**Methods**

**Observational indicators and follow-up plan**

Diagnostic criteria: within 2 years from the time the patient came to the hospital for the first mammogram, the pathological diagnosis of the same lesion as malignant was defined as malignant lesion. Definition of benign lesions :(1) pathological diagnosis of the same lesion within 2 years was benign;(2) the patients were followed up for more than 2 years, and mammography more than 2 years after the first mammography examination indicated benign, without pathological diagnosis.

Follow-up plan: (1) retrospective study part: a. Collect mammography from the image workstation for patients with suspicious breast lesions and collect pathological diagnosis information. b. For patients without pathological diagnosis, all mammography information of the patient was collected from the image workstation to confirm whether the patient had mammography more than 2 years after the first mammography in our hospital, and the diagnostic tendency was benign. (2) prospective study part: a. Mammography was performed for suspicious breast lesions, and biopsies were performed in category of BI-RADS4 and 5 patients to obtain pathological diagnosis and terminate follow-up. Patients with BI-RADS 3 were followed up by mammography every six months for more than 2 years. During the follow-up period, patients diagnosed as BI-RADS 4 and 5 by mammography received biopsy, and obtained pathological diagnosis and terminated the follow-up. Patients in the BI-RADS 1 and 2 categories received mammography 2 years after the first mammography, and those with BI-RADS 4 and 5 categories received biopsy.

**Statistical analysis**

We evaluated the diagnostic accuracy, sensitivity, specificity, and positive predictive value (PPV) and negative predictive value (NPV) of the model in the differential diagnosis of cancerous lesions.(1) detection rate of lesions: the detection rate of lesions in the model was calculated based on the results of pathological examination and the results of imaging follow-up of more than 2 years, and the detection rates of different types of lesions were further calculated respectively.

(2) Diagnostic accuracy of lesions: the sensitivity, specificity, positive predictive value, negative predictive value and overall accuracy of the model for breast lesions were calculated, and the diagnostic accuracy of different types of lesions was calculated respectively.

(3) AUC for patient level diagnosis: to measure the diagnosis performance of the model, we use the predicted score of the most malignant detected lesion as the malignant score of a patient, and calculate AUC with pathological examination results.

**Additional Tables**

**Table S1：Study sites**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Institution** | **No. of patients enrolled** | **Manufacturer, Model Name** |
| Center A | Beijing Cancer Hospital | 1450 | GE MEDICAL SYSTEMS，Senographe Essential VERSION ADS\_54.20;SIEMENS,Mammomat Novation DR |
| Center B | Shunyi maternal & child health hospital | 1454 | HOLOGIC Inc.Selenia Dimensions |
| Center C | DaxingQu hospital | 958 | Philips Medical Systems,MammoDiagnost DR |
| Center D | Beijing Chaoyang Women and Children Hospital | 1098 | SIEMENS,Mammomat Inspiration |
| Center E | Shunyi district hospital | 279 | HOLOGIC Inc.Selenia Dimensions |
| Center F | BJ SJS HOSP | 507 | GE MEDICAL SYSTEMS, Senograph DS VERSION ADS\_54.20 |

**Table S2: The basic characteristics of patients in developing the model**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Training set(n=3389) | Verification set(n=730) | P value |
| **Age, years** | 52.45 (19-88) | 53.23 (26-85) | 0.070 |
| **Breast density** |  |  | 0.028 |
| a | 177(5.2%) | 36(4.9%) |  |
| b | 641(18.9%) | 153(21.0%) |  |
| c | 2323(68.6%) | 509(69.7%) |  |
| d | 248(7.3%) | 32(4.4%) |  |
| **Lesion type** |  |  | <0.001 |
| **Malignant type patients** | 2001 | 453 | 0.190 |
| Mass | 1665(58.1%) | 388(57.0%) |  |
| Calcification | 1122(39.1%) | 270(39.6%) |  |
| Distortion | 16(0.6%) | 1(0.2%) |  |
| Asymmetry | 64(2.2%) | 22(3.2%) |  |
| **Benign type patients** | 1388 | 277 | 0.199 |
| Mass | 1318(64.5%) | 251(62.9%) |  |
| Calcification | 619(30.3%) | 135(33.8%) |  |
| Distortion | 4(0.2%) | 1(0.3%) |  |
| Asymmetry | 102(5.0%) | 12(3.0%) |  |
| Data are n (%) unless otherwise stated. | | | |

**Table S3: The basic clinical information of 200 test patients**

|  |  |
| --- | --- |
| Variable | Patients（N=200） |
| Patient age (y) |  |
| Mean | 59 |
| Median | 59 |
| Range | 33-85 |
| Interquartile range | 46-58 |
| BI-RADS breast density† | |
| a | 12(6%) |
| b | 37(18.5%) |
| c | 82(41%) |
| d | 69(34.5%) |
| † Data are numbers of examinations, with percentages in parentheses. | |

**Table S4: pathological results and morphological features of lesions in 70 malignant patients**

|  |  |
| --- | --- |
| Characteristics | No. of examinations |
| **Histological type** |  |
| Invasive ductal carcinoma | 53 |
| Ductal carcinoma in situ | 10 |
| Invasive papillary carcinoma | 6 |
| Others | 1 |
| **Lesion type** |  |
| Mass† | 49 |
| Calcification | 20 |
| Asymmetry | 6 |
| Architectural distortion | 5 |
| † 10 cases presented with mass with calcification. | |

**Table S5: The basic information of prospective application cases of the model**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Center A (n=1450) | Center B (n=1454) | Center C (n=958) | Center D (n=1098) | Center E (n=279) | Center F (n=507) | P value |
| Age, years | 50.35(26-85) | 50.57(25-86) | 50.84(29-82) | 49.99(26-79) | 51.50(30-77) | 50.61(33-85) | 0.66 |
| Breast density |  |  |  |  |  |  | <0.001 |
| a | 38 | 75 | 68 | 56 | 41 | 31 |  |
| b | 101 | 261 | 246 | 253 | 64 | 161 |  |
| c | 1000 | 1062 | 551 | 732 | 157 | 300 |  |
| d | 111 | 56 | 93 | 57 | 17 | 15 |  |
| Lesion type |  |  |  |  |  |  | <0.001 |
| Malignant type patients | 228 | 86 | 72 | 42 | 31 | 36 |  |
| Mass | 195 | 71 | 61 | 36 | 28 | 31 |  |
| Calcification | 144 | 47 | 40 | 26 | 16 | 21 |  |
| Distortion | 1 | 0 | 0 | 0 | 0 | 0 |  |
| Asymmetry | 8 | 5 | 2 | 2 | 2 | 3 |  |
| Benign type patients | 144 | 69 | 41 | 36 | 17 | 30 |  |
| Mass | 132 | 69 | 30 | 32 | 17 | 26 |  |
| Calcification | 54 | 34 | 24 | 14 | 7 | 24 |  |
| Distortion | 0 | 0 | 0 | 0 | 0 | 1 |  |
| Asymmetry | 11 | 4 | 1 | 2 | 0 | 1 |  |
| Negative Patients | 1078 | 1299 | 845 | 1020 | 231 | 441 |  |

**Table S6: classification performance of the model(by lesions)**

|  |  |
| --- | --- |
| Variable | Hospital validation |
| **Malignant mass** | n=397 |
| Accuracy (95% CI) | 0.784(0.752, 0.816) |
| Sensitivity (95% CI) | 0.743(0.700, 0.786) |
| Specificity (95% CI) | 0.853(0.808, 0.899) |
| Positive predictive value (95% CI) | 0.897(0.864, 0.930) |
| Negative predictive value (95% CI) | 0.660(0.606, 0.714) |
| **Malignant calcification** | n=264 |
| Accuracy (95% CI) | 0.769(0.726, 0.812) |
| Sensitivity (95% CI) | 0.788(0.739, 0.837) |
| Specificity (95% CI) | 0.722(0.638, 0.807) |
| Positive predictive value (95% CI) | 0.874(0.832, 0.916) |
| Negative predictive value (95% CI) | 0.582(0.499, 0.666) |
| **Total malignant lesions** | n=468 |
| Accuracy (95% CI) | 0.769(0.740, 0.799) |
| Sensitivity (95% CI) | 0.726(0.686, 0.767) |
| Specificity (95% CI) | 0.836(0.795, 0.878) |
| Positive predictive value (95% CI) | 0.872(0.839, 0.905) |
| Negative predictive value (95% CI) | 0.666(0.619, 0.713) |
| †Patients with calcification, not including typical benign calcification. | |

**Table S7：AUC for Each Radiologist and Reader-averaged AUCs for Reading Mammograms Unaided and with AI Support**

|  |  |  |
| --- | --- | --- |
| Radiologist | Read alone | Read with the model |
| A | 0.781 | 0.836 |
| B | 0.765 | 0.824 |
| C | 0.829 | 0.877 |
| D | 0.821 | 0.794 |
| E | 0.775 | 0.865 |
| F | 0.793 | 0.828 |
| G | 0.891 | 0.905 |
| H | 0.796 | 0.852 |
| I | 0.889 | 0.891 |
| J | 0.777 | 0.893 |
| K | 0.788 | 0.846 |
| L | 0.758 | 0.812 |
| Average | 0.805 | 0.852 |

**Table S8：Mean Sensitivity and Specificity across Radiologists**

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Radiologists alone | Radiologist with the model | P value |
| Sensitivity | 68.70% | 68.78% | 0.937 |
| Specificity | 82.05% | 88.34% | 0.005 |

**Table S9: The sensitivity and specificity of 12 radiologists read alone and read with the model**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Radiologist Alone | | Radiologist with the model | |
| Radiologist | Sensitivity(%) | Specificity(%) | Sensitivity(%) | Specificity(%) |
| A | 72.9 | 76.9 | 58.5 | 98.5 |
| B | 38.6 | 91.5 | 35.7 | 93.8 |
| C | 62.9 | 86.9 | 77.1 | 96.2 |
| D | 70 | 80 | 47.1 | 84.6 |
| E | 60 | 85.4 | 67.1 | 96.2 |
| F | 67.1 | 81.5 | 60 | 85.4 |
| G | 94.3 | 82.3 | 95.7 | 82.3 |
| H | 70 | 80 | 81.4 | 81.5 |
| I | 98.6 | 78.5 | 100 | 78.5 |
| J | 51.4 | 86.2 | 65.7 | 88.5 |
| K | 64.3 | 80 | 77.1 | 81.5 |
| L | 74.3 | 75.4 | 60 | 93.1 |

**Table S10: The performance of the model in prospective clinical application in each center (by patient)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | prospective application performance(n=5746) | | | | | |
|  | Center A(n=1450) | Center B(n=1454) | Center C(n=958) | Center D(n=1098) | Center E(n=279) | Center F(n=507) |
| Accuracy (95% CI) | 0.959（0.948，0.969） | 0.959（0.949，0.969） | 0.986（0.979，0.994） | 0.970（0.960，0.980） | 0.941（0.913，0.969） | 0.989（0.979，0.998） |
| Sensitivity (95% CI) | 0.754（0.699，0.810） | 0.674（0.575，0.773） | 0.806（0.714，0.897） | 0.690（0.551，0.830） | 0.742（0.588，0.896） | 0.778（0.642，0.914） |
| Specificity (95% CI) | 0.965（0.954，0.975） | 0.979（0.971，0.986） | 0.984（0.976，0.992） | 0.985（0.977，0.992） | 0.976（0.957，0.995） | 0.996（0.990，1.002） |
| Positive predictive value (95% CI) | 0.800（0.747，0.853） | 0.667（0.568，0.766） | 0.806（0.714，0.897） | 0.644（0.505，0.784） | 0.793（0.646，0.941） | 0.933（0.844，1.023） |
| Negative predictive value (95% CI) | 0.955（0.943，0.966） | 0.980（0.972，0.987） | 0.984（0.976，0.992） | 0.988（0.981，0.994） | 0.968（0.946，0.990） | 0.983（0.972，0.995） |

**Additional Figure Captions**

Figure S1. Selection of 200 patients.

Figure S2. The diagnostic module of the algorithm.

Figure S3: Flowchart of cross testing between the model and 12 doctors

Figure S4: The FROC curve of detection in calcification and mass in 6 centers

Figure S5 :(a) the detection performance of the model in prospective data in multicenter (b) the classification performance of the model in prospective data in multicenter

Figure S6：ROC curves of the model’s diagnosing performance in center (a)~(f).

Figure S7. A 40-year-old woman with a lobulated mass (red arrow) in the left external breast quadrant on the CC view, and an ambiguous lesion on the MLO view, was pathologically confirmed as invasive apocrine carcinoma by biopsy. Only 4 of the 12 doctors detected the lesion when they read the film alone. When doctors read with the model, 9 of 12 doctors detected the disease. (a) CC view, with local zoom image in the upper right corner.(b) MLO view.(c) CC view, the yellow line is the detection result of the model .

Figure S8: A 51-year-old woman with clustered calcification (red arrow) in the upper quadrant of her left breast. No significant changes were observed in the 2-year follow-up, which was considered to be benign. When the doctors read alone, 9 of the 12 doctors misjudged it as malignant lesions. The model accurately detected the lesion and suggested that the lesion type was calcification, and the possibility of malignancy was 1.8%. Bi-rads3 was recommended. With the model, three of the original nine doctors changed right.(a) CC view of the first mammogram.(b) MLO view of the first mammogram.(c) CC view, detection results of the model, and yellow lines delineated the lesion range for the model.(d) MLO view, detection results of the model, and yellow lines delineated the lesion range for the model.(e) mammogram images at 1 year follow-up.(f) mammographic images at two-year follow-up.