



Digital Versus Optical Diagnosis of Follicular Patterned Thyroid Lesions

Ayat Aloqaily^{1,2} · Antonio Polonia^{1,5} · Sofia Campelos¹ · Nusaiba Alrefae^{1,3} · Joao Vale¹ · Ana Caramelo¹ · Catarina Eloy^{1,4,5}

Received: 2 September 2020 / Accepted: 21 October 2020 / Published online: 31 October 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Objectives To study the concordance between pathologists in the diagnosis of follicular patterned thyroid lesions using both digital and conventional optical settings.

Material and Methods Five pathologists reviewed 50 hematoxylin and eosin-stained slides of follicular patterned thyroid lesions using both digital (the D-Sight 2.0 scanner and navigator viewer) and conventional optical instruments with washout interval time.

Results The mean concordance rate with the ground truth (GT) was similar between conventional optical and digital observation (83.2 and 85.2%, respectively). The most frequent reason for diagnostic discordance with GT on both systems was the evaluation of nuclear features (69.1% for conventional optical observation and 59.4% for digital observation). The intraobserver diagnostic concordance mean was 86.8%. Time for digital observation (mean time per case = 2.9 ± 0.8 min) was higher than that for conventional optical observation (mean time per case = 2.0 ± 0.7 min). Interobserver correlation of measurements was higher in the digital observation than the conventional optical observation.

Conclusion Conventional optical and digital observation settings showed a comparable accuracy for the diagnosis of follicular patterned thyroid nodules, as well as substantial intraobserver agreement and a significant improvement in the reproducibility of the measurements that support the use of digital diagnosis in thyroid pathology. The origins underlying the variability of the diagnosis were the same in both conventional optical microscopy and digital pathology systems.

Keywords Digital pathology · Papillary thyroid carcinoma · Noninvasive follicular thyroid neoplasm with papillary-like nuclear features · Thyroid · Thyroid cancer

Introduction

Digital systems in the pathologist workstation offer a different environment for evaluation and reporting in comparison with the one provided in the setting of traditional optical microscopy. The benefit of having a digital pathology

workflow classically includes better ergonomics, immediate access to archive slides, distance reporting, and consultations [1]. Pathologists are still discovering new ways to take advantage of digital systems, searching for achievements in their work quality as well as in the quality of the input they provide for patient orientation. Numerous validation studies in digital pathology had confirmed its value as an observation tool and proved that it can be interpreted with comparable diagnostic accuracy to the conventional optical tools [2]. However, a longer time to diagnosis remains a significant barrier for its full adoption in routine diagnosis [3].

For thyroid pathology, in particular, digital microscopy potentials have not yet been fully explored, and there is some skepticism regarding its use in contrast with the conventional optical method in terms of diagnostic accuracy and efficiency. The skepticism that involves the use of digital systems for the diagnosis of malignancy in thyroid nodules is probably due to low reproducible morphological criteria

✉ Ayat Aloqaily
Ayat.aloqaily@gmail.com

¹ Institute of Molecular Pathology and Immunology, Ipatimup Diagnostics, University of Porto (IPATIMUP)/i3S, Rua Júlio Amaral de Carvalho 45, 4200-135 Porto, Portugal

² King Abdullah University Hospital (KAUH), Jordan, University of Science and Technology (JUST), Irbid, Jordan

³ Kuwait Institute for Medical Specializations, Kuwait, Kuwait

⁴ Medical Faculty, Porto University, Porto, Portugal

⁵ Instituto de Investigação E Inovação Em Saúde (i3S), University of Porto, Porto, Portugal

used that includes the appreciation of invasion (capsular and vascular) and the presence of papillary thyroid carcinoma type nuclear features (PTC-nuclei) [4–6].

These criteria underlie the distinction between benign follicular lesions, papillary thyroid carcinoma (PTC), follicular thyroid carcinoma (FTC), and noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP), a neoplasm with very low malignant potential [7]. The assessment of nuclei is performed at high magnification and considers nuclear shape and size, nuclear membrane irregularities, and chromatin features [8]. The nuclear alterations can be diffuse, multifocal, or distributed in a hazardous “sprinkling” pattern [9], and can be more or less evident among different areas of the nodule. Vascular invasion, capsular invasion, and invasion of the thyroid parenchyma are nowadays more relevant for the diagnosis in the distinction between benign and malignant lesions, since the introduction of NIFTP as an entity [8, 10].

The evaluation of these complex morphological criteria used for the diagnosis of malignancy in thyroid nodules has an already known inter and intraobserver discordance that is quite evident in the diagnosis of the follicular variant of PTC [4]. Even the experts in endocrine pathology showed poor agreement among themselves in the identification of the diagnostic nuclear features of PTC [11, 12]. As such, the introduction of digital solutions may eventually affect the degree of concordance in the diagnosis of follicular patterned thyroid lesions. In this work, we aim to study the concordance between pathologists in the diagnosis of follicular patterned lesions using both the digital and the conventional optical settings.

Materials and Methods

A series of 48 consecutive thyroid surgical excisional specimens with follicular patterned lesions was retrieved from Ipatimup Diagnostics archives from August 2017 to February 2019. The series included 50 hematoxylin and eosin (HE) stained slides (two cases had two slides) representative of follicular patterned thyroid lesions that encompassed follicular adenomas/adenomatous nodules ($n = 10$), nodular transformation of lymphocytic thyroiditis ($n = 9$), NIFTPs ($n = 3$), PTCs ($n = 23$) (including papillary microcarcinomas and PTCs larger than 10 mm with predominant follicular pattern), and FTCs ($n = 5$). The lesions were classified according to the fourth edition of the WHO [8]. Lesions with NIFTP features and 10 mm or less in the greatest dimension were classified as papillary microcarcinomas [10, 13]. The nodules labeled as “malignant” for staging purposes

were FTCs and PTCs. The ground truth (GT) diagnosis was considered as the diagnosis of the majority of pathologists under the conventional light microscope (at least three pathologists).

Five pathologists, from three different countries (P1–P5), reviewed all the slides using both digital and conventional optical instruments, and provided the diagnosis and the largest dimension of the malignant nodules, along with the distance to the nearest surgical margin. General experience in endocrine pathology was 8, 6, 6, 5, and 4 years for P1, P2, P3, P4, and P5, respectively. Regarding experience in digital diagnosis, P1 and P2 pathologists have more than 5 years of experience, while P3, P4, and P5 have less than 1 year of experience. All pathologists were equally trained to use the WSI system before the study. Pathologists were not aware of the diagnostic types included in the series, except the pathologist responsible for the selection of the cases (P1). P1 had a washout period of 4 weeks after glass slide selection and before starting the study.

The slides were scanned at 20 \times magnification lens in the D-Sight 2.0 scanner (Menarini) obtaining 50 whole slide images (WSI) in the .gxp format with a resolution of 0.5 micron per pixel. All pathologists started with the digital observation of the WSI at Ipatimup using the D-Sight Navigator viewer in a 20" DELL touch screen (1600 \times 900) and a regular mouse for navigation (Fig. 1). Then, conventional optical observation with a traditional light microscope (Leica DM200 LED) was also performed at Ipatimup after a washout period of at least 10 working days for all pathologists [14]. The conventional optical observation was done separately and on pre-assigned dates to prevent time loss in slide rotation and to confirm an adequate washout period. Time of diagnosis was defined as the time needed to make the diagnosis, the measurements, and to enter the results in an Excel worksheet, excluding break times and times of external distraction. Time assessment was done similarly for both methods by each pathologist at the time of observation.

Statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS) version 25.0 for Windows. The Pearson's Chi-squared (χ^2) test (or the Fisher's exact test, if appropriate) and McNemar (MN) test were used for comparison of qualitative variables, and the *t* test, the Mann–Whitney U-test (MW), the Wilcoxon (WC) test and the Pearson's correlation coefficient (PCC) were used for quantitative variables. The level of significance was set at $p < 0.05$. Concordance rates (intra- and inter-observer) were evaluated with kappa statistics. The Landis and Koch classification was used to interpret the values: no agreement to slight agreement (< 0.20), fair agreement (0.21–0.40), moderate agreement (0.41–0.60), substantial agreement (0.61–0.80), and excellent agreement (> 0.81).

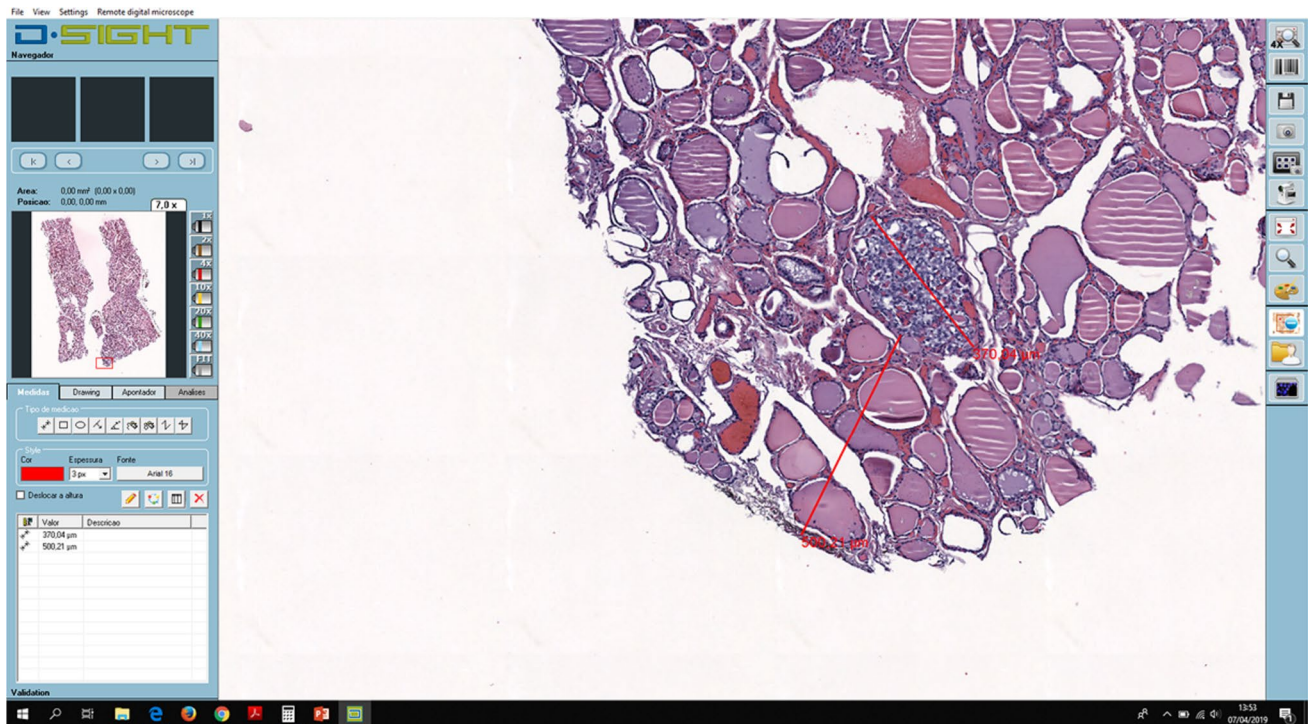


Fig. 1 D.Sight navigator viewer. Size of the malignant nodule and its distance from the nearest margin were measured

Results

The series included 48 patients with predominantly follicular pattern thyroid nodules, 13 males (27.1%), and 35 females (72.9%), with ages ranging from 19 to 80 years, and a median age at the time of diagnosis of 49.5 years.

Concerning the diagnosis of the thyroid nodules, the mean concordance rate with the GT in the conventional optical observation was 83.2% (68–100%), similar to the mean concordance rate with the GT in the digital observation of 85.2% (76–96%) (MN $p = 0.209$). The results of the diagnostic concordance rates with the GT in both conventional optical and digital observations are summarized in Table 1.

The most frequent type of diagnostic discordance with GT on both systems was the evaluation of nuclear features that occurred in 29 diagnoses (11.6%) in conventional optical observations and in 22 diagnoses in digital observations (8.8%). In the conventional optical observation, the nuclei features tended to be overestimated, with consequent overdiagnosis of PTC and NIFTP (40.5% of discordant cases). In the digital observation, comparable percentages of over and underestimation of PTC-nuclei were observed (27.0 and 32.5%, respectively). The evaluation of features that underlie the distinction between PTC and NIFTP was the second more frequent type of discordance in both types of observations (2.0 and 2.8% for conventional optical and digital observations, respectively). In contrast, the evaluation of

invasion appears to be a minor source of discordant diagnosis in both types of observation (0.4% of diagnosis for conventional optical and digital observations). The causes of diagnostic discordance with GT are summarized in Table 2.

When comparing the digital with the conventional optical diagnosis for each pathologist, the intraobserver diagnostic concordance rate varied from 80 to 96%, with a mean value of 86.8% (mean $k = 0.80$) (Table 1). Most of the intraobserver diagnostic discordances were attributable to the assessment of nuclear features (8.4%). In cases with discordance between conventional optical and digital appreciation of nuclear features, the presence of PTC-nuclei was more frequently detected in the conventional optical setting (42.4%) than in the digital setting (21.2%). Causes of intraobserver diagnostic discordance are listed in Table 3.

Pathologists needed more time for digital observation (mean time per case = 2.9 ± 0.8 min) than for conventional optical observation (mean time per case = 2.0 ± 0.7 min). The digital observation required, on average, an additional 0.9 min per case, corresponding to 33.2% of time to add on the conventional optical observation time (Table 4).

The additional mean time needed for digital diagnosis was similar for the experienced group of pathologists (P1 and P2 = 0.9 min) and the less experienced group (P3, P4, and P5 = 0.93 min). The percentage of extra time consumed for digital diagnosis was 42.9% for the experienced pathologists and 27.2% for the less experienced pathologists.

Table 1 Summary of diagnoses by pathologists (P1, P2, P3, P4 and P5) and diagnostic concordance rates

Diagnoses	GT	P1		P2		P3		P4		P5	
		Optical	Digital	Optical	Digital	Optical	Digital	Optical	Digital	Optical	Digital
FA/AD	10	10	11	8	11	20	18	8	11	9	12
LT	9	9	9	8	9	6	8	5	8	7	9
NIFTP	3	3	4	6	4	1	3	3	1	0	0
PTC	23	23	21	22	22	17	17	30	27	34	29
FTC	5	5	5	6	4	5	3	2	2	0	0
Others	0	0	0	0	0	1	1	2	1	0	0
Concordance with the GT (k statistics)	–	100% (1.00)	96% (0.94)	92% (0.89)	90% (0.86)	74% (0.64)	82% (0.76)	82% (NC)	82% (0.73)	68% (NC)	76% (NC)
Intraobserver concordance (k statistics)	–	96% (0.95)	–	86% (0.81)	–	80% (0.72)	–	84% (0.75)	–	88% (0.78)	–

FA/AD follicular adenoma/adenomatous nodule, LT nodular transformation of lymphocytic thyroiditis, NIFTP non-invasive follicular tumour with papillary like nuclear features, PTC papillary thyroid carcinoma, FTC follicular thyroid carcinoma, GT ground truth, NC not calculated

The size of the malignant lesions measured between 0.5 and 18 mm (mean = 7 ± 3.6 mm) in the conventional optical observation, and between 1.4 and 14.4 mm (Mean = 7 ± 3.3 mm) in the digital observation. We observed a higher interobserver correlation of this measurement in the digital observation (PCC from 0.970 to 0.997; $p < 0.001$) than in the conventional optical observation (PCC from 0.745 to 0.942; $p < 0.001$). The distance of the malignant lesions to the nearest surgical margin measured between 0 and 10 mm (1.75 ± 1.36) in the conventional optical observation, and between 0 and 10.2 mm (1.71 ± 2.26) in the digital observation. We also observed a higher interobserver correlation of this measurement in the digital observation (PCC from 0.882 to 0.989; $p < 0.001$) than in the conventional optical observation (PCC from 0.547 to 0.886; $p < 0.001$). Additionally, the intraobserver correlation of both measurements was higher in the size of the malignant lesions (PCC = 0.917; $p < 0.001$) than in the distance of the malignant lesion to the nearest surgical margin (PCC = 0.725; $p < 0.001$).

Discussion

The implementation of digital solutions represents new possibilities in the way of performing diagnosis in pathologists' daily practice. Moving forward with this implementation requires a strong effort to replicate in the digital setting the conditions that allow the pathologists to do the same diagnoses as in the conventional optical setting. In the context of thyroid lesions, namely those with a follicular pattern, the conventional optical diagnosis is, per se, a major source of disagreement with low reproducibility rates in the diagnosis of malignancy [15–17].

In the present work, five pathologists diagnosed a series of follicular patterned thyroid lesions using both conventional optical and digital observations with washout interval time. The diagnostic concordance with the ground truth was similar in both conventional optical and digital settings, indicating that the method is not a major additional factor in contributing to the diagnosis variability in thyroid lesions.

Previous reports on concordance using the light microscopy showed similar to lower agreement rates. In Elsheikh et al. [11] study, the majority of experts in thyroid pathology had an interobserver agreement of 53% for malignancy, and 40% for follicular variant PTC diagnosis. In the same study, intraobserver agreement for diagnosing malignancy was 60–100%, and for diagnosing follicular variant PTC was 17–100%.

The major source of diagnostic discordance with the ground truth was related to the classification of the nuclei, in both conventional optical and digital settings, as already reported [11, 12]. It was kept in mind that consideration of the ground truth as a standard reference here, as well as in

Table 2 Causes for diagnostic discordance with ground truth

Evaluated features	Optical observation		Digital observation	
	Distribution of discordance events according to type % (n)	Percentage of discordances per total number of events % (n)	Distribution of discordance events according to type % (n)	Percentage of discordances per total number of events % (n)
Nuclear features	69.1% (29)	11.6% (29)	59.4% (22)	8.8% (22)
<i>Overestimation of PTC-nuclei</i>	40.5% (17)	6.8% (17)	27.0% (10)	4.0% (10)
<i>Underestimation of PTC-nuclei</i>	28.6% (12)	4.8% (12)	32.5% (12)	4.8% (12)
NIFTP/PTC distinction	11.9% (5)	2.0% (5)	18.9% (7)	2.8% (7)
Inflammation	9.5% (4)	1.6% (4)	13.5% (5)	2.0% (5)
Invasion	2.4% (1)	0.4% (1)	2.7% (1)	0.4% (1)
Others	7.1% (3)	1.2% (3)	5.4% (2)	0.8% (2)
Total	100.0% (42)	16.8% (42)	100.0% (37)	14.8% (37)

Table 3 Causes for intraobserver diagnostic discordancy

Evaluated features	Proportion of type of discordancy (n)	Proportion of discordancy per total number of events (n)
Nuclei features	63.6% (21)	8.4% (21)
<i>Optical evaluation of PTC-nuclei</i>	42.4% (14)	5.6% (14)
<i>Digital evaluation of PTC-nuclei</i>	21.2% (7)	2.8% (7)
NIFTP/PTC distinction	6.1% (2)	0.8% (2)
Inflammation	15.1% (5)	2.0% (5)
Invasion	9.1% (3)	1.2% (3)
Others	6.1% (2)	0.8% (2)
Total	100.0% (33)	13.2% (33)

Table 4 Time for diagnosis in both optical and digital observation

Time for diagnosis (minutes)	P1					P2					P3					P4					P5					Mean time ± SD
	P1					P2					P3					P4					P5					
Optical observation Total time (per case)	60(1.2)					60(1.2)					85(1.7)					170(3.4)					117(2.3)					98.4 ± 41.5 (2.0 ± 0.8)
Digital observation Total time (per case)	88(1.8)					118(2.4)					135(2.7)					225(4.5)					153(3.1)					143.8 ± 45.9 (2.9 ± 0.9)
Extra time needed for digital diagnosis Total time (per case)	28(0.6)					58(1.2)					50(1.0)					55(1.1)					36(0.7)					45.4 ± 11.5 (0.9 ± 0.2)
Percentage of extra time needed for digital diagnosis	31.8%					49.2%					37%					24.4%					23.5%					33.2%

other similar studies, is subjective and not highly accurate, due to the lack of validation against more objective measures, such as immunohistochemical and molecular profiles, and the absence of correlation with outcome.

Of notice, the tendency to estimate PTC nuclei is different in the two methods. In this work, the pathologists tended to overestimate PTC nuclei in the conventional optical environment, over-diagnosing NIFTP and PTC, in comparison

with the digital setting. The digital observation had, in this study, the overall advantage to result in no situations of overdiagnosis of NIFTP, and the overdiagnosis of PTC was less frequent specifically in nodules classified as adenomatous/follicular adenomas or thyroiditis in the ground truth. This is particularly important to prevent further surgical and medical overtreatment in these cases. The underestimation of PTC nuclei was similar in both settings. It was empirically

noticed that participants were less likely to go for a higher resolution observation on the screen display in comparison with light microscopy. However, the number of high-resolution scanning per nodule was not recorded and cannot be confirmed at this level.

A large noninferiority study was recently performed to compare WSI diagnosis and microscopic diagnosis in surgical pathology [18], and it provided a detailed organ system analysis for the major discordance rates with the reference standard. Interestingly, the endocrine system had the highest difference in major discordance rates for digital and conventional optical diagnosis, and the cases were mostly of thyroid pathology (7 out of 9 discordances). Two out of four participant pathologists underdiagnosed PTC with the digital diagnosis in most of the discordances (6 out of 7 discordances). This result correlates with our finding that pathologists tend to appreciate PTC nuclear features more on the conventional optical diagnosis. This tendency in our investigation resulted in the overdiagnosis of PTC and NIFTP after conventional optical observation. The slides in the noninferiority study [18] were scanned at 40× magnification, while in our study were scanned at 20× magnification becoming a possible source of interference. However, the final resolution of a digitized slide is not only dependent on the scanner objective lens magnification, but also the size and number of individual pixels within the digital camera's sensor, and the size and amount of individual pixels within the monitor [19]. A comparison between final image resolutions used in both studies is not possible, as the resolution in the noninferiority study was not mentioned. Scanning at higher magnification in our study (40×) will increase the resolution of images; however, it is not clear how this will influence the correlation results. The results may not necessarily improve; another study could be beneficial to analyze the possible effects.

Other causes of disagreement with the ground truth appear to be not so relevant for the diagnostic performance, namely the evaluation of invasion and other features that underlie the distinction between PTC and NIFTP, at variance with previous observations regarding the interobserver variability of invasion in thyroid nodules [20].

The PTC nuclei features evaluation was also the major reason of intraobserver discordance. Only one pathologist had a digital versus conventional optical diagnostic concordance above 95%. For those with lower than 95% concordance rates, two pathologists had no major variations of the overall diagnostic concordance with the ground truth, and two pathologists were discordant with themselves (optical versus digital) because they improved their concordance with the ground truth by 8% in the digital observation. The results presented here raise the possibility that intraobserver concordance between conventional optical versus digital diagnosis should not be considered as the single measure

of validation for new digital systems. Other measurements must be added to evaluate the digital performance, like the concordance with the ground truth. The intraobserver variation in intrinsically challenging cases, such as those included in this study, with a difficult differential diagnosis that has been also described in the conventional optical setting alone [11, 12] may be an additional reason for the low rates of intraobserver concordance recorded in this study. Furthermore, the method that was used for the initial selection of cases, and determination of the ground truth, can affect the results of correlation, mostly in favor of the method used for this purpose (conventional light microscopy in our study).

The digital system adopted in this study was not fast enough to allow navigation with the same speed as in the conventional optical observation, resulting in a 0.9 min delay per case, and 45.4 min delay per 50 slides. This difference is an important factor for discouraging pathologists from implementing digital systems in routine diagnosis of large nodules that are present on several slides. However, other digital pathology systems already available in the market allow a faster time for navigation and, consequently, for diagnosis. Some systems were able to achieve a shorter time with the digital diagnosis in comparison with the conventional optical ones. For example, in a study by Vodovnik [3] this achievement was possible mainly through providing adequate and stable network speeds (median 299 Mbps), a fully integrated laboratory information management system, double displays (23") with advanced larger digital viewers (1920 × 1200), and an adequate processor (i3-3240M 3.4Ghz). The improvement of digital diagnosis time will encourage more implementation of digital pathology systems in pathology laboratories. Another consideration is that the washout period in this study was a bit shorter than that recommended [14], which may slightly lead to recall bias and a decrease in time for diagnosis in favor of the second method used for observation.

Similar extra time needed for digital diagnosis between experienced and less experienced pathologists indicates that the level of experience is a minor factor for the delay in digital diagnosis. Generally, pathologists got used to the system and its tools rapidly and developed confidence in making digital diagnosis easily, especially for younger pathologists who are in close relation with new and advanced technologies. This proves that the major role for the digital delay is rather related to the technical obstacles of the digital system used.

A substantial improvement in the interobserver reproducibility of the measurements (tumor size and distance between the tumor and the nearest surgical margin) was observed in the digital observation in comparison with the conventional optical one. This advantage is relatively of lower significance for thyroid tumors in comparison with other tumor types. In thyroid, a higher level of measurement precision will unlikely affect staging, also specific margin status will

not lead to deep changes in management, unlike in some other organs. However, measurements are still much easier and faster to perform for large and multicentric tumors that are present on several slides.

Novel applications of deep learning technologies to detect and diagnose cancers on WSI are in progress. Potential improvements in these algorithms could help in facilitating the pathology practice and increasing the efficiency and accuracy of diagnosis [21]. One of the recent studies showed promising results for detecting and localizing three types of thyroid carcinomas on digitized WSI using convolutional neural networks [22].

Conclusion

Comparable agreement with the ground truth between conventional optical and digital observation of follicular patterned thyroid nodules, as well as substantial intraobserver agreement and a significant improvement in the reproducibility of the measurements, are legit reasons to support the use of digital observation for diagnosis in thyroid pathology. Additionally, the origins underlying the variability of the diagnosis of thyroid nodules are the same in both conventional optical and digital systems. Finally, digital observation of follicular patterned thyroid nodules should be performed only under optimal speed and resolution conditions so that conventional optical environment can be reproduced.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Vodovnik A. Distance reporting in digital pathology: a study on 950 cases. *J Pathol Inform.* 2015;6:18.
- Mills AM, Gradecki SE, Horton BJ, Blackwell R, Moskaluk CA, Mandell JW, et al. Diagnostic efficiency in digital pathology: a comparison of optical versus digital assessment in 510 surgical pathology cases. *Am J Surg Pathol.* 2018;42(1):53–9.
- Vodovnik A. Diagnostic time in digital pathology: a comparative study on 400 cases. *J Pathol Inform.* 2016;7:4.
- Hirokawa M, Carney JA, Goellner JR, DeLellis RA, Heffess CS, Katoh R, et al. Observer variation of encapsulated follicular lesions of the thyroid gland. *Am J Surg Pathol.* 2002;26(11):1508–14.
- Thompson LDR, Poller DN, Kakudo K, Burchette R, Nikiforov YE, Seethala RR. An international interobserver variability reporting of the nuclear scoring criteria to diagnose noninvasive follicular thyroid neoplasm with papillary-like nuclear features: a validation study. *Endocr Pathol.* 2018;29(3):242–9.
- Liu Z, Bychkov A, Jung CK, Hirokawa M, Sui S, Hong S, et al. Interobserver and intraobserver variation in the morphological evaluation of noninvasive follicular thyroid neoplasm with papillary-like nuclear features in Asian practice. *Pathol Int.* 2019;69(4):202–10.
- Amendoeira I, Maia T, Sobrinho-Simões M. Non-invasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP): impact on the reclassification of thyroid nodules. *Endocr Relat Cancer.* 2018;25(4):R247–58.
- Lloyd RV, Osamura RY, Klöppel G, Rosai J. WHO Classification of Tumours of Endocrine Organs. 4th ed. Lyon: International Agency for Research on Cancer; 2017.
- Vanzati A, Mercalli F, Rosai J. The “sprinkling” sign in the follicular variant of papillary thyroid carcinoma: a clue to the recognition of this entity. *Arch Pathol Lab Med.* 2013;137(12):1707–9.
- Nikiforov YE, Seethala RR, Tallini G, Baloch ZW, Basolo F, Thompson LD, et al. Nomenclature revision for encapsulated follicular variant of papillary thyroid carcinoma: a paradigm shift to reduce overtreatment of indolent tumors. *JAMA Oncol.* 2016;2(8):1023–9.
- Elsheikh TM, Asa SL, Chan JK, DeLellis RA, Heffess CS, LiVolsi VA, et al. Interobserver and intraobserver variation among experts in the diagnosis of thyroid follicular lesions with borderline nuclear features of papillary carcinoma. *Am J Clin Pathol.* 2008;130(5):736–44.
- Lloyd RV, Erickson LA, Casey MB, Lam KY, Lohse CM, Asa SL, et al. Observer variation in the diagnosis of follicular variant of papillary thyroid carcinoma. *Am J Surg Pathol.* 2004;28(10):1336–40.
- Seethala RR, Baloch ZW, Barletta JA, Khanafshar E, Mete O, Sadow PM, et al. Noninvasive follicular thyroid neoplasm with papillary-like nuclear features: a review for pathologists. *Mod Pathol.* 2018;31(1):39–55.
- Pantanowitz L, Sinard JH, Henricks WH, Fatheree LA, Carter AB, Contis L, et al. Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med.* 2013;137(12):1710–22.
- LiVolsi VA, Baloch ZW. Follicular-patterned tumors of the thyroid: the battle of benign vs. malignant vs. so-called uncertain. *Endocr Pathol.* 2011;22(4):184–9.
- Baloch ZW, LiVolsi VA. Follicular-patterned afflictions of the thyroid gland: reappraisal of the most discussed entity in endocrine pathology. *Endocr Pathol.* 2014;25(1):12–20.
- Suster S. Controversies regarding the interpretation of follicular thyroid nodules. *Arch Pathol Lab Med.* 2019;143(12):1472–6.
- Mukhopadhyay S, Feldman MD, Abels E, Ashfaq R, Beltaifa S, Cacciabeve NG, et al. Whole slide imaging versus microscopy for primary diagnosis in surgical pathology: a multicenter blinded randomized noninferiority study of 1992 cases (pivotal study). *Am J Surg Pathol.* 2018;42(1):39–52.
- Sellaro TL, Filkins R, Hoffman C, Fine JL, Ho J, Parwani AV, et al. Relationship between magnification and resolution in digital pathology systems. *J Pathol Inform.* 2013;4:21.
- Zhu Y, Li Y, Jung CK, Song DE, Hang JF, Liu Z, et al. Histopathologic assessment of capsular invasion in follicular thyroid neoplasms—an observer variation study. *Endocr Pathol.* 2020;31(2):132–40.
- Serag A, Ion-Margineanu A, Qureshi H, McMillan R, Saint Martin MJ, Diamond J, et al. Translational AI and deep learning in diagnostic pathology. *Front Med (Lausanne).* 2019;6:185.
- Halicek M, Shahedi M, Little JV, Chen AY, Myers LL, Sumer BD, et al. Head and neck cancer detection in digitized whole-slide histology using convolutional neural networks. *Sci Rep.* 2019;9(1):14043.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.