**Supplementary Data 2.** Supplementary methods

**Imaging acquisition and interpretation of dynamic contrast-enhanced (DCE) magnetic resonance imaging (MRI)**

The MRI sequences include an axial T2-weighted imaging (T2WI) sequence, an axial non-fat-suppressed T1-weighted imaging (T1WI) sequence, and an axial fat-suppressed dynamic 3D T1-weighted spoiled gradient-echo sequence of one unenhanced and 2–5 contrast-enhanced acquisitions, with a temporal resolution of 60 seconds. The DCE series are acquired with a maximum acquisition voxel size of $1 \times 1 \times 1$ mm and the first series is acquired before 120 seconds.

DCE MRI interpretation is based on both qualitative morphology and quantitative kinetic pattern analysis according to the BI-RADS lexicon [21]. A dedicated breast MRI computer-aided diagnosis system is used for kinetic pattern analysis.

**MR imaging data collection**

For the quality control test of the participating institutions and later computer-assisted analysis and artificial intelligence research, the MRI data are transmitted to the central lab of Asan Medical Center.

**Supplementary Table.** DWI sequences of 3 vendors used in DWIST

| Equipment          | Siemens          | Philips          | GE         |
|--------------------|------------------|------------------|------------|
| DWI technique      | 3T Skyra         | 3T Ingenia CX    | 3T MR750w  |
|                    | Simultaneous     | ss-EPI (SENSE)   | ss-EPI (ASSET) |
|                    | multislice rs-EPI (SMS RESOLVE) |               |
| Fat suppression    | CHESS            | SPAIR            | SPAIR      |
| Repetition time (TR)/echo time (TE) (msec) | 6,440/61 | >9,000/61.5 | 8,500/76 |
| Field of view (mm$^2$) | $340 \times 207$ | $340 \times 212$ | $340 \times 204$ |
| Flip angle (excitation/refocusing) | 90/180 | 90/180 | 90/180 |
| Averages           | 2/2/4            | 6                | 2/4/4      |
| Thickness (mm) | 2.5 | 3 | 3 |
|---------------|-----|---|---|
| Intersection gap (%) | 0 | 0 | 0 |
| Matrix | $256 \times 156$ | $256 \times 160$ | $256 \times 152$ |
| Acquisition/reconstruction voxel size (mm$^3$) | $1.3 \times 1.3 \times 2.5$ | $1.3 \times 1.3 \times 3$ | $1.3 \times 1.3 \times 3$ |
| Slices | 60 | 50 | 50 |
| b values (sec/mm$^3$) | 0, 800, 1,200 | 0, 800, 1,200 | 0, 800, 1,200 |
| Acquisition time (min) | 6:54 | 6:31 | 7:20 |

CHESS = chemical shift (spectral) selective; DCE = dynamic contrast-enhanced; DW = diffusion-weighted; EPI = echo-planar imaging; SPAIR = spectral adiabatic inversion recovery and spectral attenuated inversion recovery.

**Image interpretation and management**

Image interpretation is performed on a per-breast basis and cases of bilateral lesions are considered as two different breasts. In cases with multiple lesions in the breast, the most suspicious finding is chosen as the final category. On each modality, a negative result is defined as a BI-RADS category of 1 (negative) or 2 (benign). All participants with a BI-RADS category of 4 (suspicious) or 5 (highly suggestive of malignancy) are recalled for additional workup and biopsy. On mammography or ultrasonography, participants with a BI-RADS category of 3 (probably benign) are followed in 6 months using the corresponding modality according to standard practice. In participants with a BI-RADS category of 3 on DCE or DW MRI, double reading is performed; if there is consensus on a category of 3, follow-up imaging with MRI after 6 months is planned. The results of the follow-up MRI are reported as either negative (BI-RADS category of 1 or 2; return to the regular screening program) or positive (BI-RADS category of 4 or 5; recall for additional workup and biopsy). Biopsy of the lesion seen only at DCE or DW MRI is at first performed by means of targeted ultrasonography. A marker clip is placed after biopsy if the attending radiologist determines it necessary. In cases of imaging-histologic discordant results, the biopsy location is confirmed in post-biopsy non-fat-saturated T1WI, with immediate MRI-guided rebiopsy in cases of incorrect location.
**DW MRI interpretation**

In the qualitative assessment, unique areas of high signal intensity are identified on the b value of 1,200 sec/mm² DW images in MIPs, and the lesion type is classified as mass or focus versus non-mass [24]. Masses are classified as irregular or oval/round and homogeneous or heterogeneous based on their morphology and internal characteristics. Non-mass lesions are classified as either segmental/linear or focal/regional/diffuse based on their distribution. In quantitative assessment, the ADC values are measured in small (3–10 mm²) circular regions of interest (ROI) within the darkest part of the lesion on the ADC map, thus avoiding both artifact and necrotic or hemorrhagic parts of the lesion. Previous studies suggested that this method for the selection of the lowest ADC part within the lesion may reduce inter- and intra-reader variability and improve breast DW MRI consistency and comparability between sites [23]. Diffusion level is classified into very low, low, intermediate, high, and very high according to the EUSOBI guideline [23] and an ADC value of 1.3 mm²/sec is used as a cutoff for the differentiation between benign and malignant lesions. The focus is evaluated based on both signal intensity at a b value of 0 sec/mm² DW images and ADC map. Of note, lesion classification is not based on diffusion level alone, but rather with all the information available in the DW MRI data. Finally, we classify negative finding as BI-RADS category 1, typical benign findings as BI-RADS category 2, lesions with only one suspicious criterion as BI-RADS category 3, lesions with two suspicious criteria as BI-RADS category 4, and those with all three suspicious criteria as BI-RADS category 5.

**Definition of outcome measures**

Sensitivity is estimated as the fraction of participants with cancer (invasive or DCIS) in whom the imaging modality result is positive (BIRADS category of 3, 4, or 5) for a location that matches the location of the cancer indicated by the reference standard [21]. Specificity is estimated as the fraction of participants without cancer by the reference standard in whom the imaging modality result is negative (BI-RADS category of 1 or 2). Diagnostic accuracy is assessed by the area under the receiver operating characteristic (ROC) curve (AUC). PPV1 is defined as the percentage of women with detected cancers among those with positive imaging results. PPV2 is calculated based on the recommendation for tissue
diagnosis. The rate of invasive cancer detection is estimated as the fraction of participants with invasive cancer and a positive test result at the location of the cancer indicated by core or surgical biopsy. The AIR or recall rate is defined as the proportion of women interpreted as having a BI-RADS category of 3, 4, or 5. In addition, lesion characteristics of false-negative and false-positive findings on DW MRI will be analyzed. An interval cancer is defined as breast cancers diagnosed due to clinical symptoms after the last negative examination result in all imaging modalities but before the next scheduled examination [21]. Biologic characteristics of the breast cancers will be evaluated with respect to tumor size, histologic type, tumor grade, molecular subtype, and lymph node metastasis on surgery.