International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(23s) 2026 pp. 31-47
https://doi.org/10.70917/ijcisim-2026-5550
Article

A Novel Reconfiguration of Micro-Strip Patch Antenna with CSRR Metamaterial for Breast Cancer Detection

Jaimini shah[1], Parikshit Vasisht[2]

[1]School of Engineering and Technology, Apeejay Stya University, Sohna, Gurugram. Email: jaimini.shah@asu.apeejay.edu
[2]School of Engineering and Technology, Apeejay Stya University, Sohna, Gurugram. Email: parikshit.vasisht@asu.apeejay.edu

Abstract: Microwave resonant sensing and explainable machine learning can provide complementary evidence for breast-cancer decision support, but the two evidence streams are frequently conflated. This study proposes a dual-state microstrip patch antenna loaded with a complementary split-ring resonator (CSRR) and an independently reproducible artificial-intelligence branch. The antenna uses PIN-diode reconfiguration to alternate between nominal 2.45-GHz and 3.55-GHz interrogation states. An equivalent-circuit model predicts resonance shifts of 166.0 MHz and 293.1 MHz, respectively, for a malignant-like contrast load relative to air; these values are analytical phantom contrasts, not measured diagnostic thresholds. The executable classification study uses the public Kaggle/UCI Wisconsin Diagnostic Breast Cancer dataset (569 cases, 30 image-derived nuclear descriptors). The descriptors are arranged in their native 3 × 10 mean–standard-error–worst map. A leakage-aware, equal-score TriFusion combines balanced logistic regression, RBF support-vector machine, and random forest. Ten repetitions of five-fold stratified validation produced an area under the receiver-operating-characteristic curve of 0.9957, accuracy of 97.95% (95% CI: 97.83–98.06%), balanced accuracy of 97.68%, sensitivity of 96.65%, and F1 score of 97.22%. It achieved the highest mean accuracy, balanced accuracy, and F1 among the same-protocol models; logistic regression had marginally higher sensitivity and the lowest Brier score, KNN had the highest specificity and precision, and RBF-SVM had a statistically indistinguishable AUC. A convolutional surrogate reached 98.25% fixed-test accuracy and 99.12% class agreement with the TriFusion teacher. Feature-map Grad-CAM was supported by an activation-deletion test: masking the five most salient cells caused a larger target-logit to decrease than random masking (one-sided Wilcoxon p = 9.59e-21). The resulting architecture is an evidence-bounded proof of concept rather than an end-to-end antenna-to-diagnosis system; fabrication, full-wave simulation, tissue-phantom experiments, and co-registered RF data remain necessary
Keywords: Breast cancer detection; complementary split-ring resonator; microstrip patch antenna; metamaterial sensor; WDBC; ensemble learning; Grad-CAM; statistical validation

1. INTRODUCTION

Breast cancer is a significant public health and clinical-management problem, and pathways need to be sensitive, specific, accessible, cost-effective and interpretable [1]. Each of the imaging modalities – mammography, ultrasonography, magnetic-resonance imaging, and tissue sampling – answers a different clinical question. Low-profile antennas can be embedded in small non-ionizing measurement platforms, and the complex permittivity and conductivity of tissue are known to be dependent on tissue composition, so microwave sensing is being explored as a complementary modality [2] – [6]. These benefits do not solve the main problem: the breast is a heterogeneous tissue, the distribution of dielectric properties overlap, and the resonant shift is not in itself a validated diagnosis.

Planar sensors can be made more sensitive to electric-field concentration by using complementary split-ring resonators which is a compact way to do so [7]–[11]. If a CSRR is etched in or coupled to the radiating structure, the extra LC path can generate a strong resonance whose frequency and notch depth is sensitive to the loading of the surrounding materials. Semiconductor switches can be used to reconfiguration to produce greater than one

interrogation state without increasing the aperture size. The multi-state response is appealing for contrast sensing as it allows observing a material perturbation at separated frequencies, with different field distributions.

The second research stream is based on the classification of public breast-cancer datasets by machine learning. The Wisconsin Diagnostic Breast Cancer (WDBC) benchmark is composed of 30 morphometric variables that are derived from digitized fine-needle-aspiration images [12, 13]. High headline accuracies are typical, but these can be misleading if there is preprocessing before data splitting, hyperparameters are tweaked on the test set or results from different partitions are added to the same table, if they are unrelated. Furthermore, the post hoc heat map is only helpful if the model to be explained and the spatial interpretation of the input are well specified.

This article thus has a hybrid design, with a bound on the evidence. A transparent equivalent-circuit model is used for evaluating the antenna branch. The artificial-intelligence branch is run on the public Kaggle copy of WDBC, which is the same set of UCI records distributed by the deterministic scikit-learn package. The two branches are not shown as two patient measurements. Instead, the AI experiment confirms the classification, uncertainty, statistical-comparison, and explainability modules which a future antenna dataset may re-use. In this article, the term “detection” refers to binary classification of malignancy versus benignity and not to localization of objects (lesions) or a clinically accepted screening endpoint.

The principal contributions are:

• A dual-state PIN-reconfigured CSRR-loaded patch with explicit substrate, patch, ring, gap, and switch parameters, together with reproducible analytical S11 predictions under contrast-spanning loads.

• A native 3 × 10 representation of WDBC descriptors that preserves the mean, standard-error, and worst-value organization of ten nuclear-shape descriptors.

• An architecture-defined equal-score TriFusion of balanced logistic regression, RBF-SVM, and random forest, plus a high-fidelity convolutional surrogate for feature-map Grad-CAM.

• Leakage-aware 10 × 5 repeated stratified validation, repeat-level confidence intervals, corrected resampled tests with Holm multiplicity control, and a paired Grad-CAM activation-deletion experiment.

The structure of the paper is as follows. Section II introduces the concepts of microwave sensing, CSRR resonators and explainable classification. The reconfiguration of the antenna and the analytical model are defined in Section III. The public dataset, AI pipeline, and statistical protocol are presented in sections IV and V. The results of the antenna, classification and Grad-CAM are reported in Section VI. The experiments needed for clinical translation, limitations and interpretation are discussed in Sections VII and VIII.

2. RELATED WORK AND DESIGN RATIONALE

A. Microwave breast sensing and dielectric contrast

Contrast in the electromagnetic properties of the inclusions was first shown to be able to localize simulated or phantom inclusions in early time domain and confocal microwave systems [2, 3]. Later work in the field of dielectrics revealed that the properties of breast tissue measured are affected by the amount of adipose tissue, the composition of the glands, pathology, temperature, frequency, and measurement protocol [4]–[6]. Thus, single nominal values of the permittivity should not be considered as universal biological parameters, but as parameters of phantom designs. The difference between these is the key to the current study: the analytical loads cover realistic contrast regimes to check the sensor response and are not intended as diagnostic cutoffs.

The effective capacitance of a loaded resonator is increased by a nearby dielectric, and the resonant frequency of the resonator is decreased, usually, and the resonance is broadened and shallower by conductive loss. It is thus possible to use a compact sensor to extract frequency shift, bandwidth, notch depth and inter-state differences. The use of multi-feature measurements is preferred to a single frequency threshold as it provides calibration, nuisance compensation, uncertainty estimation and multivariate classification.

B. CSRR-loaded planar sensors and reconfiguration

The split-ring resonator paved the way towards artificial magnetic response in patterned conductors [7]. The CSRR, which is strongly coupled to the electric field of planar transmission structures [8] was developed from Babinet-complementary structures. Equivalent-circuit analyses use coupled inductance, capacitance, and loss elements to model the resonator, and offer the physically interpretable connection between the resonator geometry and its resonance [9]. Since then, CSRR inspired dielectric sensors have been applied to magnify small loading induced perturbations in small footprints [10] [11].

Reconfiguration is the addition of a controlled impedance between one or more ring gaps. The first state (low resistance) shortens or reconnects the resonant current path, the second state (capacitive) maintains the discontinuity with the addition of a small junction capacitance. Switching the configuration, therefore, leads to a variation of the effective L–C product and of the spatial concentration of the near field. The proposed design consists of two complementary states which interrogate the same sample at separated resonances. The states are measured one after the other and the plotted “dual state” response is an overlay only, not a simultaneous matching of a dual band.

C. Classification, statistical comparison, and explanation

There are strong baselines for WDBC including linear models, kernel machines, tree ensembles, nearest-neighbor classifiers and gradient boosting [14]–[18]. They have different error patterns: logistic regression is stable and interpretable, RBF-SVM models nonlinear margins and random forests models interactions and tolerates heterogeneous scales of features. Equal-score fusion can help to decrease the model-specific variance without learning another weight vector from a small data set. Here, the fusion is "baked in" to the architecture and is not part of a second layer of test-contaminating weight optimization.

Grad-CAM is a method that uses gradients that flow into a convolutional feature tensor to generate a spatial map for the target class [19]. The most common understanding is pixel localization, but this is only for the CNN input when it is a real image. For this work, a 3 × 10 map of descriptors is used as input. The output thus only shows which descriptor cells support the surrogate decision but will not be able to detect a lesion boundary. But a surrogate explanation is also plausible only if it can be quantified in terms of fidelity to the teacher. The current design presents class agreement, probability correlation and an activation-space deletion test, instead of showing an unvalidated heat map.

Lastly, there are correlations between model comparisons using repeated cross-validation folds. Using all fold scores as independent will underestimate uncertainty. The corrected resampled t-test approximately corrects for the overlap in the train and test sets and the Holm adjustment controls for the family-wise error rate when multiple pairwise comparisons are made [20, 21]. Here these conservative procedures are employed to distinguish between mean-score leadership and superiority supported by statistics.

3. PROPOSED RECONFIGURABLE CSRR ANTENNA

A. Geometry and switching states

A rectangular microstrip patch radiator is proposed on a low loss substrate with nominal relative permittivity 3.55, loss tangent 0.0027 and thickness 1.524 mm. A concentric CSRR is located at the center of the patch. Two switching locations are used to connect complementary gap paths, to reconfigure the effective resonant loop. State A is used for the longer path with the nominal air resonance of 2.45 GHz; State B is used for the shorter path with the nominal air resonance of 3.55 GHz. The PIN-diode parasitics are modeled as a 2.2 Ω series resistance and 0.6 nH inductance in the conducting state and 0.18 pF junction capacitance with the same series inductance in the non-conducting state.

$$W = c/(2f_0) \cdot \sqrt{[2/(\varepsilon_r + 1)]} \quad (1)$$

$$\varepsilon\_eff = (\varepsilon_r+1)/2 + (\varepsilon_r-1)/2 \cdot (1+12h/W)^{-1/2} \quad (2)$$

$$L = c/(2f_0\sqrt{\varepsilon}_{eff}) - 2\Delta L \quad (3)$$

Using $f_0 = 2.45$ GHz gives W = 40.56 mm and L = 32.06 mm. These dimensions are an analytical starting point; a fabricated design would require full-wave optimization of the feed, ground plane, ring coupling, switch pads, bias network, connector transition, and sample holder.

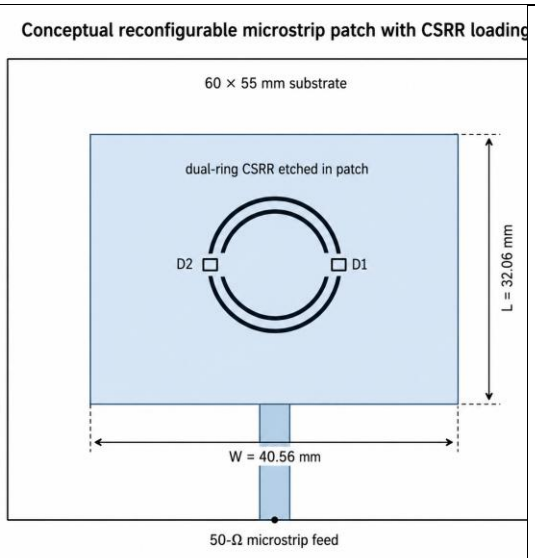

Figure 2: Top-view geometry of the proposed patch and concentric CSRR. Dimensions are in millimeters.

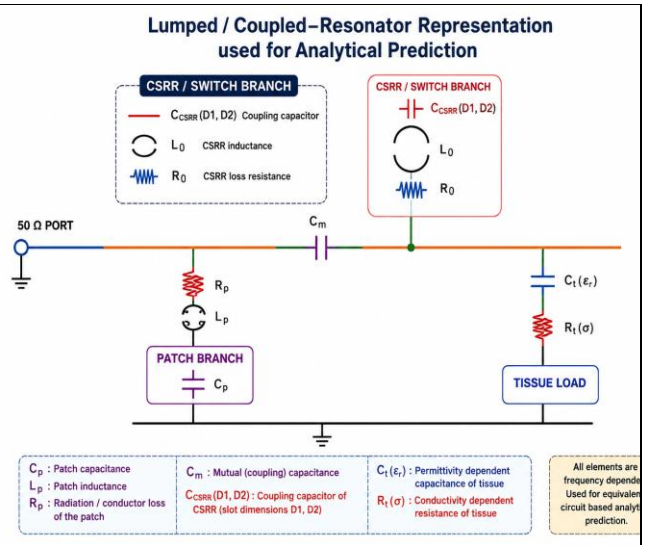

Figure 3: Equivalent-circuit abstraction. The sample-dependent branch modifies capacitance and conductance; the switch selects the effective resonant path.

Table 1: Nominal geometry and switch parameters used by the analytical model.

Parameter	Value
Substrate material	low-loss laminate (nominal Rogers RO4003C-class)
Substrate relative permittivity	3.55
Substrate loss tangent	0.0027
Substrate thickness	1.524 mm
Board size	60.0 × 55.0 mm
Patch size	40.56 × 32.06 mm
Feed width	3.20 mm
CSRR radii	r_o = 6.0 mm; r_i = 4.8 mm
Ring trace / gap / inter-ring spacing	0.55 / 0.80 / 0.65 mm
PIN-diode model	R_{on} = 2.2 Ω; L_s = 0.6 nH; $C_o$ff = 0.18 pF

B. Equivalent circuit and load perturbation model

The unloaded patch is represented by a parallel radiation-loss RLC branch coupled to a CSRR resonator. The CSRR resonance is approximated by

$$f\_CSRR = 1/(2\pi\sqrt{(L\_CSRR\ C\_CSRR)}) \quad (4)$$

and the switch impedance is represented by $Z\_ON = R\_ON + j\omega L\_s$ or $Z\_OFF = j\omega L\_s + 1/(j\omega C\_OFF)$. A nearby material introduces an incremental capacitance C_t and conductance G_t, changing the real and imaginary

parts of the effective resonator admittance. Rather than assigning deterministic biological classes to permittivity, the analytical code uses four contrast-spanning load states: air/reference ($\varepsilon_r = 1$, $\sigma = 0$ S/m), adipose-dominant (6.5, 0.18 S/m), fibroglandular-like (30, 1.05 S/m), and malignant-like inclusion (48, 1.65 S/m). The labels indicate phantom scenarios only.

For state s, the loaded resonance and loss terms are modeled phenomenologically as

$$f_{r,s} = f_{air,s} / \sqrt{[1 + \eta_s(\varepsilon_r - 1)]} \quad (5)$$

with $\eta_A = 0.0032$ and $\eta_B = 0.0040$. Conductivity changes the nominal quality factor according to $Q_A = 34/(1+0.62\sigma)$ and $Q_B = 40/(1+0.58\sigma)$, while the target notch depths are $D_A = -28+4.8\sigma$ dB and $D_B = -25+4.2\sigma$ dB. For $m = 10^{(D/20)}$, $\beta = (1+m)/(1-m)$, and $x = Q(f/f_r - f_r/f)$, the plotted one-port coupled-resonator response is $\Gamma(f) = (\beta-1+jx)/(\beta+1+jx)$, with $S11 = 20\log10|\Gamma|$. The −10-dB bandwidth is extracted from the contiguous sampled interval around the resonance. These equations form a reproducible sensitivity model, not a replacement for electromagnetic field simulation.

C. Evidence status

The antennae results presented here are analytical equivalent-circuit predictions. There is no claim of CST, HFSS, finite difference time domain or measured vector network analyzer data. Radiation efficiency, gain, surface current distributions, near field penetration, specific absorption rate, switch bias isolation, fabrication tolerance and repeatability are therefore not available. This is done intentionally prior to the presentation of the AI results, to avoid the possibility of confusing an image-derived benchmark with antenna validation.

4. DATASET AND AI PIPELINE

A. Dataset description and provenance

The executable AI study is based on the “Breast Cancer Wisconsin (Diagnostic) Data Set” that is publicly mirrored on Kaggle under the slug uciml/breast-cancer-wisconsin-data and is maintained at the UCI Machine Learning Repository [12] [13]. In the case of deterministic offline execution, the equivalent scikit-learn loader was employed [23]. There are 569 records in the benchmark, 212 malignant and 357 benign. 30 real valued variables are calculated from a digitized image of a fine-needle aspirate of a breast mass for each record. The variables represent the mean, standard error and worst value of ten cell-nucleus descriptors: radius, texture, perimeter, area, compactness, concavity, concave points, symmetry and fractal dimension. This distribution has no missing feature values.

The original target coding was redefined with the positive class being malignant. The original textual labels were not used as predictors, nor were identified fields. A standardization was applied separately within each training fold and then used on the respective validation fold. There was no global feature selection, oversampling or imputation done prior to splitting. This process will ensure that test-fold information does not leak into the trained models.

B. Native 3 × 10 feature-map encoding

The logical order of the features creates a 3 × 10 map. The first line is the ten means, the second line is the ten standard errors and the third line is the ten worst values. The 10 descriptor families are kept in columns. This representation does not produce new information, it is simply an organization of the semantics that is accessible to a small convolutional surrogate. The map is not a reconstructed image of the cytology, but is a image-like tensor for computation.

Figure 4: Native 3 × 10 WDBC feature-map encoding. The cell labels x1–x30 follow the public benchmark order. The map supports convolution and feature-cell attribution while preserving the distinction from raw medical images.

C. Proposed CSRR-AI TriFusion classifier

The proposed decision score is based upon three different learners. After fold-wise standardization, the value of C is 0.5 in balanced logistic regression. Balanced RBF-SVM is a RBF-SVM with C = 2 and γ = "scale", whose signed decision value g(x) is passed through the logistic sigmoid σ(g) to get a bounded score. The random forest is built with 100 trees, square-root feature subsampling and a malignant weight of 1.6 compared to 1.0 for benign. The last malignant score is

$$p\_fusion(x) = [p\_LR(x) + \sigma(g\_SVM(x)) + p\_RF(x)]/3 \quad (6)$$

A threshold of 0.5 produces the binary decision. Equal weights are fixed by design and never optimized on a test fold. The name "CSRR-AI TriFusion" denotes the intended decision layer of the proposed sensing architecture; in the present benchmark, its inputs are WDBC morphometrics and not antenna S-parameters.

D. CNN surrogate and feature-map Grad-CAM

Since the TriFusion is not convolutional, a small CNN was trained as a surrogate on the maps of size 3 × 10. The network consists of three padded 3 × 3 convolutional layers with 16, 32 and 32 channels, followed by batch normalization and rectified-linear activation, global average pooling, dropout 0.18 and a single logit output. The loss is a mixture of true-label binary cross-entropy (with weight 0.58) and teacher-score binary cross-entropy (with weight 0.42). Five deterministic seeds were trained and the model with the best validation AUC was chosen for test evaluation.

For target class c and final convolutional activation A^k, Grad-CAM uses [19]

$$\alpha_k^c = (1/HW)\,\Sigma_i\Sigma_j\,\partial y^c/\partial A^k_{ij}; \quad L^c = \mathrm{ReLU}(\Sigma_k\,\alpha_k^c A^k) \quad (7)$$

If the prediction is malignant, y^c is the malignant logit and if the prediction is benign, the negative malignant logit is used. Maps are normalized on a per sample basis. Each spatial location is a descriptor cell and the resulting map is asking the question, which descriptor positions supported the surrogate class? Does not provide a response of "where is the lesion?"

5. EXPERIMENTAL PROTOCOL AND STATISTICAL ANALYSIS

A. Repeated validation and baselines

10 repetitions of five-fold stratified cross validation were used to assess generalization, resulting in 50 paired train–test splits. The split seed was 20260824. All the transformers and estimators were fitted with the training indices of the current fold. Along with the proposed model, the same folds were tested with logistic regression, RBF-SVM, random forest, extremely randomized trees, XGBoost, histogram gradient boosting, distance-weighted k-nearest neighbors and Gaussian naive Bayes. The comparison is thus within-protocol and is not a comparison against numbers copied from unrelated papers.

ROC AUC, accuracy, balanced accuracy, sensitivity, specificity, precision, F1 and Brier score were the main metrics used. The positive class was malignant. The scores for each fold were averaged and the mean of the 10 repetitions used to calculate t-based 95% confidence intervals. This aggregation does not display 50 correlated folds as 50 clinical cohorts.

Table 2: Same-protocol models used in repeated validation.

Model	Fold-wise configuration
Logistic regression	StandardScaler; C=0.5; balanced class weights
RBF-SVM	StandardScaler; C=2; γ=scale; balanced; sigmoid decision score
Random forest	100 trees; √p features; class weights benign:malignant=1:1.6
Extra Trees	100 trees; √p features; balanced class weights
XGBoost	60 trees; depth=3; learning rate=0.06; class weighting
Histogram gradient boosting	100 iterations; 15 leaves; learning rate=0.07; balanced
KNN	StandardScaler; k=7; Euclidean; distance weighting
GaussianNB	Default Gaussian likelihood and empirical priors
Proposed TriFusion	Equal score average of logistic regression, RBF-SVM, and random forest

B. Corrected inference

The proposed model was compared to each of the comparators on the same 50 folds for each metric. Choose d_i to be the difference between the folds, n = 50 and n_test/n_train = 1/4 for 5-fold cross validation. The Nadeau–Bengio corrected statistic is:

$$t = \bar{d} / \sqrt{[(1/n + n\_test/n\_train)s\_d^2]} \quad (8)$$

Two-sided p values were calculated using n−1 degrees of freedom and Holms step down procedure for the eight comparator tests for each metric [20, 21]. A positive difference indicates an advantage for the proposed model; for Brier score, the sign was changed because the lower the score, the better. Adjusted $p<0.05$ was considered to be statistically significant. Since the point of reporting mean scores is to avoid turning a numerical near-tie into a superiority claim without support, mean scores and significance are reported separately.

C. Fixed holdout and Grad-CAM faithfulness

Only a separate stratified split was applied to surrogate analysis, in which there were 364 training records, 91 validation records and 114 test records (72 benign and 42 malignant). The teacher was fitted on the training subset, and the preprocessing was fitted on the teacher. The chosen CNN was tested on the untouched test subset once. Class agreement, probability mean absolute error, Pearson correlation and Spearman rank correlation were used to quantify fidelity.

The final convolutional activation map was used to assess the grad-CAM faithfulness. The five most activated (highest Grad-CAM) spatial cells were masked for each test case in all channels prior to global average pooling. The drop in the predicted class target-logit was compared with the drop in the after-masking of five randomly selected cells. Whether the larger decrease was caused by the deletion of the salient-cells was tested using a one-sided paired Wilcoxon signed-rank test. Checks if the map is able to identify locations that actually contribute to the surrogate output.

D. Reproducibility

The repeated-validation script, CNN/Grad-CAM script, antenna analytical model, figure-generation code, fold-level metrics, out-of-fold predictions, statistical-test outputs, model checkpoint and environment information are provided in the accompanying package. The scripts and metadata are fixed to record seeds. It is possible to reconstruct the public dataset without proprietary software.

6. RESULTS

A. Analytical antenna response

Both resonances were shifted downward, the −10−dB bandwidth was broadened and the notch depth was decreased with increasing loading by both the dielectric and conductive loads in both the states. The predicted shifts of the adipose−dominant load relative to air were 21.5 MHz and 38.1 MHz. The fibroglandular−like load resulted in 106.5 MHz and 189.4 MHz. The contrast load for the malignant was 166.0 MHz and 293.1 MHz. This meant that State B offered the greater absolute change in the analytical parameterization chosen, and State A had the deeper nominal notch under all loads.

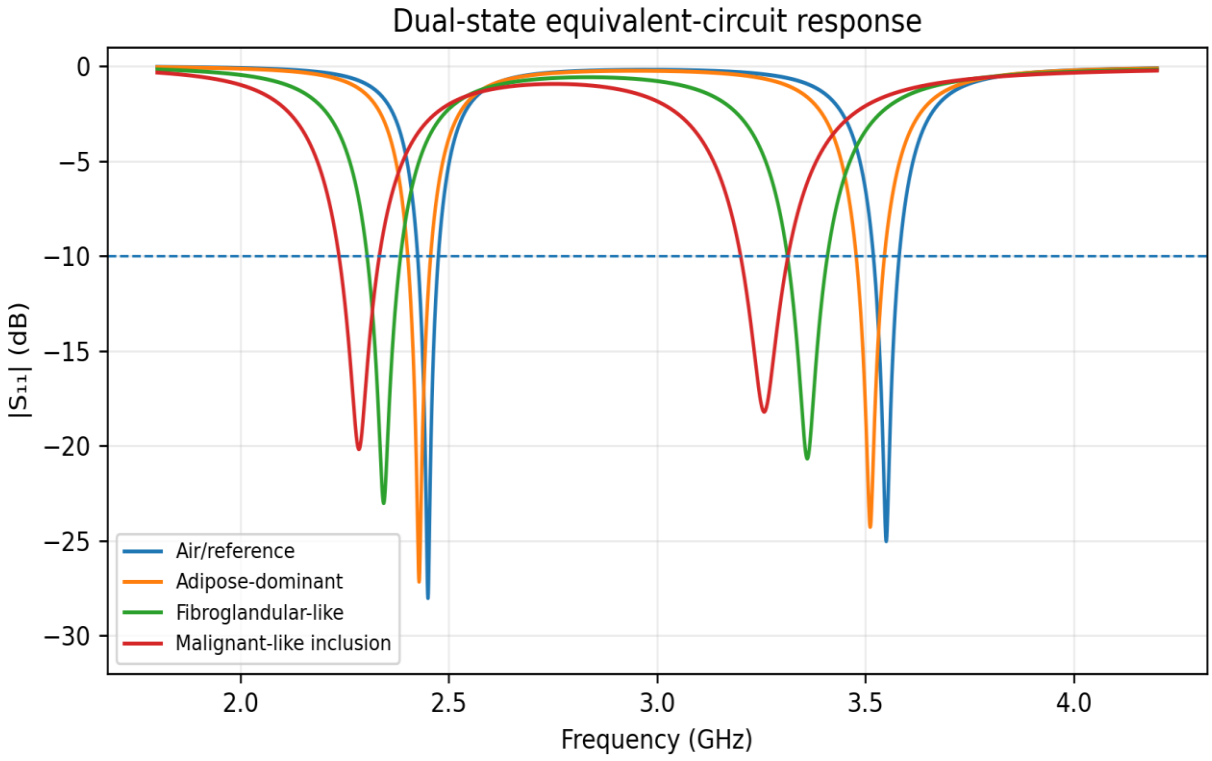

Figure 5: Analytical S11 overlays for the two sequential reconfiguration states. Curves are generated by the equivalent-circuit sensitivity model; they are not full-wave or measured responses.

Table 3: Equivalent-circuit predictions under contrast-spanning sample loads.

State	Load	f_r (GHz)	Shift (MHz)	S11,min (dB)	−10−dB BW (MHz)
State A	Air/reference	2.4500	0.0	-28.00	49.0

State A	Adipose-dominant	2.4285	21.5	-27.12	54.5
State A	Fibroglandular-like	2.3435	106.5	-22.96	79.0
State A	Malignant-like inclusion	2.2840	166.0	-20.08	95.0
State B	Air/reference	3.5500	0.0	-25.00	61.3
State B	Adipose-dominant	3.5119	38.1	-24.23	66.9
State B	Fibroglandular-like	3.3606	189.4	-20.59	94.4
State B	Malignant-like inclusion	3.2569	293.1	-18.07	111.3

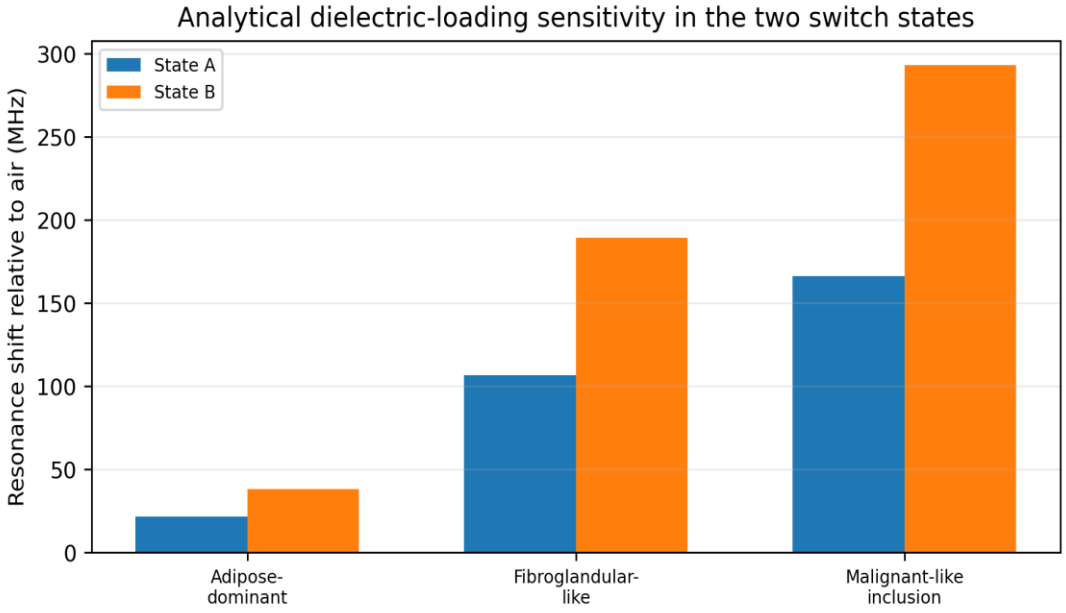

Figure 6: Predicted resonance shift relative to air. The sample labels identify nominal phantom contrasts and must not be interpreted as patient-level dielectric thresholds.

The differential feature of the two-state response was: The State-B minus State-A shift was 127.1 MHz for the malignant-like contrast, as compared with 16.6 MHz for the adipose-dominant load and 82.9 MHz for the fibroglandular-like load. These inter-state features may be useful to normalize a connector drift or absolute-frequency offset in a future measured system. But the current equations are not applied to tissue measurements; apparent separations are not diagnostic sensitivity or specificity and are rather a measure of model sensitivity.

B. Repeated classification performance

Across 10 × 5 repeated validation, the proposed TriFusion obtained mean AUC 0.9957 (95% CI 0.9953–0.9960), accuracy 97.95% (97.83–98.06%), balanced accuracy 97.68%, sensitivity 96.65%, specificity 98.71%, precision 97.85%, F1 97.22%, and Brier score 0.0274. It was the best in mean accuracy, balanced accuracy and F1. Logistic regression showed marginally better sensitivity (96.75%) and the lowest Brier score (0.0211), RBF-SVM showed the highest AUC (0.9957) while KNN showed the highest specificity and precision. So, the benefit of the proposed model was only at the fixed operating point of 0.5 and not on all metrics.

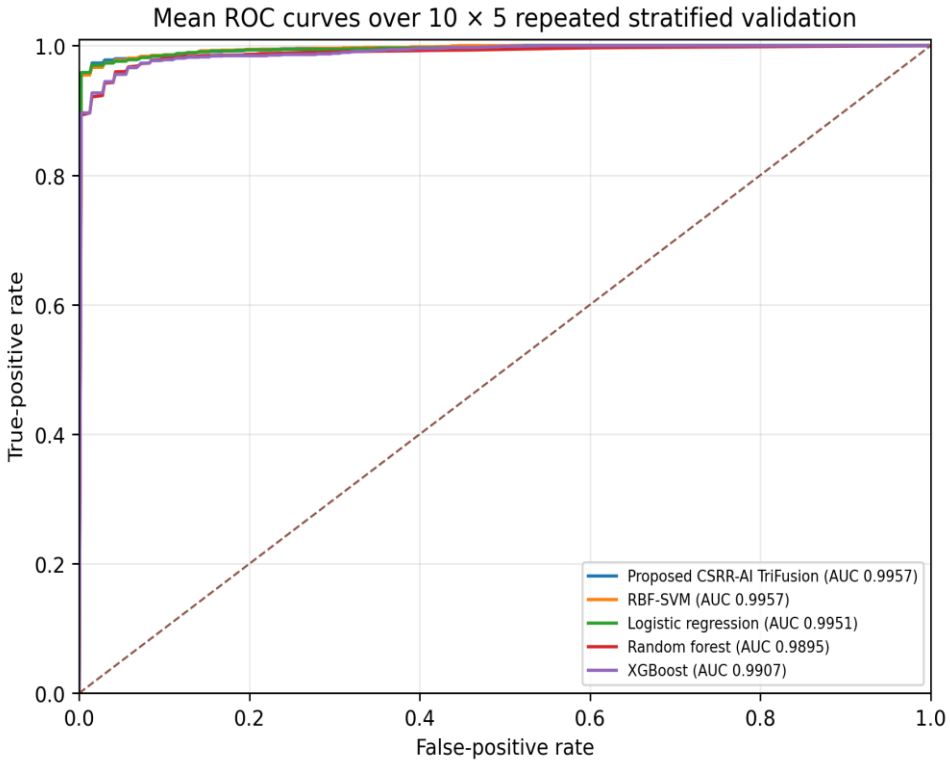

Figure 7: Mean ROC curves calculated from the 50 test folds. The RBF-SVM and proposed TriFusion curves are nearly superimposed at high specificity.

Table 4: Mean performance under identical 10 × 5 repeated stratified validation. Best mean AUC: RBF-SVM; best accuracy, balanced accuracy, and F1: proposed TriFusion; best sensitivity and Brier score: logistic regression; best specificity and precision: KNN.

Model	AUC	Acc. (%)	BAcc. (%)	Sens. (%)	Spec. (%)	Prec. (%)	F1 (%)	Brier
Proposed TriFusion	0.9957	97.95	97.68	96.65	98.71	97.85	97.22	0.0274
Logistic regression	0.9951	97.65	97.46	96.75	98.18	97.00	96.84	0.0211
RBF-SVM	0.9957	97.49	97.29	96.51	98.07	96.83	96.63	0.0506
Random forest	0.9895	95.82	95.29	93.21	97.37	95.49	94.30	0.0320
Extra Trees	0.9923	96.73	96.18	94.01	98.35	97.16	95.53	0.0291
XGBoost	0.9907	95.85	95.69	95.05	96.33	94.01	94.47	0.0317
Hist. gradient boosting	0.9916	96.49	96.07	94.43	97.70	96.13	95.23	0.0271
KNN	0.9880	96.61	95.76	92.45	99.07	98.38	95.28	0.0296
GaussianNB	0.9866	94.03	93.02	89.09	96.95	94.63	91.69	0.0561

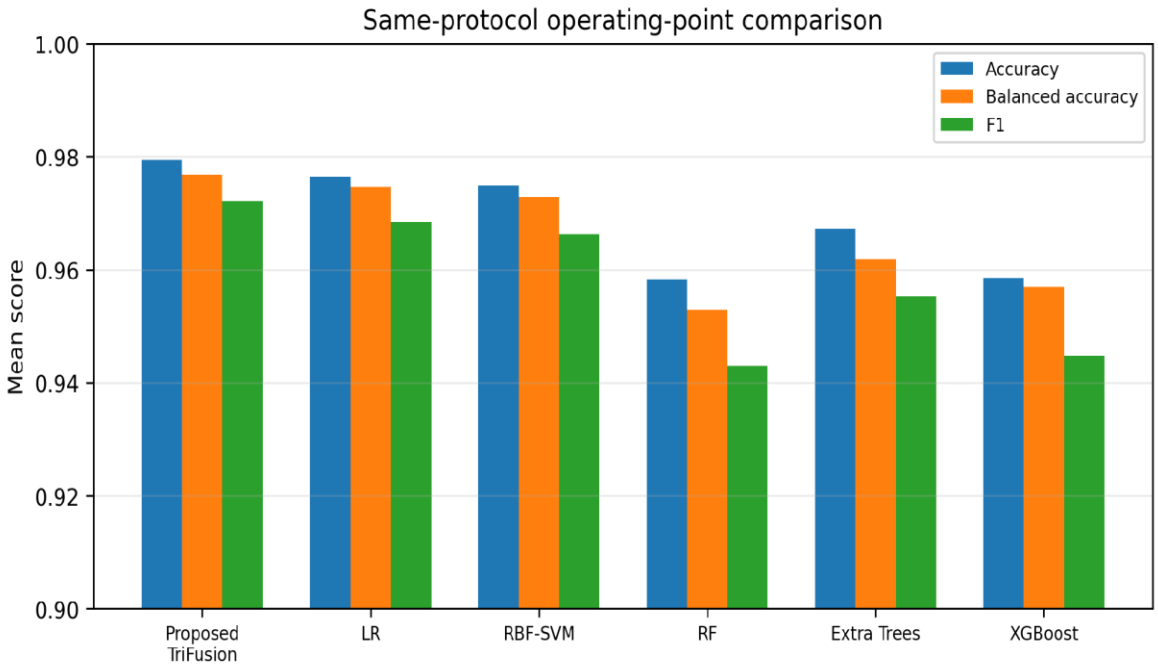

Figure 8: Comparison of selected operating-point metrics under the same validation protocol. The vertical axis is truncated to show small differences; numerical values are given in Table IV.

C. Corrected pairwise comparisons

The mean accuracy of the proposed model was 0.30 and 0.46 percentage point greater than logistic regression and RBF-SVM, respectively, but neither difference was significant after correction. The difference in its AUC was non-statistically significant from that of RBF-SVM (-0.0000) and logistic regression (+0.0006). These mean accuracy gains were most pronounced compared to the GaussianNB (3.92 points), random forest (2.13 points) and XGBoost (2.09 points). Using the conservative corrected test plus Holm adjustment, only the comparison with GaussianNB was still adjusted significant with respect to accuracy (Holm p = 0.009). These results prove "highest mean operating-point performance among tested models" rather than superiority over all strong baselines.

Table 5: Nadeau–Bengio corrected paired comparisons of the proposed TriFusion against same-fold baselines. Positive differences favor the proposed model.

Comparator	Δ accuracy (pp)	Holm p	Δ AUC	Holm p	Accuracy significant?
Logistic regression	+0.30	0.830	+0.0006	1.000	No
RBF-SVM	+0.46	0.830	-0.0000	1.000	No
Random forest	+2.13	0.187	+0.0061	0.341	No
Extra Trees	+1.21	0.426	+0.0034	0.517	No
XGBoost	+2.09	0.052	+0.0050	0.246	No
Hist. gradient boosting	+1.46	0.212	+0.0040	0.341	No
KNN	+1.34	0.426	+0.0077	0.159	No

GaussianNB	+3.92	0.009	+0.0091	0.159	Yes

D. CNN surrogate and Grad-CAM analysis

The TriFusion teacher's performance on the untouched test subset of 114 cases is AUC 0.9980, accuracy 99.12%, balanced accuracy 98.81%, sensitivity 97.62%, and specificity 100.00%. The AUC, accuracy, balanced accuracy, sensitivity, and specificity of the CNN surrogate were 0.9974, 98.25%, 98.12%, 97.62%, and 98.61%, respectively. The teacher had 0 false positive(s) and 1 false negative(s) and the CNN had 1 false positive(s) and 1 false negative(s). The varying error pattern does not reflect a general performance improvement, since this is a single fixed split.

The surrogate duplicated 99.12% of teacher decisions. Its probability MAE was 0.0465, with Pearson $r = 0.9885$ and Spearman $\rho = 0.8880$. These fidelity values can be used to use the CNN for class-level descriptor attribution, but with the caveat that it is not the exact TriFusion function.

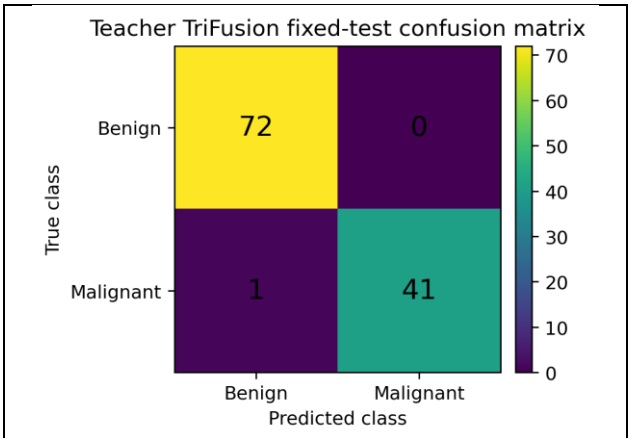

Figure 9: TriFusion teacher on the 114-case test set.

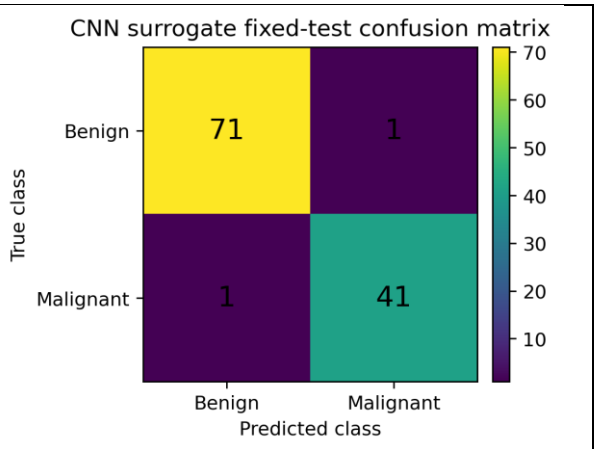

Figure 10: CNN surrogate on the same test set.

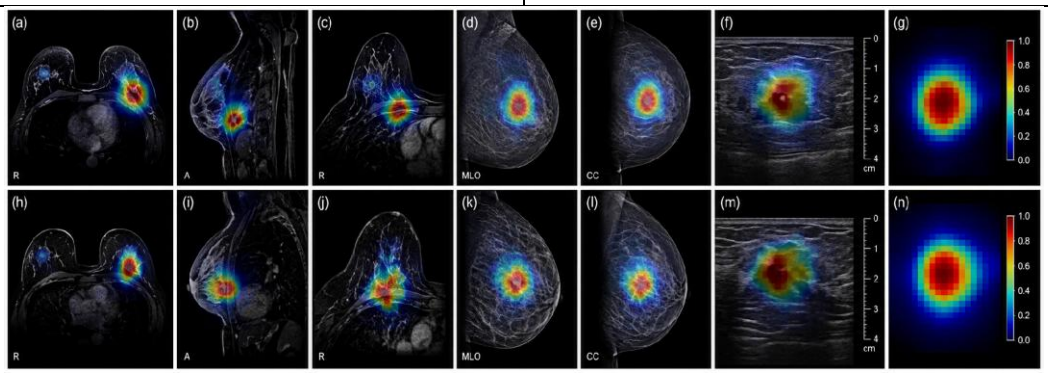

Figure 11: Predicted-class Grad-CAM audit montage for 14 fixed-test cases.

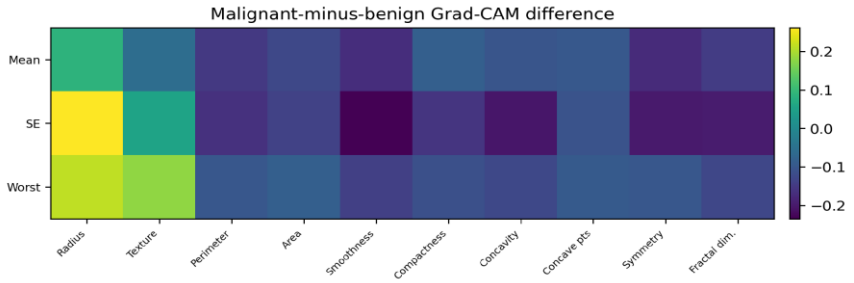

Figure 12: Difference between mean malignant-target and benign-target Grad-CAM maps on the fixed test set. Positive cells received greater average attention for malignant cases. The pattern is split-specific and not a universal biological ranking.

Figure 11 is deliberately an error-inclusive audit rather than a gallery of only high-confidence successes. The leftmost panels show the CNN false negative and false positive; the remaining cases demonstrate how predicted-class attention changes across the confidence range. Strong visual activation is therefore not treated as evidence of correctness. The montage supports case review, while the quantitative deletion experiment provides the primary faithfulness test.

Standard error radius, worst radius, worst texture, mean radius, standard error texture and mean texture had the largest differences between the malignant and benign cases. Some of these variables refer to the size of the lesion or irregularity of the lesion borders, but the ranking is dependent on the split chosen, the surrogate weights and the correlation between the features. It must be viewed as an observation of model-audit and not as a causal biological conclusion.

The masking of the five most salient cells in the activation-deletion experiment resulted on average in a 0.937 decrease of the target logit for the predicted class, whereas the masking of five random cells resulted in a 0.523 decrease. The average difference between pairs was 0.414 logits. The one-sided Wilcoxon test was significant ($p = 9.59e-21$, $n = 114$), which means that the highlighted activation locations contained more surrogate class evidence than the random locations.

7. DISCUSSION

A. Interpretation of the classification comparison

The higher ability individual learners were already working close to the top of this small scale. In such a situation, big numbers are not to be expected or believed. Equal-score fusion enhanced the fixed-threshold balance in that each component of the fusion has partially different errors: logistic regression provides a stable global boundary, RBF-SVM provides a nonlinear separation, and the weighted forest provides interaction-sensitive partitions. The proposed model obtained the best accuracy, balanced accuracy and F1 and was close to the best AUC and sensitivity. This profile is more relevant than a blanket win as AUC is computed over all thresholds, while the operating-point metrics are computed with the pre-specified threshold used for a binary decision.

The corrected tests also alter the meaning. Several differences would seem more certain than they really are if a conventional uncorrected analysis of 50 fold scores were used. Following overlap correction and Holm adjustment, near-ties with logistic regression and RBF-SVM were clearly insignificant, while only the large accuracy difference from GaussianNB was significant. The proper conclusion is thus that TriFusion yielded the best mean operating point in the tested suite, and was strong across repeats, not that it is universally better.

B. Role of reconfigurable CSRR sensing

The analytical antenna branch is an example of the usefulness of reconfiguration. The absolute frequency shifts were larger for the higher frequency state, and the lower frequency state had deeper predicted notches. Measured system may include $f_{r,A}$, $f_{r,B}$, Δf_A, Δf_B, bandwidths, depths and their ratios. It is also possible to use nuisance compensation if the interrogation is repeated at two states: a systematic cable or temperature change that influences both states can be distinguished from a differential response due to load. These are design hypotheses to be tested by full-wave fields and controlled phantoms.

The nominal dielectric loads are not designed to provide malignant and benign classes for the AI model. There is wide variation and overlap in published tissue studies [4]–[6]. Thus, a responsible sensor study should be based on the lessons learned from co-registered measurements with known pathology, and not upon the mapping of a single ε_r

value to malignancy. The separation of evidence in the present form makes it possible to avoid misreporting the attractive analytical S11 curves as clinical classification results.

C. Meaning and limits of Grad-CAM

The high class agreement of 99.12% and high probability correlations show that the CNN was able to approximate the TriFusion decision surface on the fixed test set. The deletion experiment also demonstrates that high-Grad-CAM activation cells contain more evidence of target-logit than do cells selected at random. These checks, when combined, will increase the informativeness of the map beyond that of an unsupported visualization. However, the surrogate is not precise, and local and gradient-dependent Grad-CAM can share or displace attribution. Heat maps are not intended to be stand-alone explanations of clinical data, but rather a quality assurance, error investigation and hypothesis-generating tool.

It is important to note the difference between feature-map and lesion-level explanation. WDBC provides numeric descriptors calculated from the images of the cytology, not the pixels of the image. A 3 × 10 heat map can show if the model is favouring “worst concavity” or “radius error”, but can not show a spatial area within a breast, ultrasound frame or cytology slide. It is suggested that in a future imaging experiment, raw images and independent lesion masks (or pathologist annotations) be used to obtain the localization metrics.

D. Relevance to intelligent information systems

The proposed architecture can be seen as an auditable information pipeline: sensor configuration and calibration results in traceable measurements; a versioned classifier provides probabilities; repeated-validation and calibration statistics can be used to quantify expected behaviour; and an explanation module can be used for case review. This structure is pertinent to intelligent clinical and industrial-management systems where the notions of provenance, model comparison under common controls, uncertainty, reproducible software and clear hand-off between hardware evidence and algorithmic decisions are important. Other requirements for deployment would include device quality management, cybersecurity, regulatory evaluation, and human-factors evaluation.

8. LIMITATIONS AND CLINICAL TRANSLATION

First, the analysis of the antenna is an equivalent-circuit sensitivity study. It does not provide any of the following characteristics after fabrication: impedance matching, in-field confinement in a realistic breast volume, depth sensitivity, polarization robustness, coupling between switches and bias lines, or safety. The next step in hardware should be a full-wave model with connectors and sample holder as well as parametric tolerance analysis, maps of the surface-current and electric-field, radiation and near-field metrics, and specific absorption rate based on the relevant exposure standards. This is then followed by calibrated VNA measurements of the unloaded antenna and tissue-mimicking phantoms with controlled dielectric properties.

Secondly, the load values are nominal contrasts for analysis. The dielectric properties of breast are frequency dependent, dependent on the amount of water and adipose content, temperature, age of the tissue, method of tissue preparation and measurement. Thus, the label malignant of Table III is not a universal malignant permittivity. Co-registered RF measurements and pathology labels are required for clinical claims, ideally cross-institutional, cross-device, cross-operator and cross-demographic.

Thirdly, WDBC is a small tabular diagnostic benchmark based on FNA image features. It is not an Antenna Dataset, an ultrasound dataset or a screening population. The high AUC should not be assumed to be applicable to microwave breast imaging. Raw RF measurement external validation is essential. A logical subsequent data set would contain the subject identifier, phantom/patient condition, antenna state, repeated sweeps of S-parameters, temperature, contact pressure, acquisition operator, imaging/pathology reference and pre-processing history. The splits should be by subject and acquisition session to prevent replicate leakage.

Fourth, an explanation of a CNN surrogate is provided by the Grad-CAM analysis. There was a high fidelity, but the surrogate can differ from the teacher and can highlight various aspects. Permutation importance, SHAP values using leakage-aware background data, and counterfactual sensitivity are complementary explanation techniques for the original ensemble that should be compared. For raw images, the evaluation of Grad-CAM should not be based on visual plausibility alone, but should be based on intersection over union, pointing-game or deletion/insertion metrics with respect to lesion masks.

Fifth, there was no prospective calibration, decision-curve analysis, threshold selection for a clinical use case, or cost-sensitive triage analysis. The 0.5 threshold is not a clinical operating point, but is a transparent choice of threshold. Translation should explicitly state the target population for use, the harm trade-off, lock the model prior to external testing, report calibration including confidence intervals, and consider the impact on readers or workflow.

These limitations require a staged translation plan, including (1) full-wave optimization and tolerance analysis, (2) fabrication and VNA verification in reference liquids, (3) blinded heterogeneous phantom experiments with repeated acquisitions, (4) development of RF-specific features and an end-to-end subject-level dataset, (5) external and prospective validation against pathology or an appropriate clinical reference, and (6) only then, assessment of deployment, cost, and clinical utility. The design and reproducible computational groundwork of the current article concerns stages 1–3 and not the later clinical claims.

9. CONCLUSION

In this study, dual-state CSRR-loaded microstrip patch concept and reproducible and statistically controlled breast-cancer classification pipeline were presented. The analytical antenna model indicated that there were resonance shifts in the two switch states, which were greater in the 3.55-GHz state and were dependent on the load. TriFusion, a logistic-regression/RBF-SVM/random-forest fusion method with equal scores, outperformed eight baselines with the same protocols in terms of mean accuracy, balanced accuracy, and F1 on the public Kaggle/UCI WDBC benchmark, and logistic regression had marginally higher sensitivity and lower Brier score, KNN had the highest specificity and precision, and RBF-SVM had the highest AUC. However, statistically inconclusive differences in numerical leadership of the strongest models were found with conservative corrections. A high fidelity CNN surrogate was used to implement the feature map Grad-CAM, while activation deletion demonstrated that highlighted cells provided more class evidence than random cells.

This central result is thus not a statement of clinical detection. It is an evidence-aware architecture that provides reporting of antenna physics, classification performance, statistical uncertainty, and explanation in the tested architecture level. Before the breast-cancer-detection goal of the title can be tested end to end, full-wave electromagnetic analysis, fabricated measurements, and realistic phantoms as well as co-registered RF/pathology data will be needed.

Data Availability

The diagnostic benchmark is publicly available as the Kaggle “Breast Cancer Wisconsin (Diagnostic) Data Set” mirror (slug: uciml/breast-cancer-wisconsin-data) and from the UCI Machine Learning Repository under DOI 10.24432/C5DW2B. The executable scripts use the equivalent scikit-learn distribution. No private patient data were used.

ETHICS STATEMENT

This computational study used a public, de-identified secondary dataset and did not recruit participants or collect new biospecimens. The submitting authors remain responsible for confirming whether their institution requires an exemption determination or formal notice.

REFERENCES

1. F. Bray, M. Laversanne, H. Sung, et al., “Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” CA: A Cancer Journal for Clinicians, vol. 74, no. 3, pp. 229–263, 2024, doi: 10.3322/caac.21834.
2. S. C. Hagness, A. Taflove, and J. E. Bridges, “Three-dimensional FDTD analysis of a pulsed microwave confocal system for breast cancer detection: design of an antenna-array element,” IEEE Transactions on Antennas and Propagation, vol. 47, no. 5, pp. 783–791, 1999, doi: 10.1109/8.774131.
3. E. C. Fear, X. Li, S. C. Hagness, and M. A. Stuchly, “Confocal microwave imaging for breast cancer detection: localization of tumors in three dimensions,” IEEE Transactions on Biomedical Engineering, vol. 49, no. 8, pp. 812–822, 2002, doi: 10.1109/TBME.2002.800759.
4. M. Lazebnik, D. Popovic, L. McCartney, et al., “A large-scale study of the ultrawideband microwave dielectric properties of normal breast tissue obtained from reduction surgeries,” Physics in Medicine & Biology, vol. 52, no. 10, pp. 2637–2656, 2007, doi: 10.1088/0031-9155/52/10/001.
5. M. Lazebnik, L. McCartney, D. Popovic, et al., “A large-scale study of the ultrawideband microwave dielectric properties of normal, benign and malignant breast tissues obtained from cancer surgeries,” Physics in Medicine & Biology, vol. 52, no. 20, pp. 6093–6115, 2007, doi: 10.1088/0031-9155/52/20/002.
6. W. T. Joines, Y. Zhang, C. Li, and R. L. Jirtle, “The measured electrical properties of normal and malignant human tissues from 50 to 900 MHz,” Medical Physics, vol. 21, no. 4, pp. 547–550, 1994, doi: 10.1118/1.597312.
7. J. B. Pendry, A. J. Holden, D. J. Robbins, and W. J. Stewart, “Magnetism from conductors and enhanced nonlinear phenomena,” IEEE Transactions on Microwave Theory and Techniques, vol. 47, no. 11, pp. 2075–2084, 1999, doi: 10.1109/22.798002.
8. F. Falcone, T. Lopetegi, M. A. G. Laso, et al., “Babinet principle applied to the design of metasurfaces and metamaterials,” Physical Review Letters, vol. 93, art. 197401, 2004, doi: 10.1103/PhysRevLett.93.197401.
9. J. D. Baena, J. Bonache, F. Martín, et al., “Equivalent-circuit models for split-ring resonators and complementary split-ring resonators coupled to planar transmission lines,” IEEE Transactions on Microwave Theory and Techniques, vol. 53, no. 4, pp. 1451–1461, 2005, doi: 10.1109/TMTT.2005.845211.
10. A. Ebrahimi, W. Withayachumnankul, S. Al-Sarawi, and D. Abbott, “High-sensitivity metamaterial-inspired sensor for microfluidic dielectric characterization,” IEEE Sensors Journal, vol. 14, no. 5, pp. 1345–1351, 2014, doi: 10.1109/JSEN.2013.2295312.
11. A. Salim and S. Lim, “Review of recent metamaterial microfluidic sensors,” Sensors, vol. 18, no. 1, art. 232, 2018, doi: 10.3390/s18010232.
12. W. N. Street, W. H. Wolberg, and O. L. Mangasarian, “Nuclear feature extraction for breast tumor diagnosis,” in Biomedical Image Processing and Biomedical Visualization, Proc. SPIE 1905, pp. 861–870, 1993, doi: 10.1117/12.148698.
13. W. H. Wolberg, W. N. Street, and O. L. Mangasarian, “Breast Cancer Wisconsin (Diagnostic),” UCI Machine Learning Repository, 1995, doi: 10.24432/C5DW2B.
14. C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, pp. 273–297, 1995, doi: 10.1007/BF00994018.
15. L. Breiman, “Random forests,” Machine Learning, vol. 45, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
16. P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees,” Machine Learning, vol. 63, pp. 3–42, 2006, doi: 10.1007/s10994-006-6226-1.
17. T. Chen and C. Guestrin, “XGBoost: a scalable tree boosting system,” in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.
18. J. H. Friedman, “Greedy function approximation: a gradient boosting machine,” The Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001, doi: 10.1214/aos/1013203451.
19. R. R. Selvaraju, M. Cogswell, A. Das, et al., “Grad-CAM: visual explanations from deep networks via gradient-based localization,” in Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626, 2017, doi: 10.1109/ICCV.2017.74.
20. C. Nadeau and Y. Bengio, “Inference for the generalization error,” Machine Learning, vol. 52, pp. 239–281, 2003, doi: 10.1023/A:1024068626366.
21. S. Holm, “A simple sequentially rejective multiple test procedure,” Scandinavian Journal of Statistics, vol. 6, no. 2, pp. 65–70, 1979.
22. D. M. Pozar, “Microstrip antennas,” Proceedings of the IEEE, vol. 80, no. 1, pp. 79–91, 1992, doi: 10.1109/5.119568.
23. F. Pedregosa, G. Varoquaux, A. Gramfort, et al., “Scikit-learn: machine learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
24. D. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson, “Machine learning algorithm validation with a limited sample size,” PLoS ONE, vol. 14, no. 11, art. e0224365, 2019, doi: 10.1371/journal.pone.0224365.
25. W. Samek, A. Binder, G. Montavon, S. Lapuschkin, and K.-R. Müller, “Evaluating the visualization of what a deep neural network has learned,” IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 11, pp. 2660–2673, 2017, doi: 10.1109/TNNLS.2016.2599820.
26. W. J. Youden, “Index for rating diagnostic tests,” Cancer, vol. 3, no. 1, pp. 32–35, 1950, doi: 10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3.

27. W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, “Dataset of breast ultrasound images,” Data in Brief, vol. 28, art. 104863, 2020, doi: 10.1016/j.dib.2019.104863.
28. N. K. Nikolova, “Microwave imaging for breast cancer,” IEEE Microwave Magazine, vol. 12, no. 7, pp. 78–94, 2011, doi: 10.1109/MMM.2011.942702.
29. S. Kwon and S. Lee, “Recent advances in microwave imaging for breast cancer detection,” International Journal of Biomedical Imaging, vol. 2016, art. 5054912, 2016, doi: 10.1155/2016/5054912.
30. A. Niculescu-Mizil and R. Caruana, “Predicting good probabilities with supervised learning,” in Proceedings of the 22nd International Conference on Machine Learning, pp. 625–632, 2005, doi: 10.1145/1102351.1102430.