AI deep learning tumour detection directly on ER, PR and Ki-67 IHC slides yields a single slide automated workflow with high concordance to manual scoring

Henrik Holten-Rossing1, Henrik Klingberg1

1 Visiopharm A/S, Hørsholm, Denmark

INTRODUCTION

Assessment of ER, PR and Ki-67 provides essential prognostic information in the classification of breast carcinomas.1 Conventional manual assessment of these biomarkers has shown inter- and intra-observer variation and are both tedious and time demanding when following the ASCO/CAP guidelines which have paved the way for more accurate digital image analysis (DIA) methods2,3,4. The development of artificial intelligence (AI) has made it possible to automatically detect tumour in IHC slides without the need of cytokeratin (CK) stains. Earlier DIA techniques such as VirtualDoubleStaining™ (VDS) has relied on sequential CK stains in tumour regions to identify CK+ tumour cells. AI in DIA was used in the assessment of ER, PR and Ki-67 will thus eliminate the need for additional serial sectioning and CK staining and this will reduce both costs and valuable time that could be distributed to less prioritized diagnostic areas.

The aim of this study was to assess the performance of AI deep learning tumour detection and subsequent automated DIA ER, PR or Ki-67 scoring in comparison to conventional manual assessment.

MATERIALS AND METHODS

632 breast cancer tissue samples collected from 5 hospitals were included in this study. Slides were stained with site specific IHC stainer systems and clones and digitized by 4 whole slide-scanner systems (table 1). Visiopharm’s image analysis software (VIS) platform with the AI Deep Learning module was used to detect tumour regions by the 2016-12 - IHC, Tumor Detection, AI APP (translational APP). The AI APP was developed using the DeepLabv3+ neural network. The VirtualDoubleStaining™ (VDS) method was used to identify tumour regions based on a CK marker and was corrected manually where needed, a process that limits the subjectivity of manual annotations. The nuclear biomarkers was quantified using Visiopharm’s site specific calibrated ER, PR or Ki-67 CE-IVD APPs. The AI DIA scores were compared to the manual biomarker assessments. All ER, PR and Ki-67 breast cancer samples were assessed manually by experienced pathologists from each hospital and reported as positive percentage (ER, PR) and proliferation index (Ki-67). The manual assessment included both scoring on conventional microscope and scoring on digitalized slides according to national and international guidelines and compared to AI DIA.

Table 1. This multi-site study included four different whole slide-scanner systems and three various clones for each biomarker (ER, PR and Ki-67).

<table>
<thead>
<tr>
<th>Whole slide-scanner systems</th>
<th>ER clones</th>
<th>PR clones</th>
<th>Ki-67 clones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Omnisys VL4</td>
<td>Agilent Dako 105</td>
<td>Agilent Dako Ppt 635</td>
<td>Agilent Dako MIB-1</td>
</tr>
<tr>
<td>Ventana Ioscan PT</td>
<td>Ventana SP1</td>
<td>Ventana 1E2</td>
<td>Ventana 30-9</td>
</tr>
<tr>
<td>Hamanatsu Nanozoomer</td>
<td>Leica EF11</td>
<td>Leica 16</td>
<td>Cell-Marquise SP6</td>
</tr>
<tr>
<td>Leica SCN 400</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RESULTS

The combined results from all 5 hospitals showed a high agreement percentage (90%) and a substantial kappa agreement (0.67) even though individual results showed variation. Bland-Altman plotting showed a small mean bias of 1.25% of AI DIA when compared to manual scoring. Figure 1 to 3 shows examples of AI tumour detection in ER, PR and Ki-67 IHC images. Inter-operator variability was assessed on ER, PR and Ki-67 datasets (n = 126) and a variation coefficient of 0.02% ± 0.05% was observed with a kappa agreement of 1. Results of the individual IHC indicators are presented in table 2. Images scanned on the Leica SCN 400 was removed from this analysis (table 2) due to poor overall performance of the tumour detection (visual inspection), the same was the case for biopsy image data, both sources were not represented in the AI training dataset.

REFERENCES