Can Artificial Intelligence (AI) completely replace human reader in mammography screening program?: A retrospective evaluation with Digital Mammography (DM) and Digital Breast Tomosynthesis (DBT)


*Reina Sofia University Hospital (Córdoba, Spain).
**ScreenPoint Medical (Nijmegen, Holanda)
PURPOSE

- To retrospectively evaluate an AI system as a stand-alone reader in a breast cancer screening program with digital mammography (DM) and digital breast tomosynthesis (DBT).

HYPOTHESIS

- The AI system could achieve a similar sensitivity to radiologists with an acceptable recall rate.
ARTIFICIAL INTELLIGENCE SYSTEM

The AI system (Transpara 1.7.0, ScreenPoint Medical BV) automatically detects lesions suspicious of breast cancer in DM and DBT.

• **REGION SCORE**: score from 1 to 100 in each region marked according to the probability of malignancy.

• **EXAM SCORE**: each exam is given an overall score from 1 to 100 which corresponds to the region assigned the highest score.

Exam score: 95
Potentially eligible population:
16067 women
(Tomosynthesis Cordoba screening Trial)
113 cancers
(98 screen-detected cancers and 15 interval cancers)

Study population:
15999 exams (DM y DBT)
113 cancers
(98 screen-detected cancers and 15 interval cancers)

Excluded exams:
(Images not retrieved from PACS)
68 exams
(no cancers)

✓ All of the included exams were analyzed using our AI system.
✓ Only the cancers correctly localized and marked with the highest score were considered true positives.
We compared the AI system performance as a stand-alone reader for DM and DBT with the original human readings.

AI stand-alone performance was measured using the area under the receiver operating characteristic curve (AUC). The highest region score found in the exam was used as exam score. 95% CIs were computed using bootstrapping.

We investigated whether AI as a single independent reader in DM or DBT screening could achieve similar sensitivity with an acceptable recall rate. The recall rate of AI and the human readings were compared using a McNemar test.

(*)Different operating points selected to have non-inferior sensitivity as the human readings (non-inferiority margin 5%)
A) Receiver operating characteristic curve (ROC) of AI system for DM and DBT exams. The AI system achieved an AUC of 0.93 (95%CI :0.89-0.96) for DM and 0.94 (95%CI :0.91-0.97) for DBT.

B) Zoomed image of the ROC curve of AI system shows the range when radiologist operated.
## RESULTS

<table>
<thead>
<tr>
<th></th>
<th>DM</th>
<th>Cut-off</th>
<th>P</th>
<th></th>
<th>DBT</th>
<th>Cut-off</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>58.4% (66/113)</td>
<td>62.8% (71/113)</td>
<td>0.458</td>
<td>Sensitivity</td>
<td>77% (87/113)</td>
<td>80.5% (91/113)</td>
<td>0.648</td>
</tr>
<tr>
<td>Recall rate</td>
<td>3.1%</td>
<td><strong>1.7%</strong></td>
<td>&lt;0.001</td>
<td>Recall rate</td>
<td>3%</td>
<td><strong>9.2%</strong></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>67.3% (76/113)</td>
<td>70.8% (80/113)</td>
<td>0.523</td>
<td>Sensitivity</td>
<td>81.4% (92/113)</td>
<td>85% (96/113)</td>
<td>0.481</td>
</tr>
<tr>
<td>Recall rate</td>
<td>5.1%</td>
<td><strong>3.1%</strong></td>
<td>&lt;0.001</td>
<td>Recall rate</td>
<td>4.4%</td>
<td><strong>16.7%</strong></td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

- **AI in DM** achieved non-inferior sensitivity as a single or double reading with a significantly reduction in recall rate.
- **AI in DBT** achieved non-inferior sensitivity as a single or double reading but with a higher recall rate.
A 67-year-old woman recalled only by the original DBT readings. This study would be recalled both in DM and DBT by the AI. Grade II invasive lobular carcinoma of 18mm was diagnosed at percutaneous biopsy.
CONCLUSIONS

- In breast cancer screening with digital mammography, artificial intelligence could replace human readings, reducing the recall rate with non inferior sensitivity.

- In breast cancer screening with digital breast tomosynthesis, however, to maintain the sensitivity of radiologists, the recall rate would be substantially increased.
THANK YOU FOR YOUR ATTENTION