Journal of Pathology Informatics 14 (2023) 100309

Contents lists available at ScienceDirect

Journal of Pathology Informatics

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

Review Article

Current status of machine learning in thyroid cytopathology

Charles M. Wong, Brie E. Kezlarian, Oscar Lin *

Memorial Sloan-Kettering Cancer Center, New York, NY, USA

ARTICLE INFO

Keywords:
Computational pathology
Digital pathology
Machine learning algorithms
Thyroid
Cytology

ABSTRACT

The implementation of Digital Pathology has allowed the development of computational Pathology. Digital image-based applications that have received FDA Breakthrough Device Designation have been primarily focused on tissue specimens. The development of Artificial Intelligence-assisted algorithms using Cytology digital images has been much more limited due to technical challenges and a lack of optimized scanners for Cytology specimens. Despite the challenges in scanning whole slide images of cytology specimens, there have been many studies evaluating CP to create decision-support tools in Cytopathology. Among different Cytology specimens, thyroid fine needle aspiration biopsy (FNAB) specimens have one of the greatest potentials to benefit from machine learning algorithms (MLA) derived from digital images. Several authors have evaluated different machine learning algorithms focused on thyroid cytology in the past few years. The results are promising. The algorithms have mostly shown increased accuracy in the diagnosis and classification of thyroid cytology specimens. They have brought new insights and demonstrated the potential for improving future cytopathology workflow efficiency and accuracy. However, many issues still need to be addressed to further build on and improve current MLA models and their applications. To optimally train and validate MLA for thyroid cytology specimens, larger datasets obtained from multiple institutions are needed. MLAs hold great potential in improving thyroid cancer diagnostic speed and accuracy that will lead to improvements in patient management.

Contents

Introduction

Pathology is facing revolutionary shifts with advancements and applications of slide digitization and machine learning-based algorithms for image analysis. Although the light microscope is still the most prevalent method used by pathologists to make a morphologic diagnosis, this method is slowly being impacted by the advent of digital pathology. Digital images of glass slides (Fig. 1) offer several advantages over analog review, including easier image sharing for peer consultation, remote review of images, decreased physical storage space, and image analysis, while the major disadvantage is the initial cost required to digitize the images. The digitization of glass slides was particularly helpful during the COVID-19 pandemics when pathologists had to quarantine. Pathology departments validated their workflow and were able to adopt a digital sign-out during this period. Pathological diagnosis could be made remotely without the need for the pathologist to use a microscope or commute to the laboratory.[1–3]

More importantly, the digitization of glass slides creating whole slide images (WSI) has allowed the development of the field of Computational Pathology (CP). It uses the massive amount of data embedded in the Pathology digital images to facilitate computer-assisted diagnostics.[4] The images

* Corresponding author at: Department of Pathology and Laboratory Medicine, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA.
E-mail address: lino@mskcc.org (O. Lin).

http://dx.doi.org/10.1016/j.jpi.2023.100309
Received 24 February 2023; Received in revised form 24 March 2023; Accepted 27 March 2023
Available online 1 April 2023

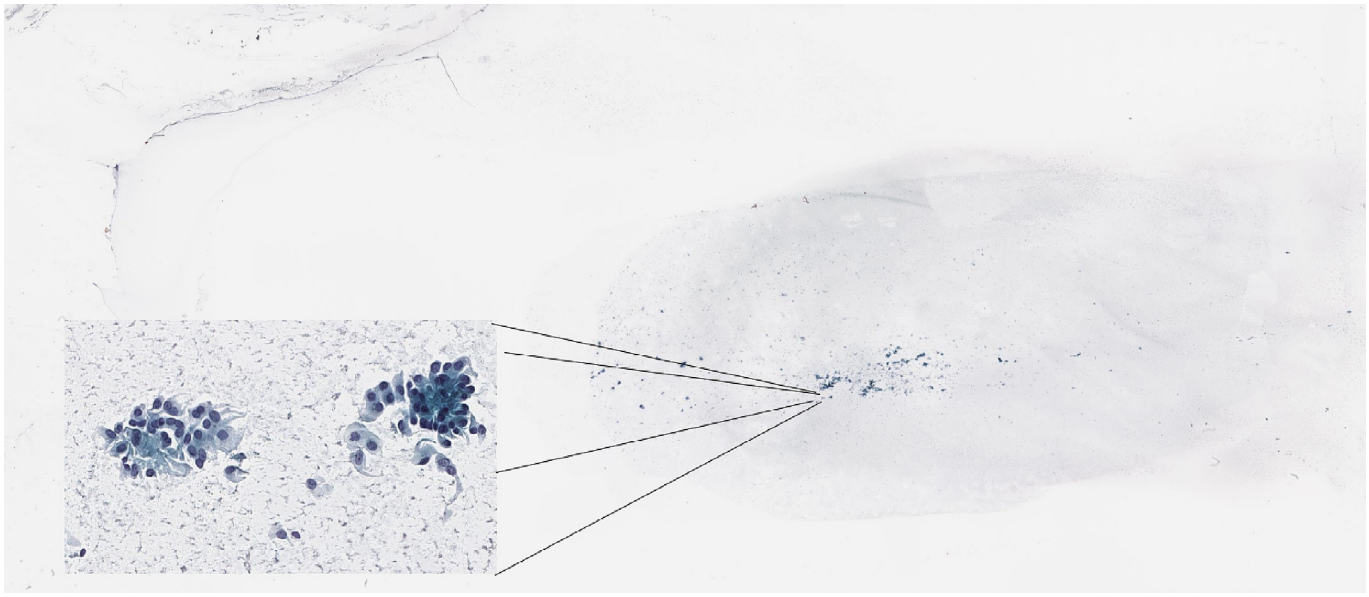

Fig. 1. Whole slide image of a fine needle aspiration biopsy of a thyroid papillary carcinoma smear.

are converted into millions of pixels that can be subjected to image analysis, allowing the development of image-based screening algorithms in Pathology similar to algorithms already implemented in Radiology.[5] CP has an enormous potential to revolutionize the way pathologic diagnoses are made but its development is not without challenges. Digital images used to develop the algorithms have to be of good quality, reliable, and require validation of both scanning process and digital images and reviewed by a pathologist.[6–8] The variability of morphologic patterns that can be seen in Pathology is nearly infinite which makes it difficult for computer algorithms to independently process without the use of a large dataset.[9–11] The success of machine learning in Pathology requires properly annotated data, ideally by several experts in the field to decrease potential selection bias. This data can be used initially to train machine learning algorithms (MLA), which are critical in the development of clinically useful computer-assisted diagnostics. Only by addressing these challenges, CP will broaden its application in digital pathology for clinical purposes and impact patient management.

Currently, digital image-based applications that have received FDA Breakthrough Device Designation are focused on tissue specimens. These MLA are primarily used to screen prostate and breast adenocarcinoma.[12–14] Conversely, no digital image applications using Cytology specimens have been approved by the FDA. This disparity can be attributed to the differences between cytology and surgical pathology specimens. The development of AI assisted algorithms using Cytology digital images has been much more limited due to technical challenges and a lack of optimized scanners for Cytology specimens. Whole slide image scanners available are optimized to create digital images from Hematoxylin and Eosin (H&E) formalin-fixed paraffin-embedded (FFPE) tissue specimens. The images from FFPE tissue specimens can be completely reproduced with only one scanning layer as the tissue sections are thin (approximately 4–5 μm). Cytology specimens contain whole cells and might contain clusters that might be several cells thick. These intact cells and clusters require multiple layers of scanning to reproduce a complete image of the Cytology slide. The need to scan multiple layers leads to increased scanning time, storage space, and impact image quality.[15]

Despite the challenges in scanning WSI of cytology specimens, there have been many studies evaluating CP to create decision support tools in Cytopathology.[16] The types of Cytology specimens that have been evaluated by CP include cervical vaginal, urine, effusions cytopathology, breast FNAB, thyroid FNAB, and lymph node FNAB specimens among others.[17] Among different cytology specimens, thyroid FNAB specimens have one of the greatest potentials to benefit from MLA derived from digital images.

MLA in thyroid specimens

Currently, thyroid FNAB is a common diagnostic procedure with an estimated 600 000 biopsies performed each year in the USA. It is part of the standard of care in the evaluation of thyroid nodules that meet certain radiologic criteria according to the American Thyroid Association guidelines.[18] While thyroid FNAB is an integral part of the workup of a thyroid nodule, there is great variability in the assessment of thyroid cytology specimens among different pathologists and institutions. The reported sensitivity in the literature ranges from 68% to 98% while the specificity ranges from 56% to 100%. In addition, approximately 15%–30% of the FNAB are classified as indeterminate of which 25% are later found to be malignant.[19] It is in this setting that MLA could be used to increase accuracy and help in the standardization of the diagnosis of thyroid FNAB specimens.

The evaluation of MLA using thyroid FNAB specimens started more than a decade ago.[20,21] One of the earliest studies was performed by Karakitsos et al,[21] who attempted to classify benign and malignant follicular and Hurtle cell lesions using a neural network evaluating geometric and densimetric nuclear features. Their method was able to classify correctly 91.12% and 89.64% of the benign and malignant nuclei, respectively, with an overall accuracy of 90.61% on test data. The two-class classifier was able to distinguish benign from malignant lesions with 94.9% sensitivity and 98.9% specificity. Gopinath et al published several studies using images taken from the Papanicolaou Society of Cytopathology online atlas. Statistical textural features from the images were input into one or more classifiers, including decision tree, K nearest neighbor, Elman neural network, and support vector machine.[22–24] As the classifier was trained by and tested on textbook images, it was unlikely that the results would be widely applicable to real-world applications because the ground truth to train the system was too narrow. This is an example of overfitting, which most commonly arises in the context of small sample size or when the training data have an insufficient variation to capture the variation seen in practice and highlights the need for a large dataset for training. All these studies were performed using selected photographic images from thyroid FNAB specimens using a limited number of specimens.

Current application of MLA for cytopathology

Most recent studies have taken advantage of technological advances in artificial intelligence (AI) and digital image acquisition. These authors have focused on the use of different MLAs of thyroid FNAB specimens. Sanyal et al reported in 2018 the use of an artificial neural network developed in

the Python programming language.[25] In the training phase, 186 microphotographs from smears of papillary carcinomas and 184 microphotographs from smears of other thyroid lesions were used. Performance was evaluated with a test set of 174 microphotographs (66 non-papillary carcinomas and 21 papillary carcinomas). Using the FNAB diagnosis as the gold-standard, results showed good sensitivity (90.48%), specificity (83.33%), negative-predictive value (96.49%), and diagnostic accuracy (85.06%). However, vague papillary formations by benign follicular cells were identified incorrectly as papillary carcinoma, revealing the limitations of the MLA used. Although this is a relatively small study using relatively small training and test datasets, this was one of the first reports to demonstrate promising results using Python programming language to create a MLA focused on thyroid cytopathology.

MLA using WSI was described by Dov et al, who developed and validated a MLA to screen ROIs from thyroid │FNAB WSIs of smear preparations.[26,27] Examining 1 representative slide from each of 109 thyroid FNABs, the MLA was able to screen and select a subset of the 100 most informative ROIs from each WSI and form an image gallery. To train the screening MLA to identify follicular cells, the investigators used a training set of labeled WSIs containing ROIs annotated for follicular cells. They also used a random selection of areas to train the MLA to identify negative regions of interest (NROIs). To assess the adequacy of these ROIs, the investigators had a cytopathologist first read all the glass slides. After 117 days to eliminate recall, the same cytopathologist reread the slides using the image gallery of ROIs that were selected by the MLA instead of the glass slide. This method allowed a comparison between the MLA gallery of ROIs and the entire original slide with respect to a cytopathologist interpretation. One potential limitation of this study is that only 1 cytopathologist was involved in the analysis.

This MLA was used by Elliot Range et al in the development of a MLA aimed at predicting malignancy in thyroid FNAB specimens by utilizing non-preprocessed WSIs.[28] They used 908 different WSIs of FNAB specimens as their input data and the MLA was broken into 2 different convolutional neural networks (CNNs) to complete these tasks. They first used a screening method similar to the one described by Dov et al followed by a second MLA designated as "classifier" to predict the Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) category for each thyroid nodule based on the WSI. To train the classifier MLA, the investigators employed 1000 different ROIs to train the MLA on each TBSRTC category. The MLA was then used to generate first local predictions of final pathology for each

Table 1
Summary of methods evaluated.

Author / Method	Features	Strengths/Performance	Weaknesses/Limitations
Karakitsos et al[21]/neural network MLA	Evaluated geometric and densimetric nuclear features in benign and malignant follicular and Hurthle cell lesions	The 2-class classifier was able to distinguish benign from malignant lesions with 94.9% sensitivity and 98.9% specificity.	Reliable discrimination of the cytologic types of the lesions was not obtained.
Gopinath et al[22–24]/classifiers	Thyroid cell regions are extracted from the auto-cropped sub-image by implementing mathematical morphology segmentation method. Subsequently, statistical features are extracted by 2-level wavelet decomposition based on texture characteristics of the thyroid cells. After that, decision tree (DT), k-nearest neighbor (k-NN), Elman neural network (ENN), and support vector machine (SVM) classifiers are used separately to classify thyroid nodules into benign and malignant. The 4 individual classifier outputs are then fused together using majority voting rule and linear combination rules.	Highest diagnostic accuracy (96.66 %) distinguishing benign form malignant lesions was obtained by multiple classifier fusion with majority voting rule and linear combination rules.	Images were taken from an online atlas. Textbook images make it unlikely that the results would be widely applicable to real-world applications because the ground truth to train the system was too narrow.
Sanyal et al[25]/artificial neural network	Evaluated microphotographs from smears of papillary carcinomas and other thyroid lesions. Artificial neural network (ANN) developed using Python programming language.	The ANN sensitivity was 90.48% with a specificity of 83.33% with an overall diagnostic accuracy of 85.06%.	Study focused on papillary carcinomas. Vague papillary formations by benign follicular cells were identified incorrectly as papillary carcinoma.
Dov et al[27]/machine learning algorithm	Employed machine learning algorithm that can identify regions of interest (ROIs) on thyroid fine-needle aspiration biopsy whole slide images (WSI) of smear preparations.	Almost perfect concordance between cytopathologist diagnosis of WSI and diagnosis using MLA generated ROI image gallery	Small test set Inherent bias in this study, as the training of the screening and classifier MLAs was partially based on supervised learning performed by the study reviewer. Inability to evaluate particular follicular groups in the context of surroundings, view additional ROIs, and identify subtle examples of colloid/lymphocytes located between ROIs.
Elliot Range et al[28]/machine learning algorithm	Machine learning algorithm (MLA) based on 2 convolutional neural networks (CNNs), one to identify follicular groups (screening MLA) and the other to simultaneously predict the TBSRTC category and the final pathology (classifier MLA)	Achieved a sensitivity of 92% and specificity of 90.5% and areas under the curve for the prediction of malignancy by MLA was 0.932	Selection bias
Hirokawa et al[29]/EfficientNetV2-L.	Used EfficientNetV2 convolutional neural network–based model to distinguish a benign from a malignant lesion.	The sensitivity, specificity, positive-predictive value, and negative-predictive values were 94.4%, 15.4%, 60.7%, and 66.7%, respectively. Precision-recall AUC was over 0.95.	Precision-recall AUC was worse for poorly differentiated thyroid cancer (0.49) and medullary thyroid cancer (0.91).
Yao et al[30]/supervised machine learning-based digital image analysis	Used ImageJ and Python scikit-learn models to evaluate cases originally classified as atypical and later diagnosed as benign or follicular adenomas on histologic sections.	The area under the curve for receiver operating characteristics was 0.75 (0.74–0.82) for the model based on low-power images and 0.74 (0.69–0.79) for the model based on high-power images.	Low number of cases in the dataset. Cytology diagnosis was rendered by a single cytopathologist. Scope of the study was limited to lesions that were later diagnosed as follicular adenoma or benign thyroid nodules.

ROI, and then to aggregate these local predictions into a single and final pathology prediction for the entire WSI using multiple instance learning. This MLA was able to identify and predict malignancy in final surgical resection, with a sensitivity of 92% and specificity of 90.5%. The area under this receiver operating curve (AUC) was 0.932, which is comparable to the AUC of 0.931 achieved by the cytopathologist. According to this study, when aggregating medical record information into the MLA, the MLA was able to perform even better, with the combined AUC increasing to 0.962. In summary, this study demonstrated that a novel MLA was able to: (1) screen a WSI to create a gallery of significant ROIs, and (2) classify the TBSRTC status and malignancy status for a WSI, both with excellent sensitivity and specificity equivalent to that of a human cytopathologist.

Hirokawa et al reported the results of a study demonstrating the efficiency and accuracy of their AI-image based analysis in identifying thyroid lesions with FNAB.[29] This group used 148 395 images of FNAB smear slides from 393 thyroid nodules to train and validate an AI image-classification system called EfficientNetV2-L. The dataset was divided into 5 training and validation sets for 5-fold cross-validation (80%) and a separate test set (20%). The investigators reported that EfficientNetV2-L had a precision-recall AUC of over 0.95. The performance was notably worse for poorly differentiated thyroid cancer (0.49) and medullary thyroid cancer (0.91). Poorly differentiated thyroid cancer had the lowest recall (35.4%) and the system had difficulty distinguishing it from papillary thyroid carcinoma, medullary thyroid carcinoma, and follicular thyroid carcinoma. However, when given 35 nodules that were originally reported as atypia of undetermined significance (AUS), 16 were estimated correctly. For determining benign or malignant nodules, the sensitivity, specificity, positive-predictive value, and negative-predictive values were 94.4%, 15.4%, 60.7%, and 66.7%, respectively. Of the 11 papillary thyroid cancers, 8 were correctly estimated and the remaining 3 were speculated to be follicular thyroid cancers. This study, therefore, demonstrates promising performance by this MLA, but also points out its limitations and deficiencies, including reduced specificity for AUS nodules and worse performance for poorly differentiated carcinomas. One limitation of this study was that a cytopathologist was required to select the photographs used for the image set from non-degenerative and representative areas. This deficiency could be addressed by designing MLAs to select ROIs as described above by Dov et al. The image set used in the study was also derived from cytologically typical cases from a single institution. AI should optimally be trained on cases from multiple institutions to teach it a more accurate interpretation of atypical cases, degenerate cells, and differences in smear and staining.

In another study, Yao et al evaluated a feature engineering and supervised machine learning-based digital image analysis method using ImageJ and Python scikit-learn.[30] They focused on cases originally classified as atypical and were later diagnosed as benign or follicular adenomas on histologic sections. The method was trained and validated on 400 low power (100x magnification) and 400 high power (400x magnification) images generated from 40 thyroid FNAB cases Thinprep slides. The area under the curve (AUC) for receiver operating characteristics (ROC) was 0.75 (0.74–0.82) for model based on low-power images and 0.74 (0.69–0.79) for the model based on high-power images. The authors suggested that the MLA performed better than a cytopathologist to subclassify atypical follicular lesions. The limitations the study included the low number of cases in the dataset, the cytology diagnosis was rendered by a single cytopathologist and the scope of the study was limited to lesions that were later diagnosed as follicular adenoma or benign thyroid nodules. A summary of the different studies evaluated is listed in Table 1.

Conclusions

The development and implementation of AI in thyroid cytology have greatly advanced in just the past few years, bringing new insights and potential for improving future cytopathology workflow efficiency and accuracy. However, many issues still need to be addressed to further build on and improve current AI models and their applications. The image quality needs to be standardized, including the resolution, the number of levels scanned, and the distance between each level. The standardization is necessary for real comparison of the different methods. To optimally train and validate MLA, larger datasets obtained from multiple institutions are needed. The continued development of MLAs capable of selecting ROIs and performing more consistent and accurate diagnoses will enable cytopathologists to focus their attention on the ROIs, allowing more accurate and potentially faster interpretations. The field will undoubtedly continue to improve current MLAs to achieve better AUCs for sensitivity, specificity, and precision recall. Importantly, the development of MLAs is not restricted to thyroid cytopathology. MLAs are now being created that aggregate thyroid ultrasound imaging and thyroid cancer clinical data in determining a diagnosis of thyroid cancer.[31] Future MLAs may integrate cytopathology, radiology, and clinical information, creating an even more powerful algorithm. Ultimately, with further improvement, AI and MLAs hold great future promise in improving thyroid cancer diagnostic speed and accuracy.

Funding/Disclosure

This study was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA008748.

Declaration of interests

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Oscar Lin reports a relationship with Hologic Inc that includes: consulting or advisory. Oscar Lin reports a relationship with Janssen Biotech Inc that includes: consulting or advisory.

References

1. Hanna MG, et al. Validation of a digital pathology system including remote review during the COVID-19 pandemic. Mod Pathol 2020;33(11):2115–2127.
2. Scarl RT, Parwani A, Yearsley M. From glass-time to screen-time: a pathology resident's experience with digital sign-out during the Coronavirus 2019 pandemic. Arch Pathol Lab Med 2021;145(6):644–645.
3. Lujan GM, et al. Digital pathology initiatives and experience of a large academic institution during the Coronavirus disease 2019 (COVID-19) pandemic. Arch Pathol Lab Med 2021;145(9):1051–1061.
4. Parwani AV, Amin MB. Convergence of digital pathology and artificial intelligence tools in anatomic pathology practice: current landscape and future directions. Adv Anat Pathol 2020;27(4):221–226.
5. Dikici E, et al. Integrating AI into radiology workflow: levels of research, production, and feedback maturity. J Med Imaging (Bellingham) 2020;7(1), 016502.
6. Goacher E, et al. The diagnostic concordance of whole slide imaging and light microscopy: a systematic review. Arch Pathol Lab Med 2017;141(1):151–161.
7. Saco A, et al. Validation of whole-slide imaging for histolopathogical diagnosis: current state. Pathobiology 2016;83(2-3):89–98.
8. Hanna MG, et al. Whole slide imaging equivalency and efficiency study: experience at a large academic center. Mod Pathol 2019;32(7):916–928.
9. Colling R, et al. Artificial intelligence in digital pathology: a roadmap to routine use in clinical practice. J Pathol 2019;249(2):143–150.
10. Acs B, Rantalainen M, Hartman J. Artificial intelligence as the next step towards precision pathology. J Intern Med 2020;288(1):62–81.
11. Rashidi HH, et al. Artificial intelligence and machine learning in pathology: the present landscape of supervised methods. Acad Pathol 2019;6.p. 2374289519873088.
12. da Silva LM, et al. Independent real-world application of a clinical-grade automated prostate cancer detection system. J Pathol 2021;254(2):147–158.
13. Raciti P, et al. Novel artificial intelligence system increases the detection of prostate cancer in whole slide images of core needle biopsies. Mod Pathol 2020;33(10):2058–2066.
14. Sandbank J, et al. Validation and real-world clinical application of an artificial intelligence algorithm for breast cancer detection in biopsies. NPJ Breast Cancer 2022;8(1):129.
15. Kim D, et al. 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(8):630–639.
16. Pouliakis A, et al. Artificial neural networks as decision support tools in cytopathology: past, present, and future. Biomed Eng Comput Biol 2016;7:1-18.
17. Alrafiah AR. Application and performance of artificial intelligence technology in cytopathology. Acta Histochem 2022;124(4), 151890.
18. Haugen BR. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: what is new and what has changed? Cancer 2017;123(3):372–381.
19. Kezlarian B, Lin O. Artificial Intelligence in Thyroid Fine Needle Aspiration Biopsies. Acta Cytol 2021;65(4):324–329.

20. Cochand-Priollet B, et al. Discriminating benign from malignant thyroid lesions using artificial intelligence and statistical selection of morphometric features. Oncol Rep 2006;15 Spec no:1023–1026.
21. Karakitsos P, et al. Learning vector quantizer in the investigation of thyroid lesions. Anal Quant Cytol Histol 1999;21(3):201–208.
22. 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(5):653–662.
23. Gopinath B, Shanthi N. Support Vector Machine based diagnostic system for thyroid cancer using statistical texture features. Asian Pac J Cancer Prev 2013;14(1):97-102.
24. Gopinath B, Shanthi N. Computer-aided diagnosis system for classifying benign and malignant thyroid nodules in multi-stained FNAB cytological images. Australas Phys Eng Sci Med 2013;36(2):219–230.
25. Sanyal P, et al. Artificial intelligence in cytopathology: a neural network to identify papillary carcinoma on thyroid fine-needle aspiration cytology smears. J Pathol Inform 2018;9:43.
26. Dov D, et al. Weakly supervised instance learning for thyroid malignancy prediction from whole slide cytopathology images. Med Image Anal 2021;67, 101814.
27. Dov D, et al. Use of machine learning-based software for the screening of thyroid cytopathology whole slide images. Arch Pathol Lab Med 2022;146(7):872–878.
28. Elliott Range DD, et al. Application of a machine learning algorithm to predict malignancy in thyroid cytopathology. Cancer Cytopathol 2020;128(4):287–295.
29. Hirokawa M, et al. Application of deep learning as an ancillary diagnostic tool for thyroid FNA cytology. Cancer Cytopathol 2023 Apr;131(4):217–225.
30. Yao K, et al. A study of thyroid fine needle aspiration of follicular adenoma in the “atypia of undetermined significance” Bethesda category using digital image analysis. J Pathol Inform 2022;13, 100004.
31. Xi NM, Wang L, Yang C. Improving the diagnosis of thyroid cancer by machine learning and clinical data. Sci Rep 2022;12(1):11143.