the field, with a need for caution when relying on AI-driven tools in clinical applications to ensure they are verifiable and applicable across diverse contexts and settings.

(h) Challenges in clinical applications: The integration of ML and transformer models for TCD has faced several challenges, leading to negative results and failed clinical implementations. Issues such as limited and imbalanced datasets, overfitting, lack of model interpretability, and bias in training data have hindered their accuracy and generalization across diverse patient populations. Moreover, the complexity and computational demands of these models, combined with regulatory hurdles and privacy concerns, have slowed their adoption in clinical settings. The failure to seamlessly integrate these technologies into existing workflows, coupled with insufficient validation in real-world conditions, has led to difficulties in achieving reliable, widespread use. Overcoming these barriers requires improved data quality, more transparent models, and greater collaboration between researchers, clinicians, and regulatory bodies.

(i) Regulatory barriers: Applications of AI in TC diagnosis face several critical challenges in clinical practice. First, there are regulatory barriers, such as the need for approval from the Food and Drug Administration (FDA) or European Medicines Agency (EMA), which hinders the widespread adoption of these technologies. Furthermore, integrating AI systems into current clinical workflows is complex, as healthcare environments are often resistant to change and require substantial adjustments to existing practices. Additionally, physicians encounter issues related to trust and acceptance, as AI systems are often viewed with skepticism, and their decision-making processes lack transparency, making it harder for doctors to embrace them. Cost-effectiveness is another crucial factor, as AI technologies require significant investments, which may not be justified in many resource-limited healthcare settings. Moreover, AI systems may experience performance degradation in real-world environments compared to controlled settings, leading to a reliability gap. Finally, issues of responsibility and medical negligence arise, as the role of AI in clinical decision-making could lead to complications if the systems fail or produce incorrect results. Addressing these issues comprehensively is essential to bridge the gap between research and effective clinical application.

(j) Clinical workflows: The integration of modern systems and technologies, such as AI, into current clinical workflows presents significant challenges in hospitals and healthcare centers. Clinical workflows encompass all the daily procedures followed by doctors and nurses in diagnosing and treating patients, relying on specific tools and software developed and refined over years. When new technologies, like AI tools for medical image analysis or decision-support systems, are introduced, it can be difficult to seamlessly integrate them with existing systems. These new technologies often require substantial changes to both hardware and software, leading to coordination difficulties between the new and old tools. Furthermore, healthcare professionals need training and education to effectively use these modern systems, adding an additional burden to the daily workflow. These challenges can delay the actual implementation of these innovations in the clinical setting, affecting the effectiveness of healthcare delivery and limiting patients' ability to benefit from technological advancements.

(k) Physician trust and acceptance issues: Doctors' trust in AI technologies and their acceptance of them are significant challenges in applying these systems to clinical practice. Physicians often have concerns about the accuracy and reliability of AI systems, particularly if the technology relies on data or algorithms that are not entirely transparent or well-understood. Some doctors prefer traditional methods based on their extensive clinical experience, making them hesitant to adopt new systems. There is also a fear of losing control over treatment decisions, as some worry that AI might make decisions that are inaccurate or poorly thought out, raising concerns about legal liability in the event of an error. Additionally, there may be a perception that AI could replace the human role of the doctor, leading to a reduction in personal interaction between doctors and patients. To ensure the acceptance of these technologies, continuous training should be provided to explain how these systems work, how they can be improved, and to strengthen the understanding of how these systems integrate with human work, rather than replacing it. This approach would help build trust in the ability of these systems to improve the quality of healthcare.

(l) Performance degradation in real-world vs. controlled settings: In real-world environments, such as hospitals or clinics, AI faces significant challenges that affect its performance compared to controlled environments like laboratories or research studies. In real-world settings, the data collected is typically more complex and varied, making it difficult for pre-trained models to adapt effectively. For instance, the data may be incomplete, noisy, or heterogeneous due to differences in devices used for data collection or varying recording methods across sites. Moreover, in controlled environments, most factors such as the type of equipment used, the surrounding environment, and even patient interaction are controlled, ensuring more accurate and consistent results. However, in real-world environments, the data may be influenced by factors like environmental changes (lighting, noise), patient behavior variations, or shifts in administrative policies. Real-world work settings are also often under pressure, which may lead to the use of lower-quality devices, delays in data collection, or even slower model responses, all of which can hinder the model's ability to achieve optimal performance. Ultimately, this disparity between controlled and real-world environments can lead to a significant decline in AI's ability to deliver accurate and reliable results, limiting its effectiveness in clinical practice.

(m) Liability and malpractice considerations: Medical liability and negligence considerations in AI applications are related to determining responsibility when errors or harm occur due to the use of these systems in healthcare. When AI is used for diagnosing diseases or providing treatment recommendations, it can be difficult to pinpoint responsibility in the event of a diagnostic or treatment error. For example, if an AI algorithm misdiagnoses TC and leads to inappropriate treatment, who is accountable? Is it the system designer, the hospital that implemented the system, or the doctor who followed the recommendations without thorough verification? This raises legal and ethical questions about medical negligence liability. In many cases, AI may be viewed as a complementary tool to the human doctor, complicating the determination of whether the issue stemmed from a fault in the system or a human error in decision-making. Additionally, the use of smart systems requires legal and ethical safeguards to protect patients from potential harm and to establish how to address cases of negligence in utilizing these systems or their misalignment with optimal clinical practices.

(n) Absence of standardized statistical reporting: While most research works reported performance metrics such as accuracy, sensitivity, specificity, AUC, and F1-score, very few provided statistical measures such as P-values, effect sizes, or confidence intervals. This lack of consistency makes it difficult for researchers to conduct formal meta-analyses or meta-regressions, as pooling and comparing results requires uniform statistical outputs. As a result, evidence synthesis in this field often remains descriptive rather than quantitative. This limitation reduces the strength of generalizable conclusions and highlights the need for future studies to adopt more rigorous and standardized statistical reporting practices.

6. Future research directions

This segment delves into the anticipated developments of AI in identifying TC, scrutinizing forthcoming trends and advancements alongside their ethical repercussions. Ethical concerns encompass more than the immediate area of focus; issues regarding data privacy, responsibility, and fairness are also discussed. This part underscores research avenues poised to significantly improve TCD classification, and prediction.

(a) Using cloud, fog, and edge computing: The concept of edge networks merges edge computing with AI, allowing AI algorithms to operate closer to where data originates [214], a discussion brought forward by Sayed et al. [215]. This method enhances efficiency and cost-effectiveness for data-intensive applications, minimizing the requirement for extensive communication between patients and healthcare providers. By positioning data and storage closer to users in the healthcare field, this approach enables direct and swift access, a point highlighted by Alsalemi et al. [216]. To further improve the detection of TC within edge networks, fog computing is integrated. Fog computing introduces a distributed framework that bridges cloud computing and data generation sources, offering a versatile distribution of computational and storage capacities at key locations to boost overall system performance [217]. Cloud computing acts as a pivotal facilitator for the efficient functioning of AI-driven TCD systems, offering readily available access to data storage, servers, databases, networks, and applications for healthcare professionals, contingent upon internet connectivity. This integrated approach has proven its worth in medical scenarios, such as in the TC detection, as corroborated by several researches [218,219].

(b) Reinforcement learning (RL): RL, a branch of ML enables agents to navigate and make decisions in evolving environments by engaging in a learning cycle of trial and error, observation, and interaction [220, 221]. The interest in leveraging RL for diagnosing untreatable diseases and enhancing the support for medical decision-making processes has grown recently. For example, Balaprakash et al. [222] apply RL in cancer data classification, whereas Li et al. [223] explore the use of deep RL for lymph node dataset segmentation. In this approach, pseudo-ground truths are created using RECIST-slices, facilitating the simultaneous tuning of lymph node bounding boxes through the collaborative efforts of a segmentation network and a policy network.

(c) Transfer learning (TL): TL is recognized as an effective approach to reduce overfitting and improve the precision of diagnostic tools [95, 224–227]. This technique applies knowledge acquired from one domain to help solve related issues in another, such as shortening the duration of training and minimizing the amount of data needed [228,229]. It is particularly useful in diagnosing TG. For example, the Enhance-Net model, described in [230], could act as a foundational model to boost the efficacy of a targeted DL model aimed at analyzing medical images in real-time. Furthermore, in [117], the research focuses on identifying pertinent characteristics of nodules using CNNs. By transferring insights from generic data to a dataset of US images, they achieve a fusion of hybrid semantic deep features. The application of transfer learning has also proven beneficial in categorizing images of TNs, as shown in [120].

(d) Federated learning (FL): FL has gained traction in healthcare applications due to its capacity to enhance patient data privacy across different healthcare settings [197]. The influence of environmental factors on human health, which can subsequently impact economic stability, is substantial. An increase in the incidence of thyroid gland disorders has been observed across diverse populations. ML plays a crucial role in addressing such health issues by leveraging collected data to train models capable of foreseeing severe health conditions. Considering the critical need for maintaining the confidentiality of patient information among various health institutions, FL stands out as an optimal framework for these purposes. Lee et al. [231] conducted a study comparing the effectiveness of FL with five traditional DL techniques in analyzing and detecting TC.

(e) Panoptic segmentation (PS): Accurately identifying and segmenting objects with varied and intersecting features continues to be a significant hurdle, especially in the medical field. To tackle this issue, several scholars have developed holistic and unified segmentation methods [232]. PS has received considerable attention, merging the principles of instance and semantic segmentation to detect and delineate objects efficiently. It involves the classification of each pixel into distinct categories, whereas instance segmentation focuses on delineating individual object instances. AI has been applied to this framework through either supervised or unsupervised instance segmentation learning techniques, making it highly applicable to medical scenarios [233].

(f) IoMT images and 3D-TCD: The IoMT images has gained substantial interest in the healthcare industry in recent years. IoMT images seeks to advance the quality of healthcare services and minimize treatment expenses by facilitating the exchange of medical information between patients and healthcare providers via interconnected devices equipped with wireless communication technology. An instance of such integration is showcased in [234], where an AI-enhanced solution for the preemptive identification of TC within the IoMT images paradigm is introduced. This method employs CNN to refine the distinction between nodules, aiming ultimately at life preservation. Additional investigations pertinent to IoMT images [235]. 2D US is a prevalent technique for evaluating TNs, yet its static imagery might not fully capture the nodules' complex structures. Consequently, there is a growing interest in utilizing three-dimensional (3D) imaging and/or object reconstruction [236], which offers a holistic view of the lesion by reconstructing nodule characteristics, thereby facilitating enhanced discrimination between different diagnostic categories [237]. The capability of 3D US to analyze intricate growth patterns, edges, and forms from various perspectives and depths allows for a more accurate assessment of TNs morphological features compared to 2D US.

(g) AI-based thyroid surgical techniques: As surgical practices face complex challenges, the essential role of AI-driven robotic assistance is becoming increasingly recognized. AI has the capability to navigate clinical intricacies by processing and leveraging large volumes of data, offering decision-making support with precision that rivals that of medical experts [238]. Businesses are AI into surgical operations through the development of AI systems and the deployment of robots to aid surgeons in the operating room. These robots fulfill various functions, such as managing surgical tools, handling potentially contaminated materials and medical waste, conducting remote patient monitoring, and compiling patient information including electronic health records, vital signs, lab results, and video documentation [239]. It is therefore vital for surgeons to develop a comprehensive understanding of AI and its potential impacts on healthcare. While AI-enabled robotic surgery is still emerging, fostering interdisciplinary collaboration can accelerate the progress of AI technology, thereby improving surgical outcomes [240].

(h) Recommender systems (RSs): The vast amount of information produced by online medical platforms and electronic health records presents a challenge for TC patients seeking specific and accurate data [241]. Additionally, the substantial costs associated with healthcare data management can complicate the task for physicians handling a broad spectrum of patients and treatment alternatives. The implementation of RS has been suggested as a solution to improve decision-making within healthcare, reducing the load on both patients and oncologists [242]. Incorporating RS into digital health contributes to providing tailored and accurate recommendations through the analysis of large data sets using AI and ML techniques. By applying these advanced technologies, health data can be analyzed more quickly and accurately, leading to improved healthcare quality and more personalized decisions for patients. Additionally, this integration enhances privacy measures, as intelligent algorithms can be used to ensure the protection and secure handling of personal data [243].

(i) Image and video compression and denoising for TCD: The use of medical image and video compression plays a pivotal role in enhancing the detection and diagnosis of cancer, leveraging the advancements in digital imaging and telecommunications. This technological advancement allows for the efficient storage and transmission of high-resolution diagnostic images such as X-rays, MRIs, and CT scans, which are critical in identifying malignant tumors at early stages. Compression algorithms, both lossless and lossy, are meticulously designed to ensure that the integrity of the diagnostic information is maintained, making it possible for radiologists and oncologists to discern fine details crucial for accurate diagnosis. Furthermore, the reduced file sizes facilitate quicker transfer speeds across networks, enabling real-time collaboration and consultation among healthcare professionals worldwide, thereby significantly improving the speed and accuracy of cancer diagnosis. This is especially vital in remote or resource-limited settings where access to high-quality healthcare and specialist consultations might be restricted, thus democratizing the access to crucial diagnostic services and improving patient outcomes [244–248]. Moreover, employing DL-based preprocessing techniques for denoising, as demonstrated in [249], could further enhance classification performance and improve TC diagnosis. In addition, Transformer-based methods, which have been successfully applied to image tasks such as jet images backgound noise removal [250], should be adapted to medical denoising, particularly for TCD and classification.

(j) Features selection: The information gain (IG) method is useful in simplifying the classification of medical images. Researchers are encouraged to explore its usefulness for detecting TC by identifying the most informative features that distinguish between malignant and benign TNs. The process begins with data collection, where a comprehensive dataset containing relevant features, such as patient demographics, ultrasound characteristics, biopsy results, genetic markers, and blood test results, is gathered. Each instance in the dataset is labeled as benign or malignant based on definitive diagnostic methods. In the preprocessing stage, the data is cleaned by handling missing values and outliers and normalizing if necessary. Feature engineering may also be performed to create new features that enhance predictive power. The IG for each feature is then calculated, measuring how much it contributes to the classification. Features with high IG are considered more informative and are used to build a predictive model for TC detection. This method helps in selecting the most relevant features, thereby improving the accuracy and efficiency of the diagnostic process.

(k) Generating synthetic datasets: To advance TCD, future research should focus on enhancing dataset quality and diversity. Developing well-annotated datasets remains a challenge, which can be addressed through innovative techniques such as synthetic data generation, data augmentation, and multi-modal integration. Future work could explore more effective synthetic data generation methods, including enhanced GANs and variational DAEs, to create diverse, high-quality datasets that improve diagnostic accuracy and represent rare cancer subtypes [251, 252]. Further investigation into advanced data augmentation strategies, such as complex image transformations and domain adaptation, could enhance model generalization by expanding dataset variability [253]. Additionally, multi-modal integration, combining imaging, genomics, and clinical data, holds promise for improving robustness and predictive performance through deep learning models and novel fusion strategies.

(l) Employing SSL: self-supervised learning (SSL) offers a promising avenue for feature extraction in TCDs, particularly when working with large volumes of unannotated medical images. By leveraging unlabeled data, SSL techniques can learn meaningful and discriminative representations, reducing the dependency on extensive manual annotation efforts. This approach has shown potential in improving the robustness and generalization of diagnostic models by capturing complex patterns within medical imaging data. Incorporating SSL strategies could further enhance the development of automated diagnostic tools, especially in data-limited scenarios. Therefore, a dedicated discussion of SSL methods, including their applications and potential benefits in the TCD field, has been added to provide a comprehensive overview of emerging advancements. Recent studies have highlighted the effectiveness of SSL in medical image classification. For instance, [254] discusses various SSL strategies and their applications in medical imaging, emphasizing their potential to improve diagnostic performance. Additionally, research published in [255] explores SSL pre-training approaches, such as contrastive and masked modeling, demonstrating their superiority over traditional supervised methods in medical imaging tasks.

(m) Randomized controlled trials (RCT)s: AI systems are innovative tools that enhance healthcare, especially in diagnosis and treatment, such as in TC. However, their effectiveness and safety must be validated before clinical use. The validation process includes training models on diverse medical datasets, external validation in experimental settings, and continuous monitoring during real-world clinical use. The most crucial phase is RCTs, which compare AI systems with traditional treatments to ensure their safety and efficacy. For instance, [256] highlights the importance of external validation, continual monitoring, and RCTs for real-world AI deployment in healthcare. Additionally, [257] discusses a randomized controlled trial for AI-assisted diagnostic methods in papillary thyroid microcarcinoma (PTMC) using CT imaging, aiming to improve accuracy and reduce errors in diagnosis and treatment planning through AI techniques like GANs.

(n) Architectural complexity, computational requirements, and clinical integration: A key future direction for Transformer- and DL-based TC diagnosis lies in addressing the substantial computational complexity associated with these models. Despite their impressive diagnostic capabilities, most current research does not sufficiently tackle the challenges posed by the high computational demands of deep neural networks. Effective implementation requires advanced hardware resources, including high-performance processors, extensive memory capacity, and prolonged training and inference times. These demands are particularly acute when processing large-scale medical datasets, such as high-resolution ultrasound or histopathological images. To enable practical deployment in clinical settings, future work should prioritize the development of computationally efficient architectures, lightweight Transformer variants, and model compression techniques that maintain diagnostic accuracy while significantly reducing computational burden. Furthermore, advancements in computational efficiency will be crucial for enabling the integration of AI with real-time imaging modalities, such as live ultrasound or histopathology analysis. Such integration can support personalized medicine approaches, allowing rapid, patient-specific diagnostic and treatment decisions, and ultimately improving clinical workflow and patient outcomes.

7. Conclusion

This investigation delves deeply into ML and DL, highlighting their growing prominence due to their enhanced precision over other methods. It comprehensively reviews various algorithms and training models, discussing their benefits and drawbacks. Specifically, DL techniques are celebrated for their application in a myriad of real-world scenarios, notably for their generalization capabilities and resilience to noise. Nevertheless, significant challenges obstruct the full adoption of DL in detecting TC, with the lack of clean data and appropriate platforms being primary concerns. Tackling these data challenges with detailed precision is essential for creating effective and robust models for detecting more complex cancer stages.

Future research should aim at overcoming these hurdles and improving TCD classification and prediction methods. This study highlights the urgent need for increased research focus on TC diagnostics to match the high precision expectations of healthcare practitioners. While cancer detection in two or three dimensions is progressing, the limited expertise in handling various geometric transformations and multi-dimensional data compromises the accuracy in diagnosing life-threatening diseases. Therefore, it is vital to innovate in distinguishing between cancerous nodule sizes. Such innovations could significantly speed up treatment, improve diagnostic precision, foster proactive epidemiological tracking, and reduce death rates. Novel technologies like XAI, edge computing, TL, RL, FL for privacy-preserving mechanisms, and remote sensing are paving new paths in AI-based TCD research. These developments are crucial for medical professionals, simplifying the diagnostic process, reducing detection times, and enhancing patient confidentiality. Future research will explore the impact of these advanced technologies further. The objective is to create a major transformation in cancer detection approaches by crafting advanced, privacy-focused technologies for the identification of TC and extending into domains like Tele-health.

CRediT authorship contribution statement

Yassine Habchi: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Hamza Kheddar:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Yassine Himeur:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Mohamed Chahine Ghanem:** Writing – review & editing, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Ethical approval

This research was deemed not to require ethical approval.

Funding

The APC and Open Access fees for this work are funded by the University of Liverpool, United Kingdom.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] M. Castellana, A. Piccardo, C. Virili, L. Scappaticcio, G. Grani, C. Durante, L. Giovanella, P. Trimboli, Can ultrasound systems for risk stratification of thyroid nodules identify follicular carcinoma? Cancer Cytopathol. 128 (4) (2020) 250–259.
[2] L. Hitu, K. Gabora, E.-A. Bonci, A. Piciu, A.-C. Hitu, P.-A. Ştefan, D. Piciu, MicroRNA in papillary thyroid carcinoma: A systematic review from 2018 to June 2020, Cancers 12 (11) (2020) 3118.
[3] L. Giovanella, G. Treglia, I. Iakovou, J. Mihailovic, F.A. Verburg, M. Luster, EANM practice guideline for PET/CT imaging in medullary thyroid carcinoma, Eur. J. Nucl. Med. Mol. Imaging 47 (1) (2020) 61–77.
[4] S.M. Ferrari, G. Elia, F. Ragusa, I. Ruffilli, C. La Motta, S.R. Paparo, A. Patrizio, R. Vita, S. Benvenga, G. Materazzi, et al., Novel treatments for anaplastic thyroid carcinoma, Gland. Surg. 9 (Suppl 1) (2020) S28.
[5] M. Khammari, A. Chouchane, A. Ouamane, M. Bessaoudi, Y. Himeur, M. Hassaballah, et al., High-order knowledge-based Discriminant features for kinship verification, Pattern Recognit. Lett. 175 (2023) 30–37.
[6] A. Hamza, B. Lekouaghet, Y. Himeur, Hybrid Whale-Mud-Ring optimization for precise color skin cancer image segmentation, in: 2022 6th International Conference on Signal Processing and Information Security, ICSPIS, IEEE, 2023, pp. 1–6.
[7] F.N. Tessler, W.D. Middleton, E.G. Grant, Thyroid imaging reporting and data system (TI-RADS): a user's guide, Radiology 287 (1) (2018) 29–36.
[8] H. Zhou, Y. Jin, L. Dai, M. Zhang, Y. Qiu, J. Tian, J. Zheng, et al., Differential diagnosis of benign and malignant thyroid nodules using deep learning radiomics of thyroid ultrasound images, Eur. J. Radiol. 127 (2020) 108992.
[9] R. Nayak, N. Nawar, J. Webb, M. Fatemi, A. Alizad, Impact of imaging cross-section on visualization of thyroid microvessels using ultrasound: Pilot study, Sci. Rep. 10 (1) (2020) 1–9.
[10] V. Kumar, J. Webb, A. Gregory, D.D. Meixner, J.M. Knudsen, M. Callstrom, M. Fatemi, A. Alizad, Automated segmentation of thyroid nodule, gland, and cystic components from ultrasound images using deep learning, IEEE Access 8 (2020) 63482–63496.
[11] S.Y. Hahn, J.H. Shin, Y.L. Oh, K.W. Park, Y. Lim, Comparison between fine needle aspiration and core needle biopsy for the diagnosis of thyroid nodules: Effective indications according to US findings, Sci. Rep. 10 (1) (2020) 1–7.
[12] Y. Himeur, S. Al-Maadeed, I. Varlamis, N. Al-Maadeed, K. Abualsaud, A. Mohamed, Face mask detection in smart cities using deep and transfer learning: lessons learned from the COVID-19 pandemic, Systems 11 (2) (2023) 107.
[13] A. Chouchane, A. Ouamane, Y. Himeur, W. Mansoor, S. Atalla, A. Benzaibak, C. Boudellal, Improving CNN-based Person Re-identification using score normalization, in: 2023 IEEE International Conference on Image Processing, ICIP, IEEE, 2023, pp. 2890–2894.
[14] S.D. Khan, H. Ullah, A survey of advances in vision-based vehicle re-identification, Comput. Vis. Image Underst. 182 (2019) 50–63.
[15] S.D. Khan, S. Basalamah, A. Lbath, A novel deep learning framework for retinal disease detection leveraging contextual and local features cues from retinal images, Med. Biol. Eng. Comput. (2025) 1–18.
[16] S.D. Khan, S. Basalamah, A. Lbath, Multi-module attention-guided deep learning framework for precise gastrointestinal disease identification in endoscopic imagery, Biomed. Signal Process. Control. 95 (2024) 106396.
[17] M.A. Tahiri, B. Boudaaoua, H. Karmouni, H. Tahiri, H. Oufettoul, H. Amakdouf, H. Qjidaa, M. Abouhawwash, S. Askar, M. Sayyouri, Octonion-based transform moments for innovative stereo image classification with deep learning, Complex Intell. Syst. 10 (3) (2024) 3493–3511.
[18] S.S. Sohail, F. Farhat, Y. Himeur, M. Nadeem, D.Ø. Madsen, Y. Singh, S. Atalla, W. Mansoor, Decoding ChatGPT: A taxonomy of existing research, current challenges, and possible future directions, J. King Saud Univ.-Comput. Inf. Sci. (2023) 101675.
[19] Y. Himeur, M. Elnour, F. Fadli, N. Meskin, I. Petri, Y. Rezgui, F. Bensaali, A. Amira, AI-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives, Artif. Intell. Rev. 56 (6) (2023) 4929–5021.
[20] B. Zhang, J. Tian, S. Pei, Y. Chen, X. He, Y. Dong, L. Zhang, X. Mo, W. Huang, S. Cong, et al., Machine learning–assisted system for thyroid nodule diagnosis, Thyroid 29 (6) (2019) 858–867.
[21] C.Q. Yang, L. Gardiner, H. Wang, M.T. Hueman, D. Chen, Creating prognostic systems for well-differentiated thyroid cancer using machine learning, Front. Endocrinol. 10 (2019) 288.
[22] Y. Liu, J. Liang, S. Peng, W. Wang, H. Xiao, A deep-learning model to assist thyroid nodule diagnosis and management–Authors' reply, Lancet Digit. Health 3 (7) (2021) e411–e412.
[23] A. Iesato, C. Nucera, Role of regulatory non-coding RNAs in aggressive thyroid cancer: Prospective applications of neural network analysis, Molecules 26 (10) (2021) 3022.
[24] Y. Sharifi, M.A. Bakhshali, T. Dehghani, M. DanaiAshgzari, M. Sargolzaei, S. Eslami, Deep learning on ultrasound images of thyroid nodules, Biocybern. Biomed. Eng. 41 (2) (2021) 636–655.
[25] E.J. Ha, J.H. Baek, Applications of machine learning and deep learning to thyroid imaging: where do we stand? Ultrasonography 40 (1) (2021) 23.
[26] Y.-J. Lin, T.-K. Chao, M.-A. Khalil, Y.-C. Lee, D.-Z. Hong, J.-J. Wu, C.-W. Wang, Deep learning fast screening approach on cytological whole slides for thyroid cancer diagnosis, Cancers 13 (15) (2021) 3891.
[27] X.L. Wu, M. Li, X.-w. Cui, G. Xu, Deep multimodal learning for lymph node metastasis prediction of primary thyroid cancer, Phys. Med. Biol. (2022).
[28] S. Pavithra, G. Yamuna, R. Arunkumar, Deep learning method for classifying thyroid nodules using ultrasound images, in: 2022 International Conference on Smart Technologies and Systems for Next Generation Computing, ICSTSN, IEEE, 2022, pp. 1–6.
[29] R. Paul, A. Juliano, W. Faquin, A.W. Chan, An artificial intelligence ultrasound platform for screening and staging of thyroid cancer, Int. J. Radiat. Oncol. Biol. Phys. 112 (5) (2022) e8.
[30] M. Ilyas, H. Malik, M. Adnan, U. Bashir, W.A. Bukhari, M.I.A. Khan, A. Ahmad, Deep learning based classification of thyroid cancer using different medical imaging modalities: A systematic review, 2022.
[31] H. Kheddar, M. Hemis, Y. Himeur, D. Megias, A. Amira, Deep learning for steganalysis of diverse data types: A review of methods, taxonomy, challenges and future directions, Neurocomputing 581 (2024) 127528.
[32] M. Lasseck, Audio-based Bird Species Identification with Deep Convolutional Neural Networks, CLEF (Working Notes), 2018.
[33] W. Zhou, Z. Wang, W. Xie, Weighted signal-to-noise ratio robust design for a new double sampling npx chart, Comput. Ind. Eng. 139 (2020) 106124.
[34] X. Ma, B. Sun, W. Liu, D. Sui, J. Chen, Z. Tian, AMSeg: A novel adversarial architecture based Multi-scale fusion framework for Thyroid Nodule Segmentation, IEEE Access (2023).
[35] C. Liu, Y. Huang, J.A. Ozolek, M.G. Hanna, R. Singh, G.K. Rohde, SetSVM: an approach to set classification in nuclei-based cancer detection, IEEE J. Biomed. Health Inform. 23 (1) (2018) 351–361.
[36] S. Zhang, H. Du, Z. Jin, Y. Zhu, Y. Zhang, F. Xie, M. Zhang, Z. Jiao, X. Tian, J. Zhang, et al., Integrating clinical knowledge in a Thyroid Nodule Classification model based on, in: 2019 IEEE International Ultrasonics Symposium, IUS, IEEE, 2019, pp. 2334–2336.
[37] H. Zhang, C. Zhao, L. Guo, X. Li, Y. Luo, J. Lu, H. Xu, Diagnosis of Thyroid nodules in ultrasound images using two combined classification modules, in: 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), IEEE, 2019, pp. 1–5.
[38] D. Chen, J. Zhang, W. Li, Thyroid nodule classification using two levels Attention-Based bi-directional LSTM with ultrasound reports, in: 2018 9th International Conference on Information Technology in Medicine and Education, ITME, IEEE, 2018, pp. 309–312.
[39] V. Sajeev, A.H. Vyshnavi, P.K. Namboori, Thyroid cancer prediction using gene expression profile, pharmacogenomic variants and quantum image processing in deep learning Platform-A theranostic approach, in: 2020 International Conference for Emerging Technology, INCET, IEEE, 2020, pp. 1–5.
[40] B. Shankarlal, P. Sathya, Performance analysis of thyroid tumor detection and segmentation using PCA-Based random classification method, in: Innovations in Electrical and Electronics Engineering, Springer, 2020, pp. 833–841.
[41] A. Soulaymani, H. Aschawa, Epidemiological study of thyroid carcinoma using principal component analysis, J. Clin. Epigenetics 4 (1) (2018) 9.
[42] L. Cui, L. Ge, H. Gan, X. Liu, Y. Zhang, Ovarian cancer identification based on feature weighting for High-throughput mass spectrometry data, J. Syst. Biol. 1 (1) (2018) 1.
[43] M. Al-Batah, B. Zaqaibeh, S.A. Alomari, M.S. Alzboon, Gene microarray cancer classification using correlation based feature selection algorithm and rules classifiers, Int. J. Online Biomed. Eng. 15 (8) (2019).
[44] I. Jain, V.K. Jain, R. Jain, Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification, Appl. Soft Comput. 62 (2018) 203–215.
[45] Z. Rustam, N. Maghfirah, Correlated based SVM-RFE as feature selection for cancer classification using microarray databases, in: AIP Conference Proceedings, Vol. 2023, AIP Publishing LLC, 2018, 020235.
[46] S. Bhalla, H. Kaur, R. Kaur, S. Sharma, G.P. Raghava, Expression based biomarkers and models to classify early and late-stage samples of Papillary Thyroid Carcinoma, PLoS One 15 (4) (2020) e0231629.
[47] D. O'Dea, M. Bongiovanni, G.P. Sykiotis, P.G. Ziros, A.D. Meade, F.M. Lyng, A. Malkin, Raman spectroscopy for the preoperative diagnosis of thyroid cancer and its subtypes: An in vitro proof-of-concept study, Cytopathology 30 (1) (2019) 51–60.
[48] S.O. Haji, R.Z. Yousif, A novel Run-length based wavelet features for Screening Thyroid Nodule Malignancy, Braz. Arch. Biol. Technol. 62 (2019).

[49] B. Yu, Z. Wang, R. Zhu, X. Feng, M. Qi, J. Li, R. Zhao, L. Huang, R. Xin, F. Li, et al., The transverse ultrasonogram of thyroid papillary carcinoma has a better prediction accuracy than the longitudinal one, IEEE Access 7 (2019) 100763–100770.

[50] D.T. Nguyen, T.D. Pham, G. Batchuluun, H.S. Yoon, K.R. Park, Artificial intelligence-based thyroid nodule classification using information from spatial and frequency domains, J. Clin. Med. 8 (11) (2019) 1976.

[51] P. Poudel, A. Illanes, C. Arens, C. Hansen, M. Friebe, Active contours extension and similarity indicators for improved 3D segmentation of thyroid ultrasound images, in: Medical Imaging 2017: Imaging Informatics for Healthcare, Research, and Applications, Vol. 10138, International Society for Optics and Photonics, 2017, 1013803.

[52] P. Poudel, C. Hansen, J. Sprung, M. Friebe, 3D segmentation of thyroid ultrasound images using active contours, Curr. Dir. Biomed. Eng. 2 (1) (2016) 467–470.

[53] H.A. Nugroho, A. Nugroho, L. Choridah, Thyroid nodule segmentation using active contour bilateral filtering on ultrasound images, in: 2015 International Conference on Quality in Research (QiR), IEEE, 2015, pp. 43–46.

[54] J. Xie, L. Guo, C. Zhao, X. Li, Y. Luo, L. Jianwei, A Hybrid Deep Learning and handcrafted features based approach for thyroid nodule classification in ultrasound images, in: Journal of Physics: Conference Series, Vol. 1693, IOP Publishing, 2020, 012160.

[55] X. Mei, X. Dong, T. Deyer, J. Zeng, T. Trafalis, Y. Fang, Thyroid nodule benignty prediction by deep feature extraction, in: 2017 IEEE 17th International Conference on Bioinformatics and Bioengineering, BIBE, IEEE, 2017, pp. 241–245.

[56] M. Dinčić, J. Todorović, J.N. Ostojić, S. Kovačević, D. Dunđerović, S. Lopičić, S. Spasić, S. Radojević-Škodrić, D. Stanisavljević, A.Ž. Ilić, The fractal and GLCM textural parameters of chromatin may be potential biomarkers of papillary thyroid carcinoma in Hashimoto's thyroiditis specimens, Microsc. Microanal. 26 (4) (2020) 717–730.

[57] I. Kalaimani, Analysis for the prediction of Thyroid Disease by using ICA and Optimal Kernel SVM Approach, Int. J. Emerg. Technol. Innov. Eng. 5 (3) (2019).

[58] W. Ahmad, L. Huang, A. Ahmad, F. Shah, A. Iqbal, A. Saeed, Thyroid diseases forecasting using a hybrid decision support system based on ANFIS, k-NN and information gain method, J. Appl. Environ. Biol. Sci. 7 (10) (2017) 78–85.

[59] H.A. Nugroho, A. Nugroho, E.L. Frannita, I. Ardiyanto, et al., Classification of thyroid ultrasound images based on shape features analysis, in: 2017 10th Biomedical Engineering International Conference (BMEiCON), IEEE, 2017, pp. 1–5.

[60] M. Mourad, S. Moubayed, A. Dezube, Y. Mourad, K. Park, A. Torreblanca-Zanca, J.S. Torrecilla, J.C. Cancilla, J. Wang, Machine Learning and feature Selection Applied to SeeR Data to Reliably Assess thyroid cancer prognosis, Sci. Rep. 10 (1) (2020) 1–11.

[61] H. Song, C. Dong, X. Zhang, W. Wu, C. Chen, B. Ma, F. Chen, C. Chen, X. Lv, Rapid identification of papillary thyroid carcinoma and papillary microcarcinoma based on serum Raman spectroscopy combined with machine learning models, Photodiagnosis Photodyn. Ther. 37 (2022) 102647.

[62] U.R. Acharya, O. Faust, S.V. Sree, F. Molinari, J.S. Suri, ThyroScreen system: high resolution ultrasound thyroid image characterization into benign and malignant classes using novel combination of texture and discrete wavelet transform, Comput. Methods Programs Biomed. 107 (2) (2012) 233–241.

[63] H.A. Nugroho, E.L. Frannita, I. Ardiyanto, L. Choridah, et al., Computer aided diagnosis for thyroid cancer system based on internal and external characteristics, J. King Saud Univ.-Comput. Inf. Sci. 33 (3) (2021) 329–339.

[64] C. Sun, Y. Zhang, Q. Chang, T. Liu, S. Zhang, X. Wang, Q. Guo, J. Yao, W. Sun, L. Niu, Evaluation of a deep learning-based computer-aided diagnosis system for distinguishing benign from malignant thyroid nodules in ultrasound images, Med. Phys. 47 (9) (2020) 3952–3960.

[65] C. Liu, S. Chen, Y. Yang, D. Shao, W. Peng, Y. Wang, Y. Chen, Y. Wang, The value of the computer-aided diagnosis system for thyroid lesions based on computed tomography images, Quant. Imaging Med. Surg. 9 (4) (2019) 642.

[66] N. Liu, A. Fenster, D. Tessier, S. Gou, J. Chong, Self-supervised learning enhanced ultrasound video thyroid nodule tracking, in: Medical Imaging 2023: Image Processing, Vol. 12464, SPIE, 2023, pp. 683–687.

[67] D.C. Yadav, S. Pal, Prediction of thyroid disease using decision tree ensemble method, Hum.-Intell. Syst. Integr. 2 (1) (2020) 89–95.

[68] R.-N. Zhao, B. Zhang, X. Yang, Y.-X. Jiang, X.-J. Lai, X.-Y. Zhang, Logistic regression analysis of contrast-enhanced ultrasound and conventional ultrasound characteristics of sub-centimeter thyroid nodules, Ultrasound Med. Biol. 41 (12) (2015) 3102–3108.

[69] J. Yazdani-Charati, O. Akha, F. Khosravi, Factors affecting thyroid cancer in patients with thyroid nodules using logistic regression in interval censored data, Int. J. Cancer Manag. 11 (3) (2018).

[70] R. Pavithra, L. Parthiban, Optimal deep learning with kernel extreme learning machine based thyroid disease diagnosis and classification model, J. Comput. Theor. Nanosci. 18 (3) (2021) 639–649.

[71] B.N. Rao, D.L.S. Reddy, G. Bhaskar, Thyroid diagnosis using multilayer perceptron, in: International Conference on E-Business and Telecommunications, Springer, 2019, pp. 452–459.

[72] M. Hosseinzadeh, O.H. Ahmed, M.Y. Ghafour, F. Safara, S. Ali, B. Vo, H.-S. Chiang, et al., A multiple multilayer perceptron neural network with an adaptive learning algorithm for thyroid disease diagnosis in the internet of medical things, J. Supercomput. (2020) 1–22.

[73] V. Chandran, M. Sumithra, A. Karthick, T. George, M. Deivakani, B. Elakkiya, U. Subramaniam, S. Manoharan, Diagnosis of cervical cancer based on ensemble deep learning network using colposcopy images, BioMed Res. Int. 2021 (2021).

[74] D. Chen, J. Hu, M. Zhu, N. Tang, Y. Yang, Y. Feng, Diagnosis of thyroid nodules for ultrasonographic characteristics indicative of malignancy using random forest, BioData Min. 13 (1) (2020) 1–21.

[75] Y. Himeur, K. Ghanem, A. Alsalemi, F. Bensaali, A. Amira, Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives, Appl. Energy 287 (2021) 116601.

[76] P. Mehta, M. Bukov, C.-H. Wang, A.G. Day, C. Richardson, C.K. Fisher, D.J. Schwab, A high-bias, low-variance introduction to machine learning for physicists, Phys. Rep. 810 (2019) 1–124.

[77] Q. Pan, Y. Zhang, M. Zuo, L. Xiang, D. Chen, Improved ensemble classification method of thyroid disease based on random forest, in: 2016 8th International Conference on Information Technology in Medicine and Education, ITME, IEEE, 2016, pp. 567–571.

[78] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.

[79] J. Guo, L. Yang, R. Bie, J. Yu, Y. Gao, Y. Shen, A. Kos, An XGBoost-based physical fitness evaluation model using advanced feature selection and Bayesian hyper-parameter optimization for wearable running monitoring, Comput. Netw. 151 (2019) 166–180.

[80] Y. Chen, D. Li, X. Zhang, J. Jin, Y. Shen, Computer aided diagnosis of thyroid nodules based on the devised small-datasets multi-view ensemble learning, Med. Image Anal. 67 (2020) 101819.

[81] M.S. Nobile, G. Capitoli, V. Sowirono, F. Clerici, I. Piga, K. van Abeelen, F. Magni, F. Pagni, S. Galimberti, P. Cazzaniga, et al., Unsupervised neural networks as a support tool for pathology diagnosis in MALDI-MSI experiments: A case study on thyroid biopsies, Expert Syst. Appl. 215 (2023) 119296.

[82] X. Yu, G. Yu, J. Wang, Clustering cancer gene expression data by projective clustering ensemble, PLoS One 12 (2) (2017) e0171429.

[83] K. Chandel, V. Kunwar, A.S. Sabitha, A. Bansal, T. Choudhury, Analysing thyroid disease using density-based clustering technique, Int. J. Bus. Intell. Data Min. 17 (3) (2020) 273–297.

[84] D.K. Katikireddy Srinivas, Performa analysis of clustering of thyroid drug data using fuzzy and m-clust, J. Crit. Rev. 7 (11) (2020) 2128–2141.

[85] B. Venkataramana, L. Padmasree, M.S. Rao, D. Latha, G. Ganesan, Comparative Study on performance of Fuzzy clustering algorithms on Liver and Thyroid Data, J. Fuzzy Set Valued Anal. 2018 (1) (2018) 1–9.

[86] K.K. Mahurkar, D. Gaikwad, Normalization using improvised K-means applied in diagnosing thyroid disease with ANN, in: 2017 International Conference on Trends in Electronics and Informatics, ICEI, IEEE, 2017, pp. 579–583.

[87] Y. Yang, Y. Song, B. Cao, An information Entropy-based method to detect microRNA regulatory module, IPSJ Trans. Bioinform. 12 (2019) 1–8.

[88] H. Kheddar, Y. Himeur, A.I. Awad, Deep transfer learning for intrusion detection in industrial control networks: A comprehensive review, J. Netw. Comput. Appl. 220 (2023) 103760.

[89] S.P. Canton, E. Dadashzadeh, L. Yip, R. Forsythe, R. Handzel, Automatic detection of thyroid and adrenal incidentals using radiology reports and deep learning, J. Surg. Res. 266 (2021) 192–200.

[90] S. Peng, Y. Liu, W. Lv, L. Liu, Q. Zhou, H. Yang, J. Ren, G. Liu, X. Wang, X. Zhang, et al., Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study, Lancet Digit. Health 3 (4) (2021) e250–e259.

[91] M.F. Ferreira, R. Camacho, L.F. Teixeira, Autoencoders as weight initialization of deep classification networks applied to papillary thyroid carcinoma, in: 2018 IEEE International Conference on Bioinformatics and Biomedicine, BIBM, IEEE, 2018, pp. 629–632.

[92] V. Teixeira, R. Camacho, P.G. Ferreira, Learning influential genes on cancer gene expression data with stacked denoising autoencoders, in: 2017 IEEE International Conference on Bioinformatics and Biomedicine, BIBM, IEEE, 2017, pp. 1201–1205.

[93] T. Liu, Q. Guo, C. Lian, X. Ren, S. Liang, J. Yu, L. Niu, W. Sun, D. Shen, Automated detection and classification of thyroid nodules in ultrasound images using clinical-knowledge-guided convolutional neural networks, Med. Image Anal. 58 (2019) 101555.

[94] E.J. Ha, J.H. Baek, D.G. Na, Deep convolutional neural network models for the diagnosis of thyroid cancer, Lancet Oncol. 20 (3) (2019) e130.

[95] Y. Himeur, S. Al-Maadeed, H. Kheddar, N. Al-Maadeed, K. Abualsaud, A. Mohamed, T. Khattab, Video surveillance using deep transfer learning and deep domain adaptation: Towards better generalization, Eng. Appl. Artif. Intell. 119 (2023) 105698.

[96] X. Li, S. Zhang, Q. Zhang, X. Wei, Y. Pan, J. Zhao, X. Xin, C. Qin, X. Wang, J. Li, et al., Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study, Lancet Oncol. 20 (2) (2019) 193–201.

[97] S. Xie, J. Yu, T. Liu, Q. Chang, L. Niu, W. Sun, Thyroid nodule detection in ultrasound images with convolutional neural networks, in: 2019 14th IEEE Conference on Industrial Electronics and Applications, ICIEA, IEEE, 2019, pp. 1442–1446.
[98] J. Koh, E. Lee, K. Han, E.-K. Kim, E.J. Son, Y.-M. Sohn, M. Seo, M.-r. Kwon, J.H. Yoon, J.H. Lee, et al., Diagnosis of thyroid nodules on ultrasonography by a deep convolutional neural network, Sci. Rep. 10 (1) (2020) 1–9.
[99] X. Liang, J. Yu, J. Liao, Z. Chen, Convolutional neural network for breast and thyroid nodules diagnosis in ultrasound imaging, BioMed Res. Int. 2020 (2020).
[100] D. Chen, C. Shi, M. Wang, Q. Pan, Thyroid nodule classification using hierarchical recurrent neural network with multiple ultrasound reports, in: International Conference on Neural Information Processing, Springer, 2017, pp. 765–773.
[101] T.K. Yoo, J.Y. Choi, H.K. Kim, A generative adversarial network approach to predicting postoperative appearance after orbital decompression surgery for thyroid eye disease, Comput. Biol. Med. 118 (2020) 103628.
[102] Q. Guan, Y. Wang, J. Du, Y. Qin, H. Lu, J. Xiang, F. Wang, Deep learning based classification of ultrasound images for thyroid nodules: a large scale of pilot study, Ann. Transl. Med. 7 (7) (2019).
[103] T. Qiao, S. Liu, Z. Cui, X. Yu, H. Cai, H. Zhang, M. Sun, Z. Lv, D. Li, Deep learning for intelligent diagnosis in thyroid scintigraphy, J. Int. Med. Res. 49 (1) (2021) 0300060520982842.
[104] Q. ZHANG, J. HU, S. ZHOU, The detection of hyperthyroidism by the modified LeNet-5 network, Indian J. Pharm. Sci. (2020) 108–114.
[105] H. Tekchandani, S. Verma, N.D. Londhe, R.R. Jain, A. Tiwari, Severity assessment of cervical lymph nodes using modified VGG-Net, and squeeze and excitation concept, in: 2021 IEEE 11th Annual Computing and Communication Workshop and Conference, CCWC, IEEE, 2021, pp. 0709–0714.
[106] Y. Zhao, L. Zhao, T. Mao, L. Zhong, Assessment of risk based on variant pathways and establishment of an artificial neural network model of thyroid cancer, BMC Med. Genet. 20 (1) (2019) 1–10.
[107] J.M. Dolezal, A. Trzcinska, C.-Y. Liao, S. Kochanny, E. Blair, N. Agrawal, X.M. Keutgen, P. Angelos, N.A. Cipriani, A.T. Pearson, Deep learning prediction of BRAF-RAS gene expression signature identifies noninvasive follicular thyroid neoplasms with papillary-like nuclear features, Mod. Pathol. (2020) 1–13.
[108] C. Ma, J. Guan, W. Zhao, C. Wang, An efficient diagnosis system for Thyroid disease based on enhanced Kernelized Extreme Learning Machine Approach, in: International Conference on Cognitive Computing, Springer, 2018, pp. 86–101.
[109] D.C. Yadav, S. Pal, Decision tree ensemble techniques to predict thyroid disease, Int. J. Recent. Technol. Eng. 8 (3) (2019) 8242–8246.
[110] D.T. Nguyen, J.K. Kang, T.D. Pham, G. Batchuluun, K.R. Park, Ultrasound image-based diagnosis of malignant thyroid nodule using artificial intelligence, Sensors 20 (7) (2020) 1822.
[111] L. Ma, C. Ma, Y. Liu, X. Wang, Thyroid diagnosis from SPECT images using convolutional neural network with optimization, Comput. Intell. Neurosci. 2019 (2019).
[112] K. Chandel, V. Kunwar, S. Sabitha, T. Choudhury, S. Mukherjee, A comparative study on thyroid disease detection using K-nearest neighbor and Naive Bayes classification techniques, CSI Trans. ICT 4 (2–4) (2016) 313–319.
[113] J. Xia, H. Chen, Q. Li, M. Zhou, L. Chen, Z. Cai, Y. Fang, H. Zhou, Ultrasound-based differentiation of malignant and benign thyroid Nodules: An extreme learning machine approach, Comput. Methods Programs Biomed. 147 (2017) 37–49.
[114] J. Thomas, T. Haertling, AIBx, artificial intelligence model to risk stratify thyroid nodules, Thyroid 30 (6) (2020) 878–884.
[115] P. Sanyal, T. Mukherjee, S. Barui, A. Das, P. Gangopadhyay, Artificial intelligence in cytopathology: a neural network to identify papillary carcinoma on thyroid fine-needle aspiration cytology smears, J. Pathol. Inform. 9 (2018).
[116] J. Yoon, E. Lee, J.S. Koo, J.H. Yoon, K.-H. Nam, J. Lee, Y.S. Jo, H.J. Moon, V.Y. Park, J.Y. Kwak, Artificial intelligence to predict the BRAFV600E mutation in patients with thyroid cancer, PLoS One 15 (11) (2020) e0242806.
[117] T. Liu, S. Xie, J. Yu, L. Niu, W. Sun, Classification of thyroid nodules in ultrasound images using deep model based transfer learning and hybrid features, in: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2017, pp. 919–923.
[118] F. Abdolali, J. Kapur, J.L. Jaremko, M. Noga, A.R. Hareendranathan, K. Punithakumar, Automated thyroid nodule detection from ultrasound imaging using deep convolutional neural networks, Comput. Biol. Med. 122 (2020) 103871.
[119] X. Li, S. Wang, X. Wei, J. Zhu, R. Yu, M. Zhao, M. Yu, Z. Liu, S. Liu, Fully convolutional networks for ultrasound image segmentation of thyroid nodules, in: 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), IEEE, 2018, pp. 886–890.
[120] J. Song, Y.J. Chai, H. Masuoka, S.-W. Park, S.-j. Kim, J.Y. Choi, H.-J. Kong, K.E. Lee, J. Lee, N. Kwak, et al., Ultrasound image analysis using deep learning algorithm for the diagnosis of thyroid nodules, Medicine 98 (15) (2019).
[121] M. Barczyński, M. Stopa-Barczyńska, B. Wojtczak, A. Czarniecka, A. Konturek, Clinical validation of S-DetectTM mode in semi-automated ultrasound classification of thyroid lesions in surgical office, Gland. Surg. 9 (Suppl 2) (2020) S77.
[122] Y.J. Choi, J.H. Baek, H.S. Park, W.H. Shim, T.Y. Kim, Y.K. Shong, J.H. Lee, A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of thyroid nodules on ultrasound: initial clinical assessment, Thyroid 27 (4) (2017) 546–552.
[123] C. Fragopoulos, A. Pouliakis, C. Meristoudis, E. Mastorakis, N. Margari, N. Chroniaris, N. Koufopoulos, A.G. Delides, N. Machairas, V. Ntomi, et al., Radial basis function artificial neural network for the investigation of thyroid cytological lesions, J. Thyroid. Res. 2020 (2020).
[124] R. Savala, P. Dey, N. Gupta, Artificial neural network model to distinguish follicular adenoma from follicular carcinoma on fine needle aspiration of thyroid, Diagn. Cytopathol. 46 (3) (2018) 244–249.
[125] B. Wildman-Tobriner, M. Buda, J.K. Hoang, W.D. Middleton, D. Thayer, R.G. Short, F.N. Tessler, M.A. Mazurowski, Using artificial intelligence to revise ACR TI-RADS risk stratification of thyroid nodules: diagnostic accuracy and utility, Radiology 292 (1) (2019) 112–119.
[126] L. Wang, S. Yang, S. Yang, C. Zhao, G. Tian, Y. Gao, Y. Chen, Y. Lu, Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the YOLOv2 neural network, World J. Surg. Oncol. 17 (1) (2019) 1–9.
[127] J.A. Ozolek, A.B. Tosun, W. Wang, C. Chen, S. Kolouri, S. Basu, H. Huang, G.K. Rohde, Accurate diagnosis of thyroid follicular lesions from nuclear morphology using supervised learning, Med. Image Anal. 18 (5) (2014) 772–780.
[128] Y. Zhu, Q. Sang, S. Jia, Y. Wang, T. Deyer, Deep neural networks could differentiate Bethesda class III versus class IV/V/VI, Ann. Transl. Med. 7 (11) (2019).
[129] K. Daniels, S. Gummadi, Z. Zhu, S. Wang, J. Patel, B. Swendseid, A. Lyshchik, J. Curry, E. Cottrill, J. Eisenbrey, Machine learning by ultrasonography for genetic risk stratification of thyroid nodules, JAMA Otolaryngol.–Head Neck Surg. 146 (1) (2020) 36–41.
[130] K. Dharmarajan, K. Balasree, A. Arunachalam, K. Abirmai, Thyroid disease classification using decision tree and SVM, Exec. Ed. 11 (03) (2020) 3234.
[131] M.H. Tran, O. Gomez, B. Fei, A video transformer network for thyroid cancer detection on hyperspectral histologic images, in: Medical Imaging 2023: Digital and Computational Pathology, Vol. 12471, SPIE, 2023, pp. 32–41.
[132] J. Gu, J. Zhu, Q. Qiu, Y. Wang, T. Bai, Y. Yin, Prediction of immunohistochemistry of suspected thyroid nodules by use of machine learning–based radiomics, Am. J. Roentgenol. 213 (6) (2019) 1348–1357.
[133] B. Colakoglu, D. Alis, M. Yergin, Diagnostic value of machine learning-based quantitative texture analysis in differentiating benign and malignant thyroid nodules, J. Oncol. 2019 (2019).
[134] V.Y. Park, E. Lee, H.S. Lee, H.J. Kim, J. Yoon, J. Son, K. Song, H.J. Moon, J.H. Yoon, G.R. Kim, et al., Combining radiomics with ultrasound-based risk stratification systems for thyroid nodules: an approach for improving performance, Eur. Radiol. 31 (4) (2021) 2405–2413.
[135] B. Mughal, M. Sharif, N. Muhammad, T. Saba, A novel classification scheme to decline the mortality rate among women due to breast tumor, Microsc. Res. Tech. 81 (2) (2018) 171–180.
[136] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).
[137] Y. Lu, Y. Yang, W. Chen, Application of deep learning in the prediction of benign and malignant thyroid nodules on ultrasound images, IEEE Access 8 (2020) 221468–221480.
[138] S.W. Kwon, I.J. Choi, J.Y. Kang, W.I. Jang, G.-H. Lee, M.-C. Lee, Ultrasonographic thyroid nodule classification using a deep convolutional neural network with surgical pathology, J. Digit. Imaging 33 (2020) 1202–1208.
[139] W.-K. Chan, J.-H. Sun, M.-J. Liou, Y.-R. Li, W.-Y. Chou, F.-H. Liu, S.-T. Chen, S.-J. Peng, Using deep convolutional neural networks for enhanced ultrasonographic image diagnosis of differentiated thyroid cancer, Biomedicines 9 (12) (2021) 1771.
[140] X. Wang, E.A. Agyekum, Y. Ren, J. Zhang, Q. Zhang, H. Sun, G. Zhang, F. Xu, X. Bo, W. Lv, et al., A radiomic nomogram for the ultrasound-based evaluation of extrathyroidal extension in papillary thyroid carcinoma, Front. Oncol. 11 (2021) 625646.
[141] H. Zhou, K. Wang, J. Tian, Online transfer learning for differential diagnosis of benign and malignant thyroid nodules with ultrasound images, IEEE Trans. Biomed. Eng. (2020).
[142] P. Tsou, C.-J. Wu, Mapping driver mutations to histopathological subtypes in papillary thyroid carcinoma: applying a deep convolutional neural network, J. Clin. Med. 8 (10) (2019) 1675.
[143] J. Ma, F. Wu, J. Zhu, D. Xu, D. Kong, A pre-trained convolutional neural network based method for thyroid nodule diagnosis, Ultrasonics 73 (2017) 221–230.
[144] J. Ma, F. Wu, T. Jiang, Q. Zhao, D. Kong, Ultrasound image-based thyroid nodule automatic segmentation using convolutional neural networks, Int. J. Comput. Assist. Radiol. Surg. 12 (11) (2017) 1895–1910.
[145] N.T. Duc, Y.-M. Lee, J.H. Park, B. Lee, An ensemble deep learning for automatic prediction of papillary thyroid carcinoma using fine needle aspiration cytology, Expert Syst. Appl. 188 (2022) 115927.
[146] J.H. Lee, E.J. Ha, J.H. Kim, Application of deep learning to the diagnosis of cervical lymph node metastasis from thyroid cancer with CT, Eur. Radiol. 29 (2019) 5452–5457.

[147] F.-s. Ouyang, B.-l. Guo, L.-z. Ouyang, Z.-w. Liu, S.-j. Lin, W. Meng, X.-y. Huang, H.-x. Chen, H. Qiu-Gen, S.-m. Yang, Comparison between linear and nonlinear machine-learning algorithms for the classification of thyroid nodules, Eur. J. Radiol. 113 (2019) 251–257.
[148] C.-K. Zhao, T.-T. Ren, Y.-F. Yin, H. Shi, H.-X. Wang, B.-Y. Zhou, X.-R. Wang, X. Li, Y.-F. Zhang, C. Liu, et al., A comparative analysis of two machine learning-based diagnostic patterns with thyroid imaging reporting and data system for thyroid nodules: diagnostic performance and unnecessary biopsy rate, Thyroid 31 (3) (2021) 470–481.
[149] V.V. Vadhiraj, A. Simpkin, J. O'Connell, N. Singh Ospina, S. Maraka, D.T. O'Keeffe, Ultrasound image classification of thyroid nodules using machine learning techniques, Medicina 57 (6) (2021) 527.
[150] M.L. Gild, M. Chan, J. Gajera, B. Lurie, Z. Gandomkar, R.J. Clifton-Bligh, Risk stratification of indeterminate thyroid nodules using ultrasound and machine learning algorithms, Clin. Endocrinol. 96 (4) (2022) 646–652.
[151] J. Zhu, S. Zhang, R. Yu, Z. Liu, H. Gao, B. Yue, X. Liu, X. Zheng, M. Gao, X. Wei, An efficient deep convolutional neural network model for visual localization and automatic diagnosis of thyroid nodules on ultrasound images, Quant. Imaging Med. Surg. 11 (4) (2021) 1368.
[152] V.Y. Park, K. Han, E. Lee, E.-K. Kim, H.J. Moon, J.H. Yoon, J.Y. Kwak, Association between radiomics signature and disease-free survival in conventional papillary thyroid carcinoma, Sci. Rep. 9 (1) (2019) 4501.
[153] J. Chi, E. Walia, P. Babyn, J. Wang, G. Groot, M. Eramian, Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network, J. Digit. Imaging 30 (4) (2017) 477–486.
[154] Y. Zhu, Z. Fu, J. Fei, An image augmentation method using convolutional network for thyroid nodule classification by transfer learning, in: 2017 3rd IEEE International Conference on Computer and Communications, ICCC, IEEE, 2017, pp. 1819–1823.
[155] L. Gao, R. Liu, Y. Jiang, W. Song, Y. Wang, J. Liu, J. Wang, D. Wu, S. Li, A. Hao, et al., Computer-aided system for diagnosing thyroid nodules on ultrasound: A comparison with radiologist-based clinical assessments, Head Neck 40 (4) (2018) 778–783.
[156] D. Zuo, L. Han, K. Chen, C. Li, Z. Hua, J. Lin, Extraction of calcification in ultrasonic images based on convolution neural network, Sheng Wu Yi Xue Gong Cheng Xue Za Zhi=J. Biomed. Eng.=Shengwu Yixue Gongchengxue Zazhi 35 (5) (2018) 679–687.
[157] G. Kim, E. Lee, H. Kim, J. Yoon, V. Park, J. Kwak, Convolutional neural network to stratify the malignancy risk of thyroid nodules: diagnostic performance compared with the American college of radiology thyroid imaging reporting and data system implemented by experienced radiologists, Am. J. Neuroradiol. 42 (8) (2021) 1513–1519.
[158] G.-G. Wu, W.-Z. Lv, R. Yin, J.-W. Xu, Y.-J. Yan, R.-X. Chen, J.-Y. Wang, B. Zhang, X.-W. Cui, C.F. Dietrich, Deep learning based on ACR TI-RADS can improve the differential diagnosis of thyroid nodules, Front. Oncol. 11 (2021) 575166.
[159] Z. Jin, Y. Zhu, S. Zhang, F. Xie, M. Zhang, Y. Zhang, X. Tian, J. Zhang, Y. Luo, J. Cao, Ultrasound computer-aided diagnosis (CAD) based on the thyroid imaging reporting and data system (TI-RADS) to distinguish benign from malignant thyroid nodules and the diagnostic performance of radiologists with different diagnostic experience, Med. Sci. Monit.: Int. Med. J. Exp. Clin. Res. 26 (2020) e918452–e918451.
[160] Y.-J. Kim, Y. Choi, S.-J. Hur, K.-S. Park, H.-J. Kim, M. Seo, M.K. Lee, S.-L. Jung, C.K. Jung, Deep convolutional neural network for classification of thyroid nodules on ultrasound: Comparison of the diagnostic performance with that of radiologists, Eur. J. Radiol. 152 (2022) 110335.
[161] H. Kheddar, Y. Habchi, M.C. Ghanem, M. Hemis, D. Niyato, Recent advances in transformer and large language models for UAV applications, 2025, arXiv preprint arXiv:2508.11834.
[162] H. Kheddar, Transformers and large language models for efficient intrusion detection systems: A comprehensive survey, Inf. Fusion (2025) 103347.
[163] N. Djeffal, H. Kheddar, D. Addou, A.C. Mazari, Y. Himeur, Automatic speech recognition with BERT and CTC transformers: A review, in: 2023 2nd International Conference on Electronics, Energy and Measurement (IC2EM), Vol. 1, IEEE, 2023, pp. 1–8.
[164] H. Kheddar, M. Hemis, Y. Himeur, Automatic speech recognition using advanced deep learning approaches: A survey, Inf. Fusion (2024) 102422.
[165] R. Jiang, X. Yin, P. Yang, L. Cheng, J. Hu, J. Yang, Y. Wang, X. Fu, L. Shang, L. Li, et al., A transformer-based weakly supervised computational pathology method for clinical-grade diagnosis and molecular marker discovery of gliomas, Nat. Mach. Intell. 6 (8) (2024) 876–891.
[166] J. Chi, Z. Li, Z. Sun, X. Yu, H. Wang, Hybrid transformer UNet for thyroid segmentation from ultrasound scans, Comput. Biol. Med. 153 (2023) 106453.
[167] R. Sharma, G.K. Mahanti, G. Panda, A. Rath, S. Dash, S. Mallik, R. Hu, A framework for detecting thyroid cancer from ultrasound and histopathological images using deep learning, Meta-Heuristics, and MCDM algorithms, J. Imaging 9 (9) (2023) 173.
[168] A. Pathak, Z. Yu, D. Paredes, E.P. Monsour, A.O. Rocha, J.P. Brito, N.S. Ospina, Y. Wu, Extracting thyroid nodules characteristics from ultrasound reports using transformer-based natural language processing methods, 2023, arXiv preprint arXiv:2304.00115.
[169] H. Bi, C. Cai, J. Sun, Y. Jiang, G. Lu, H. Shu, X. Ni, BPAT-UNet: Boundary preserving assembled transformer UNet for ultrasound thyroid nodule segmentation, Comput. Methods Programs Biomed. 238 (2023) 107614.
[170] F. Chen, H. Han, P. Wan, H. Liao, C. Liu, D. Zhang, Joint segmentation and differential diagnosis of thyroid nodule in Contrast-Enhanced ultrasound images, IEEE Trans. Biomed. Eng. (2023).
[171] D. Dov, S. Assaad, S. Si, R. Wang, H. Xu, S.Z. Kovalsky, J. Bell, D.E. Range, J. Cohen, R. Henao, et al., Affinitention nets: kernel perspective on attention architectures for set classification with applications to medical text and images, in: Proceedings of the Conference on Health, Inference, and Learning, 2021, pp. 14–24.
[172] F. Jerbi, N. Aboudi, N. Khlifa, Automatic classification of ultrasound thyroids images using vision transformers and generative adversarial networks, Sci. Afr. 20 (2023) e01679.
[173] T. Jiang, W. Xing, M. Yu, D. Ta, A hybrid enhanced attention transformer network for medical ultrasound image segmentation, Biomed. Signal Process. Control. 86 (2023) 105329.
[174] G. Li, R. Chen, J. Zhang, K. Liu, C. Geng, L. Lyu, Fusing enhanced Transformer and large kernel CNN for malignant thyroid nodule segmentation, Biomed. Signal Process. Control. 83 (2023) 104636.
[175] Q. Liu, F. Ding, J. Li, S. Ji, K. Liu, C. Geng, L. Lyu, DCA-Net: Dual-branch contextual-aware network for auxiliary localization and segmentation of parathyroid glands, Biomed. Signal Process. Control. 84 (2023) 104856.
[176] J. Nam, J.-W. Choi, Y.-G. Shin, S. Park, A BERT-based artificial intelligence to analyze Free-Text clinical notes for binary classification in papillary thyroid carcinoma recurrence, in: 2023 IEEE International Conference on Consumer Electronics, ICCE, IEEE, 2023, pp. 1–2.
[177] J. Sun, B. Wu, T. Zhao, L. Gao, K. Xie, T. Lin, J. Sui, X. Li, X. Wu, X. Ni, Classification for thyroid nodule using ViT with contrastive learning in ultrasound images, Comput. Biol. Med. 152 (2023) 106444.
[178] I.E. Tampu, A. Eklund, K. Johansson, O. Gimm, N. Haj-Hosseini, Diseased thyroid tissue classification in OCT images using deep learning: Towards surgical decision support, J. Biophotonics 16 (2) (2023) e202200227.
[179] Z. Tao, H. Dang, Y. Shi, W. Wang, X. Wang, S. Ren, Local and context-attention adaptive LCA-Net for thyroid nodule segmentation in ultrasound images, Sensors 22 (16) (2022) 5984.
[180] Z. Wang, L. Yu, X. Ding, X. Liao, L. Wang, Shared-specific feature learning with bottleneck fusion transformer for Multi-modal whole slide image analysis, IEEE Trans. Med. Imaging (2023).
[181] Z. Wang, L. Yu, X. Ding, X. Liao, L. Wang, Lymph node metastasis prediction from whole slide images with transformer-guided multiinstance learning and knowledge transfer, IEEE Trans. Med. Imaging 41 (10) (2022) 2777–2787.
[182] N. Xiao, Z. Li, S. Chen, L. Zhao, Y. Yang, H. Xie, Y. Liu, Y. Quan, J. Duan, Contrast-enhanced CT image synthesis of thyroid based on transfomer and texture branching, in: 2022 5th International Conference on Artificial Intelligence and Big Data, ICAIBD, IEEE, 2022, pp. 94–100.
[183] P. Yin, B. Yu, C. Jiang, H. Chen, Pyramid Tokens-to-Token vision transformer for thyroid pathology image classification, in: 2022 Eleventh International Conference on Image Processing Theory, Tools and Applications, IPTA, IEEE, 2022, pp. 1–6.
[184] B. Yu, P. Yin, H. Chen, Y. Wang, Y. Zhao, X. Cong, J. Dijkstra, L. Cong, Pyramid multi-loss vision transformer for thyroid cancer classification using cytological smear, Knowl.-Based Syst. (2023) 110721.
[185] R. Raghunathan, A.R. Jacobs, V.R. Sant, L.J. King, G. Rothberger, J. Prescott, J. Allendorf, C.D. Seib, K.N. Patel, I. Suh, Can large language models address unmet patient information needs and reduce provider burnout in the management of thyroid disease? Surgery 177 (2025) 108859.
[186] D.T. Lee, A. Vaid, K.M. Menon, R. Freeman, D.S. Matteson, M.P. Marin, G.N. Nadkarni, Development of a privacy preserving large language model for automated data extraction from thyroid cancer pathology reports, MedRxiv (2023) 2023-2011.
[187] S.-H. Wu, W.-J. Tong, M.-D. Li, H.-T. Hu, X.-Z. Lu, Z.-R. Huang, X.-X. Lin, R.-F. Lu, M.-D. Lu, L.-D. Chen, et al., Collaborative enhancement of consistency and accuracy in US diagnosis of Thyroid Nodules using large language models, Radiology 310 (3) (2024) e232255.
[188] M. Shah, E.J. Kuo, J.H. Kuo, S. Hsu, C. McManus, R. Liou, J.A. Lee, T.S. Sathe, EndoGPT: A Proof-of-concept large language model based assistant for the management of thyroid nodules, MedRxiv (2024) 2024-2005.
[189] J. Yao, Y. Wang, Z. Lei, K. Wang, X. Li, J. Zhou, X. Hao, J. Shen, Z. Wang, R. Ru, et al., AI-Generated content enhanced Computer-Aided diagnosis model for thyroid nodules: A ChatGPT-Style assistant, 2024, arXiv preprint arXiv:2402.02401.
[190] M. Zhang, Q. Cheng, Z. Wei, J. Xu, S. Wu, N. Xu, C. Zhao, L. Yu, W. Feng, BertTCR: a Bert-based deep learning framework for predicting cancer-related immune status based on T cell receptor repertoire, Brief. Bioinform. 25 (5) (2024) bbae420.
[191] Y. Habchi, H. Kheddar, Y. Himeur, A. Belouchrani, E. Serpedin, F. Khelifi, M.E. Chowdhury, Advanced deep learning and large language models: Comprehensive insights for cancer detection, Image Vis. Comput. 157 (2025) 105495.

[192] Z. Wang, Z. Zhang, A. Traverso, A. Dekker, L. Qian, P. Sun, Assessing the role of GPT-4 in thyroid ultrasound diagnosis and treatment recommendations: enhancing interpretability with a chain of thought approach, Quant. Imaging Med. Surg. 14 (2) (2024) 1602.
[193] T. Zhou, Y. Xu, Y. Zhu, C. Xiao, H. Bian, L. Wei, X. Zhang, DrVD-Bench: Do Vision-Language models reason like human doctors in medical image diagnosis?, 2025, arXiv preprint arXiv:2505.24173.
[194] O.C. Thawakar, A.M. Shaker, S.S. Mullappilly, H. Cholakkal, R.M. Anwer, S. Khan, J. Laaksonen, F. Khan, XrayGPT: Chest radiographs summarization using large medical vision-language models, in: Proceedings of the 23rd Workshop on Biomedical Natural Language Processing, 2024, pp. 440–448.
[195] J. Clusmann, S.J. Schulz, D. Ferber, I.C. Wiest, A. Fernandez, M. Eckstein, F. Lange, N.G. Reitsam, F. Kellers, M. Schmitt, et al., Incidental prompt injections on Vision–Language models in Real-Life histopathology, NEJM AI 2 (6) (2025) AIcs2500078.
[196] H. Rasaee, T. Koleilat, H. Rivaz, GroundingDINO-US-SAM: Text-prompted Multi-Organ segmentation in ultrasound with LoRA-Tuned Vision-Language models, 2025, arXiv preprint arXiv:2506.23903.
[197] Y. Himeur, I. Varlamis, H. Kheddar, A. Amira, S. Atalla, Y. Singh, F. Bensaali, W. Mansoor, Federated learning for computer vision, 2023, arXiv preprint arXiv:2308.13558.
[198] A.A. Shah, H.A.M. Malik, A. Muhammad, A. Alourani, Z.A. Butt, Deep learning ensemble 2D CNN approach towards the detection of lung cancer, Sci. Rep. 13 (1) (2023) 2987.
[199] J. Salazar-Vega, E. Ortiz-Prado, P. Solis-Pazmino, L. Gómez-Barreno, K. Simbaña-Rivera, A.R. Henriquez-Trujillo, J.P. Brito, T. Toulkeridis, M. Coral-Almeida, Thyroid cancer in ecuador, a 16 years population-based analysis (2001–2016), BMC Cancer 19 (2019) 1–8.
[200] L.W. Elmore, S.F. Greer, E.C. Daniels, C.C. Saxe, M.H. Melner, G.M. Krawiec, W.G. Cance, W.C. Phelps, Blueprint for cancer research: Critical gaps and opportunities, CA: Cancer J. Clin. 71 (2) (2021) 107–139.
[201] S.H. Park, J. Choi, J.-S. Byeon, Key principles of clinical validation, device approval, and insurance coverage decisions of artificial intelligence, Korean J. Radiol. 22 (3) (2021) 442.
[202] Y.-C. Zhu, A. AlZoubi, S. Jassim, Q. Jiang, Y. Zhang, Y.-B. Wang, X.-D. Ye, D. Hongbo, A generic deep learning framework to classify thyroid and breast lesions in ultrasound images, Ultrasonics 110 (2021) 106300.
[203] J. Yao, Z. Lei, W. Yue, B. Feng, W. Li, D. Ou, N. Feng, Y. Lu, J. Xu, W. Chen, et al., DeepThy-Net: A multimodal deep learning method for predicting cervical lymph node metastasis in papillary thyroid cancer, Adv. Intell. Syst. 4 (10) (2022) 2200100.
[204] A.N. Sayed, Y. Himeur, F. Bensaali, From time-series to 2d images for building occupancy prediction using deep transfer learning, Eng. Appl. Artif. Intell. 119 (2023) 105786.
[205] C.-W. Wang, K.-Y. Lin, Y.-J. Lin, M.-A. Khalil, K.-L. Chu, T.-K. Chao, A soft label deep learning to assist breast cancer target therapy and thyroid cancer diagnosis, Cancers 14 (21) (2022) 5312.
[206] Z. Al-Qurayshi, G.W. Randolph, E. Kandil, Cost-effectiveness of computed tomography nodal scan in patients with papillary thyroid carcinoma, Oral Oncol. 118 (2021) 105326.
[207] H. Yuan, Overcoming computational resource limitations in deep learning for healthcare: Strategies targeting data, model, and computing, Med. Adv. 3 (1) (2025) 42–45.
[208] D. Dov, S. Kovalsky, J. Cohen, D. Range, R. Henao, L. Carin, Thyroid cancer malignancy prediction from whole slide cytopathology images, 2019, arXiv preprint arXiv:1904.00839.
[209] M. Halicek, M. Shahedi, J.V. Little, A.Y. Chen, L.L. Myers, B.D. Sumer, B. Fei, Head and neck cancer detection in digitized whole-slide histology using convolutional neural networks, Sci. Rep. 9 (1) (2019) 1–11.
[210] A. Karsa, S. Punwani, K. Shmueli, An optimized and highly repeatable MRI acquisition and processing pipeline for quantitative susceptibility mapping in the head-and-neck region, Magn. Reson. Med. 84 (6) (2020) 3206–3222.
[211] C. Sardianos, I. Varlamis, C. Chronis, G. Dimitrakopoulos, A. Alsalemi, Y. Himeur, F. Bensaali, A. Amira, The emergence of explainability of intelligent systems: Delivering explainable and personalized recommendations for energy efficiency, Int. J. Intell. Syst. 36 (2) (2021) 656–680.
[212] A. Gueriani, H. Kheddar, A.C. Mazari, M.C. Ghanem, A robust cross-domain IDS using BiGRU-LSTM-attention for medical and industrial IoT security, ICT Express (2025).
[213] T. Masuda, T. Nakaura, Y. Funama, K. Sugino, T. Sato, T. Yoshiura, Y. Baba, K. Awai, Machine learning to identify lymph node metastasis from thyroid cancer in patients undergoing contrast-enhanced CT studies, Radiography 27 (3) (2021) 920–926.
[214] M.C. Ghanem, S. Salloum, Integrating AI-driven deep learning for energy-efficient smart buildings in Internet of Thing-based industry 4.0, Babylon. J. Internet Things 2025 (7) (2025) 121–130.
[215] A.N. Sayed, F. Bensaali, Y. Himeur, M. Houchati, Edge-Based Real-Time occupancy detection system through a Non-Intrusive sensing system, Energies 16 (5) (2023) 2388.
[216] A. Alsalemi, Y. Himeur, F. Bensaali, A. Amira, An innovative edge-based internet of energy solution for promoting energy saving in buildings, Sustain. Cities Soc. 78 (2022) 103571.
[217] A. Sayed, Y. Himeur, A. Alsalemi, F. Bensaali, A. Amira, Intelligent edge-based recommender system for internet of energy applications, IEEE Syst. J. 16 (3) (2021) 5001–5010.
[218] A. Sufian, A. Ghosh, A.S. Sadiq, F. Smarandache, A survey on deep transfer learning to edge computing for mitigating the COVID-19 pandemic, J. Syst. Archit. 108 (2020) 101830.
[219] J.P. Rajan, S.E. Rajan, R.J. Martis, B.K. Panigrahi, Fog computing employed computer aided cancer classification system using deep neural network in internet of things based healthcare system, J. Med. Syst. 44 (2) (2020) 1–10.
[220] A.S. Basnet, M.C. Ghanem, D. Dunsin, H. Kheddar, W. Sowinski-Mydlarz, Advanced persistent threats (apt) attribution using deep reinforcement learning, Digit. Threat.: Res. Pract. (2025).
[221] H. Kheddar, D.W. Dawoud, A.I. Awad, Y. Himeur, M.K. Khan, Reinforcement-Learning-Based intrusion detection in communication networks: A review, IEEE Commun. Surv. Tutor. (2024).
[222] P. Balaprakash, R. Egele, M. Salim, S. Wild, V. Vishwanath, F. Xia, T. Brettin, R. Stevens, Scalable reinforcement-learning-based neural architecture search for cancer deep learning research, in: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 2019, pp. 1–33.
[223] Z. Li, Y. Xia, Deep reinforcement learning for weakly-supervised lymph node segmentation in ct images, IEEE J. Biomed. Health Inform. (2020).
[224] H. Kheddar, Y. Himeur, A. Amira, BreathAI: Transfer Learning-Based thermal imaging for automated breathing pattern recognition, in: 2025 IEEE International Conference on Image Processing, ICIP, IEEE, 2025, pp. 2612–2617.
[225] J. Yang, G. Wang, X. Xiao, M. Bao, G. Tian, Explainable ensemble learning method for OCT detection with transfer learning, PLoS One 19 (3) (2024) e0296175.
[226] Y. Habchi, H. Kheddar, Y. Himeur, M.C. Ghanem, A. Boukabou, S. Atalla, W. Mansoor, H. Al-Ahmad, Deep transfer learning for kidney cancer diagnosis, 2024, arXiv preprint arXiv:2408.04318.
[227] Y. Habchi, H. Kheddar, Y. Himeur, Ultrasound images classification of thyroid cancer using deep transfer learning, in: 2024 International Conference on Telecommunications and Intelligent Systems, ICTIS, IEEE, 2024, pp. 1–6.
[228] H. Kheddar, Y. Himeur, S. Al-Maadeed, A. Amira, F. Bensaali, Deep transfer learning for automatic speech recognition: Towards better generalization, Knowl.-Based Syst. 277 (2023) 110851.
[229] S. Lachenani, H. Kheddar, M. Ould Zmirli, Advancing bearing fault diagnosis using deep transfer learning for wireless sensor networks, Arab. J. Sci. Eng. (2025) 1–21.
[230] V. Narayan, P.K. Mall, A. Alkhayyat, K. Abhishek, S. Kumar, P. Pandey, et al., Enhance-Net: An approach to boost the performance of deep learning model based on Real-Time medical images, J. Sens. 2023 (2023).
[231] H. Lee, Y.J. Chai, H. Joo, K. Lee, J.Y. Hwang, S.-M. Kim, K. Kim, I.-C. Nam, J.Y. Choi, H.W. Yu, et al., Federated learning for thyroid ultrasound image analysis to protect personal information: Validation study in a real health care environment, JMIR Med. Inform. 9 (5) (2021) e25869.
[232] O. Elharrouss, S. Al-Maadeed, N. Subramanian, N. Ottakath, N. Almaadeed, Y. Himeur, Panoptic segmentation: A review, 2021, arXiv preprint arXiv:2111.10250.
[233] X. Yu, B. Lou, D. Zhang, D. Winkel, N. Arrahmane, M. Diallo, T. Meng, H. von Busch, R. Grimm, B. Kiefer, et al., Deep attentive panoptic model for prostate cancer detection using biparametric MRI scans, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 594–604.
[234] D. Ivanova, Artificial intelligence in internet of medical imaging things: The power of thyroid cancer detection, in: 2018 International Conference on Information Technologies (InfoTech), IEEE, 2018, pp. 1–4.
[235] P. Borovska, D. Ivanova, I. Draganov, Internet of medical imaging things and analytics in support of precision medicine for the case study of thyroid cancer early diagnostics, Serdica J. Comput. Bulg. Acad. Sci. Inst. Math. Inform. Accept. Pap. (2018).
[236] O. Rabab, M.A. Tahiri, A. Bencherqui, H. Amakdouf, M.O. Jamil, H. Qjidaa, Efficient localization and reconstruction of 3D objects using the new hybrid squire moment, in: 2022 International Conference on Intelligent Systems and Computer Vision, ISCV, IEEE, 2022, pp. 1–8.
[237] P. Seifert, S.-L. Ullrich, C. Kühnel, F. Gühne, R. Drescher, T. Winkens, M. Freesmeyer, Optimization of thyroid volume determination by stitched 3D-Ultrasound data sets in patients with structural thyroid disease, Biomedicines 11 (2) (2023) 381.
[238] N. Pakkasjärvi, T. Luthra, S. Anand, Artificial intelligence in surgical learning, Surgeries 4 (1) (2023) 86–97.
[239] S. Bodenstedt, M. Wagner, B.P. Müller-Stich, J. Weitz, S. Speidel, Artificial intelligence-assisted surgery: potential and challenges, Visc. Med. 36 (6) (2020) 450–455.

[240] D. Lee, H.W. Yu, H. Kwon, H.-J. Kong, K.E. Lee, H.C. Kim, Evaluation of surgical skills during robotic surgery by deep learning-based multiple surgical instrument tracking in training and actual operations, J. Clin. Med. 9 (6) (2020) 1964.

[241] Y. Himeur, A. Alsalemi, A. Al-Kababji, F. Bensaali, A. Amira, C. Sardianos, G. Dimitrakopoulos, I. Varlamis, A survey of recommender systems for energy efficiency in buildings: Principles, challenges and prospects, Inf. Fusion 72 (2021) 1–21.

[242] Q.M. Areeb, M. Nadeem, S.S. Sohail, R. Imam, F. Doctor, Y. Himeur, A. Hussain, A. Amira, Filter bubbles in recommender systems: Fact or fallacy—A systematic review, Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. (2023) e1512.

[243] S. Atalla, M. Daradkeh, A. Gawanmeh, H. Khalil, W. Mansoor, S. Miniaoui, Y. Himeur, An intelligent recommendation system for automating academic advising based on curriculum analysis and performance modeling, Mathematics 11 (5) (2023) 1098.

[244] Y. Habchi, R. Bouddou, A.F. Aimer, M. Beladgham, A new cell spreading video coding based on bandelet transform and lifting scheme for medical applications, Prz. Elektrotech. 99 (3) (2023).

[245] M. Beladgham, Y. Habchi, A. Taleb-Ahmed, et al., Medical video compression using bandelet based on lifting scheme and SPIHT coding: In search of high visual quality, Inform. Med. Unlocked 17 (2019) 100244.

[246] Y. Habchi, A.F. Aimer, J. Baili, M. Inc, Y. Menni, G. Lorenzini, Improving medical video coding using multi scale quincunx lattice: From low bitrate to high quality, 2022.

[247] Y. Habchi, A.F. Aimer, M. Beladgham, R. Bouddou, Ultra low bitrate retinal image compression using integer lifting scheme and subband encoder, Indonesian J. Electr. Eng. Comput. Sci. (IJEECS) 24 (1) (2021) 295–307.

[248] Y. Habchi, M. Beladgham, A. Taleb-Ahmed, RGB medical video compression using geometric wavelet and SPIHT coding, Int. J. Electr. Comput. Eng. 6 (4) (2016) 1627–1636.

[249] I. Boucherit, H. Kheddar, Reinforced residual Encoder–Decoder network for image denoising via deeper encoding and balanced skip connections, Big Data Cogn. Comput. 9 (4) (2025) 82.

[250] H. Kheddar, Y. Himeur, A. Amira, R. Soualah, Image and point-cloud classification for jet analysis in high-energy physics: A survey, Front. Phys. 20 (3) 035301.

[251] A. Gangwal, A. Ansari, I. Ahmad, A.K. Azad, W.M.A.W. Sulaiman, Current strategies to address data scarcity in artificial intelligence-based drug discovery: A comprehensive review, Comput. Biol. Med. 179 (2024) 108734.

[252] C. Luo, Y. Yu, J. Zhu, L. Chen, D. Li, X. Peng, Z. Liu, Q. Li, Q. Cao, K. Huang, et al., Deubiquitinase PSMD7 facilitates pancreatic cancer progression through activating Nocth1 pathway via modifying SOX2 degradation, Cell Biosci. 14 (1) (2024) 35.

[253] T. Kumar, R. Brennan, A. Mileo, M. Bendechache, Image data augmentation approaches: A comprehensive survey and future directions, IEEE Access (2024).

[254] S.-C. Huang, A. Pareek, M. Jensen, M.P. Lungren, S. Yeung, A.S. Chaudhari, Self-supervised learning for medical image classification: a systematic review and implementation guidelines, NPJ Digit. Med. 6 (1) (2023) 74.

[255] D. Wolf, T. Payer, C.S. Lisson, C.G. Lisson, M. Beer, M. Götz, T. Ropinski, Self-supervised pre-training with contrastive and masked autoencoder methods for dealing with small datasets in deep learning for medical imaging, Sci. Rep. 13 (1) (2023) 20260.

[256] H. Yuan, Toward real-world deployment of machine learning for health care: external validation, continual monitoring, and randomized clinical trials, Health Care Sci. 3 (5) (2024) 360.

[257] R.E. Teiu, T.F. Ursuleanu, H.C. Stefan, R. Grigorovici, A.R. Luca, M.P. Comanescu, A.I. Calin, The use of artificial intelligence in the therapeutic management of Papillary Thyroid Microcarcinoma: a randomized controlled trial protocol, Romanian J. Oral Rehabil. 16 (4) (2024).