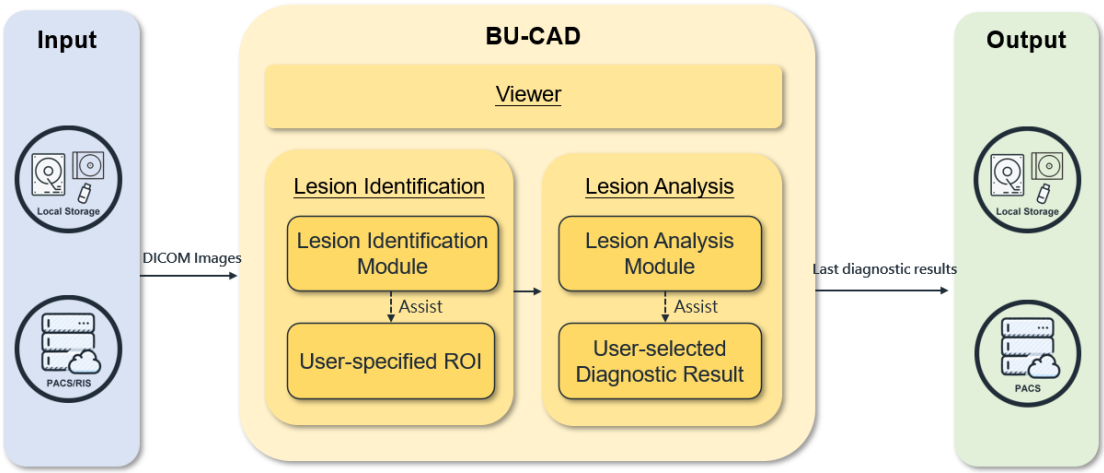
Evaluation of TaiHao Breast Ultrasound Diagnosis Software

NCT number: NCT04551105

Document date: March 2, 2022

I. Device Description

BU-CAD developed by TaiHao Medical Inc. is a software system designed to assist users in analyzing breast ultrasound images including identification of regions suspicious for breast cancer and assessment of their malignancy. The following figure shows the architecture chart of BU-CAD which consists of a Viewer, a Lesion Identification Module, and a Lesion Analysis Module.



Architecture chart of BU-CAD

The Viewer is able to load breast ultrasound and mammography images (FDA-cleared full-field digital mammography only) from local storage or a picture archiving and communication system (PACS) for review. The Viewer also includes tools that allow users to measure lesion size and adjust the image (such as window level and window width adjustment). Additionally, the report may be saved in local storage or uploaded to PACS. BU-CAD also supports exporting CAD results to third-party reporting software to facilitate the reporting process.

The Lesion Identification Module identifies regions of interest (automated ROIs) of a single suspicious soft tissue lesion in up to two orthogonal views of breast ultrasound images for assisting users in detecting soft tissue lesions. Additionally, the Lesion Identification Module generates an ROI and a lesion contour on each breast ultrasound image. The lesion contour on each image will be automatically delineated by the given ROI. The Lesion Analysis Module analyzes given ROIs of a breast lesion on ultrasound images, and generates a score of lesion characteristics (SLC) in terms of malignancy or benignity of a lesion, BI-RADS category, and BI-RADS descriptors (with limitations as described in the User Manual) for the concurrent read. The users are able to replace the automated ROIs with re-delineated rectangular ROIs for

analysis by Lesion Analysis Module. Only the last analysis results will be displayed on the user interface and are modifiable by the user. Note that the SLC is analyzed based on the rectangular ROIs, unless the user re-delineates the ROIs, the SLC will not be changed.

In clinical practice, after opening multi-modality digital images including ultrasound and mammography on the Viewer, the users may identify and analyze lesions with the assistance of the Lesion Identification Module and Lesion Analysis Module on the breast ultrasound images. Finally, the user confirms the diagnostic results (output from Lesion Analysis Module or modified by the user) shown on the user interface and saves them to the report.

Output of BU-CAD analysis

Region-based Analysis Item	Range
Score of lesion characteristics (SLC)	[0,100]
BI-RADS category	2 / 3 / 4a / 4b / 4c / 5
BI-RADS descriptors (mass)	Shape, Orientation, Margin, Echo Pattern, Posterior Features

II. Indications for Use

BU-CAD is a software application indicated to assist trained interpreting physicians in analyzing the breast ultrasound images of patients with soft tissue breast lesions suspicious for breast cancer who are being referred for further diagnostic ultrasound examination.

Output of the device includes regions of interest (ROIs) and lesion contours placed on breast ultrasound images assisting physicians to identify suspicious soft tissue lesions from up to two orthogonal views of a single lesion, and region-based analysis of lesion malignancy upon the physician's query. The region-based analysis indicates the score of lesion characteristics (SLC), and corresponding BI-RADS categories in user-selected ROIs or ROIs automatically identified by the software. In addition, BU-CAD also automatically classifies lesion shape, orientation, margin, echo pattern, and posterior features according to BI-RADS descriptors.

BU-CAD may also be used as an image viewer of multi-modality digital images, including ultrasound and mammography. The software includes tools that allow users to adjust, measure and document images, and output into a structured report (SR).

Patient management decisions should not be made solely on the basis of analysis by BU-CAD.

Limitations: BU-CAD is not to be used on sites of post-surgical excision, or images with Doppler, elastography, or other overlays present in them. BU-CAD is not intended for the

primary interpretation of digital mammography images. BU-CAD is not intended for use on mobile devices.

III. Clinical Performance Data

◆ Summary of the Reader Study

The performance of physicians without and with the aid of BU-CAD decision support in interpreting breast ultrasound images was compared by using a fully crossed multi-reader multi-case receiver operating characteristic (MRMC-ROC) retrospective study (also known as Obuchowski-Rockette Dorfman-Berbaum-Metz MRMC-ROC or OR-DBM MRMC-ROC).

The study consisted of 628 cases, of which 456 cases (189 malignant and 267 benign) were collected from the United States and 172 cases (65 malignant and 107 benign) were collected from Taiwan. Sixteen readers participated in the study. Each reader was asked to identify the lesion, provide a linear score of lesion characteristics (SLC), select a BI-RADS category and select BI-RADS descriptors for an ultrasound breast lesion with or without the aid of BU-CAD.

Dataset Demographic

A total of 628 cases collected from two institutions were used in the reader study. The source of cases is listed below.

- U.S.: 456 cases
- Taiwan: 172 cases

The BI-RADS category distribution included in this study were listed below:

- BI-RADS 2: 5 cases
- BI-RADS 3: 123 cases
- BI-RADS 4A: 204 cases
- BI-RADS 4B: 111 cases
- BI-RADS 4C: 105 cases
- BI-RADS 5: 80 cases

The number of benign and malignant cases included in this study were listed below.

- Benign cases
    - Pathology proof benign: 197 cases
    - Two-year follow-up benign: 177 cases
- Malignant cases

- Ductal carcinomas in situ (DCIS): 17 cases
- invasive ductal carcinoma (IDC): 193 cases
- Invasive lobular carcinoma (ILC): 40 cases
- Other cancer types: 4 cases

The imaging hardware distribution included in this study were listed below:

- GE: 451 cases
- Acuson: 5 cases
- Philips: 100 cases
- Canon/Toshiba: 72 cases

Reader Experience

Study Reader	Specialty	MQSA	Received Breast Image Fellowship	Year of experience as a radiologist
Dr. X01	Radiologist	Yes	No	24
Dr. X02	Radiologist	Yes	Yes	3
Dr. X03	Radiologist	Yes	No	13
Dr. X04	Radiologist	Yes	No	14
Dr. X05	Radiologist	Yes	No	8
Dr. X06	Radiologist	Yes	Yes	5
Dr. X07	Radiologist	Yes	Yes	2
Dr. X08	Radiologist	Yes	No	10
Dr. X09	Radiologist	Yes	Yes	12
Dr. X10	Radiologist	Yes	No	11
Dr. X11	Breast Surgeon	No	No	> 30 (breast surgeon)
Dr. X12	Breast Surgeon	No	No	> 30 (breast surgeon)
Dr. X13	Radiologist	Yes	No	21
Dr. X14	Radiologist	Yes	No	1
Dr. X15	Radiologist	Yes	No	13
Dr. X16	Radiologist	Yes	No	5

Primary Objective

The primary objective of this clinical study is to prove that the user's performance (AUC of location-specific ROC) aided by the BU-CAD software is superior to the unaided performance. The aided AUC of the location-specific ROC for BU-CAD was superior to that of the unaided scenario for the diagnosis of breast ultrasound images. The mean AUC of location-specific ROC shift of 0.0374.

Primary Results of the Pivotal Study

Reading Scenario	AUC_LROC	95% CI	*p*-value
Unaided	0.7786	(0.7463, 0.8109)	
Aided	0.8160	(0.7862, 0.8458)	
Aided – Unaided	0.0374	(0.0190, 0.0557)	0.0001

Subgroup Analysis

Subgroup of reader specialty (with and without MQSA certification), with and without breast image fellowship training, ultrasound systems (GE, Acuson, Philips, and Canon/Toshiba), benign types (pathology proof benign and two-year follow-up benign), cancer types (DCIS, IDC, ILC, and others), lesion sizes (less than 1 cm, between 1 cm and 2 cm, and larger than 2 cm), lesion locations (center and not in center), ages (≤ 50 years, > 50 years, ≤ 55 years, and >55 years), and source of cases (U.S. and Taiwan) were performed. Except for the subgroup of Acuson ultrasound system, where the sample size was relatively low, the readers aided by the BU-CAD achieved higher performance than unaided reading in the other subgroups.

Secondary Objective

The secondary objective of this clinical study is to compare that the user's performance (sensitivity, specificity, PPV, and NPV) between the unaided and aided readings. Sensitivity, specificity, PPV, and NPV produced from the aided arm were higher than unaided. The specificity, unadjusted PPV, and unadjusted NPV differed significantly from zero between the aided and unaided sessions.

Sensitivity, Specificity, PPV, and NPV between Unaided and Aided Reading Scenarios

Statistical Parameter	Unaided (95% CI)	Aided (95% CI)
Sensitivity	0.9225 (0.8896, 0.9554)	0.9353 (0.9050, 0.9655)
Specificity	0.3165 (0.2694, 0.3636)	0.3611 (0.3124, 0.4098)
NPV (unadjusted)	0.8623 (0.8048, 0.9198)	0.8945 (0.8456, 0.9434)
NPV_U.S. (adjusted)	0.9982 (0.9902, 1.0000)	0.9986 (0.9918, 1.0000)
NPV_Taiwan (adjusted)	0.9969 (0.9767, 1.0000)	0.9975 (0.9809, 1.0000)
PPV (unadjusted)	0.4876 (0.4433, 0.5319)	0.5056 (0.4607, 0.5505)
PPV_U.S. (adjusted)	0.0108 (-0.0001, 0.0216)	0.0113 (0.0000, 0.0225)
PPV_Taiwan (adjusted)	0.0256 (-0.0002, 0.0514)	0.0283 (0.0006, 0.0560)

Although the specificity in the aided scenario is 36.11%, the following confusion table summarizes the event count from a false-positive (FP) unaided to a true-negative (TN) when aided by BU-CAD or a reverse for all 374 benign cases. A total of 790 FP events unaided were changed to TN events aided by BU-CAD for all 16 readers, and a total of 523 TN events unaided were changed to FP events aided by BU-CAD for all 16 readers. The overall benefit

was +267 events and shows that BU-CAD is able to assist the majority of readers in reducing false positives even for datasets where readers have a low specificity performance in the unaided scenario.

Confusion Table FP to TN Net Benefit for Benign Cases

All benign (374)	X01	X02	X03	X04	X05	X06	X07	X08	X09	X10	X11	X12	X13	X14	X15	X16	Total
FP (unaided) → TN (aided)	83	24	93	23	86	33	44	73	34	46	30	46	24	41	69	41	790
TN (unaided) → FP (aided)	33	28	18	70	16	67	16	52	38	33	22	47	4	20	17	42	523
Difference	50	-4	75	-47	70	-34	28	21	-4	13	8	-1	20	21	52	-1	267

In addition, BU-CAD software was found to significantly decrease readers' interpretation times (by ~40%) which was shown in analyses including and excluding outliers. Statistical analyses also indicated that BU-CAD improved readers' determination of BI-RADS descriptors (Shape, Orientation, Margin, Echo Pattern, and Posterior Features), where at least one or more subcategories for each descriptor demonstrated improved aided read performance, with limitations described in the User Manual.

Accuracy of BI-RADS Descriptors

Reading Scenario	Shape	Orientation	Margin	Echo Pattern	Posterior Features
Unaided	78.14%	82.15%	79.22%	76.49%	66.51%
Aided	78.92%	82.20%	77.34%	66.52%	67.53%
BU-CAD Standalone	71.91%	75.24%	73.57%	66.73%	58.03%

◆ **Summary of the Standalone Study**

A total of 1139 cases (628 reader study cases plus 511 extended cases) collected from multiple institutions were used in the standalone study.

Dataset Demographic

The source of cases is listed below.

- North America: 531 cases
- Europe: 36 cases
- Taiwan: 572 cases

The BI-RADS category distribution included in this study were listed below:

- BI-RADS 2: 31 cases
- BI-RADS 3: 223 cases

- BI-RADS 4A: 356 cases
- BI-RADS 4B: 218 cases
- BI-RADS 4C: 181 cases
- BI-RADS 5: 130 cases

The number of benign and malignant cases included in this study were listed below.

- Benign cases
    - Pathology proof benign: 465 cases
    - Two-year follow-up benign: 177 cases
- Malignant cases
    - Ductal carcinomas in situ (DCIS): 53 cases
    - invasive ductal carcinoma (IDC): 361 cases
    - Invasive lobular carcinoma (ILC): 51 cases
    - Other cancer types: 32 cases

The imaging hardware distribution included in this study were listed below:

- GE: 634 cases
- Siemens: 188 cases
- Canon/Toshiba: 90 cases
- Philips: 111 cases
- Supersonic: 24 cases
- Others: 92

Lesion Identification Module (CADe) Performance

A total of 59 benign cases (including 11 of the 20 missing cases) and 18 malignant cases (including 9 of the 20 missing cases) did not meet the objective performance criteria (automated ROI center must be within ground truth ROI with at least 50% overlap in ROI area). The accuracy of the lesion identification algorithm was 93.24% (1062/1139). For the LROC analysis, 18 malignant cases were penalized due to wrong location or undetected by BU-CAD.

Comparison between Standalone and Unaided Reading Performance

The standalone performance of BU-CAD was measured in AUC_LROC on the 628 reader study cases and the standalone study cases (combined the 628 reader study cases and 511 extended cases), a total of 1,139 cases (497 malignant and 642 benign). Table below shows the standalone AUC_LROCs in both datasets are higher than that of unaided reading performance.

Standalone and Unaided Reading Performances

Reading Scenario	AUC_LROC	95% CI
BU-CAD Standalone (628 reader study cases)	0.7987	(0.7626, 0.8348)
BU-CAD Standalone (1,139 standalone study cases)	0.8203	(0.7947, 0.8458)
Unaided Reading (628 reader study cases)	0.7786	(0.7463, 0.8109)

Summary of Subgroup Analysis

Subgroup of the different ultrasound systems (GE, Siemens, Canon/Toshiba, Philips, Supersonic, and others), benign types (pathology proof benign and two-year follow-up benign), cancer types (DCIS, IDC, ILC, and others), lesion size (less than 1 cm, between 1 cm and 2 cm, and larger than 2cm), Lesion Locations (center and not in center), view type (two view vs. single view), ages (≤ 50 years, > 50 years, ≤ 55 years, and >55 years), and sources of cases (North America, Europe, and Taiwan) were performed. The performance of distinguishing between benign and malignant in Siemens ultrasound system, DCIS and ILC cancer type, cases where the lesion is not in the center, two-orthogonal views, and source of North America and Europe achieved acceptable discrimination (AUC_LROC from 0.7 to 0.8). The remaining subgroups achieved excellent (AUC_LROC from 0.8 to 0.9) or outstanding (AUC_LROC > 0.9) discrimination.

Sensitivity, Specificity, PPV, and NPV

The standalone performances of sensitivity and specificity were assessed by using the 1,139 cases and summarized in Table 9. Results show the standalone sensitivity and specificity were 88.53% and 57.94%. In addition, the adjusted PPV of U.S. and Taiwan were 1.28% and 4.74% respectively, the adjusted NPV of U.S. and Taiwan were 99.83% and 99.67% respectively. Because both the prevalence rates of U.S. and Taiwan are relatively low, the adjusted PPVs were relatively low and the adjusted NPVs were relatively high. However, the standalone PPVs in U.S. and Taiwan were higher than those of unaided and aided scenarios.

Standalone Sensitivity, Specificity, PPV, NPV

Statistical Parameter	Standalone (Frequency)	95% CI
With Modification for Wrong-location Penalty)		
Sensitivity (%)	88.33 (439/497)	(0.8551, 0.9115)
Specificity (%)	57.94 (372/642)	(0.5413, 0.6176)
PPV (%) [unadjusted]	61.92 (439/709)	(0.5834, 0.6549)
PPV_US (%)	1.28	(0.0011, 0.0245)*
PPV_TW (%)	4.74	(0.0246, 0.0703)*
NPV (%) [unadjusted]	86.51 (372/430)	(0.8328, 0.8974)
NPV_US (%)	99.82	(0.9921, 1.0000)*
NPV_TW (%)	99.67	(0.9895, 1.0000)*

* The 95% Confidence Interval (CI) was estimated conditioning on the obtained prevalence rates of 0.72% and 1.94% in U.S. and Taiwan, respectively.

The following table showed the calculated sensitivity and specificity using each BI-RADS category as the threshold. Since the clinical decision threshold for cancer vs. non-cancer is BI-RADS 3 vs BIRADS 4a and the BI-RADS fifth edition concluded that patients with category ≥ 4a lesions are recommended to undergo biopsy, the analysis of sensitivity and specificity are still based on BI-RADS 4a as the cutoff point (i.e., a BI-RADS category of 4a or higher defines a positive call for cancer diagnosis).

Standalone Sensitivity and Specificity by Using Different Cut-Off Points

Statistical Parameter	3	4A*	4 B	4C	5
Sensitivity	0.9416 (0.9210, 0.9623)	0.8833 (0.8551, 0.9115)	0.8249 (0.7915, 0.8584)	0.6962 (0.6557, 0.7366)	0.4588 (0.4149, 0.5026)
Specificity	0.3302 (0.2938, 0.3666)	0.5794 (0.5413, 0.6176)	0.6994 (0.6639, 0.7348)	0.8271 (0.7979, 0.8564)	0.9252 (0.9049, 0.9456)

* The cut-off value used in the standalone study.

Robustness of the Lesion Analysis Module (CADx)

To evaluate the robustness of the CADx algorithm (Lesion Analysis Module) when different rectangular ROIs are drawn around the same lesion on a given single-view image or two-view images, two reproducibility experiments of the same lesion cropped by different rectangular ROIs were conducted. In the first reproducibility experiment, each corner point of an ROI was shifted by randomly changing the horizontal and vertical dimensions up to 20% respectively from the ground truth ROI defined by the expert panel. The experiment was repeated 20 times with all 1139 test cases (the original dataset was 628 cases and the

extended dataset was 511 cases). The results show that randomly enlarging the width and height of the ROIs did not affect the performance of the BU-CAD CADx algorithm (Lesion Analysis Module). The AUC remained stable between 0.840 and 0.846.

In the second reproducibility experiment, each corner point of ground truth ROI was altered by systematically shrinking the horizontal and vertical dimensions respectively from 1% to 30%. The experiment was conducted with all 1139 cases. The new ROIs and their corresponding images were then processed by the BU-CAD CADx algorithm (Lesion Analysis Module) to produce analysis outputs. The results show that as long as the shrinking percentage of the width and height of the ROIs is within 16%, the AUC remained above 0.8.

IV. Conclusions

The aided AUCs of the location-specific ROC for BU-CAD were superior to that of the un-aided reads for the diagnosis of breast ultrasound images.

BU-CAD decision support device was found to significantly decrease readers' interpretation times (by ~40%) which were shown in both analyses of including and excluding outliers.

Sensitivity, specificity, PPV, and NPV produced from the aided arm were higher than unaided. The specificity, PPV, and NPV differed significantly from zero between the aided and unaided sessions.

Statistical analyses also indicated that BU-CAD improved readers' determination of BI-RADS descriptors (shape, orientation, margin, echo pattern, and posterior features). The Orientation of the BI-RADS descriptor in both unaided and aided reading was in substantial agreement with ground truth. The Shape and Orientation rated by readers who have received breast image fellowship training in the aided scenario were in substantial agreement with ground truth.