Automated artificial intelligence-based analysis to support clinical assessments of Ki-67 immunohistochemistry-stained neuroendocrine tumors in the routine workflow

Vidya Arole MD[1]*, Jesse Kulewsky[1], Jenifer Caldara[2], Rachel Scarl DO[1], Swati Satturwar MD[1], Zaibo Li MD[1], Wei Chen MD[1], Anil Parwani MD[1], Giovanni Lujan MD[1]

1 The Ohio State University Wexner Medical Center, 2 Visiopharm Corporation

The Ohio State University
Wexner Medical Center

Background

Cell proliferation is commonly assessed to grade neuroendocrine tumors (NETs) and aid in the evaluation of patient prognosis and treatment selection. To grade NETs, a proliferation index (PI) is evaluated based on the presence of immunohistochemical staining of the Ki-67 nuclear antigen. In the gastrointestinal tract, grading is performed using WHO criteria for PI as Grade 1-3 (less than 3%, 3%-20%, and greater than 20% respectively). It is recommended to count at least 500 cells followed by calculating the percentage of Ki-67 positive nuclei for cases with the Ki-67 index close to grade cut-offs on eyeballing. Ki 67 proliferation index near cut-offs is a real challenge for practicing pathologists

Figure 1. Example of APP results in core biopsy with metastic NET

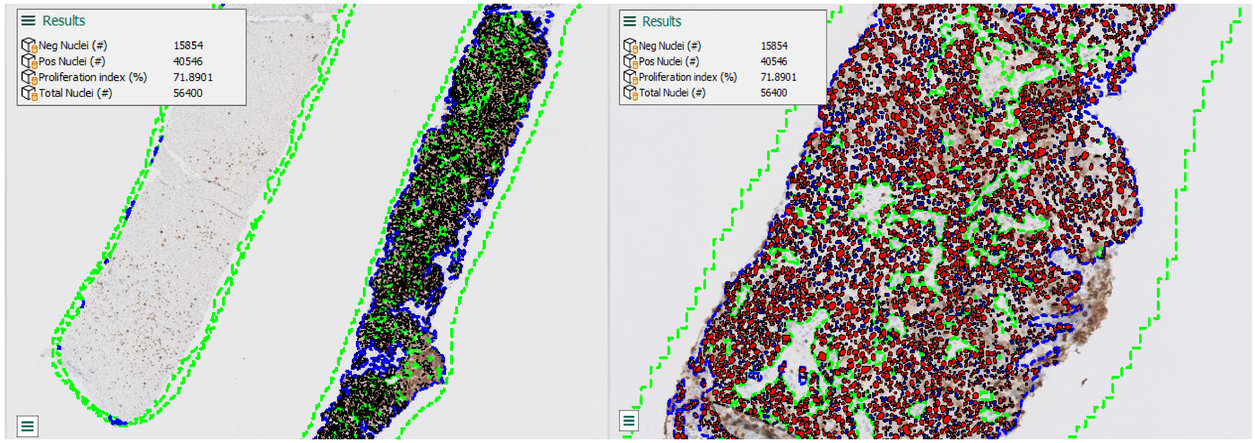

Fig1: Ki 67 APP results on a biopsy sample of metastatic NET with the green area identified as tissue, blue as the tumor, and red as Ki-67 positive nuclei

Methods

This study utilized 8 samples to assess the utility and accuracy of machine learning-based algorithms. Primary gastrointestinal or metastatic NETs, which were stained for Ki-67 using the Mib1 assay, digitized with Philips IntelliSite Ultra Fast Scanner and added to the Philips Image Management System for routine, digital sign-out. These whole slide images (WSI) were streamed directly into the Visiopharm platform and analyzed with an automated Ki-67 algorithm (APP) to produce a PI for Ki-67 nuclei positivity in select NETs based on the assessment of the entire tumor. The resulting digital image analysis (DIA) reads were compared to known, manual pathologist scores. Pathologists' reads at the time of diagnosis are based on different methods used in calculating the PI in the Hot spots (eyeballing, eyeballing complemented with digital reads or manual calculation by counting the positive nuclei in 500 cells, with hotspot region printed on a paper).

Table 1. Comparison of APP and manual reads

Sample	Neg Nuclei (#)	Pos Nuclei (#)	Proliferation index (%)	Total Nuclei (#)	Pathologists Read (%)
1	15,854	40,546	71.89%	56,400	80
2	395,216	21,022	5.05%	416,238	6
3	120,896	3,092	2.49%	123,988	10 to 15
4	9,307	24,992	72.87%	34,299	70 to 80
5	362,309	20,559	5.37%	382,868	13
6	18,241	30,708	62.73%	48,949	70
7	272,124	1,814	0.66%	273,938	2
8	1,009,931	3,611	0.36%	1,013,542	less than 2

Results

We found 87.5% concordance between the DIA and manual reads. Overall, the APP reads were lower than the pathologist reads, although the grade assigned by the pathologist was unaltered in 7/8 cases. The discrepant case (sample 3) was reviewed blindly by original pathologist (with PI read of 4%) and 3 additional pathologists to reveal variable PI reads (1-2%, 2% and 4.8%), half in concordance with APP read (within the same tumor grade). Two pathologists in concordance with APP PI calculated their PI by eyeballing the entire tumor (non-gastrointestinal pathologists) and the other two pathologists with discordant reads calculated PI in the Hotspot region (gastrointestinal pathologists).

Conclusions

There is no uniform method used for the calculation of PI in NETs amongst pathologists across the world. This limited sample study proposes the utility of integrating Visiopharm Ki 67 APP into a clinical workflow as a pathologist support tool to save labor, improve accuracy and increase reproducibility, especially in cases with borderline PI. The APP is built to generate both global as well as Hotspot PI scores. We tested the utility of APP for global scores in this phase. Incorporating, followed by validating Hot spot scores would satisfy recommended criteria in assessing the grade of NETs, which is our plan in phase 2 of this study.