The application of artificial intelligence to thyroid nodule assessment

Paola Chiara Rizzo
Stefano Marletta
Nicolò Caldonazzi
Alessia Nottegar
Albino Eccher
Fabio Pagni
Vincenzo L'Imperio
Liron Pantanowitz

Abstract

Artificial intelligence (AI) is of considerable interest in the healthcare community including its diagnostic applications for thyroid nodules in assisting both radiology and FNA assessment. Fine-needle aspiration (FNA) helps distinguishing benign from malignant thyroid nodules and is a crucial step in the initial diagnosis of cancer. The classification of some lesions can be challenging, and the use of AI in some cases may become essential in order not to give an indeterminate result to the lesion. In this review, we summarize the available evidence regarding the application of AI in thyroid imaging and cytopathology. There are now considerable applications in digital waiting to be approved that will save time and cut costs. The published literature to date has shown promising results. However, future work is required to better define how this technology can be exploited in routine cytopathology practice.

Keywords Artificial intelligence; cytology; FNA; thyroid; WSI

Paola Chiara Rizzo MD *Department of Diagnostics and Public Health, Section of Pathology, University of Verona, Verona, Italy. Conflicts of interest: none declared.*

Stefano Marletta MD *Department of Diagnostics and Public Health, Section of Pathology, University of Verona, Verona and Division of Pathology, Humanitas Istituto Clinico Catenese, Catania, Italy. Conflicts of interest: none declared.*

Nicolò Caldonazzi MD *Department of Diagnostics and Public Health, Section of Pathology, University of Verona, Verona, Italy. Conflicts of interest: none declared.*

Alessia Nottegar MD *Department of Diagnostics and Public Health, Section of Pathology, University of Verona, Verona, Italy. Conflicts of interest: none declared.*

Albino Eccher MD Professor *Section of Pathology, Department of Medical and Surgical Sciences for Children and Adults, University of Modena and Reggio Emilia, University Hospital of Modena, Modena, Italy. Conflicts of interest: none declared.*

Fabio Pagni MD Professor *Department of Medicine and Surgery, ASST Monza, San Gerardo Hospital, University of Milano-Bicocca, Monza, Italy. Conflicts of interest: none declared.*

Vincenzo L'Imperio MD *Department of Medicine and Surgery, ASST Monza, San Gerardo Hospital, University of Milano-Bicocca, Monza, Italy. Conflicts of interest: none declared.*

Liron Pantanowitz MD Professor *Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA. Conflicts of interest: none declared.*

Introduction

Fine-needle aspiration (FNA) cytology is considered the gold standard for the diagnosis of thyroid nodules because of its safety, accuracy, and cost effectiveness.[1,2] FNA cytology coupled with ultrasonographic guidance is valuable for distinguishing benign from malignant thyroid nodules, and for providing information on cancer subtypes. The Bethesda system[3] represents the classification system for reporting thyroid cytology and consists of six malignancy groups. Bethesda categories III (atypia of indeterminate significance or follicular lesion of indeterminate significance) and IV (follicular neoplasm or suspected follicular neoplasm) represent the most challenging categories, the so-called "indeterminate" nodules, which are characterized by very large malignancy risks ranging from 15.7% to 54.6% and from 16.8% to 72.4%, respectively.[4] Nowadays, immunocytochemistry and molecular tests on material aspirated from thyroid lesions can provide very useful ancillary information to guide the clinical management of affected patients, especially in the case of cytologically indeterminate thyroid FNA, allowing for better determination of which patients will benefit from conservative surveillance versus definitive surgery.[5–10] There is already abundant literature about the application of artificial intelligence (AI) to digital images of thyroid pathology.[11–13] Particularly in the field of radiology, there is ample evidence that has been published about digital ultrasound images of thyroid lesions.

Recently, digital pathology (DP) has emerged and is widely transforming the practice of pathology, especially following regulatory approvals of whole slide imaging (WSI) for primary diagnosis.[14,15] In 2013, the College of American Pathologists (CAP) published specific guidelines on validating WSI for diagnostic purposes, which were then updated in 2021. With the widespread adoption of WSI in clinical practice, many opportunities have risen to apply AI to analyze images in an attempt to automate and improve the diagnostic accuracy of rapid onsite and final cytologic interpretation.

In this review, we summarize the available evidence regarding the application of AI in thyroid imaging and cytopathology (Figure 1).

Digital pathology and AI models

There are two main computing approaches of AI: the classic machine learning and the deep learning (DL). The first one is a supervised method which creates an algorithm requiring a dataset annotated by human experts; the created algorithm then compares its results with those contained in the dataset created using human input (Figure 2). The deep learning model instead uses an unsupervised approach allowing a computer to learn to

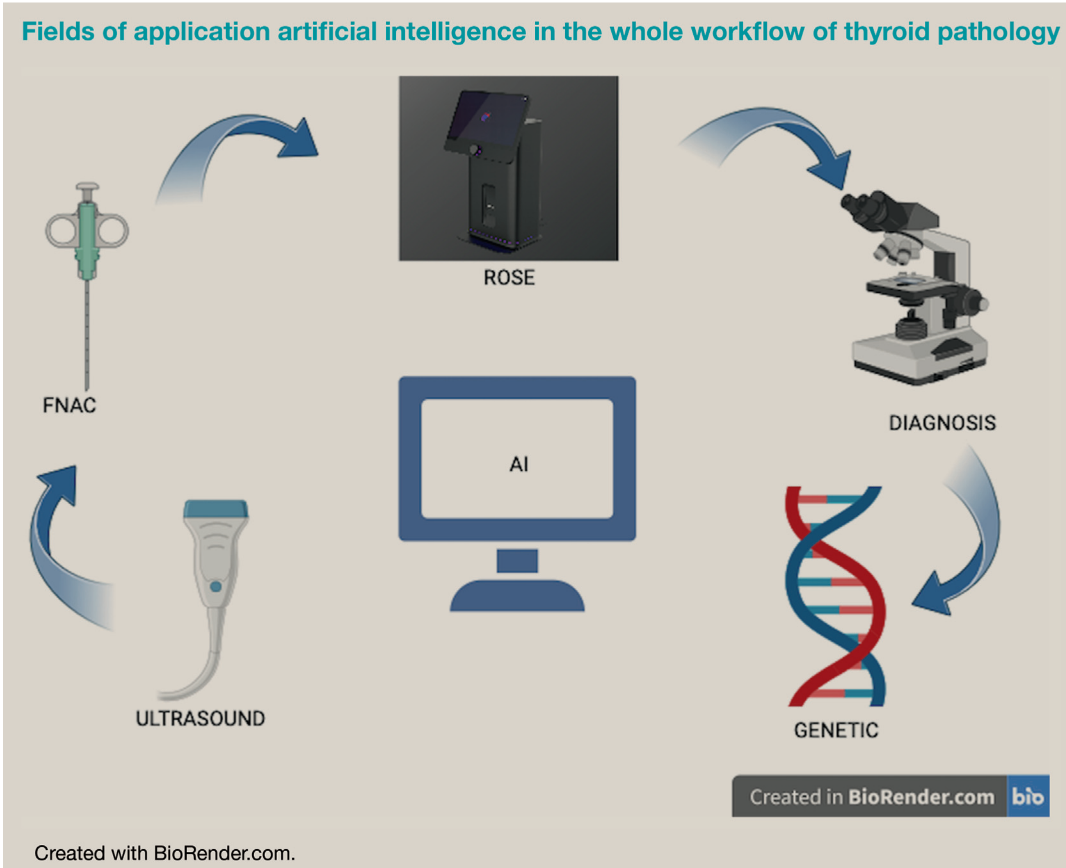


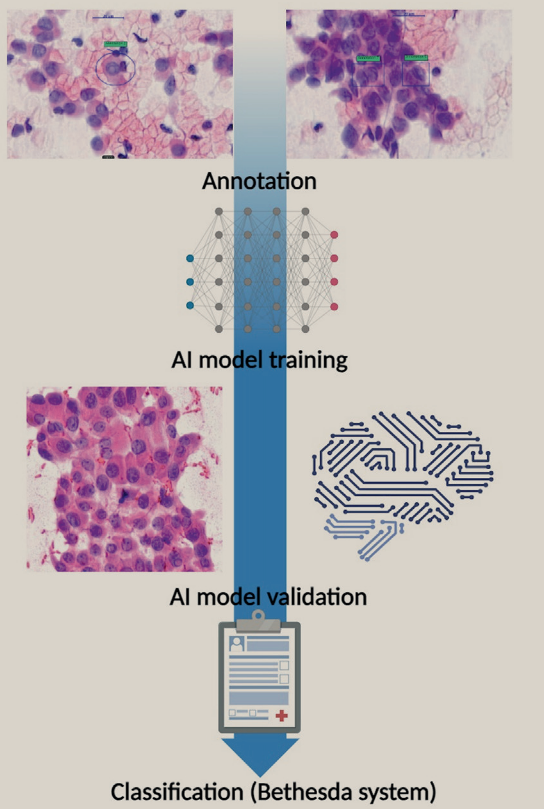
Figure 1

address specific topics without the need of human support by using and combining wide datasets for model formation. DL algorithms are nowadays mainly represented by 'artificial neural networks (ANN)' and their state-of-the art evolution 'convolutional neural network (CNN)' consisting of many connected artificial layers working similar to the networks of neurons in the human cortex.[16,17]

Radiomics is another application of AI that uses automatic or semi-automatic analytical methods, including image acquisition, image segmentation, construction of classification, and prediction models. Radiomics can either be DL-based where the features to be extracted are not predefined or rely on classic machine learning tools where the features to be extracted are predefined.[18]

AI was first applied in thyroid cytology specimens by Karakitsos et al.[19] in 1996, when they analyzed 26 different nuclear features using 2 different neural networks. They initially attempted to classify thyroid cells into one of four categories (benign follicular cell, benign oncocyte, malignant follicular cell, and malignant oncocyte), and then, relying on a different classifier, they attempted to label nuclei as benign or malignant. Their second classifier performed much better than the early four-tiered one, with an overall accuracy of 90.61%. Similar results were reported by Gopinath et al.[20] and Sanyal et al.,[21] who exploited a similar approach to classify thyroid tumors as either benign or malignant. More recently, Guan et al.[22] analyzed 407 images of papillary thyroid carcinoma (PTC) and 352 images of benign thyroid lesions and developed a novel AI cytological classification based on nuclear size and staining. In their study, Visual Geometry Group (VGG)-16 and Inception-v3 deep convolutional neural networks (CNN) were trained and tested, yielding a 97.66% accuracy for discriminating PTC from benign nodules. Both are deep convolutional neural networks (DCNNs) - machines made of levels (layers) that are most commonly applied to analyze visual imagery. The former is based on 13 convolutional layers and 3 fully connected layers, while the latter uses 3 different modules combined with each other, each of them composed of several convolutional and pooling layers in parallel. The VGG-16 obtained a better performance rather than Incption-v3 with an accuracy of 97,66% for discriminating PTC from benign nodules. Moreover, in 2022 Hirokawa et al.[23] used AI-based image analysis to analyze thyroid FNA in different prediction classes. Normally, the distinction between follicular adenoma (FA) and carcinoma (FC) requires surgical resection to look for capsular and vascular invasion. However, with their AI-based approach using EfficientNetV2-L as their image-classification model, 148,395 microscopic images of FNAC patches were divided into a training and validation set (80%) and a test set (20%). The PR curve and a confusion matrix were used to evaluate classification performance. These investigators showed that follicular adenomas and follicular thyroid carcinomas could be distinguished from other thyroid lesions in 86.7% and 93.9% respectively. Savala et al.[24] developed a promising AI-based tool

Supervised machine learning: whole slide images of thyroid smears annotated for pseudoinclusions (circle) and nuclear grooves (rectangles) are used to feed and train a fully supervised machine learning model



Afterwards, when being presented a novel unlabeled picture from the validation dataset the algorithm is able to classify it into the proper Bethesda category by comparison with the originally provided annotated patches.

Figure 2

which was proficiently able to distinguish FA from FC, albeit this study only employed 9 such cases.

Early machine-learning efforts in pathology used static images acquired by digital cameras coupled to microscopes. However, current training of AI algorithms relies mostly on large WSI files that are broken into thousands of smaller more manageable image tiles. While AI systems that analyze digital pathology files initially relied on manual annotation by pathologists for supervised learning, modern CNNs utilizing very large datasets have veered more towards weak (semi-supervised) and/or unsupervised learning. CNNs are best leveraged when dealing with WSI, and their architecture is designed to mimic the human brain cortex. Many of these narrow (task-based) algorithms have been shown to be robust, generalizable and yield highly predictive performances when trained using large, heterogeneous datasets. Studies evaluating the performance of CNNs to the gold-standard, which is a diagnosis rendered by an expert cytologist, have demonstrated much success, even when comparing AI output to patients' outcomes.[25]

AI applied to radiology

As mentioned earlier, the most commonly used radiological method for the evaluation of thyroid nodules is ultrasound (US). An intrinsic problem with this instrument is that it is highly operator dependent, with high sensitivity but low specificity and hence subject to poor interobserver agreement. The Thyroid Imaging Reporting and Data System (TIRADS) is of remarkable value in helping classify the risk of an imaged nodule harboring a PTC. However, the reliability of TIRADS to predict other entities needs improvement.[26]

In this regard, AI can help radiologists sort out challenging cases. Several studies have accordingly shown how AI can be applied to evaluate specific ultrasound features. For example, a method described by Widman Tobriner et al.[27] involved AI TIRADS prediction based on the American College of Radiology (ACR) eight-tiered category system (AI-TIRADS). These authors reported that their AI-based models had better accuracy than human derived TIRADS results, when performed by both non-dedicated (55% vs 48%) and dedicated or specialized (65% vs 47%) radiologists.

A novel application of AI is to employ texture analysis of imaged thyroid nodules. Transforming black-and-white US images into numeric matrices thereby allowing objective evaluation using texture analysis, which in turn helps circumvent inter-observer variability. Machine learning leveraged for such narrow tasks have proven to be very successful, as demonstrated by Prochazka et al.[28] who reached a diagnostic accuracy of 94.64%, or by Raghavendra et al.[29] who integrated texture with spatial information to yield an excellent area under the curve (94.45%) for diagnostic performance.

AI vs molecular testing in indeterminate nodules

Today, molecular testing of aspirated material from thyroid lesions can provide very useful ancillary information to guide the clinical management of patients, especially for cytologically indeterminate thyroid FNAs[7] Molecular diagnosis of thyroid nodules has been made possible by genomic tests due to the feasibility of analyzing small samples and the increasing cost-effectiveness of next-generation sequencing. Molecular analysis to support the diagnosis of indeterminate thyroid nodules seems to have excellent results, as in the case of Afirma GSC and Thyroseq v3.[30] However, these kinds of molecular tests are limited by the required instrumentation, highly specialized staff, and technical expertise. Moreover, the extraction of genetic material involves the destruction of tissue obtained by fine-needle aspiration though this is usually not an issue as additional FNA passes are used to collect material for molecular testing.

Recently Wang et al. explored the use of AI for prediction of BRAF mutation in thyroid cytopathology. A total of 118 slides were used to evaluate the proposed DL model. The results show an accuracy of 87%, a precision of 94%, a sensitivity of 91%, a specificity of 71%, and a mean of sensitivity and specificity of 81%.[31]

The use of AI could have several advantages over molecular tests, or at least become an additional element to support diagnosis. Firstly, the software does not require any special training by the pathologist, and in the future may also become a useful

Pros and cons of using artificial intelligence (AI) vs molecular tests in thyroid pathology

Parameters	Artificial intelligence	Molecular testing
Expertise	Medium	High
Availability	Small and medium—high volume centers	Only in medium—high volume centers
Time	Few minutes	Several hours
Space	A desk	A room
Costs	Low in a digital laboratory setting	High
Maintenance	Related to the robustness of the software	Reagents, material, instruments
Automation	Possible	Still not possible
Material	Available for other uses	Consumed in order to obtain data

Table 1

tool to support less experienced pathologists. In addition, the digital workstation requires less physical space than the instrumentation needed for molecular analysis and would also be useable for those laboratories that do not have the facilities to perform molecular tests. The use of AI allows standardization, among different centers, and automation in high-volume centers, speeding up patient diagnosis and could ease the workload in molecular laboratories (Table 1).

Conclusion and future perspectives

The current literature has shown promising results for the adoption of AI in thyroid imaging and cytopathology. However, there are some still unsolved specific issues in the application of AI to cytopathology, mainly concerning the fact that cytological material is difficult to digitalize as overlapping and obscured cells are difficult to segment. Nevertheless, we need to understand what kind of algorithms cytopathologists need, whether for screening or for diagnostics. A further concept that needs to be investigated is whether AI algorithms should be trained to replicate existing classification systems or not, and what kind of gold standard should they be tested against.

Some future applications of AI in thyroid cytology diagnostics include the possibility of ROSE (rapid on-site evaluation) evaluation and the use of algorithms to support challenging diagnoses rather than employing much more expensive molecular investigations. The use of AI for ROSE may be an important timesaver for pathologists, as there is a study underway to see if it can be performed at the patient's bedside using fresh unstained slides. Pathware [32,33] is an imaging platform, under development, for assessing adequacy on unstained specimens. This would allow the avoidance of the need for a slide staining area with all the problems of storage, space, and safety of reagents and dyes.

Young, less experienced pathologists, or particularly complex cases could benefit by using analytical tools which have a low error rate such as those proposed by Shang et al. to distinguish papillary carcinoma from benign lesions or discriminate with a certain level of confidence the follicular lesions that fall into a Bethesda group 3—4,[3] as proposed by Hirokawa.[23]

More studies are indeed warranted to further define how emerging AI-based technology can become more widely employed in routine practice. ◆

REFERENCES

1 Brito JP, Gionfriddo MR, Al Nofal A, et al. The accuracy of thyroid nodule ultrasound to predict thyroid cancer: systematic review and meta-analysis. *J Clin Endocrinol Metab* 2014; **99:** 1253—63. https://doi.org/10.1210/jc.2013-2928.

2 Singh Ospina N, Brito JP, Maraka S, et al. Diagnostic accuracy of ultrasound-guided fine needle aspiration biopsy for thyroid malignancy: systematic review and meta-analysis. *Endocrine* 2016; **53:** 651—61. https://doi.org/10.1007/s12020-016-0921-x.

3 Ali SZ, Baloch ZW, Cochand-Priollet B, et al. The 2023 Bethesda system for reporting thyroid cytopathology. *Thyroid* 2023; **33:** 1039—44. https://doi.org/10.1089/thy.2023.0141.

4 Zahid A, Shafiq W, Nasir KS, et al. Malignancy rates in thyroid nodules classified as Bethesda categories III and IV; a subcontinent perspective. *J Clin Transl Endocrinol* 2021; **23**: 100250. https://doi.org/10.1016/j.jcte.2021.100250.

5 Cibas ES, Ali SZ. The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid* 2017; **27:** 1341—6. https://doi.org/10.1089/thy.2017.0500.

6 Bose S, Sacks W, Walts AE. Update on molecular testing for cytologically indeterminate thyroid nodules. *Adv Anat Pathol* 2019; **26:** 114—23. https://doi.org/10.1097/PAP.0000000000000211.

7 Zhang M, Lin O. Molecular testing of thyroid nodules: a review of current available tests for fine-needle aspiration specimens. *Arch Pathol Lab Med* 2016; **140:** 1338—44. https://doi.org/10.5858/arpa.2016-0100-RA.

8 Nishino M, Nikiforova M. Update on molecular testing for cytologically indeterminate thyroid nodules. *Arch Pathol Lab Med* 2018; **142:** 446—57. https://doi.org/10.5858/arpa.2017-0174-RA.

9 Pisapia P, Pepe F, Sgariglia R, et al. Next generation sequencing in cytology. *Cytopathology* 2021; **32:** 588—95. https://doi.org/10.1111/cyt.12974.

10 Pisapia P, Pepe F, Iaccarino A, et al. Next generation sequencing in cytopathology: focus on non-small cell lung cancer. *Front Med* 2021; **8**. https://doi.org/10.3389/fmed.2021.633923.

11 Marletta S, Fusco N, Munari E, et al. Atlas of PD-L1 for pathologists: indications, scores, diagnostic platforms and reporting systems. *J Personalized Med* 2022; **12**. https://doi.org/10.3390/jpm12071073.

12 Girolami I, Marletta S, Pantanowitz L, et al. Impact of image analysis and artificial intelligence in thyroid pathology, with

particular reference to cytological aspects. *Cytopathology* 2020; **31:** 432–44. https://doi.org/10.1111/cyt.12828.

13 Marletta S, Salatiello M, Pantanowitz L, et al. Delphi expert consensus for whole slide imaging in thyroid cytopathology. Cytopathology. Published online 2023. doi:10.1111/cyt.13279.

14 Fraggetta F, Garozzo S, Zannoni GF, et al. Routine digital pathology workflow: the catania experience. *J Pathol Inf* 2017; **8:** 51. https://doi.org/10.4103/jpi.jpi_58_17.

15 Pantanowitz L, Wiley CA, Demetris A, et al. Experience with multimodality telepathology at the university of Pittsburgh medical center. *J Pathol Inf* 2012; **3:** 45. https://doi.org/10.4103/2153-3539.104907.

16 Ludwig M, Ludwig B, Mikuła A, et al. The use of artificial intelligence in the diagnosis and classification of thyroid nodules: an update. *Cancers* 2023; **15**. https://doi.org/10.3390/cancers15030708.

17 Tessler FN, Thomas J. Artificial intelligence for evaluation of thyroid nodules: a primer. *Thyroid* 2023; **33:** 150–8. https://doi.org/10.1089/thy.2022.0560.

18 Gao X, Ran X, Ding W. The progress of radiomics in thyroid nodules. *Front Oncol* 2023; **13**: 1109319. https://doi.org/10.3389/fonc.2023.1109319.

19 Karakitsos P, Cochand-Priollet B, Guillausseau PJ, et al. Potential of the back propagation neural network in the morphologic examination of thyroid lesions. *Anal Quant Cytol Histol* 1996; **18:** 494–500.

20 Gopinath B, Shanthi N. Development of an automated medical diagnosis system for classifying thyroid tumor cells using multiple classifier fusion. *Technol Cancer Res Treat* 2015; **14:** 653–62. https://doi.org/10.7785/tcrt.2012.500430.

21 Sanyal P, Mukherjee T, Barui S, et al. Artificial intelligence in cytopathology: a neural network to identify papillary carcinoma on thyroid fine-needle aspiration cytology smears. *J Pathol Inf* 2018; **9:** 43. https://doi.org/10.4103/jpi.jpi_43_18.

22 Guan Q, Wang Y, Ping B, et al. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study. *J Cancer* 2019; **10:** 4876–82. https://doi.org/10.7150/jca.28769.

23 Hirokawa M, Niioka H, Suzuki A, et al. Application of deep learning as an ancillary diagnostic tool for thyroid FNA cytology. *Cancer Cytopathol* 2023; **131:** 217–25. https://doi.org/10.1002/cncy.22669.

24 Savala R, Dey P, Gupta N. Artificial neural network model to distinguish follicular adenoma from follicular carcinoma on fine needle aspiration of thyroid. *Diagn Cytopathol* 2018; **46:** 244–9. https://doi.org/10.1002/dc.23880.

25 Dov D, Kovalsky S, Cohen J, et al. Thyroid cancer malignancy prediction from whole slide cytopathology images. Published online 2019.

26 Trimboli P, Castellana M, Piccardo A, et al. The ultrasound risk stratification systems for thyroid nodule have been evaluated against papillary carcinoma. *A meta-analysis. Rev Endocr Metab Disord* 2021; **22**(2)**:** 453–60. https://doi.org/10.1007/s11154-020-09592-3.

27 Wildman-Tobriner B, Buda M, Hoang JK, et al. Using artificial intelligence to revise ACR TI-RADS risk stratification of thyroid nodules: diagnostic accuracy and utility. *Radiology* 2019; **292:** 112–9. https://doi.org/10.1148/radiol.2019182128.

28 Prochazka A, Gulati S, Holinka S, et al. Classification of thyroid nodules in ultrasound images using direction-independent features extracted by two-threshold binary decomposition. *Technol Cancer Res Treat* 2019; **18**: 153303381983074. https://doi.org/10.1177/1533033819830748.

29 Raghavendra U, Rajendra Acharya U, Gudigar A, et al. Fusion of spatial gray level dependency and fractal texture features for the characterization of thyroid lesions. *Ultrasonics* 2017; **77:** 110–20. https://doi.org/10.1016/j.ultras.2017.02.003.

30 Khan TM, Zeiger MA. Thyroid nodule molecular testing: is it ready for prime time? *Front Endocrinol* 2020; **11**. https://doi.org/10.3389/fendo.2020.590128.

31 Wang CW, Muzakky H, Lee YC, et al. Annotation-free deep learning-based prediction of thyroid molecular cancer biomarker BRAF (V600E) from cytological slides. *Int J Mol Sci* 2023; **24:** 2521. https://doi.org/10.3390/ijms24032521.

32 https://www.pathware.com/.

33 https://www.sbir.gov/node/2511257-Leveraging Label-Free Digital imaging to enable rapid on-site evaluation (ROSE) during thyroid biopsy procedures.

Practice points

- The classification of some thyroid lesions can be challenging
- A novel application of AI is to employ texture analysis of imaged thyroid nodules
- The use of AI allows for standardization and automation in high-volume centers
- The use of AI may be a useful tool for less experienced pathologists
- There are still many unsolved specific issues in the application of AI to cytopathology

