Improvement of mitoses counting on whole slide images of breast cancer using artificial intelligence

Author: Stijn van Bergeijk
Medical student
Universtiteit Utrecht

Department of pathology

Abstract

Introduction: *Breast cancer (BC) prognosis is largely influenced by histopathological grade, assessed according to the Nottingham modification of Bloom-Richardson (BR). Mitotic count (MC) is a component of histopathological grading but is prone to subjectivity. This study investigated whether mitoses counting in BC using digital whole slide imaging (WSI) compares better to light microscopy (LM) when assisted by artificial intelligence (AI), and to which extent differences in digital MC (AI assisted or not) result in BR grade variations.*

Methods: *Fifty BC patients with paired core biopsies and resections were randomly selected. Component scores for BR grade were abstracted from pathology reports. MC was assessed using LM, WSI and AI. Different modalities (LM-MC, WSI-MC and AI-MC) were analyzed for correlation with scatterplots and linear regression, and for agreement in BR with Cohen's Kappa.*

Results: *MC modalities strongly correlated in both biopsies and resections: LM-MC and WSI-MC (R^2 0.737 and 0.773, respectively), LM-MC and AI-MC (R^2 0.545 and 0.706) and WSI-MC and AI-MC (R^2 0.692 and 0.760). Agreement in BR between modalities was high in both biopsies and resections: LM-MC and WSI-MC (kappa 0.93 and 0.834, respectively), LM-MC and AI-MC (kappa 0.89 and 0.825), and WSI-MC and AI-MC (kappa 0.96 and 0.732).*

Conclusion: *MC in WSI did not compare better to LM-MC when supported by AI. However, LM-MC and WSI-MC were already strongly correlated, so the expected gain from AI was inherently low. Agreement between different modalities for BR was high. WSI-MC appears as a viable alternative to LM-MC.*

Introduction

The yearly worldwide breast cancer (BC) incidence is over 2 million, which makes it the most diagnosed cancer. Female BC currently occupies the fifth place in cancer mortality worldwide, and incidence keeps rising. [1] However, when diagnosed in an early stage, the prognosis of BC can be good. [1,2] One of the strongest factors to determine BC prognosis is histological grade, usually assessed according to the Nottingham modification of Bloom-Richardson grade (BR). [3,4] BR requires the pathologist to score three features: tubule formation, nuclear pleomorphism and mitotic count (MC). Each category gets a score from 1-3. Scores 3-5 define grade I, 6-7 grade II and 8-9 make up grade III BC. Grade I cancers have a significantly better survival than grade II or III cancers. [3,5,6] Studies have shown histological grading, tumor size and lymph node status to be of equal importance for the prognosis of BC. [5,6] Furthermore, histological grade proved to be decisive in up to a third of treatment decisions.[7]

MC is, as a marker of tumor proliferation, the strongest constituent of grade, a high MC associated with poor prognosis.[8,9,10] Several studies have shown a moderate to good reproducibility for BR.[11–13] Despite this there are concerns when focusing solely on MC reproducibility also ranges from moderate to high.[14,15] However, one recent study again found substantial inter- and intra-laboratory variations in BR in more than 33,000 patients. [7] Because of significant inter- and intra-laboratory variations and the importance of MC for the prognosis of BC, higher reproducibility is required.

With the development of digital Whole Slide Imaging (WSI), breast cancer diagnostics have increasingly been performed digitally as WSI has been validated for diagnostic purposes.[16,17] It has been argued that WSI has limitations for reliable histologic

grading, as the quality of the images may not be high enough for properly assessing the MC in all cases due to lack of a z-axis, which pathologists often use when microscopically assessing MC, in standard WSI. Two studies have shown that MCs in WSI and traditional light microscopy (LM) show comparable results. [18,19] However, several studies suggest that although the inter-observer agreement on WSI is similar to LM, MC tended to be systematically lower on WSI. [16,17,20,21]

The increased usage of WSI has stimulated the rise of artificial intelligence (AI) algorithms in pathology. Several of these have been developed for assessing MC, or at least assisting the pathologist in performing MC. Most AI assisted MC has only been tested on validation cohorts. [19,22–26] Experts expect these developments to improve the reproducibility of the MC.[27] The next step is to test an AI algorithm in a clinical setting. The present study tried to answer two questions: does mitoses counting in BC using digital WSI compare better to light microscopic MC when assisted by AI, and to which extent do differences in digital MC (AI assisted or not) result in BR grade variations?

Abbreviations	
BC	*Breast Cancer*
LM	*Light microscopy*
WSI	*Whole slide imaging*
AI	*Artificial intelligence*
BR	*Nottingham modification of Bloom-Richardson*
MC	*Mitotic Count*
LM-MC	*Mitotic count assessed by light microcopy*
WSI-MC	*Mitotic count assessed by whole slide imaging*
AI-MC	*Mitotic count assessed by artificial intelligence*

Methods

Study design and population

Fifty BC patients with paired core biopsies and resections were randomly selected from the workflow of the department of pathology at the UMC Utrecht between December 2018 and February 2020. For each patient, tubular differentiation (scored 1, 2 or 3) and nuclear polymorphism scores (1, 2 or 3) according to Elston and Ellis[3] were taken from the original pathology report. An experienced Pathology Assistant trained in breast microscopy reassessed MC using LM (LM-MC) in 2 mm^2 of adjacent fields in the most cellular and proliferative area of the tumor.[14] MC was scored 1, 2 or 3 points, for respectively ≤7, 8-12 and ≥13 mitoses. After a washout period of at least 2 months, MC was assessed digitally using WSI (WSI-MC), and after another 2 months washout period MC was assessed supported by the AI algorithm (AI-MC).

Digital pathology and AI

Slides had routinely been scanned within the workflow of the UMC Utrecht at 40x magnification (resolution 0f 0.22 µm per pixel) with a Nanozoomer 2.0-XR (Hamamatsu, Japan). All WSI were viewed using standard high-resolution 4k computer screens in the Sectra PACS (Linköping, Sweden).

The MC AI algorithm was in-house developed based on the winner of the AMIDA13 challenge. In short, in the Sectra PACS, an area of interest of the appropriate size of 2 mm^2 (as described for LM-MC) is interactively drawn, after which the algorithm automatically identifies candidate mitoses and mitoses-like objects and displays them in 2 galleries. Objects are interactively reviewed and dragged to the correct gallery, resulting in a final AI MC per 2 mm^2 (Figure 1).

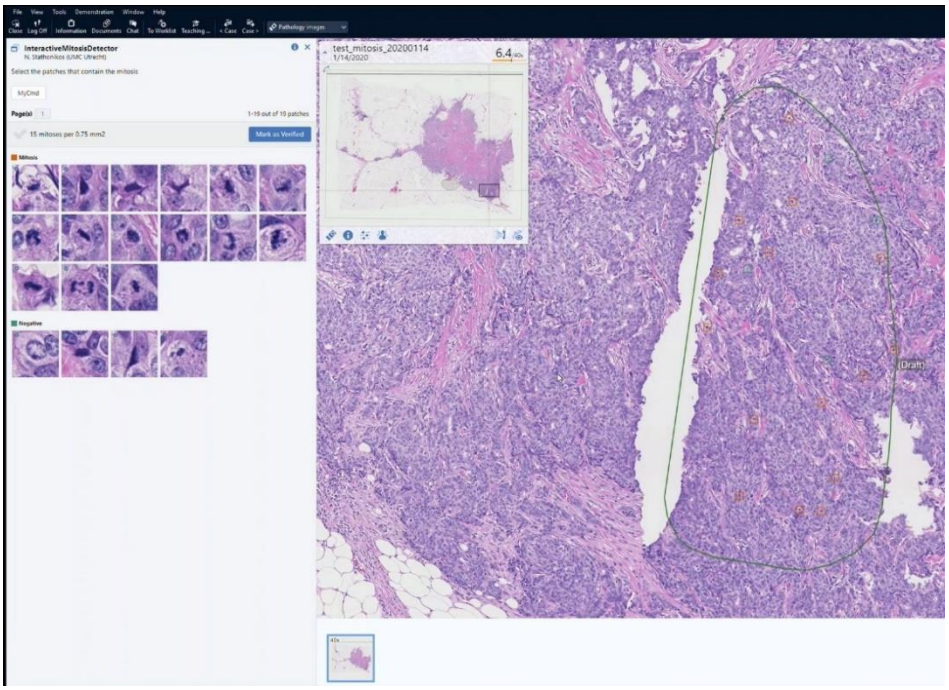

Figure 1. Screen shot of the Sectra PACS where an area of interest has interactively been drawn on the right-hand side, after which an AI algorithm has found candidate mitoses and mitosis-like objects which are displayed in the galleries in the upper left-hand side of the screen. By clicking on a thumbnail in either of the galleries, the PACS displays the candidate object in the center on the right for review, and false positives can be dragged to the negative gallery and vice versa, after which a final AI supported MC is established.

Data analysis

Using the MC from the 3 modalities, three BR grades were composed for each biopsy and resection as usual summing up the scores from tubular differentiation, nuclear polymorphism, and MC, total score 3-5 defining grade 1, scores 6-7 grade 2 and scores 8-9 grade 3.

Data for biopsies and resections were separately analyzed. MC data were pairwise displayed in logarithmic scatterplots and linear regression was used to determine the correlation between the different MC modalities (LM-MC, WSI-MC, and AI-MC), calculating R^2 and noting slope and intercept to detect systematic differences. To assess the concordance in BR resulting from the different MC modalities, crosstabs were created, using Cohen's Kappa to assess BR agreement between the different MC modalities. Scores of 0 meant no agreement, 0.01-0.20 none to slight, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.8 substantial and 0.81-1.00 almost perfect agreement.[28]

All statistics were done using SPSS version 27 for Windows (IBM Corp, Armonk, NY, USA).

Case characteristics		N (%)
	N	50 (100)
Age (years)	Range	34-74
	Mean	57
	Median	60
Estrogen receptor	Positive	48 (96)
	Negative	2 (4)
Progesterone receptor	Positive	37 (26)
	Negative	13 (74)
HER2	Positive	1 (2)
	Negative	49 (98)
BR grade	1	14 (28)
	2	28 (56)
	3	8 (16)

Tabel 1. Demographics and BR grade of collected cases

Results

Case characteristics

See table 1 for demographic data of the collected cases. There were 50 different patients, age ranged from 34-74, BR grade 2 was most common (56%), followed by 1 (28%).

Biopsies

Scatterplots for pairwise comparison between the three MC modalities are shown in figures 2, 3 and 4. All MC modalities were strongly correlated: R^2 between LM-MC and WSI-MC was 0.737, 0.545 between LM-MC and AI-MC, and 0.692 between WSI-MC and AI-MC.

The crosstabs for the BR grades resulting from the different MC modalities are shown in tables 2, 3 and 4, all showing high kappa values: 0.93 for LM-MC versus WSI-MC based BR, 0.89 for LM-MC versus AI-MC based BR, and 0.96 for WSI-MC versus AI-MC based BR.

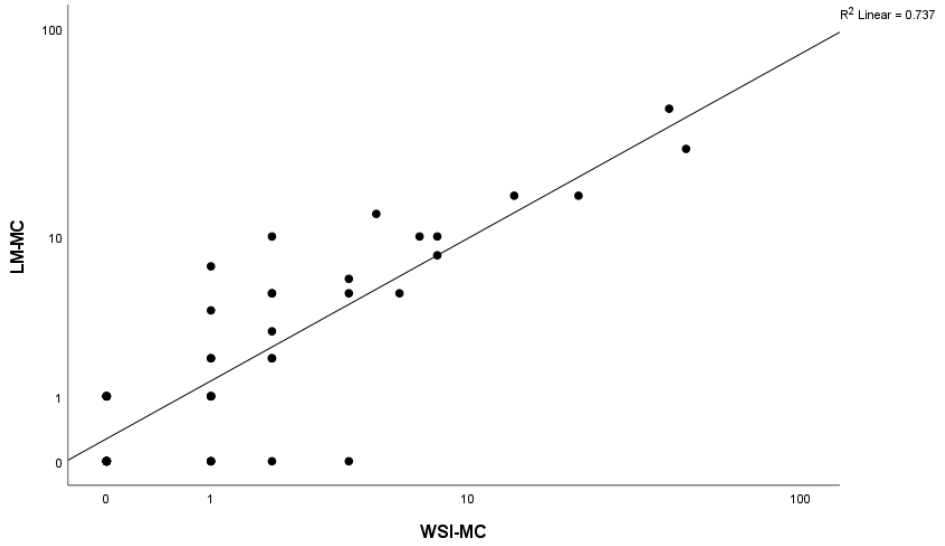

Figure 2. Scatterplot showing a high concordance between Whole Slide Image based digital Mitotic Mount (WSI-MC) and Light-Microscopic MC (LM-MC) in 50 breast cancer biopsies. Slope= 0.78, intercept= 1.2

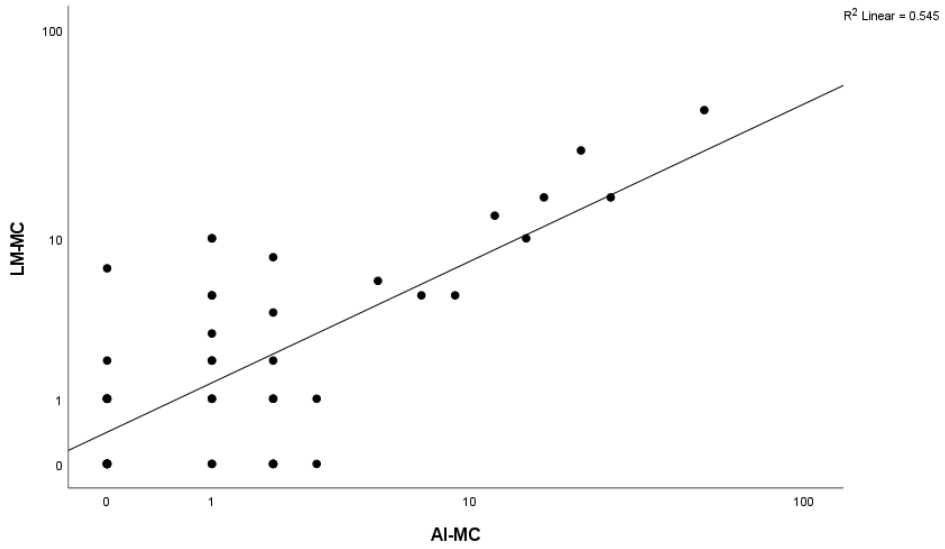

Figure 3. Scatterplot showing a high concordance between Artificial Intelligence based Mitotic Mount (AI-MC) and Light-Microscopic MC (LM-MC) in 50 breast cancer biopsies. Slope= 0.79, intercept= 1.05

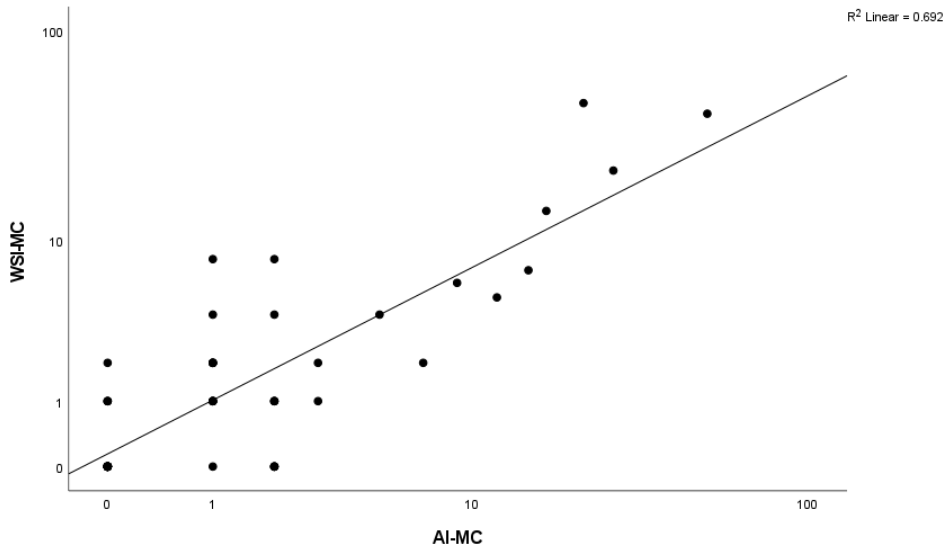

Figure 4. Scatterplot showing a high concordance between Artificial Intelligence based Mitotic Mount (AI-MC) and Whole Slide Image based digital MC (WSI-MC) in 50 breast cancer biopsies. Slope= 0.89, intercept= 0.31

BR grade		WSI-MC based			Total
		1	2	3	
LM-MC based	1	17	0	0	17
	2	1	27	0	28
	3	0	1	4	5
Total		18	28	4	50

Table 2. Crosstab between Bloom&Richardson (BR) grade based on Light Microscopic Mitotic Count (LM-MC) and Whole Slide Image based digital MC (WSI-MC) in 50 breast cancer biopsies (kappa=0.928, 95% CI 0.83-1.01).

BR grade		AI-MC based			Total
		1	2	3	
LM-MC based	1	17	0	0	17
	2	1	26	1	28
	3	0	1	4	5
Total		18	27	5	50

Table 1. Crosstab between Bloom&Richardson (BR) grade based on Light Microscopic Mitotic Count (LM-MC) and artificial intelligence supported MC (AI-MC) in 50 breast cancer biopsies (kappa = 0894, 95% CI 0.78-1.01).

BR grade		AI-MC based			Total
		1	2	3	
WSI-MC based	1	17	0	0	17
	2	1	26	1	28
	3	0	1	4	5
Total		18	27	5	50

Table 3. Crosstab between Bloom&Richardson (BR) grade based on whole slide image based digital Mitotic Count (WSI-MC) and artificial intelligence supported MC (AI-MC) in 50 breast cancer biopsies (kappa = 0.964, 95% CI 0.90-1.03).

Resections

Scatterplots for pairwise comparison between the three MC modalities are shown in figures 5, 6, and 7. All MC modalities were strongly correlated: R^2 between LM-MC and WSI-MC was 0.773, 0.706 between LM-MC and AI-MC and 0.760 between WSI-MC and AI-MC. AI-MC was systematically slightly lower than LM-MC (slope 0.82) and WSI-MC (slope 0.54).

The crosstabs for the BR grades resulting from the different MC modalities are shown in tables 5, 6 and 7, all showing high kappa values: 0.834 for LM-MC based BR versus WSI-MC, 0.825 for LM-MC versus AI-MC based BR, and 0.732 for WSI-MC versus AI-MC based BR.

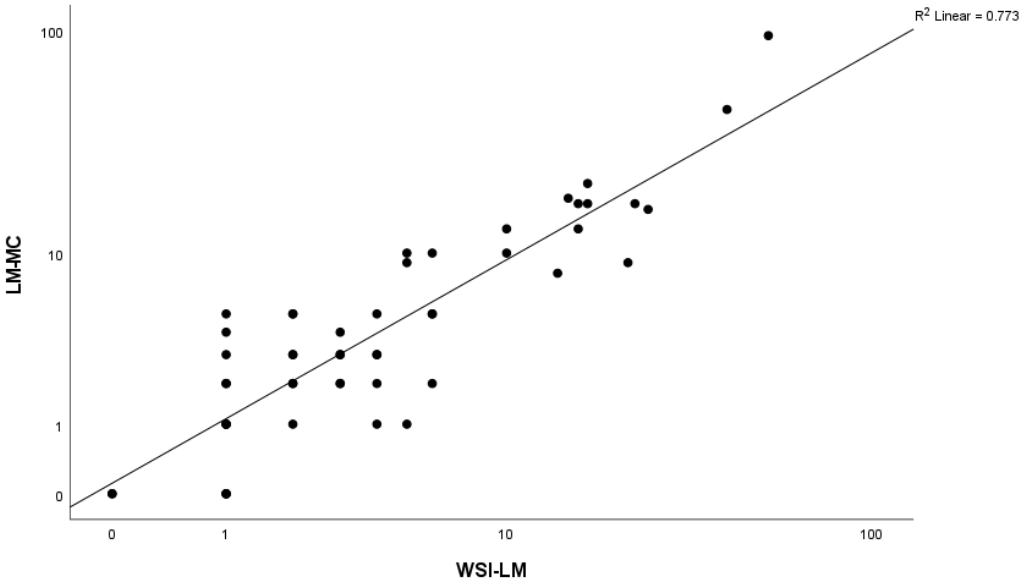

Figure 5. Scatterplot showing a high concordance between Whole Slide Image based digital Mitotic Mount (WSI-MC) and Light-Microscopic MC (LM-MC) in 50 breast cancer resections. Slope= 1.29, intercept= 1.55

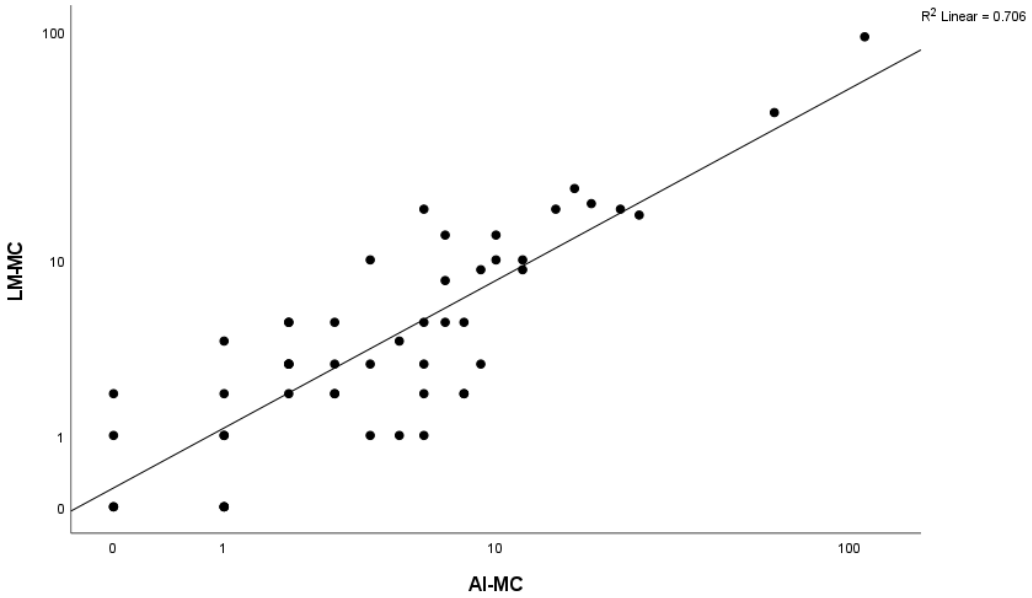

Figure 6. Scatterplot showing a high concordance between Artificial Intelligence based Mitotic Mount (AI-MC) and Light-Microscopic MC (LM-MC) in 50 breast cancer resections. Slope= 0.82, intercept= 0.6

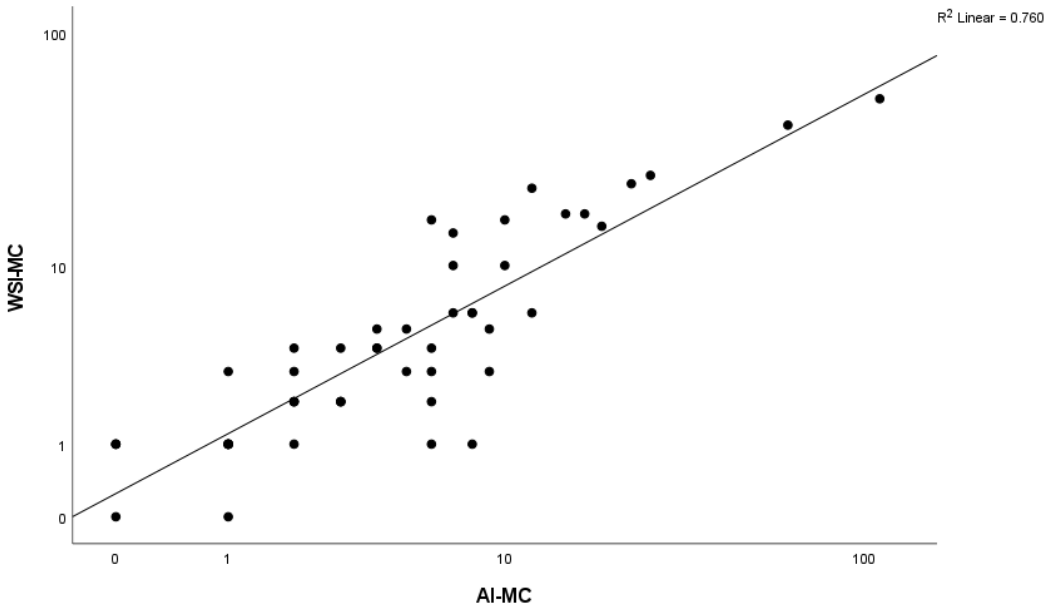

Figure 7. Scatterplot showing a high concordance between Artificial Intelligence based Mitotic Mount (AI-MC) and Whole Slide Image based digital MC (WSI-MC) in 50 breast cancer resections. Slope= 0.54, intercept= 2.55

BR grade		WSI-MC based			Total
		1	2	3	
LM-MC based	1	13	1	0	14
	2	2	24	2	28
	3	0	0	8	8
Total		15	25	10	50

Table 4. Crosstab between Bloom&Richardson (BR) grade based on Light Microscopic Mitotic Count (LM-MC) and Whole Slide Image based digital MC (WSI-MC) in 50 breast cancer resections (kappa=0.834, 95% CI 0.70-0.97).

BR grade		AI-MC based			Total
		1	2	3	
LM-MC based	1	13	1	0	14
	2	2	26	0	28
	3	0	2	6	8
Total		15	29	6	50

Table 5. Crosstab between Bloom&Richardson (BR) grade based on Light Microscopic Mitotic Count (LM-MC) and artificial intelligence supported MC (AI-MC) in 50 breast cancer resections (kappa = 0.825, 95% CI 0.68-0.97).

BR grade		AI-MC based			Total
		1	2	3	
WSI-MC based	1	13	2	0	15
	2	2	23	0	25
	3	0	4	6	10
Total		15	29	6	50

Table 6. Crosstab between Bloom&Richardson (BR) grade based on whole slide image based digital Mitotic Count (WSI-MC) and artificial intelligence supported MC (AI-MC) in 50 breast cancer resections (kappa = 0.732, 95% CI 0.56-0.90).

Discussion

In this study we investigated whether mitoses counting in BC using digital WSI compares better to LM-MC when assisted by AI, and to which extent differences in digital MC (AI assisted or not) result in BR grade variations. Both biopsies and resections showed strong and comparable correlations between LM-MC on the one hand and WSI-MC and AI-MC on the other, and also WSI-MC and AI-MC correlated well. This indicates on the one hand AI does not clearly improve digital MC on whole slide images, but on the other hand MC on digital WSI already correlated well with the gold standard LM-MC, so that clear improvements due to AI could not be expected.

It was noted that AI-MC resulted in systematically slightly lower MC values compared to LM-MC and WSI-MC. This indicates that the AI algorithm misses some mitoses and needs further improvement. On the other hand, the observers checked the results and may not have been critical enough when reviewing the AI results. This underlines the importance of careful human supervision of the output of algorithms when AI is used in daily practice, not limited to reviewing the found objects.

Several other studies showed similar results regarding the comparability between LM-MC and WSI-MC.[16,20,29,30] Noted differences between LM-MC and WSI-MC were perceived to be within the range of inter-observer differences in LM-MC. Also, studies which used 40x magnification for scanning and high-resolution displays noted that differences between WSI and LM tended to get smaller, suggesting that standard high-end technology is required for proper mitoses counting on WSI. As to AI, a recent study applying AI to select a mitoses hotspot in which to count showed improved inter-observer agreement in interactive mitoses counting on WSI, with similar inter-observer kappa values for LM-MC and AI-MC.[19]

However, one study demonstrated higher inter-observer agreement for AI-MC compared to LM-MC. Furthermore, they demonstrated a substantial saving in time.[31] So, different studies seem to point at least to non-inferiority of AI-MC compared to LM-MC in BC. The potential to save time is another reason to further explore the possibilities of AI.

Both biopsies and resections showed near perfect agreement in BR between different modalities, although the kappa for WSI-MC versus AI-MC based BR in the resection group was slightly lower. This indicates that differences in MC between different modalities hardly influence BR grade.

One study compared BR based on LM and WSI in over 1600 cases, showing a strong association (Cramer's V: 0.58) between both modalities.[16] Another study focusing on inter-observer differences in BR when using WSI, showed the concordance to be similar to inter-observer differences in BR using LM.[21] These studies substantiate our results. To the best of our knowledge, no previous study has been conducted that compares agreement of BR using LM-MC or WSI-MC and AI-MC. The high agreement in BR in this study is probably related to by two factors. Firstly, WSI-MC and AI-MC were performed on the exact same slide as LM-MC, whereas larger tumors may be heterogeneous across different tissue blocks. Secondly, grading in different modalities was assessed by the same observer, causing the criteria for mitotic figures to be interpreted singularly and increasing the chance of selecting the same hotspot.

This study has some limitations. First, the gold standard is LM-MC assessed by a single observer. Due to significant inter-observer differences for LM-MC[7], a study with multiple observers may provide a more realistic view on the added value of AI. Another option would be to use Phosphohistone H3 immunohistochemistry, which enhances recognition of mitotic figures and may make LM-MC (and perhaps even AI-MC) more reproducible.[32] Secondly, this study has a relatively small number of cases.

In daily pathology practice, digital WSI is increasingly used worldwide. This study, in combination with previous studies in this field, shows WSI-MC to be suitable for grading BC. Especially pathology laboratories which have a digital workflow could thereby incorporate WSI-MC in their daily practice of grading BC.

In general, AI algorithms show great promise in improving pathology practice. We feel encouraged by the non-inferiority of the present first-generation AI mitoses algorithm, since we expect next generation algorithms to be substantially improved. These algorithms may at least save valuable interaction time for the pathologist, especially when algorithms run in the background on WSI, providing the pathologist with mitotic hotspots.

In conclusion, WSI-MC does not compare better to LM-MC by using AI. However, LM-MC and both WSI-MC and AI-MC already strongly correlated. Agreement between different modalities for BR was high. WSI-MC appears as a viable alternative to LM-MC. Further research is required to advance our knowledge of AI-MC, but it appears at least non-inferior to LM-MC and has the potential to save time.

Bibliography

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*. 2021;71(3):209-249. doi:10.3322/CAAC.21660
2. Ahmad A. Breast Cancer Statistics: Recent Trends. In: Ahmad A, ed. *Breast Cancer Metastasis and Drug Resistance: Challenges and Progress*. Advances in Experimental Medicine and Biology. Springer International Publishing; 2019:1-7. doi:10.1007/978-3-030-20301-6_1

3. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*. 1991;19(5):403-410. doi:10.1111/j.1365-2559.1991.tb00229.x
4. Genestie C, Zafrani B, Asselain B, et al. Comparison of the prognostic value of Scarff-Bloom-Richardson and Nottingham histological grades in a series of 825 cases of breast cancer: major importance of the mitotic count as a component of both grading systems. *Anticancer research*. 1998;18(1B):571-576.
5. van Dooijeweert C, van Diest PJ, Ellis IO. Grading of invasive breast carcinoma: the way forward. *Virchows Archiv 2021*. 2021;1:1-11. doi:10.1007/S00428-021-03141-2
6. Rakha EA, Reis-Filho JS, Baehner F, et al. Breast cancer prognostic classification in the molecular era: The role of histological grade. *Breast Cancer Research*. 2010;12(4). doi:10.1186/bcr2607
7. van Dooijeweert C, van Diest PJ, Willems SM, et al. Significant inter- and intra-laboratory variation in grading of invasive breast cancer: A nationwide study of 33,043 patients in the Netherlands. *International Journal of Cancer*. 2020;146(3):769-780. doi:10.1002/ijc.32330
8. Diest PJ van, Wall E van der, Baak JPA. Prognostic value of proliferation in invasive breast cancer: a review. *Journal of Clinical Pathology*. 2004;57(7):675. doi:10.1136/JCP.2003.010777
9. Baak JPA, van Diest PJ, Voorhorst FJ, et al. Prospective multicenter validation of the independent prognostic value of the mitotic activity index in lymph node-negative breast cancer patients younger than 55 years. *Journal of Clinical Oncology*. 2005;23(25):5993–6001. doi:10.1200/JCO.2005.05.511
10. Klintman M, Strand C, Ahlin C, et al. The prognostic value of mitotic activity index (MAI), phosphohistone H3 (PPH3), cyclin B1, cyclin A, and Ki67, alone and in combinations, in node-negative premenopausal breast cancer. *PloS one*. 2013;8(12):e81902. doi:10.1371/journal.pone.0081902
11. Meyer JS, Alvarez C, Milikowski C, et al. Breast carcinoma malignancy grading by Bloom-Richardson system vs proliferation index: reproducibility of grade and advantages of proliferation index. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*. 2005;18(8):1067-1078. doi:10.1038/modpathol.3800388
12. Robbins P, Pinder S, de Klerk N, et al. Histological grading of breast carcinomas: a study of interobserver agreement. *Human pathology*. 1995;26(8):873-879. doi:10.1016/0046-8177(95)90010-1
13. F T, KD K, G H, et al. Histological Grading of Breast Cancer: Interobserver, Reproducibility and Prognostic Significance. *Pathology Research and Practice*. 1990;186(6):732-736. doi:10.1016/S0344-0338(11)80263-3
14. van Diest PJ, Baak JPA, Matze-Cok P, et al. Reproducibility of mitosis counting in 2,469 breast cancer specimens: Results from the Multicenter Morphometric Mammary Carcinoma Project. *Human Pathology*. 1992;23(6):603-607. doi:10.1016/0046-8177(92)90313-R
15. Boiesen P, PO B, L A, et al. Histologic grading in breast cancer–reproducibility between seven pathologic departments. South Sweden Breast Cancer Group. *Acta oncologica (Stockholm, Sweden)*. 2000;39(1):41–45. doi:10.1080/028418600430950
16. Rakha EA, Aleskandarani M, Toss MS, et al. Breast cancer histologic grading using digital microscopy: concordance and outcome association. *Journal of Clinical Pathology*. 2018;71(8):680-686. doi:10.1136/JCLINPATH-2017-204979
17. Williams B, Hanby A, Millican-Slater R, et al. Digital pathology for primary diagnosis of screen-detected breast lesions - experimental data, validation and experience from four centres.

Histopathology. 2020;76(7):968-975. doi:10.1111/his.14079
18. Al-Janabi S, Slooten HJ van, Visser M, et al. Evaluation of Mitotic activity index in breast cancer using whole slide digital images. *PLoS ONE*. 2013;8(12). doi:10.1371/JOURNAL.PONE.0082576
19. Balkenhol MCA, Tellez D, Vreuls W, et al. Deep learning assisted mitotic counting for breast cancer. *Laboratory investigation; a journal of technical methods and pathology*. 2019;99(11):1596-1606. doi:10.1038/s41374-019-0275-0
20. Lashen A, Ibrahim A, Katayama A, et al. Visual assessment of mitotic figures in breast cancer: a comparative study between light microscopy and whole slide images. *Histopathology*. Published online 2021. doi:10.1111/HIS.14543
21. Ginter PS, Idress R, D'Alfonso TM, et al. Histologic grading of breast carcinoma: a multi-institution study of interobserver variation using virtual microscopy. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*. 2021;34(4):701-709. doi:10.1038/s41379-020-00698-2
22. Malon C, Brachtel E, Cosatto E, et al. Mitotic Figure Recognition: Agreement among Pathologists and Computerized Detector. *Analytical Cellular Pathology (Amsterdam)*. 2012;35(2):97. doi:10.3233/ACP-2011-0029
23. Veta M, van Diest PJ, Willems SM, et al. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Medical image analysis*. 2015;20(1):237-248. doi:10.1016/j.media.2014.11.010
24. Roux L, Racoceanu D, Loménie N, et al. Mitosis detection in breast cancer histological images An ICPR 2012 contest. *Journal of Pathology Informatics*. 2013;4(1):8. doi:10.4103/2153-3539.112693
25. Nateghi R, Danyali H, Helfroush MS. A deep learning approach for mitosis detection: Application in tumor proliferation prediction from whole slide images. *Artificial intelligence in medicine*. 2021;114:102048. doi:10.1016/j.artmed.2021.102048
26. Li C, Wang X, Liu W, Latecki LJ, Wang B, Huang J. Weakly supervised mitosis detection in breast histopathology images using concentric loss. *Medical Image Analysis*. 2019;53:165-178. doi:10.1016/j.media.2019.01.013
27. van Dooijeweert C, van Diest PJ, Ellis IO. Grading of invasive breast carcinoma: the way forward. *Virchows Archiv 2021*. 2021;1:1–11. doi:10.1007/S00428-021-03141-2
28. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276-282.
29. Wei BR, Halsey CH, Hoover SB, et al. Agreement in Histological Assessment of Mitotic Activity Between Microscopy and Digital Whole Slide Images Informs Conversion for Clinical Diagnosis. *Acad Pathol*. 2019;6:2374289519859841. doi:10.1177/2374289519859841
30. Shaw EC, Hanby AM, Wheeler K, et al. Observer agreement comparing the use of virtual slides with glass slides in the pathology review component of the POSH breast cancer cohort study. *J Clin Pathol*. 2012;65(5):403-408. doi:10.1136/jclinpath-2011-200369
31. Pantanowitz L, Hartman D, Qi Y, et al. Accuracy and efficiency of an artificial intelligence tool when counting breast mitoses. *Diagnostic Pathology*. 2020;15(1):80. doi:10.1186/s13000-020-00995-z
32. van Steenhoven JEC, Kuijer A, Kornegoor R, et al. Assessment of tumour proliferation by use of the mitotic activity index, and Ki67 and phosphohistone H3 expression, in early-stage luminal breast cancer. *Histopathology*. 2020;77(4):579–587. doi:10.1111/his.14185