

Review

Artificial Intelligence in Thyroid Cytopathology: Diagnostic and Technical Insights

Mariachiara Negrelli ^{1,†}, Chiara Frascarelli ^{1,2,†}, Fausto Maffini ¹, Elisa Mangione ¹, Clementina Di Tonno ¹, Mariano Lombardi ¹, Francesca Maria Porta ¹, Mario Urso ³, Vincenzo L'Imperio ³, Fabio Pagni ³, Claudio Bellevicine ⁴, Mariantonia Nacchio ⁴, Umberto Malapelle ⁴, Giancarlo Troncone ⁴, Antonio Marra ^{2,5}, Giuseppe Curigliano ^{2,5}, Konstantinos Venetis ^{1,*}, Elena Guerini-Rocco ^{1,2,‡} and Nicola Fusco ^{1,2,‡}

- ¹ Division of Pathology, European Institute of Oncology IRCCS, 20139 Milan, Italy; mariachiara.negrelli@ieo.it (M.N.); chiara.frascarelli@ieo.it (C.F.); fausto.maffini@ieo.it (F.M.); elisa.mangione@ieo.it (E.M.); clementina.ditonno@ieo.it (C.D.T.); mariano.lombardi@ieo.it (M.L.); francescamaria.porta@ieo.it (F.M.P.); elena.guerinirocco@ieo.it (E.G.-R.); nicola.fusco@ieo.it (N.F.)
- ² Department of Oncology and Hemato-Oncology, University of Milan, 20133 Milan, Italy; antonio.marra@ieo.it (A.M.); giuseppe.curigliano@ieo.it (G.C.)
- ³ Department of Medicine and Surgery, Pathology, IRCCS Fondazione San Gerardo dei Tintori, University of Milano-Bicocca, 20900 Monza, Italy; mario.urso@unimib.it (M.U.); vincenzo.limperio@unimib.it (V.L.); fabio.pagni@unimib.it (F.P.)
- ⁴ Department of Public Health, University of Naples Federico II, 80131 Naples, Italy; claudio.bellevicine@unina.it (C.B.); mariantonia.nacchio@unina.it (M.N.); umberto.malapelle@unina.it (U.M.); giancarlo.troncone@unina.it (G.T.)
- ⁵ Division of New Drugs and Early Drug Development for Innovative Therapies, European Institute of Oncology IRCCS, Via G. Ripamonti 435, 20141 Milan, Italy
- * Correspondence: konstantinos.venetis@ieo.it; Tel.: +39-327-874-5064
- † These authors contributed equally to this work.
- ‡ These authors also contributed equally to this work.

Simple Summary

Thyroid nodules are very common, and fine-needle aspiration cytology is the main test used to decide whether a nodule is benign or not. While this test is reliable in most cases, many samples fall into an “indeterminate” category, often leading to unnecessary operations or delays in treatment. New computer-based methods, known as deep learning, can analyze digital images of thyroid cytology slides and may help reduce this uncertainty. By learning patterns that even experienced specialists may overlook, these systems could support pathologists in making faster and more accurate decisions, especially in difficult cases. In this article, we discuss how deep learning has been applied to thyroid cytology, the technical and practical challenges it faces, and how it could eventually help make thyroid cancer diagnosis more precise, consistent, and accessible worldwide.

Abstract

Fine-needle aspiration cytology (FNAC) is the cornerstone of thyroid nodule evaluation, standardized by the Bethesda System. However, indeterminate categories (Bethesda III–IV) remain a major challenge, often leading to unnecessary surgery or delayed molecular testing. Deep learning (DL) has recently emerged as a promising adjunct in thyroid cytopathology, with applications spanning triage support, Bethesda category classification, and integration with molecular data. Yet, routine adoption is limited by preanalytical variability (staining, slide preparation, Z-stack acquisition, scanner heterogeneity), annotation bias, and domain shift, which reduce generalizability across centers. Most studies remain retrospective and single-institution, with limited external validation. This article provides a technical overview of DL in thyroid cytology, emphasizing preanalytical sources of variability,



Academic Editor: Yutaka Midorikawa

Received: 22 August 2025

Revised: 22 October 2025

Accepted: 27 October 2025

Published: 31 October 2025

Citation: Negrelli, M.; Frascarelli, C.; Maffini, F.; Mangione, E.; Di Tonno, C.; Lombardi, M.; Porta, F.M.; Urso, M.; L'Imperio, V.; Pagni, F.; et al. Artificial Intelligence in Thyroid Cytopathology: Diagnostic and Technical Insights. *Cancers* **2025**, *17*, 3525. <https://doi.org/10.3390/cancers17213525>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

architectural choices, and potential clinical applications. We argue that standardized datasets, multicenter prospective trials, and robust explainability frameworks are essential prerequisites for safe clinical deployment. Looking forward, DL systems are most likely to enter practice as diagnostic co-pilots, Bethesda classifiers, and multimodal risk-stratification tools. With rigorous validation and ethical oversight, these technologies may augment cytopathologists, reduce interobserver variability, and help transform thyroid cytology into a more standardized and data-driven discipline.

Keywords: thyroid cytology; deep learning; artificial intelligence; convolutional neural networks; multiple instance learning; Bethesda system; molecular prediction; explainable AI; multimodal models

1. Introduction

Thyroid nodules are common clinical findings, palpable in ~5% of adults but detectable by ultrasound in up to 60% of the population [1]. Fine-needle aspiration cytology (FNAC) according to The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) is the gold standard for initial risk stratification [2]. However, conventional FNAC is limited by interobserver variability and diagnostic uncertainty, particularly in indeterminate categories (Bethesda III/IV) [3]. These cases frequently result in unnecessary surgery, with benign histology confirmed in 30–40% of resected nodules [4,5]. Molecular testing is a reliable integrative test in indeterminate cases, but the high cost and limited availability restrict its widespread use [6–8].

Traditional computer-aided diagnosis (CAD) methods, based on handcrafted features or shallow machine learning, have achieved only limited generalizability, mainly because cytological smears show high intra- and inter-slide variability in staining, fixation, and cell distribution. Deep learning (DL), by contrast, offers an end-to-end approach capable of automatically learning hierarchical morphological patterns directly from digitized slides, without the need for predefined features. These models have shown remarkable success in histopathology and radiology, suggesting that cytology—particularly thyroid FNAC—could be the next frontier for AI-assisted diagnosis [9]. Despite this promise, DL applications in thyroid cytopathology remain relatively underexplored compared to other fields [10]. Existing reviews have primarily offered general overviews of AI in thyroid disease or radiology, with limited focus on cytology-specific challenges and pre-analytical variability [11,12].

In contrast, the present review provides a comprehensive and technically oriented synthesis of deep learning applications in thyroid cytology, connecting the algorithmic principles—ranging from convolutional neural networks to weakly supervised and hybrid frameworks—with their diagnostic implications. Particular attention is given to how pre-analytical factors influence model robustness, how explainability tools can enhance clinical trust, and how these systems might be realistically integrated into routine diagnostic workflows as assistive “co-pilots” rather than replacements for cytopathologists. By focusing on this intersection between technical design and diagnostic feasibility, the review moves beyond descriptive enumeration to offer a framework for critical evaluation and translational readiness of DL systems in thyroid cytology.

2. Preanalytical Considerations

Preanalytical variables play a critical role in digital cytopathology, as factors related to staining, specimen preparation, and image acquisition can significantly affect the performance and reliability of downstream computational analyses.

2.1. Staining Quality

Staining variability represents one of the most critical and underappreciated sources of pre-analytical heterogeneity in thyroid cytology. Different cytological preparations, such as Diff-Quik, Papanicolaou, and hematoxylin–eosin (H&E) in cell blocks, produce markedly distinct chromatic, textural, and contrast profiles. Each stain provides complementary diagnostic information: Diff-Quik facilitating assessment of overall cellularity and cytoplasmic detail [13], Papanicolaou enabling fine evaluation of nuclear morphology and chromatin [14], and H&E reproducing a histology-like appearance useful for cyto-histologic correlation [15]. This variety poses a formidable challenge for both digital acquisition and downstream DL analysis. Even within the same staining protocol, variations in fixation time, dye concentration, reagent pH, incubation duration, rinsing procedures, and batch-to-batch reagent differences may result in substantial alterations of hue, saturation, and contrast. When digitized, these inconsistencies manifest as domain shifts in color space, whereby nuclear and cytoplasmic tones, edge definition, and background hue differ across slides or institutions. DL models that are not robust to these shifts risk learning color artifacts rather than morphological features truly associated with the underlying pathology.

Classical stain normalization algorithms have been widely employed to mitigate inter-laboratory color variability in digital pathology. Among these, the Macenko and Reinhard methods remain the most established. The Macenko approach operates in the optical density (OD) space, estimating the dominant stain vectors through singular value decomposition (SVD) and reprojecting each image onto a standardized color basis [16]. This technique effectively harmonizes hue and intensity across slides while preserving most morphological information, although it may be sensitive to noise, illumination differences, and overlapping cells [16]. In contrast, the Reinhard method works in the Lab color space, which models human color perception, and aligns the mean and standard deviation of each channel to those of a chosen reference image. It is computationally efficient and performs well when color variations are moderate, but tends to lose accuracy when staining differences or background artifacts are more pronounced [17]. Both techniques provide rapid and accessible color harmonization but rely heavily on the choice of a representative reference image and do not fully address structural or staining artifacts.

Recent deep learning–based normalization models, such as StainNet and Colour Adaptive GAN (CAGAN), have been introduced to overcome the intrinsic limitations of classical algorithms. Unlike Macenko and Reinhard, which rely on global color statistics or linear stain decomposition, these networks learn a non-linear mapping between staining domains directly from data. StainNet employs a compact pixel-wise architecture trained through knowledge distillation, achieving faster inference and improved preservation of fine nuclear detail compared to GAN-based or statistical methods [18]. CAGAN further enhances adaptability by decoupling morphological structure from color attributes, allowing the model to modify chromatic style while maintaining cytological architecture intact. In practice, these approaches better accommodate complex inter-laboratory differences, such as shifts in hue, illumination, or reagent chemistry, resulting in higher consistency across multicenter datasets. However, they remain computationally demanding and require extensive validation to ensure that color transformation does not inadvertently alter diagnostically relevant features [19]. In summary, managing stain variability is essential for developing reliable deep learning models in thyroid cytology. While normalization

methods can improve visual consistency, true reproducibility ultimately depends on combining standardized laboratory protocols with robust computational harmonization—an ambitious but achievable goal only through coordinated multicenter efforts.

2.2. Specimen Preparation and Slide Digitalization

Cytological preparations for thyroid fine-needle aspiration (FNA) are intrinsically heterogeneous, characterized by variable thickness, irregular topography, and the coexistence of follicular clusters, colloid material, and background debris. These features make both digital acquisition and downstream deep learning (DL) analysis considerably more complex than in histological sections. In conventional smears, thick colloid pools and overlapping follicular groups frequently create abrupt changes in optical density, affecting focus and illumination within the same field. Such variations can distort nuclear and cytoplasmic features, particularly in hyperplastic or cystic nodules, where refractile colloid or air-drying artifacts alter transparency and color balance [20]. Preparation method further affects digital quality. Conventional smears preserve fine cytoplasmic detail and colloid texture but display marked variability in cell overlap and background; liquid-based cytology (LBC), in contrast, produces thinner, cleaner layers that enhance overall uniformity but often disperse fragile follicular and Hürthle cells and may reduce colloid visibility, elements that are morphologically informative for DL models. Moreover, LBC monolayers are not perfectly planar: subtle height variations along the z-axis, especially in poorly cellular samples, make it difficult to maintain uniform focus. This phenomenon, observed in both thyroid and cervical cytology, can result in globally sharp but locally blurred images where diagnostically relevant cells lie slightly out of plane [21,22]. The digitization process itself introduces additional variability. Most whole-slide scanners are optimized for histologic sections and cannot consistently capture the depth of thick thyroid smears. Z-stack scanning mitigates this issue by acquiring multiple focal planes, which is particularly useful for visualizing papillary carcinoma features such as nuclear grooves and pseudoinclusions. Studies in urinary and thyroid cytology have demonstrated that multi-plane imaging improves the detection of nuclear details and interobserver concordance. However, Z-stacking increases acquisition time, file size, and computational load, so most laboratories still rely on single-plane imaging, balancing throughput and optical fidelity [23]. Additional sources of variability, such as scanner type, image compression format (e.g., SVS, TIFF, JPEG2000), and magnification level (20× vs. 40×), can also affect downstream model performance [24–26]. From a practical standpoint, it is advisable to employ scanner profiles tailored to the specific preparation type. For conventional thyroid smears, broader focal ranges help capture overlapping follicular aggregates and colloid-rich regions, whereas in LBC slides, narrower z-ranges centered within the circular deposition area yield sharper and more reproducible results. Multi-point focusing within this region, combined with metadata-based quality control, can substantially improve digital consistency and DL inference. The establishment of such preparation-specific scanning profiles represents a feasible and cost-effective approach to enhance reproducibility in digital thyroid cytology.

2.3. Human Variability in Region of Interest (ROI) Annotation

Annotation of regions of interest (ROIs) is one of the most significant sources of variability in supervised DL pipelines for thyroid cytology. Unlike histology, where lesions are often well defined, cytological material is inherently heterogeneous: diagnostic cells are scattered among debris, colloid, and non-diagnostic areas. Consequently, even experienced cytopathologists may disagree on which fields truly represent the diagnostic component. These subjective differences, whether in selecting representative follicles, atypical nuclei, or excluding artifacts, can substantially influence model performance and generalizability [27].

Weakly supervised and multiple instance learning (MIL) approaches have been introduced to mitigate these challenges by associating slide-level rather than pixel-level labels. This strategy allows models to learn from whole-slide or case-level diagnoses (e.g., Bethesda categories) without requiring exhaustive manual annotation, reducing dependence on human demarcation. However, these methods remain susceptible to dataset-level bias, including case selection and class imbalance [28,29]. From a practical perspective, improving annotation reliability requires shared standards that clearly define what constitutes a diagnostic region in thyroid cytology. Consensus labeling by multiple experts and the development of public, quality-controlled datasets can mitigate subjectivity and support reproducibility across studies. Moving forward, hybrid annotation workflows, in which AI pre-selects candidate regions for expert review through pre-segmentation or attention maps, may offer a balanced compromise between accuracy and scalability, turning annotation into a collaborative rather than purely manual task [30].

2.4. Data Quality, Inclusion Criteria, and Domain Shift

The diagnostic landscape of thyroid cytology poses unique challenges for dataset construction and quality control. Across published studies, data inclusion criteria vary substantially—some focus exclusively on unequivocal benign (Bethesda II) and malignant (Bethesda VI) cases, while others also incorporate indeterminate or non-diagnostic categories (Bethesda I, III, and IV). Although this selective inclusion simplifies model training and improves apparent accuracy, it limits the algorithm's ability to manage real-world diagnostic uncertainty. In clinical practice, it is precisely the indeterminate categories that generate the greatest need for diagnostic support; excluding them therefore undermines the translational value of such models [31]. A second limitation lies in the handling of suboptimal or low-cellularity slides. Many datasets exclude FNA smears with poor fixation, thick colloid, or extensive blood contamination, conditions frequently encountered in routine practice. This “data cleaning” artificially inflates performance metrics and may produce models that perform well on idealized slides but fail on typical daily cases. In thyroid cytology, where cellular adequacy and background composition vary widely between centers, maintaining representative heterogeneity is essential for external validity [32]. Another critical issue is domain shift, the systematic discrepancy between training and testing data caused by differences in staining, scanning equipment, preparation method, or patient population. A DL model trained on LBC slides from a single institution may underperform when applied to conventional smears or slides digitized at different resolutions. In thyroid FNA, where both preparation types coexist, this risk is particularly high. Domain adaptation and color normalization can partly mitigate the problem, but they cannot replace a diverse and representative dataset [33,34]. In conclusion, robust dataset design should balance data quality with sufficient heterogeneity to reflect the variability inherent in routine thyroid cytology, thereby supporting the development of DL models that generalize across institutions and preparation types.

3. Architectural Variables

The optimal deep learning (DL) model for thyroid fine-needle aspiration cytology (FNAC) must address both technical variability and the biological diversity of lesions. Traditional machine learning (ML) approaches based on handcrafted features—such as nuclear texture or geometric descriptors—showed limited reproducibility across centers, as cytological smears display high intra-slide variability. In contrast, DL architectures autonomously learn multilevel representations from raw image data, enabling them to identify non-linear and context-dependent morphological cues that are often difficult to quantify by eye. This property makes DL particularly suitable for thyroid cytology, where

diagnostic features such as nuclear grooves, chromatin clearing, or colloid background may appear focal and heterogeneous [35–38].

3.1. Convolutional Neural Networks (CNNs)

CNNs remain the foundation of most DL pipelines in cytopathology. By applying trainable filters across the image, they extract increasingly abstract spatial features—from edges and textures to complex cell arrangements—without predefined feature engineering [39,40]. Classical architectures combine convolutional and pooling layers with non-linear activations (e.g., ReLU) and fully connected layers for classification [41,42]. In thyroid cytology, CNN-based models have shown particular promise in distinguishing benign from malignant lesions by leveraging subtle morphological cues. Lin et al. demonstrated that a CNN-based screening approach applied to thyroid cytology WSIs could accurately differentiate benign colloid-rich nodules from papillary and microfollicular carcinomas, capturing fine variations in nuclear morphology and colloid density [43]. Recently, CNN models have evolved from early architectures such as VGGNet and ResNet to more efficient designs like EfficientNet, which offer a good balance between accuracy and computational cost. When cytology datasets are small, transfer learning—using models pre-trained on large image collections such as ImageNet—can improve stability and speed up training [44–47]. Despite these advantages, CNNs still require large, well-curated datasets and often act as “black boxes”. Visualization tools such as Grad-CAM or SHAP can help confirm that model predictions are based on true diagnostic areas rather than color or scanning artifacts [11].

3.2. Multiple Instance Learning (MIL)

MIL is a specific weakly supervised framework particularly suited for WSI classification in pathology and cytology [48]. In this setting, each slide is treated as a ‘bag’ of smaller instances (patches or tiles) [49]. Labels are assigned only at the bag level, not to individual instances: this is particularly useful in cytology, where manual annotations are time-consuming and subject to variability [50]. Classical MIL assumes a slide is positive if at least one patch is positive, but modern implementations employ attention-based pooling to learn which patches contribute most to the final prediction, improving both performance and interpretability [51]. This strategy mirrors the reasoning process of cytopathologists, who evaluate the smear integrating focal atypia, background, and colloid distribution rather than analyzing isolated cells. Recent implementations have adopted attention-based pooling, allowing the network to assign greater importance to diagnostically relevant regions and to generate interpretable heatmaps. In a study published in 2025, by combining a custom CNN backbone (TCS-CNN) with attention-based MIL to classify WSIs into Bethesda II, IV, and VI categories, a 97% accuracy was reached without pixel-level annotation [52]. MIL represents an efficient alternative to fully supervised models but remains influenced by dataset imbalance and sampling bias. When indeterminate or low-cellularity cases are underrepresented, the model may overfit to clear-cut examples, limiting generalizability. Ensuring adequate representation of all Bethesda categories and preparation types is therefore crucial for future studies [53].

3.3. Hybrid DL Platforms

Hybrid strategies combining supervised and weakly supervised signals may offer the best compromise for complex domains like thyroid cytology. In 2021, a two-stage refined CNN demonstrated accurate benign-versus-malignant classification of thyroid FNAC slides by using expert-selected cellular regions to guide the model, improving robustness and reducing misclassification. A subsequent 2023 study applied a semi-supervised Noisy Student approach to cervical cytology and achieved performance comparable to fully supervised models, highlighting the potential of this method for datasets with limited

annotations [54,55]. In thyroid cytology, such hybrid and semi-supervised strategies are especially promising for indeterminate categories (Bethesda III–IV), where explicit region-level annotations are scarce but slide-level diagnoses are available. By generating both diagnostic probabilities and attention maps, these systems provide interpretable outputs that can assist cytopathologists in verifying predictions and identifying diagnostically relevant areas, supporting the integration of AI into real-world workflows [52,56–61].

4. Toward a Reliable Digital Cytodiagnostic Pipeline

Although no DL system is yet approved for routine use in thyroid cytology, several applications have been proposed. These include triage systems that flag low-risk slides for deferred review, decision-support tools that suggest Bethesda categories with confidence scores, and telepathology platforms [23,59,62,63]. All these together can theoretically be integrated into a clinical diagnostic workflow, as portrayed in Figure 1.

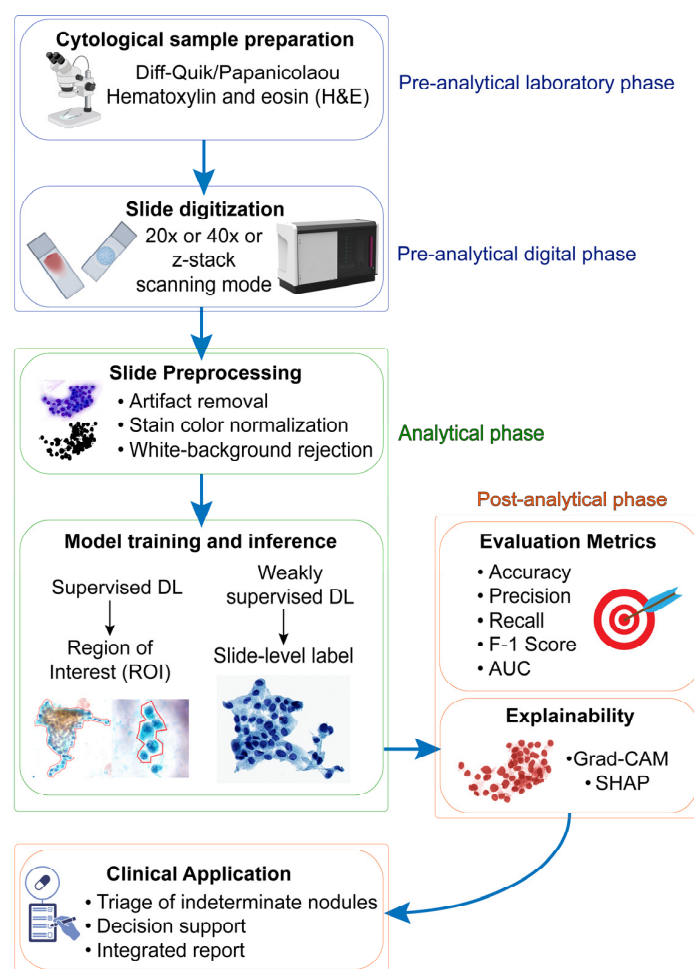


Figure 1. Deep learning workflow in thyroid cytology. The diagram outlines the end-to-end pipeline across four phases. Pre-analytical (laboratory): cytological sample preparation with common stains (Diff-Quik, Papanicolaou, H&E). Pre-analytical (digital): slide digitization at 20×/40× magnification, with or without Z-stack acquisition (improved focal coverage vs increased time, file size, and computational load). Analytical: preprocessing (artifact removal, color normalization, white-background rejection), followed by model training/inference using supervised (ROI-based) or weakly supervised (slide-level/MIL/attention) strategies. Evaluation and post-analytical: performance assessment with standard metrics (accuracy, precision, recall, F1-score, AUC) and integration of explainability tools (e.g., Grad-CAM, SHAP) to support interpretability. Clinical outputs: triage of indeterminate nodules, diagnostic decision support (e.g., Bethesda categorization), and integrated reporting. Detailed assumptions and trade-offs for each block are discussed in Sections 2–4.

4.1. Co-Pilot

The most immediate application of DL models in thyroid FNAC is as diagnostic support tools, often described as co-pilots [58,64]. Instead of replacing the cytopathologist, these systems can assist in routine thyroid workflows by prioritizing cases, flagging atypical or suspicious smears, and providing second-opinion Bethesda category suggestions, particularly for diagnostically challenging cases (Bethesda III–IV). When embedded into digital pathology viewers (e.g., QuPath or SlideViewer), DL models can deliver real-time overlays and ROI highlights, enabling cytopathologists to correlate AI outputs with key thyroid-specific cytomorphological features such as nuclear grooves, pseudoinclusions, or colloid density [65]. Although several co-pilot DL systems have been developed for thyroid cytology, none are yet validated for clinical use. These tools can prioritize challenging smears, flag atypical areas, or provide Bethesda category suggestions with confidence scores. Their true value lies in assisting, not replacing, the cytopathologist. When integrated into digital viewers, they can overlay attention maps highlighting features such as nuclear grooves or pseudoinclusions. However, their adoption still depends on standardized interfaces, explainable outputs, and demonstration of clinical benefit in workflow-based studies.

4.2. Bethesda Classifiers

DL models have been evaluated for automatic Bethesda categorization of thyroid FNAC slides. Most approaches use patch-level inference aggregated at the slide level through attention pooling or ensemble voting. CNN-based pipelines remain the most common, as in ThyroidEffi 1.0, which achieved high performance across Bethesda II, V, and VI (macro F1 = 0.897; AUC up to 0.98) [56]. More recently, transformer-based architectures have been explored: Zhu et al. developed vision–language models capable of generating Bethesda-style reports directly from digital images and textual inputs [66]. Despite these advances, fine-grained classification across all six Bethesda categories remains difficult, largely due to interobserver variability, overlapping cytomorphology, and annotation noise. The greatest challenge lies in indeterminate cases (Bethesda III–IV), which drive clinical uncertainty and often lead to unnecessary surgery or delayed molecular testing. Several studies have investigated whether DL can help refine risk stratification in this setting. For example, Zhong et al. [59] combined ultrasound radiomics with clinical features to classify Bethesda III nodules (AUC = 0.82), while Poursina et al. [67] and a 2025 meta-analysis [68] reported promising accuracy (pooled AUC $\approx$ 0.85) for AI-based reclassification of indeterminate nodules, though with considerable heterogeneity and limited external validation. Overall, DL-based classifiers show encouraging results, but evidence remains preliminary. Prospective, multi-center studies are essential to establish robustness, and future work should assess whether AI-driven risk stratification can safely guide conservative management in low-risk indeterminate cases.

4.3. Molecular Classifiers

The molecular landscape of thyroid tumors has become integral to cytological diagnosis and risk assessment [69–75]. The fifth edition of the WHO Classification of Thyroid Tumors incorporates key genetic alterations (i.e., BRAF V600E, RAS mutations, RET/PTC rearrangements, and PAX8–PPARG fusions) into the diagnostic framework for follicular-derived neoplasms [76]. With the growing use of NGS panels, cytopathologists now routinely integrate molecular data into FNAC interpretation. Commercial tools such as ThyroSeq, Afirma GSC, and ThyGeNEXT/ThyraMIR are widely applied in Bethesda III–IV nodules, providing genomic risk profiles that guide management [77,78]. In parallel, DL models have been explored for their ability to predict molecular alterations directly from cytological images. Early studies suggest that CNNs can capture genotype-associated

morphologic features, particularly in BRAF and RAS-mutated tumors [79]. While still experimental, such models could eventually be integrated into multimodal pipelines that combine cytology, molecular, and clinical data to improve risk stratification. In the long term, the integration of cytological, molecular, and clinical data through multimodal DL frameworks could help move thyroid cytology toward a precision-medicine model. Such tools are not expected to replace current molecular assays but to complement them—linking morphology, genotype, and outcome in a unified predictive continuum [80].

5. Conclusions and Future Directions

DL has emerged as a promising tool for improving diagnostic accuracy and workflow efficiency in thyroid cytology. Nevertheless, its translation from proof-of-concept to clinical reality remains constrained by technical, methodological, and organizational factors. Most published studies rely on relatively small, single-center datasets collected under heterogeneous staining, fixation, and digitization conditions: this lack of standardization limits external generalizability and hinders regulatory approval. In addition, variable annotation quality, inconsistent Bethesda categorization, and exclusion of low-cellularity or suboptimal smears often result in overoptimistic performance metrics that do not reflect real-world complexity.

Beyond quantitative accuracy, reproducibility and explainability are emerging as essential prerequisites for clinical acceptance. Models must demonstrate robustness to stain variability, scanner configuration, and preparation type, ensuring consistent behavior across institutions. While visualization methods such as Grad-CAM and SHAP have increased transparency, they remain qualitative in nature. This is particularly critical in indeterminate categories (Bethesda III–IV), where uncertainty has the greatest impact on patient management.

Legal, ethical, and regulatory frameworks will play a decisive role. Data privacy, algorithmic bias, and liability remain unresolved, and future deployment must balance innovation with accountability.

Looking ahead, DL applications most likely to reach the clinic include Bethesda classifiers, triage support, and multimodal models integrating cytology with molecular and clinical data. These systems should be evaluated not only by accuracy, but also by their ability to improve outcomes, reduce interobserver variability, and optimize resource use. With collaborative validation, ethical oversight, and a focus on clinical utility, AI can evolve from experimental prototypes into reliable co-pilots—supporting cytopathologists and transforming thyroid cytology into a more standardized, data-driven discipline.

Author Contributions: Study conception and design, N.F., E.G.-R. and K.V.; methodology, M.N. (Mariachiara Negrelli), C.F. and K.V.; writing—original draft preparation, M.N. (Mariachiara Negrelli) and C.F.; writing—review and editing, M.N. (Mariachiara Negrelli), C.F., F.M. and K.V.; revision, E.M., C.D.T., M.L., F.M.P., M.U., V.L., F.P., C.B., M.N. (Mariantonia Nacchio), U.M., G.T., A.M., G.C., E.G.-R. and N.F.; figure draft, C.F., M.N. (Mariachiara Negrelli) and K.V.; supervision, E.G.-R., N.F. and K.V.; project administration, N.F. and E.G.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the Italian Ministry of Health through Ricerca Corrente 5 × 1000 funds; the Italian Ministry of Innovations via the Sustainable Growth Fund–Innovation Agreements under the Ministerial Decree of 31 December 2021, and the Director’s Decree of 14 November 2022 (2nd Call), Project No. F/350104/01-02/X60.

Data Availability Statement: Data sharing is not applicable.

Acknowledgments: Konstantinos Venetis was supported by the Fondazione Umberto Veronesi; Antonio Marra by the ESMO José Baselga Fellowship for Clinician Scientists 2023–2025. The final

proofreading of grammar and syntax for the manuscript was conducted using ChatGPT 4 (GPT-5, OpenAI, October 2025 release) and Grammarly v.6.8.263.

Conflicts of Interest: U.M., Consulting or advisory role (unrelated to the current work): Boehringer Ingelheim, MSD, Roche, Amgen, Lilly, Thermo Fisher Scientific, Diaceutics, Merck, Glaxo Smith Kline, AstraZeneca; Speakers' Bureau (unrelated to the current work): Boehringer Ingelheim, Roche, AstraZeneca, MSD, Merck, Amgen, Thermo Fisher Scientific, Diaceutics, Lilly, Glaxo Smith Kline. A.M. has received support from Menarini Group and served on the Speakers' Bureau for Roche and AstraZeneca. G.C. has received honoraria for speaker engagements from Roche, Seattle Genetics, Novartis, Lilly, Pfizer, Foundation Medicine, NanoString, Samsung, Celltrion, BMS, and MSD; honoraria for consultancy from Roche, Seattle Genetics, and NanoString; honoraria for participation in advisory boards from Roche, Lilly, Pfizer, Foundation Medicine, Samsung, Celltrion, and Mylan; honoraria for writing engagements from Novartis and BMS; and honoraria for participation in the Ellipsis Scientific Affairs Group. He has also received institutional research funding for conducting phase I and II clinical trials from Pfizer, Roche, Novartis, Sanofi, Celgene, Servier, Orion, AstraZeneca, Seattle Genetics, AbbVie, Tesaro, BMS, Merck Serono, Merck Sharp & Dohme, Janssen-Cilag, Philogen, Bayer, Medivation, and Medimmune. K.V. Has received honoraria for speaker bureau from Merck Sharp & Dohme (MSD), Roche, and AstraZeneca; E.G-R. has received advisory fees, honoraria, travel accommodations/expenses, grants, and/or non-financial support from AstraZeneca, Exact Sciences, GSK, Illumina, MSD, Novartis, Roche, and Thermo Fisher Scientific. N.F. has received honoraria for consulting, advisory role, speaker bureau, travel, and/or research grants from Merck Sharp & Dohme (MSD), Merck, Novartis, AstraZeneca, Roche, Menarini Group, Daiichi Sankyo, GlaxoSmithKline (GSK), Gilead, Sysmex, Genomic Health, Veracyte, Sakura, Leica Biosystems, Lilly, Pfizer, ThermoFisher, Abbvie. These companies had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and/or in the decision to publish the results. All other authors declare no potential conflicts of interest.

References

1. Zamora, E.A.; Khare, S.; Cassaro, S. Thyroid nodule. In *StatPearls*; StatPearls Publishing: Treasure Island, FL, USA, 2025.
2. Haugen, B.R. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: What is new and what has changed? *Cancer* **2016**, *123*, 372–381. [\[CrossRef\]](#)
3. Bayrak, Y.B.; Erucar, A.T. Malignancy rates for Bethesda III and IV thyroid nodules: A retrospective study of the correlation between fine-needle aspiration cytology and histopathology. *BMC Endocr. Disord.* **2020**, *20*, 48. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Piticchio, T.; Russ, G.; Radzina, M.; Frasca, F.; Durante, C.; Trimboli, P. Head-to-head comparison of American, European, and Asian TIRADSs in thyroid nodule assessment: Systematic review and meta-analysis. *Eur. Thyroid. J.* **2024**, *13*, e230242. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Xu, L.; Zeng, F.; Wang, Y.; Bai, Y.; Shan, X.; Kong, L. Prevalence and associated metabolic factors for thyroid nodules: A cross-sectional study in Southwest of China with more than 120 thousand populations. *BMC Endocr. Disord.* **2021**, *21*, 175. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Angerilli, V.; Galuppini, F.; Pagni, F.; Fusco, N.; Malapelle, U.; Fassan, M. The Role of the Pathologist in the Next-Generation Era of Tumor Molecular Characterization. *Diagnostics* **2021**, *11*, 339. [\[CrossRef\]](#)
7. Cappello, F.; Angerilli, V.; Munari, G.; Cecon, C.; Sabbadin, M.; Pagni, F.; Fusco, N.; Malapelle, U.; Fassan, M. FFPE-Based NGS Approaches into Clinical Practice: The Limits of Glory from a Pathologist Viewpoint. *J. Pers. Med.* **2022**, *12*, 750. [\[CrossRef\]](#)
8. Fusco, N.; Pruneri, G.; Pagni, F.; Malapelle, U. Molecular Testing in Solid Tumors: Best Practices from the Molecular Pathology and Precision Medicine Study Group of the Italian Society of Pathology (PMMP/SIAPeC): Shaping Excellence in Molecular Diagnostics. *Pathol. J. Ital. Soc. Anat. Pathol. Diagn. Cytopathol.* **2025**, *117*, S1–S4. [\[CrossRef\]](#)
9. Hirokawa, M.; Niioka, H.; Suzuki, A.; Abe, M.; Arai, Y.; Nagahara, H.; Miyauchi, A.; Akamizu, T. Application of deep learning as an ancillary diagnostic tool for thyroid FNA cytology. *Cancer Cytopathol.* **2022**, *131*, 217–225. [\[CrossRef\]](#)
10. Wei, X.; Zhu, J.; Zhang, H.; Gao, H.; Yu, R.; Liu, Z.; Zheng, X.; Gao, M.; Zhang, S. Visual Interpretability in Computer-Assisted Diagnosis of Thyroid Nodules Using Ultrasound Images. *Med Sci. Monit.* **2020**, *26*, e927007-1–e927007-11. [\[CrossRef\]](#)
11. Hou, L.; Samaras, D.; Kurc, T.M.; Gao, Y.; Davis, J.E.; Saltz, J.H. Patch-Based Convolutional Neural Network for Whole Slide Tissue Image Classification. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2424–2433. [\[CrossRef\]](#)
12. Dimitriou, N.; Arandjelović, O.; Caie, P.D. Deep Learning for Whole Slide Image Analysis: An Overview. *Front. Med.* **2019**, *6*, 264. [\[CrossRef\]](#)

13. IHC World. Diff-Quick (Diff-Quik) Staining Protocol. 2024. Available online: <https://ihcworld.com/2024/01/26/diff-quick-diff-quick-staining-protocol/> (accessed on 15 June 2025).
14. Sathawane, P.; Kamal, M.M.; Deotale, P.R.; Mankar, H. Nuances of the Papanicolaou stain. *Cytojournal* **2022**, *19*, 43. [\[CrossRef\]](#)
15. IHC World. Hematoxylin and Eosin (H&E) Staining Protocol. 2024. Available online: <https://ihcworld.com/2024/01/25/hematoxylin-and-eosin-he-staining-protocol/> (accessed on 15 June 2025).
16. Ding, F.; Cai, C.; Li, J.; Liu, M.; Jiao, Y.; Wu, Z.; Xu, J. Classification of Whole-Slide Pathology Images Based on State Space Models and Graph Neural Networks. *Electronics* **2025**, *14*, 2056. [\[CrossRef\]](#)
17. Reinhard, E.; Adhikhmin, M.; Gooch, B.; Shirley, P. Color transfer between images. *IEEE Comput. Graph. Appl.* **2001**, *21*, 34–41. [\[CrossRef\]](#)
18. Kang, G. StainNet: A fast and robust stain normalization network. *Front. Med.* **2021**, *8*, 746307. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Zanjani, F. StainGAN: Stain Style Transfer for Digital Histological Images. *Sci. Rep.* **2023**, *13*, 13781.
20. Kim, D.; Burkhardt, R.; Alperstein, S.A.; Gokozan, H.N.; Goyal, A.; Heymann, J.J.; Patel, A.; Siddiqui, M.T. Evaluating the role of Z-stack to improve the morphologic evaluation of urine cytology whole slide images for high-grade urothelial carcinoma: Results and review of a pilot study. *Cancer Cytopathol.* **2022**, *130*, 630–639. [\[CrossRef\]](#)
21. Gu, H.; Onstott, E.; Yan, W.; Xu, T.; Wang, R.; Wu, Z.; Chen, X.A.; Haeri, M. Z-stack scanning can improve AI detection of mitosis: A case study of meningiomas. In Proceedings of the 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI), Houston, TX, USA, 14–17 April 2025.
22. Osamura, R.Y.; Matsui, N.; Kawashima, M.; Saiga, H.; Ogura, M.; Kiyuna, T. Digital/Computational Technology for Molecular Cytology Testing: A Short Technical Note with Literature Review. *Acta Cytol.* **2021**, *65*, 342–347. [\[CrossRef\]](#)
23. Yao, K.; Shen, R.; Parwani, A.; Li, Z. Comprehensive Study of Telecytology Using Robotic Digital Microscope and Single Z-Stack Digital Scan for Fine-Needle Aspiration-Rapid On-Site Evaluation. *J. Pathol. Inform.* **2018**, *9*, 49. [\[CrossRef\]](#)
24. Ji, X.; Salmon, R.; Mulliqi, N.; Khan, U.; Wang, Y.; Blilie, A.; Olsson, H.; Pedersen, B.G.; Sørensen, K.D.; Ulhøi, B.P.; et al. Physical Color Calibration of Digital Pathology Scanners for Robust Artificial Intelligence—Assisted Cancer Diagnosis. *Modern Pathol.* **2023**, *38*, 100715. [\[CrossRef\]](#)
25. Brixtel, R.; Bougleux, S.; Lezoray, O.; Caillot, Y.; Lemoine, B.; Fontaine, M.; Nebati, D.; Renouf, A. Whole Slide Image Quality in Digital Pathology: Review and Perspectives. *IEEE Access* **2022**, *10*, 131005–131035. [\[CrossRef\]](#)
26. Holub, P.; Müller, H.; Bíl, T.; Pireddu, L.; Plass, M.; Prasser, F.; Schlünder, I.; Zatloukal, K.; Nenutil, R.; Brázdil, T. Privacy risks of whole-slide image sharing in digital pathology. *Nat. Commun.* **2023**, *14*, 2577. [\[CrossRef\]](#)
27. Hossain, M.S.; Shahriar, G.M.; Syeed, M.M.M.; Uddin, M.F.; Hasan, M.; Shivam, S.; Advani, S. Region of interest (ROI) selection using vision transformer for automatic analysis using whole slide images. *Sci. Rep.* **2023**, *13*, 11314. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Liu, X.; Zhang, W.; Zhang, M.-L. Attention is not what you need: Revisiting multi-instance learning for whole slide image classification. *arXiv* **2024**, arXiv:2408.09449. [\[CrossRef\]](#)
29. Cazzaniga, G.; Del Carro, F.; Eccher, A.; Becker, J.U.; Gambaro, G.; Rossi, M.; Pieruzzi, F.; Fraggetta, F.; Pagni, F.; L’imperio, V. Improving the Annotation Process in Computational Pathology: A Pilot Study with Manual and Semi-automated Approaches on Consumer and Medical Grade Devices. *J. Imaging. Inform. Med.* **2024**, *38*, 1112–1119. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Corradi, A.; Bonizzi, G.; Sajjadi, E.; Pavan, F.; Fumagalli, M.; Molendini, L.O.; Monturano, M.; Cassi, C.; Musico, C.R.; Leoni, L.; et al. The Regulatory Landscape of Biobanks in Europe: From Accreditation to Intellectual Property. *Curr. Genom.* **2025**, *26*, 15–23. [\[CrossRef\]](#)
31. Lee, B.; Smola, B.; Roh, M.H.; Hughes, D.T.; Miller, B.S.; Jing, X. The impact of using the Bethesda System for reporting thyroid cytology diagnostic criteria on the follicular lesion of undetermined significance category. *J. Am. Soc. Cytopathol.* **2014**, *3*, 131–136. [\[CrossRef\]](#)
32. Barthe, P.; Brixtel, R.; Caillot, Y.; Lemoine, B.; Renouf, A.; Thurotte, V.; Beniken, O.; Bougleux, S.; Lézoray, O. Assessing the quality of whole slide images in cytology from nuclei features. *J. Pathol. Informatics* **2025**, *17*, 100420. [\[CrossRef\]](#)
33. Tellez, D.; Litjens, G.; Bándi, P.; Bulten, W.; Bokhorst, J.-M.; Ciompi, F.; van der Laak, J. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Med. Image Anal.* **2019**, *58*, 101544. [\[CrossRef\]](#)
34. Stacke, K.; Eilertsen, G.; Unger, J.; Lundstrom, C. Measuring Domain Shift for Deep Learning in Histopathology. *IEEE J. Biomed. Health Inform.* **2020**, *25*, 325–336. [\[CrossRef\]](#)
35. Lin, Y.-J.; Chao, T.-K.; Khalil, M.-A.; Lee, Y.-C.; Hong, D.-Z.; Wu, J.-J.; Wang, C.-W. Deep Learning Fast Screening Approach on Cytological Whole Slides for Thyroid Cancer Diagnosis. *Cancers* **2021**, *13*, 3891. [\[CrossRef\]](#)
36. Gadermayr, M.; Tschuchnig, M. Multiple instance learning for digital pathology: A review of the state-of-the-art, limitations & future potential. *Comput. Med. Imaging Graph.* **2024**, *112*, 102337.
37. Pittaro, A.; Del Gobbo, A.; Iofrida, E.; Fusco, N. Chondroid Differentiation in Thyroid Nodular Hyperplasia: An Innocent Bystander? *Int. J. Surg. Pathol.* **2018**, *27*, 274. [\[CrossRef\]](#)

38. Smith, A.; Galli, M.; Piga, I.; Denti, V.; Stella, M.; Chinello, C.; Fusco, N.; Leni, D.; Manzoni, M.; Roversi, G.; et al. Molecular signatures of medullary thyroid carcinoma by matrix-assisted laser desorption/ionisation mass spectrometry imaging. *J. Proteom.* **2019**, *191*, 114–123. [\[CrossRef\]](#)
39. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Boehm, K.M.; El Nahhas, O.S.M.; Marra, A.; Waters, M.; Jee, J.; Braunstein, L.; Schultz, N.; Selenica, P.; Wen, H.Y.; Weigelt, B.; et al. Multimodal histopathologic models stratify hormone receptor-positive early breast cancer. *Nat. Commun.* **2025**, *16*, 2106. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2012; Volume 25.
42. Rawat, R.R.; Ortega, I.; Roy, P.; Sha, F.; Shibata, D.; Ruderman, D.; Agus, D.B. Deep learned tissue “fingerprints” classify breast cancers by ER/PR/Her2 status from H&E images. *Sci. Rep.* **2020**, *10*, 7275. [\[CrossRef\]](#)
43. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
44. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
45. Talo, M. Automated classification of histopathology images using transfer learning. *Artif. Intell. Med.* **2019**, *101*, 101743. [\[CrossRef\]](#)
46. Gupta, P.; Huang, Y.; Sahoo, P.K.; You, J.-F.; Chiang, S.-F.; Onthoni, D.D.; Chern, Y.-J.; Chao, K.-Y.; Chiang, J.-M.; Yeh, C.-Y.; et al. Colon Tissues Classification and Localization in Whole Slide Images Using Deep Learning. *Diagnostics* **2021**, *11*, 1398. [\[CrossRef\]](#)
47. Matias-Guiu, X.; Temprana-Salvador, J.; Lopez, P.G.; Kammerer-Jacquet, S.-F.; Rioux-Leclercq, N.; Clark, D.; Schürch, C.M.; Fend, F.; Mattern, S.; Snead, D.; et al. Implementing digital pathology: Qualitative and financial insights from eight leading European laboratories. *Virchows Arch.* **2025**, *487*, 815–826. [\[CrossRef\]](#)
48. Campanella, G.; Chen, S.; Singh, M.; Verma, R.; Muehlstedt, S.; Zeng, J.; Stock, A.; Croken, M.; Veremis, B.; Elmas, A.; et al. A clinical benchmark of public self-supervised pathology foundation models. *Nat. Commun.* **2025**, *16*, 3640. [\[CrossRef\]](#)
49. Ilse, M.; Tomczak, J.; Welling, M. Attention-based deep multiple instance learning. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; PMLR: Norfolk, MA, USA, 2018.
50. Myronenko, A.; Xu, Z.; Yang, D.; Roth, H.R.; Xu, D. Accounting for dependencies in deep learning based multiple instance learning for whole slide imaging. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021*; Springer International Publishing: Cham, Switzerland, 2021.
51. Campanella, G.; Hanna, M.G.; Geneslaw, L.; Mirafior, A.; Silva, V.W.K.; Busam, K.J.; Brogi, E.; Reuter, V.E.; Klimstra, D.S.; Fuchs, T.J. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* **2019**, *25*, 1301–1309. [\[CrossRef\]](#)
52. Oh, S.Y.; Lee, Y.M.; Kang, D.J.; Kwon, H.J.; Chakraborty, S.; Park, J.H. Breaking Barriers in Thyroid Cytopathology: Harnessing Deep Learning for Accurate Diagnosis. *Bioengineering* **2025**, *12*, 293. [\[CrossRef\]](#)
53. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *arXiv* **2016**, arXiv:1610.02391.
54. Kurita, Y.; Meguro, S.; Tsuyama, N.; Kosugi, I.; Enomoto, Y.; Kawasaki, H.; Uemura, T.; Kimura, M.; Iwashita, T. Accurate deep learning model using semi-supervised learning and Noisy Student for cervical cancer screening in low magnification images. *PLoS ONE* **2023**, *18*, e0285996. [\[CrossRef\]](#)
55. Duan, W.; Gao, L.; Liu, J.; Li, C.; Jiang, P.; Wang, L.; Chen, H.; Sun, X.; Cao, D.; Pang, B.; et al. Computer-Assisted Fine-Needle Aspiration Cytology of Thyroid Using Two-Stage Refined Convolutional Neural Network. *Electronics* **2022**, *11*, 4089. [\[CrossRef\]](#)
56. Pham-Ngoc, H.; Nguyen-Van, D.; Vu-Tien, D.; Le-Hong, P. ThyroidEffi 1.0: A cost-effective system for high-performance multi-class thyroid carcinoma classification. *arXiv* **2025**, arXiv:2504.14139.
57. Sadeghi, Z.; Alizadehsani, R.; Cifci, M.A.; Kausar, S.; Rehman, R.; Mahanta, P.; Bora, P.K.; Almasri, A.; Alkhalwaldeh, R.S.; Hussain, S.; et al. A review of Explainable Artificial Intelligence in healthcare. *Comput. Electr. Eng.* **2024**, *118*, 109370. [\[CrossRef\]](#)
58. Sorrenti, S.; Dolcetti, V.; Radzina, M.; Bellini, M.I.; Frezza, F.; Munir, K.; Grani, G.; Durante, C.; D’andrea, V.; David, E.; et al. Artificial Intelligence for Thyroid Nodule Characterization: Where Are We Standing? *Cancers* **2022**, *14*, 3357. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Zhong, L.; Shi, L.; Lai, J.; Hu, Y.; Gu, L. Combined model integrating clinical, radiomics, BRAFV600E and ultrasound for differentiating between benign and malignant indeterminate cytology (Bethesda III) thyroid nodules: A bi-center retrospective study. *Gland. Surg.* **2024**, *13*, 1954–1964. [\[CrossRef\]](#)
60. Issa, P.P.; McCarthy, C.; Hussein, M.; Albuck, A.L.; Emad, E.; Shama, M.; Moroz, K.; Toraih, E.; Kandil, E. Assessing Adequacy: A Meta-Analysis of Rapid Onsite Evaluation of Thyroid Nodules. *J. Surg. Res.* **2024**, *296*, 523–531. [\[CrossRef\]](#) [\[PubMed\]](#)
61. Frascarelli, C.; Venetis, K.; Marra, A.; Mane, E.; Ivanova, M.; Cursano, G.; Porta, F.M.; Concardi, A.; Ceol, A.G.M.; Farina, A.; et al. Deep learning algorithm on H&E whole slide images to characterize TP53 alterations frequency and spatial distribution in breast cancer. *Comput. Struct. Biotechnol. J.* **2024**, *23*, 4252–4259. [\[CrossRef\]](#)
62. Ha, E.J.; Baek, J.H. Applications of machine learning and deep learning to thyroid imaging: Where do we stand? *Ultrasonography* **2021**, *40*, 23–29. [\[CrossRef\]](#)

63. Lu, Q.; Wu, Y.; Chang, J.; Zhang, L.; Lv, Q.; Sun, H. Application progress of artificial intelligence in managing thyroid disease. *Front. Endocrinol.* **2025**, *16*, 1578455. [[CrossRef](#)]
64. Yang, Y.; Guan, S.; Ou, Z.; Li, W.; Yan, L.; Situ, B. Advances in AI-based cancer cytopathology. *Interdiscip. Med.* **2023**, *1*, e20230013. [[CrossRef](#)]
65. Bankhead, P.; Loughrey, M.B.; Fernández, J.A.; Dombrowski, Y.; McArt, D.G.; Dunne, P.D.; McQuaid, S.; Gray, R.T.; Murray, L.J.; Coleman, H.G.; et al. QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* **2017**, *7*, 16878. [[CrossRef](#)]
66. Zhu, Y.; Sang, Q.; Jia, S.; Wang, Y.; Deyer, T. Deep neural networks could differentiate Bethesda class III versus class IV/V/VI. *Ann. Transl. Med.* **2019**, *7*, 231. [[CrossRef](#)]
67. Poursina, O.; Khayyat, A.; Maleki, S.; Amin, A. Artificial Intelligence and Whole Slide Imaging Assist in Thyroid Indeterminate Cytology: A Systematic Review. *Acta Cytol.* **2025**, *69*, 161–170. [[CrossRef](#)]
68. Jassal, K.; Edwards, M.; Koohestani, A.; Brown, W.; Serpell, J.W.; Lee, J.C. Beyond genomics: Artificial intelligence-powered diagnostics for indeterminate thyroid nodules—A systematic review and meta-analysis. *Front. Endocrinol.* **2025**, *16*, 1506729. [[CrossRef](#)]
69. Gharib, H.; Papini, E.; Garber, J.R.; Duick, D.S.; Harrell, R.M.; Hegedüs, L.; Paschke, R.; Valcavi, R.; Vitti, P.; AACE/ACE/AME Task Force on Thyroid Nodules. American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi Medical Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules—2016 Update. *Endocr. Pract.* **2016**, *22*, 622–639. [[CrossRef](#)]
70. Capitoli, G.; Alviano, A.M.; Monza, N.; Pagani, L.; Piga, I.; Bernasconi, D.P.; Greco, A.; Leni, D.; Maggioni, A.; Gatti, A.-V.; et al. Biomarker identification through spatial proteomics for the characterization of indeterminate thyroid nodules. *Endocrine* **2025**, *90*, 800–809. [[CrossRef](#)]
71. Denti, V.; Greco, A.; Alviano, A.M.; Capitoli, G.; Monza, N.; Smith, A.; Pilla, D.; Maggioni, A.; Ivanova, M.; Venetis, K.; et al. Spatially Resolved Molecular Characterization of Noninvasive Follicular Thyroid Neoplasms with Papillary-like Nuclear Features (NIFTPs) Identifies a Distinct Proteomic Signature Associated with RAS-Mutant Lesions. *Int. J. Mol. Sci.* **2024**, *25*, 13115. [[CrossRef](#)]
72. L'imperio, V.; Coelho, V.; Cazzaniga, G.; Papetti, D.M.; Del Carro, F.; Capitoli, G.; Marino, M.; Ceku, J.; Fusco, N.; Ivanova, M.; et al. Machine Learning Streamlines the Morphometric Characterization and Multiclass Segmentation of Nuclei in Different Follicular Thyroid Lesions: Everything in a Nutshell. *Mod. Pathol.* **2024**, *37*, 100608. [[CrossRef](#)]
73. Piga, I.; L'imperio, V.; Principi, L.; Bellevicine, C.; Fusco, N.; Maffini, F.; Venetis, K.; Ivanova, M.; Seminati, D.; Casati, G.; et al. Spatially Resolved Molecular Approaches for the Characterisation of Non-Invasive Follicular Tumours with Papillary-like Features (NIFTPs). *Int. J. Mol. Sci.* **2023**, *24*, 2567. [[CrossRef](#)]
74. Seminati, D.; Mane, E.; Ceola, S.; Casati, G.; Putignano, P.; Garancini, M.; Gatti, A.; Leni, D.; Pincelli, A.I.; Fusco, N.; et al. An Indeterminate for Malignancy FNA Report Does Not Increase the Surgical Risk of Incidental Thyroid Carcinoma. *Cancers* **2022**, *14*, 5427. [[CrossRef](#)] [[PubMed](#)]
75. Cazzaniga, G.; Seminati, D.; Smith, A.; Piga, I.; Capitoli, G.; Garancini, M.; L'Imperio, V.; Fusco, N.; Pagni, F. Lights on HBME-1: The elusive biomarker in thyroid cancer pathology. *J. Clin. Pathol.* **2022**, *75*, 588–592. [[CrossRef](#)]
76. Baloch, Z.W.; Asa, S.L.; Barletta, J.A.; Ghossein, R.A.; Juhlin, C.C.; Jung, C.K.; LiVolsi, V.A.; Papotti, M.G.; Sobrinho-Simões, M.; Tallini, G.; et al. Overview of the 2022 WHO Classification of Thyroid Neoplasms. *Endocr. Pathol.* **2022**, *33*, 27–63. [[CrossRef](#)]
77. Vora, A.; Holt, S.; Haque, W.; Lingvay, I. Long-Term Outcomes of Thyroid Nodule AFIRMA GEC Testing and Literature Review: An Institutional Experience. *Otolaryngol. Head Neck Surg.* **2020**, *162*, 634–640. [[CrossRef](#)]
78. Chen, T.; Gilfix, B.M.; Rivera, J.A.; Sadeghi, N.; Richardson, K.; Hier, M.P.; Forest, V.-I.; Fishman, D.; Caglar, D.; Pusztaszeri, M.; et al. The Role of the ThyroSeq v3 Molecular Test in the Surgical Management of Thyroid Nodules in the Canadian Public Health Care Setting. *Thyroid* **2020**, *30*, 1280–1287. [[CrossRef](#)]
79. Wang, C.-W.; Muzakky, H.; Lee, Y.-C.; Lin, Y.-J.; Chao, T.-K. Annotation-Free Deep Learning-Based Prediction of Thyroid Molecular Cancer Biomarker BRAF (V600E) from Cytological Slides. *Int. J. Mol. Sci.* **2023**, *24*, 2521. [[CrossRef](#)]
80. Anand, B.; Ramdas, A.; Ambrose, M.M.; Kumar, N.P. The Bethesda System for Reporting Thyroid Cytopathology: A Cytohistological Study. *J. Thyroid. Res.* **2020**, *2020*, 8095378. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.