electronics

IMPACT FACTOR 2.9

CITESCORE 7.0

Article

Validation-Aware Surrogate Shortlisting for Biomedical Microwave Imaging: Analytical Performance and FDTD Transfer

Lulu Wang

Special Issue

AI-Driven Metasurfaces, Antennas, and Wireless Systems

Edited by

Dr. Zhaohui Wei, Dr. Zhao Zhou and Dr. Ming Shen

MDPI

https://doi.org/10.3390/electronics15163561

electronics

MDPI

Article

Validation-Aware Surrogate Shortlisting for Biomedical Microwave Imaging: Analytical Performance and FDTD Transfer

Lulu Wang [1,2]

[1] Department of Engineering, Reykjavík University, 101 Reykjavík, Iceland; luluw@ru.is or lwang381@hotmail.com
[2] College of Science, Engineering and Technology, University of South Africa, Pretoria 0002, South Africa

Abstract

Broadband antenna, frequency and channel selection requires efficient prioritisation of finite candidate configurations, yet strong surrogate performance within a simplified analytical model does not ensure that the learned ordering will transfer to another electromagnetic representation. This study developed a validation-aware surrogate-shortlisting framework using a controlled breast-mimetic benchmark comprising 120 scenarios and 80 antenna–frequency–channel candidates per scenario. Surrogate models were developed using grouped scenario-level validation, and the model and five-candidate shortlist policy were frozen before external evaluation. Random forest produced the lowest-regret analytical-domain shortlist and remained stable across model-initialisation seeds. The frozen ranking was then challenged on held-out scenarios using a separately implemented restricted two-dimensional transverse-magnetic finite-difference time-domain model. Although numerical-reference checks supported shortlist-level use of the operational grid, candidate ordering did not transfer reliably: optimum inclusion, shortlist agreement and rank association were weak, although a qualified candidate within 2 mm of the finite-library FDTD optimum was retained in 58.3% of cases. Transfer failure varied by frequency band and channel family, while incomplete alignment between the analytical and FDTD candidate libraries prevented attribution of the discrepancy to electromagnetic-model shift alone. The framework therefore positions analytical surrogates as auditable shortlisting tools that reduce downstream candidate-level assessment while retaining independent electromagnetic evaluation before final design selection.

Keywords: surrogate modelling; candidate shortlisting; biomedical microwave imaging; analytical-to-FDTD transfer; 2-D TM_z FDTD; antenna-array design; model shift; breast-mimetic benchmark

Academic Editors: Ming Shen, Zhaohui Wei and Zhao Zhou

Received: 3 July 2026
Revised: 7 August 2026
Accepted: 8 August 2026
Published: 11 August 2026

1. Introduction

Broadband antenna-array and interrogation design requires comparison of numerous configurations defined by antenna placement, sensing distance, frequency coverage and transmit–receive channel selection. In microwave and radiofrequency sensing, the relative performance of these configurations may also depend on tissue composition, dielectric contrast, target location and background heterogeneity. Exhaustive evaluation can therefore become computationally demanding when each candidate requires detailed electromagnetic simulation, reconstruction or experimental measurement. Surrogate modelling offers a practical alternative by learning relationships between candidate descriptors and design outcomes, thereby allowing promising configurations to be prioritised for further assessment [1–4].

Microwave breast imaging provides a demanding benchmark because lesion-associated responses are generally weak relative to the heterogeneous background and are sensitive to frequency, propagation path, antenna geometry, tissue organisation, skin-layer effects, dispersion, dielectric variability and coupling conditions. Previous studies have investigated ultra-wideband antennas, Vivaldi arrays, multistatic acquisition and microwave-imaging approaches for breast-mimetic applications [5–11]. In the present study, this setting is used as a controlled RF system-design benchmark rather than as evidence of clinical diagnostic performance.

Surrogate-assisted optimisation has been applied widely to antennas, microwave devices and electromagnetic components [12–15], while more recent electromagnetic AI studies have explored physics-informed surrogates, equation-constrained learning and deep-learning-enabled inverse design [12–20]. However, most evaluations are conducted within the same analytical or numerical representation used for model development. Such testing can establish predictive consistency and ranking performance within that representation, but it cannot determine whether the learned candidate preferences remain useful when heterogeneous propagation, dispersion, boundary interactions, phase behaviour or reconstruction responses are represented differently.

This distinction is particularly important for finite-candidate shortlisting, in which a surrogate ranks a predefined library and forwards only the top K configurations for detailed evaluation. Its practical value depends on whether the shortlist retains a qualified candidate close to the best available option; regret, optimum inclusion, shortlist overlap and rank association are therefore more informative decision-level endpoints than prediction error alone.

The present model is descriptor-based: its inputs represent frequency design, channel family, candidate geometry, tissue organisation and material properties, but it does not embed Maxwell equations, electromagnetic residuals, a neural field operator or a finite-antenna solver. Although physically motivated descriptors may improve transparency and support learning within a controlled generator, they do not make the resulting ranking representation-independent. Accordingly, the methodological contribution is not a new learning architecture but a validation-aware external-evaluation framework in which the feature scope, model family, ranking rule and shortlist budget are frozen within the analytical generator before being challenged on held-out biological scenarios using a separately implemented restricted 2-D TM_z FDTD representation. In the first stage, a controlled breast-mimetic analytical benchmark comprising 120 scenarios and 80 finite antenna–frequency–channel candidates per scenario is used to develop surrogate models through grouped scenario-level validation. A representative antipodal Vivaldi element and a 16-position, two-ring sensing arrangement define the analytical design context, and a five-candidate shortlist policy is frozen before external evaluation. The second stage tests whether this ranking retains decision value under the restricted FDTD representation, which uses ideal in-plane sources and receivers and therefore evaluates restricted electromagnetic transfer rather than complete finite-antenna or hardware performance.

The study contributes a controlled finite-candidate RF design benchmark with scenario-level leakage control; regret-first surrogate development and model-seed stability assessment; shortlist-level numerical-reference evaluation of the operational FDTD grid; explicit auditing of analytical–FDTD candidate-support alignment; held-out transfer analysis using regret, optimum inclusion, shortlist overlap, rank association and localisation ambiguity; and structured interpretation of transfer failure by frequency band and channel family. Its intended outcome is a framework for determining when an analytical surrogate can provide a provisional shortlist and when independent electromagnetic evaluation remains neces-

sary before final configuration selection. It is not intended to provide a clinically validated microwave breast-imaging system or to replace three-dimensional finite-antenna simulation.

2. Materials and Methods

2.1. Study Overview and Staged Validation Design

This study developed and evaluated a surrogate-ranking framework for antenna-array, frequency and channel selection under a finite evaluation budget in a controlled breast-mimetic RF benchmark. The evaluation comprised two stages. First, a seed-controlled analytical generator was used to construct reproducible candidate-ranking tasks, develop the surrogate models and prospectively fix a finite-candidate shortlisting policy. Second, the selected surrogate was challenged using a separately implemented restricted two-dimensional TM_z finite-difference time-domain model. The biological scenario was treated as the independent unit for model development and statistical inference because candidates from the same scenario shared tissue organisation, material properties and lesion characteristics. Candidate records were therefore treated as nested evaluations rather than independent biological samples.

The benchmark comprised 120 biological scenarios and 80 analytical candidates per scenario, yielding 9600 scenario–candidate records. Of the 120 scenarios, 105 were assigned to grouped analytical development, three to numerical-reference assessment and 12 to held-out FDTD evaluation. Six held-out scenarios were fixed before inspection of their FDTD outcomes. Six additional stress scenarios were selected from the remaining eligible pool by deterministic maximin sampling based on lesion diameter, fibroglandular fraction, tissue-organisation correlation length and radial lesion distance. After standardisation within the eligible pool, each additional scenario was selected to maximise its minimum squared Euclidean distance from those already retained; the three remaining eligible scenarios were reserved for numerical-reference assessment.

No FDTD outcome informed feature-scope selection, model-family selection, hyperparameter specification, shortlist-budget selection, scenario replacement or endpoint definition. The analytical stage evaluated candidate ranking within the controlled generator, whereas the held-out FDTD stage assessed whether the prospectively fixed ranking retained decision value under a distinct electromagnetic representation.

2.2. Breast-Mimetic Benchmark, Sensing Geometry and Candidate Parameterisation

The analytical benchmark represented a hemispherical breast-mimetic domain with radius $R_b = 60$ mm and a 2 mm skin-like outer layer. These dimensions are consistent with hemispherical and layered breast-phantom configurations reported in microwave-imaging studies [5–11,21] and were selected to provide a reproducible geometry rather than to represent the full anatomical variability of the human breast. Internal heterogeneity was represented using adipose-like and fibroglandular-like tissue classes. In this study, fibroglandular fractions ranged from 10% to 40%, and tissue-organisation correlation lengths ranged from 8 to 12 mm. The fibroglandular fraction controlled tissue-class occupancy, whereas the correlation length controlled spatial clustering. Each FDTD cell retained one tissue-class assignment; no Maxwell–Garnett, Bruggeman or other effective-medium mixing rule was applied at the cell level.

Each scenario contained one lesion-like inclusion with diameter $a \in \{5, 10, 15\}$ mm. Lesion position, depth, dielectric contrast, tissue properties and background organisation varied across scenarios to create controlled differences in localisation difficulty. These parameter ranges were not intended to reproduce a clinical population distribution.

For each biological scenario, 80 finite array–frequency–channel candidates were generated within a fixed 16-position, two-ring analytical framework. The candidates represented

alternative interrogation configurations obtained by varying antenna-to-surface distance, relative ring offset, frequency subset, centre frequency, bandwidth and transmit–receive channel family; they did not represent distinct antenna elements.

The 3–8 GHz interval was selected as a bounded ultra-wideband evaluation envelope overlapping frequency ranges used in previous Vivaldi and multistatic microwave breast-imaging studies [5,6,10,11]. The interval included lower-frequency components expected to provide greater penetration and upper-frequency components expected to provide finer spatial sensitivity, while allowing the increased attenuation and sensitivity to tissue heterogeneity associated with the upper part of the band to be examined within a common candidate library. It enabled controlled comparison of lower-, intermediate- and upper-frequency subsets and was not assumed to be universally optimal.

In this study, antenna-to-surface distance varied from 5 to 20 mm in the analytical benchmark. This deliberately broad range was used only as an analytical design variable to determine whether the generator assigned different preferences across sensing distances; it was not supported by loaded-antenna matching calculations or finite-antenna simulations. The held-out FDTD challenge was therefore restricted to stand-offs of 15 and 20 mm, consistent with separations used in reported 16-element breast-imaging configurations [22,23]. No inference was made regarding realised antenna matching, radiation efficiency or usable bandwidth at stand-offs of 5 or 10 mm.

An antipodal Vivaldi antenna provided the broadband context for analytical candidate parameterisation but was not simulated as a finite port-driven antenna. The nominal assumptions used in this study were a 3–8 GHz operating range, a 50 Ω feed impedance, a 40 mm antenna length, a 28 mm aperture width, a substrate relative permittivity of 3.55, a substrate thickness of 0.813 mm and a metallisation thickness of 35 µm. Linear polarisation directed towards the breast-mimetic domain was assumed. A nominal condition of $S_{11,\mathrm{dB}} < -10$ dB defined the intended useful band but was not treated as a simulated or measured antenna result.

Figure 1 illustrates the representative Vivaldi element, 16-position, two-ring geometry and analytical descriptor profiles; the geometry was not intended to represent a coupling-qualified finite-antenna array. For ring $r \in \{1,2\}$ and position $i \in \{1,\ldots,8\}$,

$$p_{r,i} = \begin{bmatrix} (R_b + d)\cos\phi_{r,i} \\ (R_b + d)\sin\phi_{r,i} \\ z_r \end{bmatrix}, \phi_{r,i} = \frac{2\pi(i-1)}{8} + \theta_r$$

where d is the antenna-to-surface distance, z_r is the axial ring position and θ_r is the ring offset. The analytical frequency grid was defined as

$$f_k = f_{\min}^{(\mathrm{grid})} + (k-1)\Delta f, k = 1,\ldots,N_v.$$

In this study,

$$f_{\min}^{(\mathrm{grid})} = 3\ \mathrm{GHz}, \Delta f = 0.1\ \mathrm{GHz}, N_v = 51$$

giving a uniform grid from 3 to 8 GHz. For candidate c, the selected frequency subset F_c was characterised by the number of selected frequencies N_f, lower frequency $f_{\min}$, upper frequency $f_{\max}$, centre frequency $f_c = \frac{f_{\min}+f_{\max}}{2}$, and bandwidth $B = f_{\max} - f_{\min}$.

Five analytical channel families were considered: adjacent, skip-one, opposite, inter-ring and full multistatic. Restricting selection to predefined families avoided an unconstrained channel search while retaining differences in channel diversity. Full multistatic denoted all eligible transmit–receive pairs within the analytical framework and did not imply validated matching, isolation, mutual-coupling or loaded-radiation-pattern performance

for a fabricated array. No coupling-derived minimum element separation or multiport S_{mn} threshold was imposed because the 16 positions defined an analytical sensing geometry rather than a simulated finite-antenna array. The benchmark therefore represented a controlled finite-candidate RF design problem rather than a validated finite-antenna or clinical imaging system.

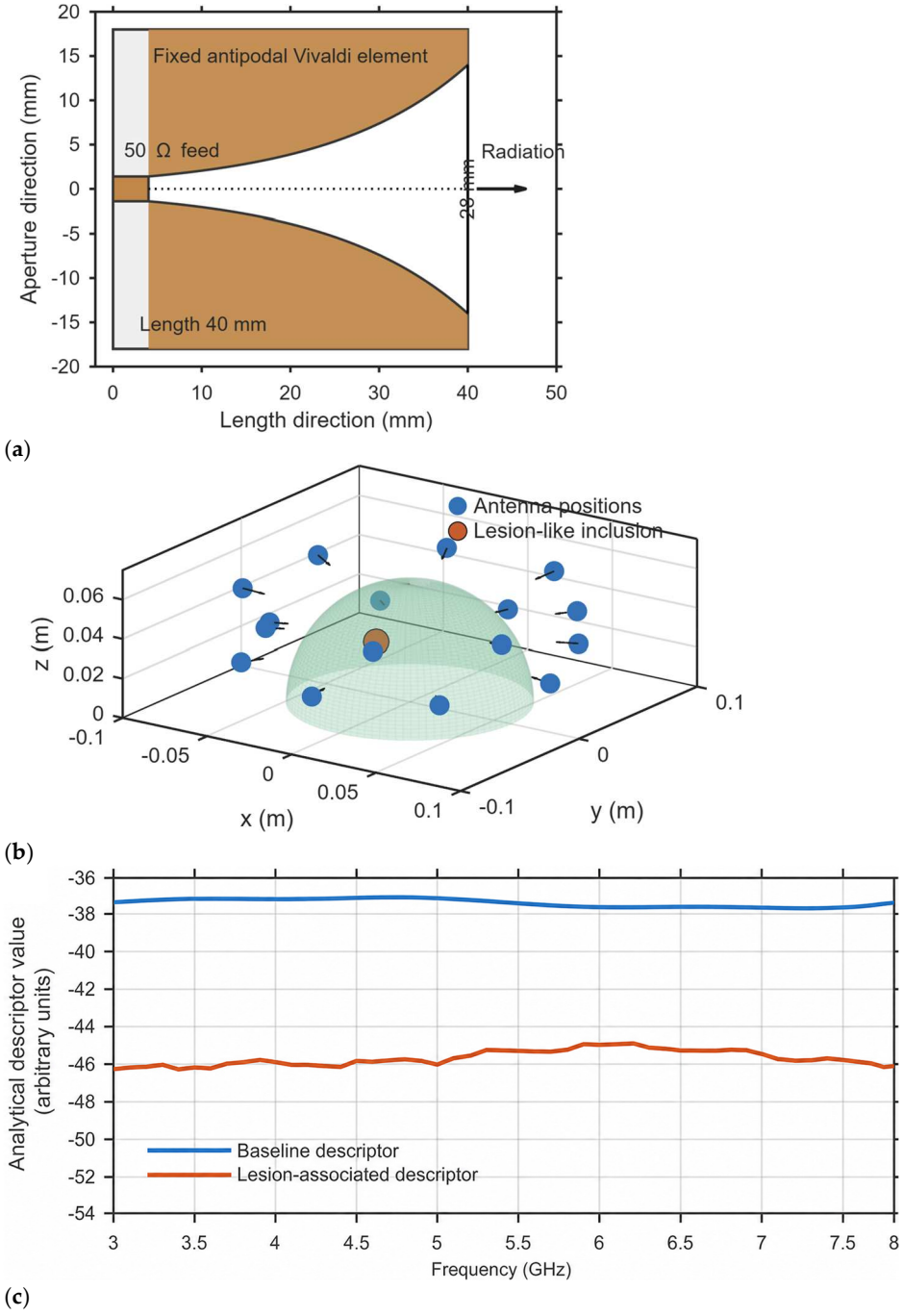


Figure 1. Controlled breast-mimetic benchmark and analytical sensing geometry. (**a**) Representative antipodal Vivaldi element defining the broadband assumptions. (**b**) Sixteen-position, two-ring analytical geometry with an illustrative lesion; spacing was not determined using mutual-coupling or multiport *S*-parameter criteria. (**c**) Representative analytical descriptor profiles over 3–8 GHz.

2.3. Controlled Analytical Benchmark Generator and Ranking Endpoint

For biological scenario s and candidate c, the complete input vector was denoted as $x_{s,c} \in \mathbb{R}^p$, where $p = 26$ in this study. The subset entering the analytical endpoint generator was

$$x_{s,c}^{(g)} = \left[d, \theta, N_f, f_c, B, \chi_c^T, h, a, z_d, C\right]^T \in \mathbb{R}^{p_g}$$

where $p_g = 14$ in this study. Here, $\chi_c \in \{0,1\}^5$ is the one-hot channel-family indicator, h is the fibroglandular fraction, a is lesion diameter, z_d is lesion depth and C is the dielectric-contrast ratio. Conductivity and tissue-correlation-length variables were retained in the complete input vector for feature-scope analysis but did not enter the analytical endpoint equations directly; tissue-permittivity variables contributed through C.

The generator encoded prespecified qualitative relationships rather than approximating an electromagnetic forward solution. Within the benchmark, larger lesion diameter and dielectric contrast increased analytical detectability, whereas greater lesion depth and fibroglandular fraction increased scenario difficulty. Candidate performance varied smoothly around scenario-dependent preferred values of centre frequency, bandwidth and stand-off distance. These relationships defined the ranking task but were not interpreted as experimentally calibrated electromagnetic laws.

Clipping to a finite interval was defined as $\mathcal{C}_{[l,u]}(x) = \min\{\max(x, l), u\}$. Candidate compatibility was represented by $S_{s,c} = \left[S_f, S_d, S_N, S_\theta, S_{\mathrm{ch}}\right]^T$, where the components represented frequency-design, stand-off, frequency-count, angular-offset and channel-family compatibility, respectively. Smooth functions were used to avoid discontinuous candidate-preference boundaries, and channel-family weights were fixed benchmark-construction constants rather than measured antenna characteristics.

Analytical field-coverage, detectability, localisation-sensitivity and robustness scores followed the common form

$$Y_{j,s,c} = \mathcal{C}_{[l_j,u_j]}\left(\beta_{j,0} + \beta_j^T v_{j,s,c} + \xi_{j,s,c}\right), j \in \{F, D, L, Q\}$$

where $v_{j,s,c}$ contains the variables entering score j, $\beta_{j,0}$ and β_j are prespecified coefficients, and $\xi_{j,s,c}$ is record-specific, seed-controlled benchmark variation. Recoverability was defined as

$$R_{s,c} = \mathcal{C}_{[R_{\min},R_{\max}]}\left(D_{s,c}^{r_D} L_{s,c}^{r_L} Q_{s,c}^{r_Q}\right)$$

A composite utility $U_{s,c}$ combined localisation sensitivity, recoverability, detectability, field coverage and robustness using fixed weights. A separate scenario-difficulty term G_s increased with smaller lesion diameter, greater lesion depth, greater fibroglandular fraction and lower dielectric contrast. The primary analytical ranking endpoint was

$$E_{s,c}^{\mathrm{loc}} = \mathcal{C}_{[E_{\min},E_{\max}]}\left(\eta_0 - \eta_U U_{s,c} + \eta_G G_s + \xi_{s,c}^E\right)$$

The analytical localisation-error label was generated directly from these benchmark rules rather than from an electromagnetic-field solution or the maximum of a Maxwell-derived localisation image. The generator also returned a frequency-dependent analytical vector

$$\ell_{s,c} = [\ell_{s,c}, (f_1), \ell_{s,c}, (f_2) \ldots \ell_{s,c}(f_{N_\nu})]^T \in \mathbb{R}^{N_\nu}$$

where $N_\nu = 51$ in this study. This vector provided reproducible frequency-dependent structure but was not interpreted as a calibrated scattering parameter, complex electromagnetic response or measured spectrum.

Appendix A reports the complete functional forms, numerical coefficients, normalisation rules, clipping bounds, channel weights, random-seed rules and baseline settings used in this study. All generator parameters were fixed before surrogate development and were neither estimated from measurements nor fitted or calibrated to the held-out FDTD outcomes. The resulting quantities were therefore interpreted as ranking variables within the analytical benchmark rather than estimates of electromagnetic-system performance.

2.4. Surrogate Development, Model Selection and Interpretability

Each surrogate predicted candidate-level analytical localisation error, after which candidates were ranked in ascending order of predicted error within each biological scenario. The primary surrogate input comprised p_{primary} target-agnostic variables available before candidate reconstruction or response evaluation, where $p_{\text{primary}} = 21$ in this study. These variables represented candidate geometry, frequency design, channel family, tissue organisation and background dielectric and conductivity properties. Five privileged variables—lesion permittivity, lesion conductivity, lesion diameter, lesion depth and dielectric-contrast ratio—were excluded from the primary scope because they would not ordinarily be available during prospective candidate selection. Reinstating these variables produced the oracle-assisted scope, for which $p_{\text{oracle}} = 26$ in this study. The oracle-assisted scope was used only to quantify the effect of privileged scenario information and was treated as a methodological upper bound. Generated response and performance quantities were excluded from both scopes.

Ridge regression, random forest, boosted trees and a random-ranking baseline were compared using grouped five-fold cross-validation repeated three times. All candidates from one biological scenario remained within the same fold. Within each repeat, the five validation folds were disjoint, and each of the 105 analytical-development scenarios appeared in validation exactly once. Fold assignments were reshuffled between repeats; each scenario therefore appeared in validation once per repeat and three times across the complete repeated-validation procedure. Validation sets were disjoint within each repeat but not across repeats. Data transformation, standardisation and model fitting were performed using the training portion of each fold and applied unchanged to the corresponding validation portion.

The model configurations were fixed before held-out FDTD evaluation. Ridge regression used a regularisation parameter of 10^{-2}. Random forest used bootstrap aggregation with 250 regression trees, a minimum leaf size of 5 and surrogate splits disabled. Boosted trees used least-squares boosting with 300 learning cycles, a learning rate of 0.05, a minimum leaf size of 8 and surrogate splits disabled. Random rankings were generated using reproducibly seeded within-scenario permutations. The model-development seed was 20260729; the separate analytical-generator seed is reported in Appendix A. These settings supported reproducible comparisons rather than exhaustive hyperparameter optimisation.

Two scenario-independent candidate templates were defined identically in every analytical scenario, a 10 mm stand-off template and a 15 mm stand-off template, each using the full 3–8 GHz band and full multistatic acquisition. These templates were evaluated at $K = 1$ as global comparators against scenario-conditioned random-forest ranking. The remaining candidate identifiers represented scenario-specific generated designs and were therefore not treated as universally shared configurations.

For the surrogate models, candidate budgets of $K \in \{1, 3, 5, 10, 20\}$ were evaluated analytically, with $K = 5$ prospectively designated as the primary shortlist budget before restricted-FDTD evaluation. Model selection followed a prespecified sequential rule: the model with the lowest mean best-in-shortlist regret was selected first; oracle recall was considered only when mean regret did not distinguish the models; and top-K overlap was

considered only when the preceding criteria remained unresolved. No FDTD outcome entered feature-scope selection, model selection, hyperparameter specification or shortlist-budget determination.

A 20-run fixed-partition stability analysis was performed for random forest and boosted trees. The grouped folds, primary input scope, preprocessing rules, hyperparameters and $K = 5$ budget remained fixed, while only the model-initialisation seed varied. These runs assessed algorithmic sensitivity and were not treated as additional biological scenarios or independent datasets.

Grouped permutation importance was evaluated using the model selected by the prespecified analytical rule and the primary $K = 5$ shortlist budget. Six feature groups were considered: frequency design, channel family, candidate geometry, tissue organisation, background conductivity and background dielectric properties. Each group was permuted over 30 iterations, and importance was defined as the resulting increase in scenario-level five-candidate regret relative to the unpermuted model. Leave-one-feature-group-out retraining provided a complementary assessment using the same grouped partitions, preprocessing rules and model settings. Oracle-assisted results were interpreted only as a methodological upper bound. These analyses assessed predictive dependence on feature groups and were not interpreted as causal physical effects.

2.5. Restricted Held-Out Two-Dimensional TM_z FDTD Challenge

The prospectively fixed analytical surrogate was challenged using a separately implemented two-dimensional TM_z FDTD model. Candidate rankings were generated without access to the corresponding FDTD outcomes. The FDTD domain and restricted candidate structure followed Table 1. Sixteen ideal in-plane source and receiver positions surrounded the breast cross-section, and a fixed spatial-organisation field assigned adipose-like or fibroglandular-like material to each interior cell.

Table 1. Relationship between the analytical benchmark and the restricted FDTD challenge.

Component	Analytical Benchmark and Surrogate Development	Restricted FDTD Challenge
Breast-mimetic representation	Hemispherical design domain with 60 mm radius and 2 mm skin-like layer	2-D TM_z in-plane cross-section with 60 mm radius and 2 mm skin layer
Scenario use	105 scenarios for grouped development after exclusions	Three numerical-reference scenarios and 12 held-out scenarios
Radiating-element representation	Representative antipodal Vivaldi element within a 16-position two-ring design framework	Sixteen ideal in-plane source and receiver positions; no finite antenna geometry
Candidate support	80 array–frequency–channel candidates per scenario	40 candidates per scenario: 20 at each of two stand-offs
Frequency parameterisation	Fifty-one samples over 3–8 GHz and candidate-specific frequency subsets	Five bands: 3–5, 3–8, 4–6, 5–7 and 6–8 GHz
Channel parameterisation	Adjacent, skip-one, opposite, inter-ring and full-multistatic families	Adjacent, skip-one, opposite and in-plane full-multistatic families
Primary shared endpoint	Candidate localisation error within the analytical benchmark	Candidate localisation error reconstructed from FDTD responses
Role	Surrogate development, model selection and within-generator evaluation	Independent evaluation of frozen shortlist transfer

For each scenario, stand-off and transmitter, matched lesion-present and lesion-absent simulations were performed. In the lesion-present state, the background material within

the prescribed circular lesion region was replaced by lesion-like tissue, whereas the lesion-absent state retained the original tissue assignment. Geometry, tissue organisation, excitation and numerical settings were otherwise identical, and the complex frequency-domain responses from the two states were differenced before reconstruction.

Candidate ranking and evaluation were performed separately within each 20-candidate scenario–stand-off unit. The same reconstruction procedure and imaging domain were used for all candidates. Target position was estimated using a log-quadratic subpixel fit around the strongest reconstructed peak, while the discrete-grid maximum was retained as a secondary diagnostic. The reconstruction-image spacing was 2 mm.

Frequency-dependent material properties were represented using the single-pole Debye model

$$\varepsilon^*(\omega) = \varepsilon_\infty + \frac{\Delta\varepsilon}{1 + j\omega\tau} - \frac{j\sigma_s}{\omega\varepsilon_0}$$

where ε^* is the complex relative permittivity, ε_∞ is the infinite-frequency relative permittivity, $\Delta\varepsilon$ is the relaxation strength, τ is the relaxation time, σ_s is the static-conductivity term, ε_0 is the vacuum permittivity and ω is the angular frequency.

Frequency-dependent FDTD tissue properties were obtained from a fixed table derived from published dielectric measurements and parametric models. Skin values were based on the Gabriel tissue data [24,25]; adipose-like values were derived from the high-adipose normal-breast group reported by Lazebnik et al. [26]; fibroglandular-like values were derived from the corresponding low-adipose group [26]; and lesion-like values were derived from malignant breast-tissue data [27]. The reference permittivity and conductivity values were tabulated at integer frequencies from 3 to 8 GHz. A positive-parameter single-pole Debye model was fitted to each scenario-specific tissue-property curve before field computation. Across the 480 scenario–material fits used in this study, the maximum normalised root-mean-square fitting error was 0.03985, below the prespecified limit of 0.05. The fitted Debye law was assigned according to the cellwise tissue label; adipose-like and fibroglandular-like properties were not combined using an effective-medium mixing equation. Debye dispersion was applied only in the restricted FDTD evaluation and did not retrospectively redefine the analytical generator as a dispersive electromagnetic model.

The operational grid spacing used in this study was 0.200 mm, the simulation duration was 12 ns, and excitation used a broadband Gaussian-modulated pulse centred at 5.5 GHz. The time step was set to 95% of the two-dimensional Courant–Friedrichs–Lewy limit. The complete tissue-property table, reference Debye fits, scenario-fit criterion, source waveform, grid, time-stepping, boundary, receiver-sampling and reconstruction settings are reported in Appendix B.

The restricted model did not include finite antennas, feed structures, substrates, ports, impedance matching, loaded radiation patterns, cable effects, inter-ring channels, out-of-plane propagation or a complete 16-port mutual-coupling solution. It therefore provided a separate in-plane electromagnetic challenge rather than complete antenna-system, hardware or experimental validation.

2.6. Numerical-Reference Assessment, Transfer Endpoints and Statistical Analysis

The operational grid was compared with a finer numerical grid using three biological scenarios at both stand-offs, yielding six numerical-reference units. In this study, the operational and finer grid spacings were 0.200 and 0.180 mm, respectively. These scenarios were excluded from analytical model development and held-out transfer evaluation, and the finer grid was treated as a numerical comparator rather than exact or fully grid-converged electromagnetic truth.

Let p_1 and p_2 denote the two highest spatially separated reconstructed peaks, with $p_1 \geq p_2$. The normalised peak margin was $m_{\text{peak}} = \frac{p_1 - p_2}{\max(p_1, \epsilon)}$, where $\epsilon > 0$ is the fixed numerical stabiliser reported in Appendix B. A candidate was classified as localisation-ambiguous when

$$m_{\text{peak}} < 0.01 \text{ and } \| r_1 - r_2 \|_2 \geq 4 \text{ mm}$$

where r_1 and r_2 are the corresponding peak positions. Candidates not meeting this ambiguity condition were considered localisation-qualified.

The relative change in the candidate-used differential response spectrum between the two grids was

$$\Delta_{\text{spec}}(\%) = 100 \frac{\| \Delta H_{\text{op}} - \Delta H_{\text{fine}} \|_2}{\max(\| \Delta H_{\text{fine}} \|_2, \epsilon)}$$

where ΔH_{op} and ΔH_{fine} are the candidate-used differential response vectors obtained on the operational and finer grids, respectively. The prespecified numerical-reference criteria used in this study required $\Delta_{\text{spec}} \leq 5\%$, best-in-qualified-shortlist regret no greater than 2 mm, qualified top-three overlap of at least 2/3, at least three qualified candidates on each grid, and at least one operational-grid shortlisted candidate that remained qualified on the finer grid.

Candidate-support alignment was assessed within each biological scenario. Direct comparison of analytical rankings with FDTD outcomes was permitted only when every restricted FDTD candidate had one unique analytical counterpart; approximate and nearest-neighbour substitutions were not used. When complete alignment was unavailable, surrogate predictions and FDTD outcomes were compared over the same restricted candidate pool within each scenario–stand-off unit. This same-support analysis assessed retained shortlist decision value but could not isolate electromagnetic-representation shift from candidate-support shift. Complete numerical-reference and alignment results are reported in Table A3.

For evaluation unit u, let $e(u, c)$ denote the realised FDTD localisation error of candidate c, $S_K(u)$ the first K candidates ranked by the prospectively fixed surrogate and $C_Q(u)$ the set of localisation-qualified candidates. When the shortlist contained at least one qualified candidate, best-in-shortlist regret was

$$R_K(u) = \min_{c \in S_K(u) \cap C_Q(u)} e(u, c) - \min_{c \in C_Q(u)} e(u, c)$$

Values below zero attributable solely to numerical tolerance were set to zero. The qualified finite-library oracle was the candidate in $C_Q(u)$ with the smallest realised localisation error, and oracle recall was the proportion of evaluation units in which this candidate appeared in $S_K(u)$.

Let $A_k(u)$ and $B_k(u)$ denote the qualified top-k sets obtained from the surrogate and realised FDTD rankings, respectively. Their overlap was

$$O_k(u) = \frac{| A_k(u) \cap B_k(u) |}{k}$$

where k was limited by the number of qualified candidates available in both rankings. Qualified top-three and top-five overlap were calculated separately at their respective shortlist sizes. Qualified Spearman association was calculated over localisation-qualified candidates within each scenario–stand-off unit, while shortlist positions occupied by ambiguous candidates remained in the accounting.

When no shortlisted candidate was qualified, penalised regret was

$$R_K^{\text{pen}}(u) = \max \left[0, e_{\text{worst},Q}(u) - e_{\text{oracle}}(u) \right]$$

where $e_{\mathrm{worst},Q}(u)$ and $e_{\mathrm{oracle}}(u)$ are the largest and smallest localisation errors among qualified candidates. When at least one shortlisted candidate was qualified, penalised regret equalled the observed best-in-shortlist regret. The proportion of evaluation units with regret no greater than 2 mm was reported as a practical secondary endpoint, and ambiguity expenditure was defined as the proportion of the five shortlisted positions occupied by localisation-ambiguous candidates.

Analytical metrics were first calculated within each biological scenario before aggregation across folds and repeated validation. Because the three cross-validation repeats reused the same 105 biological scenarios, repeat-level estimates were not treated as independent uncertainty replicates. Confidence intervals for the fixed-template contrasts, grouped-permutation effects and paired random-forest-minus-boosted-tree comparison were estimated using 2000 biological-scenario bootstrap resamples. Biological scenarios were sampled with replacement, while repeated-validation and model-seed observations associated with each sampled scenario remained nested within that scenario cluster and were not treated as independent resampling units. The pairing between compared models or candidate-selection policies was preserved within every resample. Two-sided 95% percentile-bootstrap intervals were reported.

Held-out uncertainty was estimated separately using 2000 biological-scenario cluster-bootstrap resamples. Biological scenarios were sampled with replacement, and both stand-offs associated with each sampled scenario were retained together. Two-sided 95% percentile-bootstrap intervals were reported. Analyses stratified by stand-off, frequency range or channel family were descriptive because the held-out evaluation contained only 12 independent biological scenarios and because candidate observations within each scenario–stand-off unit were nested.

The held-out library contained 480 candidate evaluations, of which the prospectively fixed $K = 5$ policy retained 120. Because candidates within each scenario–stand-off unit reused common broadband field responses, this candidate-level reduction was not interpreted as an equivalent reduction in electromagnetic propagation. The 24 scenario–stand-off configurations, 16 transmitters and two biological states required 768 transmitter-state FDTD runs.

Analytical generation, model fitting, reconstruction, statistical analysis and FDTD processing were implemented in MATLAB R2025a. The appendices report the generator equations and coefficients, numerical settings, feature definitions, random seeds, model configurations, analysis partitions and candidate-support mappings.

3. Results

3.1. Analytical Benchmark and Evaluation Partition

The controlled analytical benchmark comprised 120 breast-mimetic scenarios and 80 finite antenna–frequency–channel candidates per scenario, yielding 9600 nested scenario–candidate records. Each record contained 26 descriptors and 57 generated outputs: 51 frequency-dependent analytical descriptor values and six design-performance quantities comprising localisation error, detectability, localisation sensitivity, robustness, recoverability and field coverage. The complete analytical archive therefore preserved the multi-output structure of the original RF design benchmark, whereas the cross-representation analysis focused on localisation error because this endpoint was defined consistently in both the analytical and restricted FDTD representations.

The biological scenario, rather than the individual candidate record, was treated as the independent unit, and all candidates from the same scenario remained grouped throughout model development and evaluation. After excluding three numerical-reference scenarios and 12 held-out FDTD scenarios, 105 scenarios remained for grouped analytical develop-

ment. The primary model used 21 target-agnostic descriptors that excluded privileged lesion information. A separate 26-descriptor oracle-assisted scope, comprising the 21 primary descriptors and five additional lesion or complete-scenario variables, was evaluated as a methodological upper bound. Table 2 summarises the benchmark structure, descriptor scopes and evaluation partition.

Table 2. Analytical benchmark, descriptor scopes and evaluation partition.

Component	Count	Role
Breast-mimetic scenarios	120	Complete controlled benchmark
Analytical candidates per scenario	80	Finite antenna–frequency–channel design library
Analytical scenario–candidate records	9600	Nested candidate evaluations
Frequency-dependent analytical descriptor values	51	Reproducible analytical frequency profiles
Design-performance outputs	6	Localisation error, detectability, localisation sensitivity, robustness, recoverability and field coverage
Available descriptors	26	Complete analytical descriptor scope
Primary target-agnostic descriptors	21	Prospective analytical shortlisting
Oracle-assisted descriptors	26	Methodological upper bound containing privileged information
Analytical-development scenarios	105	Grouped model development and selection
Numerical-reference scenarios	3	Operational-grid assessment at two stand-offs
Held-out FDTD scenarios	12	Independent analytical-to-FDTD transfer challenge

The analytical benchmark generated structured variation in antenna distance, ring offset, frequency coverage, channel family, tissue organisation and lesion-like conditions. The analytical localisation-error label was generated directly from the endpoint equation defined in Section 2.3 and Appendix A. Figure 2 presents a separately generated illustrative localisation-sensitivity map and did not determine the analytical label, candidate ranking or FDTD-transfer results.

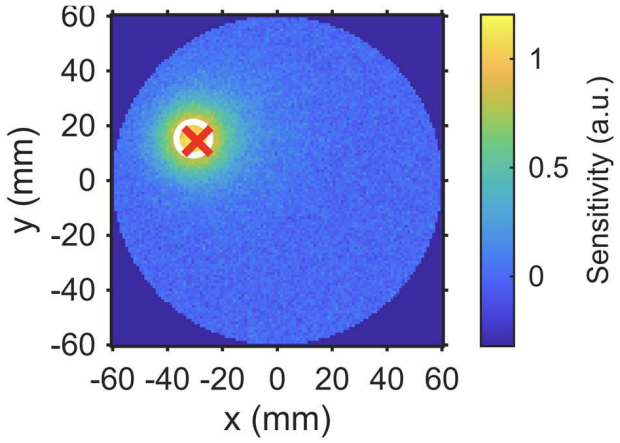


Figure 2. Representative analytical-domain localisation-sensitivity map generated for visualisation. The white circle indicates the known lesion-like inclusion centre, and the red cross indicates the maximum of the illustrative map. The colour scale represents relative localisation sensitivity in arbitrary units. The map did not determine the analytical localisation-error label, candidate ranking or FDTD-transfer results.

3.2. Random Forest Provided the Lowest-Regret Analytical Shortlist

Analytical shortlisting was evaluated using grouped five-fold cross-validation repeated three times. At the prespecified primary budget of $K = 5$, random forest achieved the lowest mean best-in-shortlist regret and was selected according to the frozen regret-first rule. Its mean and median regrets were 0.0575 and 0 mm, respectively, and every analytical validation scenario retained a candidate with regret no greater than 2 mm. Oracle recall was 0.7651, mean top-five overlap was 0.4011 and mean Spearman association was 0.7767. Boosted trees achieved higher oracle recall, top-five overlap and rank association than random forest, but their mean regret was slightly higher at 0.0661 mm.

Ridge regression produced a mean regret of 0.1057 mm, whereas random ranking performed substantially worse, with a mean regret of 0.9372 mm and near-zero rank association. These results indicate strong within-generator shortlist utility while showing that the evaluated ranking metrics did not identify the same preferred model.

The grouped analytical model performance at $K = 5$ is summarised in Table 3. The scenario-conditioned ranking was additionally compared with the two candidate templates whose parameter definitions were identical across all scenarios. At $K = 1$, random forest produced a mean regret of 0.5835 mm, compared with 0.8526 mm for the common 10 mm stand-off template and 0.9557 mm for the common 15 mm stand-off template. Both templates used the full 3–8 GHz band and full multistatic acquisition. The corresponding mean-regret reductions were 0.2691 mm (scenario-cluster 95% CI: 0.1436–0.3953 mm) and 0.3722 mm (95% CI: 0.2381–0.5018 mm), respectively. These results support an advantage of scenario-conditioned shortlisting within the analytical benchmark but do not establish superiority under FDTD, finite-antenna, hardware or clinical conditions.

Table 3. Grouped analytical model performance at $K = 5$.

Model	Mean Regret (mm)	Oracle Recall	Top-Five Overlap	Spearman Association	Fraction with Regret $\leq$2 mm
Ridge regression	0.1057	0.6476	0.2571	0.6830	1.0000
Random forest, frozen	0.0575	0.7651	0.4011	0.7767	1.0000
Boosted trees	0.0661	0.7841	0.4868	0.8367	1.0000
Random ranking	0.9372	0.0603	0.0381	−0.0095	0.9556

Model selection followed the frozen sequential hierarchy of lowest mean regret, followed, only where required, by highest oracle recall and highest top-five overlap.

The analytical advantage of model-guided ranking was maintained across the evaluated candidate budgets. Increasing K improved oracle inclusion and reduced best-in-shortlist regret for all trained models, while random ranking remained consistently weaker. The budget-dependent oracle recall and best-in-shortlist regret are shown in Figure 3. The $K = 5$ budget was retained for external evaluation because it imposed a restrictive candidate limit while preserving strong analytical-domain performance.

The 20-run fixed-partition model-seed analysis supported the stability of the regret-first selection. Across model-initialisation seeds, random forest achieved a mean regret of 0.0517 mm, with a standard deviation of 0.0049 mm and a range of 0.0418–0.0608 mm. Its mean oracle recall, top-five overlap and Spearman association were 0.775, 0.501 and 0.773, respectively. Random forest achieved lower mean regret than boosted trees in 16 of the 20 seed runs. The mean paired random-forest-minus-boosted-tree regret difference was −0.0044 mm, although the scenario-cluster 95% interval of −0.0325 to 0.0218 mm included zero. The model-seed analysis therefore showed that the selection of random forest was

not an artefact of a single initialisation seed, but it did not establish statistically definitive or universal superiority over boosted trees.

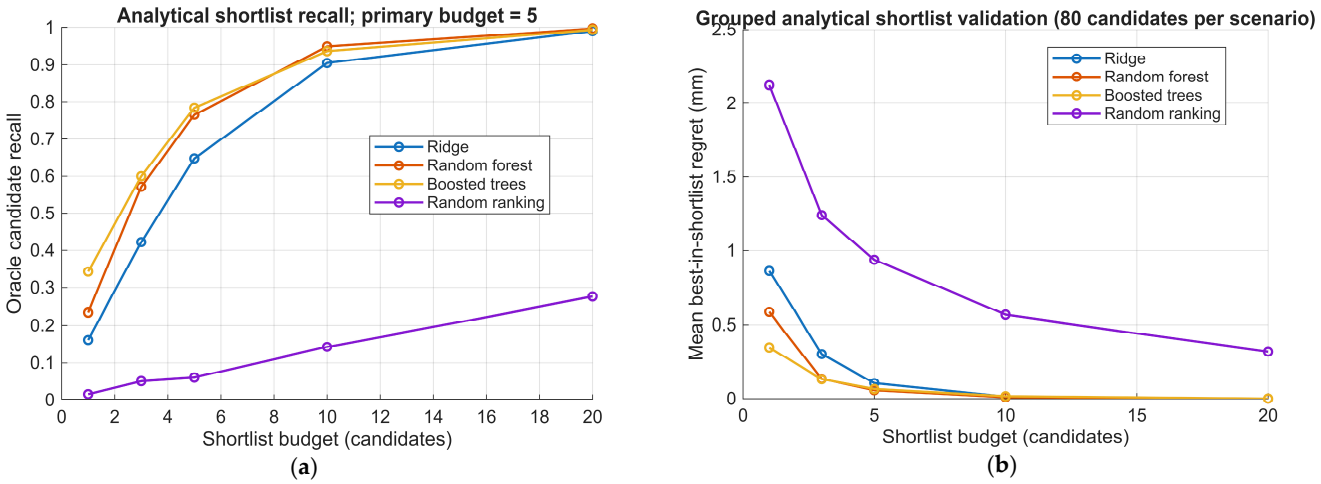


Figure 3. Analytical-domain shortlisting across candidate budgets. (**a**) Oracle recall and (**b**) mean best-in-shortlist regret for ridge regression, random forest, boosted trees and random ranking under grouped five-fold cross-validation repeated three times. The vertical dashed line marks the prespecified primary budget, $K = 5$. The three numerical-reference scenarios and 12 held-out FDTD scenarios were excluded from model development and selection.

3.3. Analytical Ranking Depended Primarily on Frequency and Channel Descriptors

Grouped permutation and leave-one-feature-group-out analyses were used to identify the descriptor groups supporting analytical-domain shortlisting. Frequency design produced the largest effect: permuting this group increased mean five-candidate regret by 0.2687 mm, with a scenario-bootstrap 95% interval of 0.1517–0.4199 mm, while removing the group and retraining the model increased regret by 0.2390 mm. Channel family produced the second-largest effects, with permutation and leave-one-feature-group-out regret increases of 0.1555 and 0.1436 mm, respectively.

Candidate geometry produced smaller but consistently positive effects. By contrast, the permutation intervals for tissue organisation, background conductivity and background dielectric properties spanned zero, and their leave-one-feature-group-out effects were negligible or negative on the prespecified split. These results quantify surrogate dependence within the analytical generator and do not establish that descriptor groups with small analytical effects are physically unimportant under FDTD, finite-antenna or experimental conditions.

On the corresponding grouped comparison split, the 21-descriptor target-agnostic model achieved a mean regret of 0.0856 mm, whereas the 26-descriptor oracle-assisted model achieved a mean regret of 0 mm. Access to privileged lesion and complete-scenario information therefore substantially simplified the analytical ranking task. Because this information would not ordinarily be available during prospective candidate selection, the oracle-assisted result was treated as a methodological upper bound rather than as deployable shortlisting performance. The grouped permutation and leave-one-feature-group-out results are summarised in Table 4.

3.4. The Operational FDTD Grid Preserved Shortlist-Level Decisions

The numerical-reference assessment compared the 0.200 mm operational grid with the finer 0.180 mm grid across three scenarios at both stand-offs, yielding six scenario–stand-off units. All six met the prespecified shortlist-level criteria. At the reference shortlist size

$K_{ref} = 3$, the maximum candidate-used response-spectrum change was 1.3035%, below the 5% criterion. The maximum best-in-qualified-shortlist regret on the finer grid was 0 mm, indicating that every operational-grid shortlist retained at least one candidate attaining the qualified finite-library optimum on the finer grid. The minimum qualified top-three overlap was 0.6667, meeting the required value of 2/3.

Table 4. Grouped permutation and leave-one-feature-group-out effects at $K = 5$.

Feature Group	Permutation (Δ) Regret (mm)	95% Interval (mm)	LOFO (Δ) Regret (mm)
Frequency design	0.2687	0.1517 to 0.4199	0.2390
Channel family	0.1555	0.0906 to 0.2514	0.1436
Candidate geometry	0.1084	0.0235 to 0.2157	0.0446
Tissue organisation	0.0098	−0.0313 to 0.0365	−0.0054
Background conductivity	−0.0011	−0.0463 to 0.0329	−0.0381
Background dielectric properties	−0.0092	−0.0398 to 0.0049	−0.0136

Positive values indicate increased regret after permutation or feature-group removal. Negative leave-one-feature-group-out estimates indicate no evidence of positive dependence on that group in this split and should not be interpreted as beneficial physical effects.

Candidate availability, tissue-map consistency, canonical propagation controls, convolutional perfectly matched layer checks, CPU–GPU agreement and candidate accounting also satisfied their respective criteria. The 0.200 mm grid was therefore retained for the held-out shortlist evaluation. Exact top-one identity was less stable: raw top-one identity was not preserved, and the maximum displacement of the reconstructed maximum was 2.3384 mm. These secondary sensitivity results indicate that the assessment supported shortlist-level decisions rather than numerical identity or exact preservation of a single nominal optimum.

Complete results are reported in Table A3. The three primary shortlist-level numerical-reference endpoints are summarised in Figure 4.

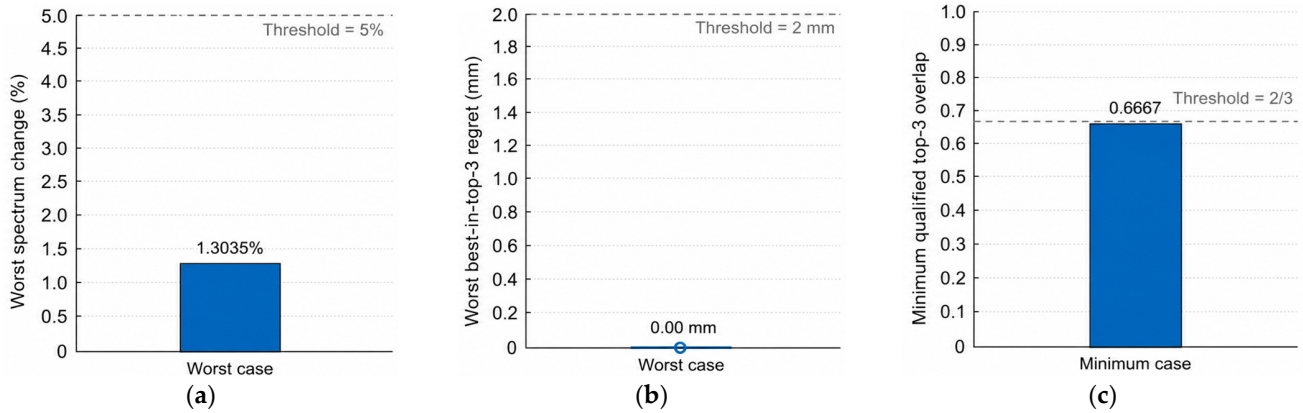


Figure 4. Numerical-reference assessment at shortlist level. The 0.200 mm operational grid was compared with the 0.180 mm grid across six numerical-reference scenario–stand-off units. (**a**) Maximum candidate-used response-spectrum change, with the prespecified threshold of 5%. (**b**) Maximum best-in-qualified-shortlist regret at the finer grid, with the 2 mm threshold. The observed value of 0 mm indicates that each operational-grid shortlist retained at least one qualified candidate attaining the finer-grid finite-library optimum. (**c**) Minimum qualified top-three overlap, with the required threshold of 2/3. Dashed lines indicate the prespecified criteria. The figure evaluates shortlist-level stability and does not imply exact top-one preservation or numerical identity between grids.

3.5. Candidate-Support Mismatch Defined the External Comparison

Exact candidate-support alignment was assessed before direct comparison of the analytical and FDTD outcomes. Only 12 of the 480 restricted FDTD scenario–candidate rows had one unique counterpart in the original analytical library within the same biological scenario; complete support alignment was therefore not achieved. Approximate and nearest-neighbour substitutions were not permitted because a direct matched analytical-to-FDTD comparison would otherwise have combined changes in electromagnetic representation with differences in candidate design. The external analysis consequently applied the frozen surrogate to the restricted FDTD candidate pool and compared its predicted ordering with the realised FDTD outcomes over the same pool. Given the incomplete alignment between the analytical and FDTD candidate libraries, the external evaluation was formulated as a decision-level transfer assessment rather than a direct field-level fidelity comparison. It quantified the extent to which the prospectively frozen analytical shortlist retained value under the restricted FDTD representation. The complete candidate-support alignment assessment is reported in Table A3.

3.6. The Frozen Shortlist Retained Partial but Unreliable Value Under FDTD Evaluation

The frozen random forest was evaluated on 12 held-out biological scenarios at stand-offs of 15 and 20 mm, yielding 24 scenario–stand-off units. Each unit contained 20 restricted candidates, of which the first five ranked by the frozen surrogate formed the retained shortlist. Every unit retained at least one localisation-qualified shortlisted candidate, so penalised regret equalled observed best-in-shortlist regret throughout the held-out evaluation. At $K = 5$, mean penalised regret was 3.2127 mm, with a scenario-cluster 95% bootstrap interval of 1.319–5.789 mm. Qualified oracle recall was 0.1667, qualified top-three overlap was 0.2222, qualified top-five overlap was 0.3250 and qualified Spearman association was −0.1444. Fourteen of the 24 units, or 58.3%, retained a qualified candidate within 2 mm of the finite-library FDTD optimum. These results were markedly weaker than the grouped analytical-domain findings: the negative mean rank association and low qualified-optimum inclusion indicated that the analytical ordering was not preserved reliably. Nevertheless, the shortlist retained a near-optimal qualified candidate in more than half of the evaluation units. The surrogate therefore retained partial shortlisting value, but the evidence did not support its use as a dependable FDTD optimiser or as a replacement for evaluation of the complete restricted candidate library. The overall and stand-off-stratified held-out transfer results are summarised in Table 5.

Table 5. Held-out same-support surrogate-to-FDTD transfer at $K = 5$.

Endpoint	Overall, 24 Units	15 mm, 12 Units	20 mm, 12 Units
Mean penalised regret, mm (95% CI)	3.2127 (1.319–5.789)	4.8796 (1.729–9.468)	1.5458 (0.497–2.965)
Qualified oracle recall	0.1667	0.1667	0.1667
Qualified top-three overlap	0.2222	0.2500	0.1944
Qualified top-five overlap	0.3250	0.3500	0.3000
Qualified Spearman association	−0.1444	−0.1418	−0.1470
Fraction with regret ≤2 mm	0.5833	0.4167	0.7500
Ambiguous fraction among selected candidates	0.3083	0.3833	0.2333

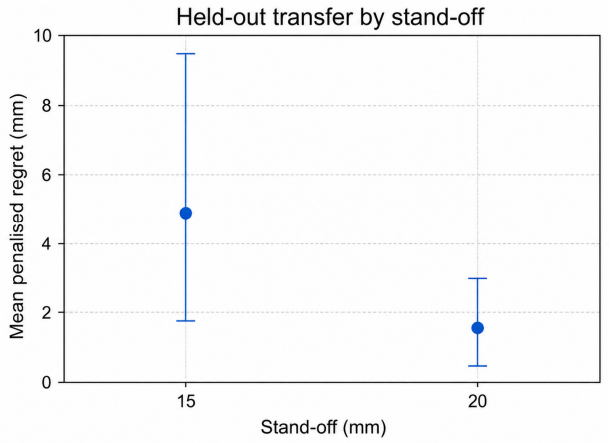
Stand-off strata are descriptive. Scenario-cluster resampling retained both stand-offs associated with each biological scenario.

Complete candidate identities, localisation errors, regret values, ranking agreement and ambiguity diagnostics for the 24 scenario–stand-off units are provided in Table A4.

3.7. Transfer Failure Was Structured by Frequency and Channel Family

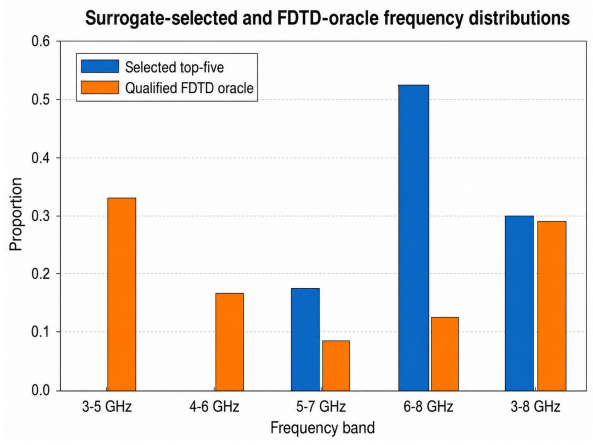
Transfer was descriptively more favourable at 20 mm than at 15 mm: mean penalised regret decreased from 4.8796 to 1.5458 mm, the proportion of units with regret no greater than 2 mm increased from 41.7% to 75.0%, and ambiguity expenditure decreased from 38.3% to 23.3%. However, the corresponding uncertainty intervals were wide and overlapping, so these differences do not establish a stand-off effect. The frequency composition of the surrogate shortlist also differed substantially from that of the qualified finite-library FDTD optima. Of the 120 selected top-five positions, 63, or 52.5%, used the 6–8 GHz band, whereas only three of the 24 qualified FDTD optima, or 12.5%, occurred in that band. No 3–5 or 4–6 GHz candidate appeared among the surrogate-selected positions, although these two bands supplied 12 of the 24 qualified FDTD optima. Channel-family preferences showed a similar mismatch. Full-multistatic candidates accounted for 69 of the 120 selected positions and eight of the 24 qualified FDTD optima. Opposite-channel candidates occupied 36 selected positions, or 30.0%, but supplied no qualified FDTD optimum. By contrast, skip-one candidates occupied only 15 selected positions, or 12.5%, but supplied 12 of the 24 qualified optima, while adjacent candidates were absent from the selected top-five positions but supplied four qualified optima. Localisation ambiguity was also strongly channel dependent.

Under the frozen ambiguity rule, the pre-held-out analytical evidence yielded ambiguity fractions of 5.3%, 19.3%, 98.7% and 3.3% for adjacent, skip-one, opposite and full in-plane multistatic candidates, respectively. The held-out FDTD library showed the same principal pattern, with corresponding fractions of 9.2%, 6.7%, 95.8% and 3.3%. The transfer gap was therefore concentrated on candidate dimensions that strongly influenced analytical selection rather than appearing solely as a uniform increase in localisation error. In particular, the surrogate allocated a substantial proportion of its shortlist to upper-frequency and opposite-channel candidates, whereas the qualified FDTD optima were distributed more broadly across frequency bands and were more frequently associated with skip-one channels. These comparisons remain descriptive because only 12 independent biological scenarios were evaluated, and the five selected positions within each scenario–stand-off unit were nested observations. Figure 5 summarises the stand-off-specific regret and the mismatch between surrogate-selected and FDTD-optimal frequency bands and channel families.

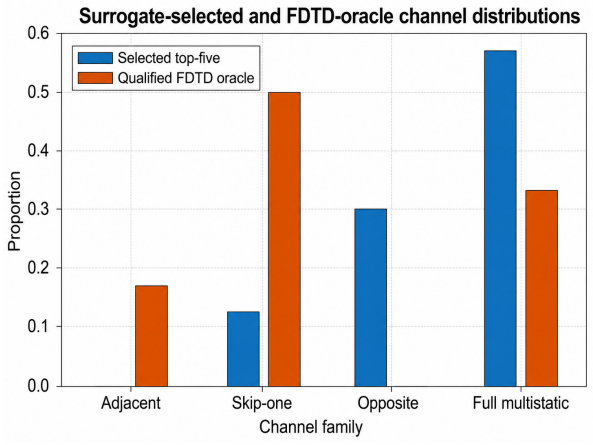


(**a**)

Figure 5. *Cont.*



(**b**)



(**c**)

Figure 5. Structured mismatch in held-out surrogate-to-FDTD transfer. (**a**) Mean penalised regret at stand-offs of 15 and 20 mm, with scenario-cluster bootstrap 95% confidence intervals. (**b**) Distribution of surrogate-selected top-five positions (n = 120) and qualified finite-library FDTD optima (n = 24) across frequency bands. (**c**) Corresponding distributions across channel families. The selected positions comprise five nested candidates from each of 24 scenario–stand-off units; panels (**b**,**c**) are therefore descriptive. The analysis included 12 independent biological scenarios, and the stand-off comparison was not treated as confirmatory.

3.8. Shortlisting Reduced Candidate-Level Assessment but Not FDTD Propagation

The complete held-out restricted library contained 480 candidate evaluations, of which the $K = 5$ policy retained 120, corresponding to a 75% reduction and a fourfold decrease in candidate-level reconstruction or configuration assessment. Candidate-level channel–frequency samples decreased from 1,036,800 to 548,496, representing a 47.1% reduction in this workload proxy. This reduction did not produce a fourfold decrease in FDTD propagation because the 20 candidates within each scenario–stand-off unit reused common broadband field responses. The held-out campaign therefore comprised 24 unique scenario–stand-off broadband configurations, each evaluated using 16 transmitter positions and two biological states, yielding 768 transmitter-state FDTD runs. All 768 recorded quality-control rows used GPU execution and met the frozen numerical criteria. The computational benefit therefore concerned candidate-level reconstruction, configuration comparison and channel–frequency processing rather than the number of Maxwell field solutions. A reliable total FDTD wall-clock time was unavailable in the preserved execution records and was not reconstructed retrospectively. Accordingly, the fourfold reduction in candidate-level assessment should not be interpreted as a fourfold FDTD speed-up.

4. Discussion

The analytical benchmark showed that the finite antenna–frequency–channel design space could be shortlisted reproducibly within the controlled generator. Comparison with the two common fixed templates further supported scenario-conditioned shortlisting within this analytical setting. However, the held-out FDTD findings showed that this analytical-domain advantage was not representation-independent. The surrogate therefore learned the preference structure encoded by the analytical model rather than a universally preserved electromagnetic ordering. Its retained value was partial rather than negligible: the frozen shortlist frequently included a candidate close to the restricted FDTD optimum, but low optimum inclusion, limited shortlist agreement and negative rank association did not support its use for final configuration selection. The surrogate should therefore be used for provisional candidate prioritisation followed by independent electromagnetic evaluation.

The transfer gap was concentrated in the candidate dimensions that most strongly influenced analytical ranking. The surrogate favoured upper-frequency and opposite-channel configurations, whereas the qualified FDTD optima were distributed more broadly across frequency bands and occurred more often among skip-one candidates. Opposite-channel configurations were also frequently localisation-ambiguous and did not supply a qualified FDTD optimum. Analytical feature importance should therefore not be interpreted directly as a transferable design rule; instead, highly influential frequency, channel and geometry descriptors identify the dimensions requiring the most stringent external validation. Performance was descriptively better at 20 mm than at 15 mm, but the small paired scenario set and overlapping uncertainty intervals do not establish a stand-off effect.

The principal contribution is a validation-aware decision framework rather than a new surrogate algorithm. The framework separates grouped analytical development from a frozen external challenge, evaluates whether the operational numerical grid preserves shortlist-level decisions, retains localisation ambiguity in the analysis and assesses support alignment between the analytical and FDTD candidate libraries. Because exact candidate correspondence was incomplete, the observed transfer gap cannot be attributed solely to differences between the analytical and FDTD forward models. This support-aware interpretation prevents combined changes in candidate design and electromagnetic representation from being presented as a pure solver comparison. It also shows that descriptor-based modelling, even when physically motivated, does not remove the need for independent evaluation under the representation relevant to the intended design claim.

The frozen shortlist reduced candidate-level reconstruction and configuration assessment, supporting its use as an initial shortlisting mechanism when downstream evaluation is costly. This saving should not be equated with an equivalent reduction in electromagnetic propagation because multiple candidates reused the same broadband field responses. The practical computational benefit therefore depends on whether the dominant cost arises from field generation, reconstruction, channel–frequency processing, configuration comparison or experimental acquisition. The concentration of selected candidates within particular frequency bands and channel families also suggests that future shortlist policies may benefit from prespecified diversity constraints rather than relying exclusively on predicted rank; any such policy should be defined prospectively and evaluated on new held-out scenarios. Operationally, the framework provides a gated design pathway: construct the analytical candidate library, develop the surrogate using grouped scenarios, freeze the model and shortlist budget, qualify the independent numerical representation and evaluate the retained candidates before final selection. Failure to preserve ranking at the external gate does not eliminate the surrogate's value for preliminary shortlisting, but it prevents the analytical shortlist from being promoted directly to a final electromagnetic configuration.

No experimental study was performed. The conclusions are restricted to a controlled analytical benchmark and a restricted 2-D TM_z FDTD challenge using ideal in-plane sources and receivers. The model omitted finite antennas, feeds, ports, loaded radiation patterns, mutual coupling, out-of-plane propagation, calibration effects and measurement noise.

Only 12 independent held-out biological scenarios and two stand-offs were evaluated, and the finer numerical grid served as a comparator rather than exact electromagnetic truth. In addition, incomplete candidate-support alignment prevented separation of electromagnetic-model shift from candidate-library shift. Future studies should use a prospectively frozen candidate manifest shared exactly across analytical and higher-fidelity representations, followed by three-dimensional finite-antenna simulation, tissue-loaded phantom measurements and hardware testing incorporating coupling, positioning, calibration, noise and repeatability. Within the present scope, the surrogate should be used to generate an auditable shortlist for subsequent validation rather than to determine the final antenna, frequency or channel configuration.

5. Conclusions

This study developed a surrogate-ranking framework for finite-budget broadband antenna, frequency and channel selection and assessed whether its analytical-domain candidate preferences transferred to a separately implemented restricted 2-D TM_z FDTD representation. Although the frozen surrogate achieved strong and reproducible shortlisting performance within the controlled analytical generator, its candidate ordering did not transfer reliably in the held-out FDTD challenge. The shortlist retained partial shortlisting value by preserving a qualified candidate within 2 mm of the finite-library FDTD optimum in more than half of the evaluated scenario–stand-off units. However, low optimum inclusion, limited shortlist agreement and negative rank association did not support its use as a final FDTD optimiser. The principal contribution is therefore a validation-aware workflow in which analytical surrogates generate provisional, auditable candidate shortlists for subsequent independent electromagnetic evaluation. The approach reduced candidate-level reconstruction and configuration assessment but did not produce an equivalent reduction in electromagnetic-field propagation because multiple candidates reused common broadband responses. Three-dimensional finite-antenna simulation, phantom measurements and hardware validation remain necessary before broader system-performance claims can be made.

Funding: This research received no external funding.

Data Availability Statement: The data supporting the principal findings of this study are included in the article tables and Sections A–C. Additional processed benchmark outputs, frozen numerical configurations and MATLAB code are available from the corresponding author upon reasonable request. No human-participant data, patient records, clinical images, animal data or patient-identifiable information were used.

Acknowledgments: During preparation of this manuscript, the author used ChatGPT for language refinement, grammar checking, formatting support and improvements to textual clarity and organisation. ChatGPT was not used to generate simulation data, perform numerical analyses, select results or draw scientific conclusions. The author reviewed and edited all AI-assisted content and takes full responsibility for the publication.

Conflicts of Interest: The author declares no conflicts of interest.

Appendix A. Complete Specification of the Controlled Analytical Benchmark Generator

Appendix A.1. Inputs and Candidate Compatibility

The analytical generator was constructed as a controlled candidate-ranking benchmark rather than an electromagnetic forward solver. Its functions encoded prespecified directional relationships: analytical detectability increased with lesion diameter and dielectric contrast; scenario difficulty increased with lesion depth and tissue heterogeneity; candidate performance varied smoothly around preferred frequency, bandwidth and stand-off values; and greater frequency and channel diversity increased analytical coverage and robustness. The coefficients were fixed benchmark-construction settings and were not physical constants or parameters estimated from measurements or electromagnetic simulations.

For biological scenario s and candidate c, the complete archived input vector was denoted $x_{s,c} \in \mathbb{R}^p$, where $p = 26$ in this study. Its ordered variables were stand-off distance d, angular offset θ, normalised angular offset, frequency count N_f, lower and upper frequencies $f_{\min}$ and $f_{\max}$, centre frequency f_c, bandwidth B, five channel-family indicators, four relative-permittivity descriptors, four conductivity descriptors, fibroglandular fraction h, tissue-correlation length, lesion diameter a, lesion depth z_d, and dielectric-contrast ratio C. The five channel indicators represented adjacent, skip-one, opposite, inter-ring and full multistatic acquisition.

The dielectric-contrast ratio was

$$C = \frac{\varepsilon_{\text{lesion}} - \varepsilon_{\text{adipose}}}{\max(\varepsilon_{\text{fib}}, 1)} \tag{A1}$$

Only the active subset

$$x_{s,c}^{(g)} = \left[d, \theta, N_f, f_c, B, \chi_c^T, h, a, z_d, C\right]^T \in \mathbb{R}^{p_g} \tag{A2}$$

entered the analytical endpoint generator, where $p_g = 14$ in this study, and $\chi_c \in \{0,1\}^5$ is the one-hot channel-family indicator. Conductivity and tissue-correlation-length variables remained in the complete archived record but did not enter the analytical endpoint equations directly. The dimension $p = 26$ refers to the complete archived record rather than the primary target-agnostic surrogate input described in Section 2.4.

Clipping to the interval $[l, u]$ was defined as

$$\mathcal{C}_{[l,u]}(x) = \min\{\max(x, l), u\} \tag{A3}$$

The normalised scenario variables were

$$q_a = \mathcal{C}_{[0,1]}\left(\frac{a-5}{10}\right), q_z = \mathcal{C}_{[0,1]}\left(\frac{z_d - 8}{42}\right) \tag{A4}$$

$$q_h = \mathcal{C}_{[0,1]}\left(\frac{h-0.10}{0.30}\right), q_C = \mathcal{C}_{[0,1]}\left(\frac{C-0.35}{1.20}\right) \tag{A5}$$

The scenario-dependent preferred centre frequency, stand-off and bandwidth were

$$f_c^* = 6.9 - 1.1q_a + 0.45q_z \tag{A6}$$

$$d^* = 7.5 + 7.5q_z + 2.0q_h, B^* = 3.6 - 0.8q_h \tag{A7}$$

Here, f_c^* and B^* are expressed in GHz, and d^* is expressed in millimetres. Frequency, stand-off, frequency-count and angular compatibility were defined as

$$S_f = \exp\left[-\left(\frac{f_c - f_c^*}{1.25}\right)^2\right]\left[0.65 + 0.35\min\left(\frac{B}{B^*}\middle|1\right)\right] \tag{A8}$$

$$S_d = \exp\left[-\left(\frac{d - d^*}{5.5}\right)^2\right] \tag{A9}$$

$$S_N = 0.55 + 0.45\min\left(\frac{N_f}{20}, 1\right) \tag{A10}$$

$$S_\theta = 0.85 + 0.15\cos^2(\theta - 11.25^\circ) \tag{A11}$$

The channel-family score was

$$S_{\text{ch}} = (0.48\chi_{c,1} + 0.60\chi_{c,2} + 0.72\chi_{c,3} + 0.86\chi_{c,4} + 0.93\chi_{c,5})(0.92 + 0.08I_{\text{broad}}) \tag{A12}$$

where $I_{\text{broad}} = 1$ for inter-ring or full multistatic acquisition and zero otherwise.

These channel weights were benchmark-construction settings and were not measured coupling, isolation or scattering-parameter values.

Appendix A.2. Analytical Scores and Localisation Endpoint

Unless otherwise stated, each Gaussian random variable below denotes a separate sequential draw from the record-specific random stream. The generator-level variation terms were

$$b_i = a_{\text{num}}\xi_i, \xi_i \sim \mathcal{N}(0,1) \tag{A13}$$

The synthetic variation amplitude was

$$a_{\text{num}} = a_{\text{grid}} + \frac{0.018}{\max(s_{\text{PML}}, 0.2)} + \frac{0.014}{\max\left(s_{\text{boundary}}, 0.2\right)} \tag{A14}$$

The primary analytical dataset used the medium setting, $a_{\text{grid}} = 0.020$, with $s_{\text{PML}} = s_{\text{boundary}} = 1$, giving

$$a_{\text{num}} = 0.052 \tag{A15}$$

The historical implementation labels associated with the latter two terms referred to PML and boundary scales; however, these terms controlled only synthetic benchmark variation. They did not implement absorbing boundaries or electromagnetic boundary conditions.

Optional measurement-noise terms were

$$m_i = \sigma_n\tilde{\xi}_i, \tilde{\xi}_i \sim \mathcal{N}(0,1) \tag{A16}$$

with $\sigma_n = 0$ in the primary analysis.

The analytical field-coverage score was

$$F_{s,c} = \mathcal{C}_{[0.18,0.98]}(0.20 + 0.36S_{\text{ch}} + 0.24S_d + 0.12S_N - 0.16q_h + 0.04\xi_F + 0.60b_6 + 0.25m_6) \tag{A17}$$

The analytical detectability score was

$$D_{s,c} = \mathcal{C}_{[0.01,0.95]}(0.18 + 0.34q_C + 0.22S_f + 0.14S_{\text{ch}} + 0.13q_a - 0.12q_h - 0.07q_z + 0.03\xi_D + 0.55b_2 + 0.30m_2) \tag{A18}$$

The analytical localisation-sensitivity score was

$$L_{s,c} = \mathcal{C}_{[0.02,0.96]}\left(0.12 + 0.34S_d + 0.24S_f + 0.20S_{\mathrm{ch}} + 0.08S_\theta + 0.06q_a - 0.13q_h - 0.08q_z + 0.03\xi_L + 0.55b_3 + 0.30m_3\right) \quad \text{(A19)}$$

Robustness was evaluated after field coverage:

$$Q_{s,c} = \mathcal{C}_{[0.05,0.98]}\left(0.25 + 0.34F_{s,c} + 0.23S_{\mathrm{ch}} + 0.16S_N - 0.26q_h + 0.02\xi_Q + 0.50b_4 + 0.25m_4\right) \quad \text{(A20)}$$

Recoverability was

$$R_{s,c} = \mathcal{C}_{[0.005,0.99]}\left(D_{s,c}^{0.52} L_{s,c}^{0.33} Q_{s,c}^{0.15}\right) \quad \text{(A21)}$$

The composite utility was

$$U_{s,c} = 0.34L_{s,c} + 0.24R_{s,c} + 0.18D_{s,c} + 0.14F_{s,c} + 0.10Q_{s,c} \quad \text{(A22)}$$

and scenario difficulty was

$$G_s = 0.35 + 0.30(1 - q_a) + 0.24q_h + 0.20q_z + 0.10(1 - q_C) \quad \text{(A23)}$$

The analytical localisation-error endpoint was

$$E_{s,c}^{\mathrm{loc}} = \mathcal{C}_{[2.4,16.0]}\left(15.4 - 10.6U_{s,c} + 3.0G_s + 0.35\xi_E + 8.0b_1 + 2.2m_1 + 0.03 \mid \delta_d \mid\right) \quad \text{(A24)}$$

where $E_{s,c}^{\mathrm{loc}}$ is expressed in millimetres, and δ_d is the antenna-position perturbation setting. The primary analysis used $\delta_d = 0$ mm.

The analytical localisation-error label was generated directly from Equation (A24). It was not obtained from an electromagnetic-field solution or from the maximum of a localisation image. Any illustrative analytical-domain map was produced separately and did not determine the endpoint or candidate ranking.

Appendix A.3. Frequency-Dependent Analytical Descriptor

For each scenario–candidate record, the generator returned

$$\ell_{s,c} = [\ell_{s,c}, (f_1), \ldots \ell_{s,c}(f_{N_\nu})]^{\mathrm{T}} \in \mathbb{R}^{N_\nu} \quad \text{(A25)}$$

The analytical frequency grid was defined generally as

$$f_n = f_{\min}^{(\mathrm{grid})} + (n-1)\Delta f, n = 1, \ldots, N_\nu \quad \text{(A26)}$$

In this study,

$$f_{\min}^{(\mathrm{grid})} = 3\ \mathrm{GHz}, \Delta f = 0.1\ \mathrm{GHz}, N_\nu = 51 \quad \text{(A27)}$$

giving a uniform grid from 3 to 8 GHz.

The profile centre and smooth spectral term were

$$f_p = f_c^* + 0.15\xi_p \quad \text{(A28)}$$

$$P_{s,c}(f_n) = \exp\left[-\left(\frac{f_n - f_p}{1.7}\right)^2\right] \quad \text{(A29)}$$

Broadband roughness was

$$r_{s,c}(f_n) = 0.15\sin\left[\frac{2\pi(f_n - 3)}{2.5 + 0.4u_1}\right] + 0.08\xi_n \tag{A30}$$

and the generator-level ripple was

$$\rho_{s,c}(f_n) = 3a_{\text{num}}\sin\left[\frac{2\pi(f_n - 3)}{0.55 + 0.2u_2} + u_3\right] \tag{A31}$$

where $u_1, u_2, u_3 \sim \mathcal{U}(0,1), and\ \xi_n \sim \mathcal{N}(0,1)$.

The final analytical descriptor was

$$\ell_{s,c}(f_n) = -51 + 7D_{s,c} + 1.2P_{s,c}(f_n) + r_{s,c}(f_n) + \rho_{s,c}(f_n) + 2.5\sigma_n\widetilde{\xi}_n \tag{A32}$$

The constants −51, 7.0 and 1.2 were benchmark-shaping values. Although the archived variable name contained the term "log", this quantity was not calculated from a field, voltage, power or scattering-parameter ratio and was not assigned a calibrated physical unit. It was used only as a reproducible frequency-dependent analytical descriptor.

Appendix A.4. Reproducibility, Calibration Status and Scope

The primary settings for measurement noise, antenna-position perturbation, material-property perturbation and coupling perturbation were 0, 0 mm, 0 and 0, respectively. The lesion-absent, lesion-diameter-override and contrast-override options were also inactive in the primary dataset. The analytical dataset was generated using the MT19937 algorithm with fixed study seed $S_0 = 20260701$.

For scenario identifier s and archived record index r, the record-specific seed was

$$S_{s,r} = \text{mod}\left(7919 + 104729s + 37r + S_0,\ 2^{31} - 1\right) \tag{A33}$$

All random quantities associated with each record were generated sequentially from this stream. The downstream model-development and validation seed is reported separately in Section 2.4 because it controls a different workflow stage.

All normalisation constants, compatibility widths, channel weights, score coefficients, exponents, clipping limits and spectral-shaping values were fixed before surrogate-model development. No parameter was estimated from patient data, experimental measurements, antenna measurements, full-wave simulations or held-out FDTD outcomes.

The generator did not calculate complex propagation phase, frequency-dependent constitutive dispersion, finite-antenna loading, impedance matching, mutual coupling, multipath scattering, physical boundary reflections or out-of-plane propagation. Its outputs quantified candidate-ranking behaviour within the specified analytical generator and candidate space; they did not quantify electromagnetic-system performance. The separately implemented restricted two-dimensional TM_z FDTD model provided the held-out electromagnetic challenge described in Section 2.5.

Appendix B. Numerical Specification of the Restricted Two-Dimensional TM_z FDTD Model

Appendix B.1. Geometry, Material Assignment and Acquisition

The held-out electromagnetic challenge used a two-dimensional TM_z Yee finite-difference time-domain model. Each biological scenario was represented as a circular breast cross-section with radius $R_b = 60$ mm and a 2 mm skin layer. The region exterior to the breast was represented as free space. A fixed scenario-specific spatial-organisation

field assigned each interior cell to either an adipose-like or fibroglandular-like tissue class. The spatial-organisation field was thresholded to preserve the prescribed fibroglandular fraction. Each cell retained one tissue-class assignment; no Maxwell–Garnett, Bruggeman or other effective-medium mixing equation was applied.

The lesion-present state was created by replacing the background material assignment within the prescribed circular lesion region with lesion-like material. The corresponding lesion-absent state retained the original background assignment. Geometry, background organisation, source location, excitation, receiver sampling and numerical settings were otherwise identical between the two states. Sixteen ideal in-plane source and receiver positions were distributed uniformly around the breast cross-section. For position $i \in \{1, \ldots, 16\}$,

$$\phi_i = \frac{2\pi(i-1)}{16}, r_i = (R_b + d)\begin{bmatrix}\cos\phi_i \\ \sin\phi_i\end{bmatrix} \tag{A34}$$

where $d \in \{15, 20\}$ mm was the stand-off used in the restricted FDTD evaluation. Ideal line-current source injection and receiver sampling were implemented using bilinear spatial interpolation. Self-channels were excluded.

The restricted library contained 40 candidates per biological scenario. At each stand-off, five frequency designs—3–5, 3–8, 4–6, 5–7 and 6–8 GHz—were crossed with four in-plane channel families: adjacent, skip-one, opposite and full in-plane multistatic. Each scenario–stand-off unit therefore contained 20 candidates.

For every scenario, stand-off and transmitter, matched lesion-present and lesion-absent field solutions were generated. Their complex frequency-domain responses were differenced before candidate-specific frequency and channel selection:

$$\Delta H_{mn}(f) = H_{mn}^{\text{les}}(f) - H_{mn}^{\text{ref}}(f) \tag{A35}$$

where $H_{mn}^{\text{les}}(f)$ and $H_{mn}^{\text{ref}}(f)$ denote the lesion-present and lesion-absent responses, respectively, for transmitter m, receiver n and frequency f.

Appendix B.2. Tissue Properties and Single-Pole Debye Fitting

The target complex relative permittivity was constructed from the tabulated relative permittivity $\varepsilon_r(\omega)$ and conductivity $\sigma(\omega)$:

$$\varepsilon_{\text{target}}^*(\omega) = \varepsilon_r(\omega) - \frac{j\sigma(\omega)}{\omega\varepsilon_0} \tag{A36}$$

A single-pole Debye representation was fitted:

$$\hat{\varepsilon}^*(\omega) = \varepsilon_\infty + \frac{\Delta\varepsilon}{1 + j\omega\tau} - \frac{j\sigma_s}{\omega\varepsilon_0} \tag{A37}$$

Skin values were based on the Gabriel tissue data [24,25]. Adipose-like and fibroglandular-like values were derived from the high-adipose-content and low-adipose-content normal-breast groups, respectively, reported by Lazebnik et al. [26]. Lesion-like values were derived from the malignant-tissue data reported by Lazebnik et al. [27].

Table A1. Reference tissue properties used for single-pole Debye fitting.

Tissue Class	Frequency (GHz)	Relative Permittivity	Conductivity (S/m)
Skin-like	3, 4, 5, 6, 7, 8	37.450986, 36.588413, 35.775241, 34.948088, 34.086621, 33.187385	1.740239, 2.339811, 3.059981, 3.890049, 4.816045, 5.822370
Adipose-like	3, 4, 5, 6, 7, 8	4.672494, 4.582929, 4.488374, 4.393259, 4.300713, 4.212750	0.109934, 0.156565, 0.208209, 0.262076, 0.316059, 0.368682
Fibroglandular-like	3, 4, 5, 6, 7, 8	46.881529, 45.524729, 44.000803, 42.368030, 40.677123, 38.970145	2.099538, 3.048790, 4.161792, 5.392169, 6.697995, 8.043231
Lesion-like	3, 4, 5, 6, 7, 8	53.899607, 52.280565, 50.468669, 48.531367, 46.527105, 44.504278	2.451693, 3.583558, 4.910415, 6.377879, 7.936871, 9.545170

Positivity of all fitted parameters was enforced through logarithmic parameterisation. Optimisation used the Nelder–Mead algorithm implemented by MATLAB fminsearch, with a maximum of 3000 iterations, a maximum of 8000 function evaluations, a parameter tolerance of 10^{-11} and an objective-function tolerance of 10^{-13}.

The fitting objective was

$$\mathrm{NRMSE} = \left[\frac{1}{N_m} \sum_{n=1}^{N_m} \left| \frac{\hat{\varepsilon}^*(\omega_n) - \varepsilon^*_{\mathrm{target}}(\omega_n)}{\max\left(|, \varepsilon^*_{\mathrm{target}}(\omega_n) \mid 1 \right)} \right|^2 \right]^{1/2} \tag{A38}$$

where $N_m = 6$ reference frequencies were used.

Table A2. Reference single-pole Debye fits over 3–8 GHz.

Tissue Class	ε_∞	$\Delta\varepsilon$	τ (ps)	σ_s (S/m)	Fit NRMSE
Skin-like	9.174191	28.915837	9.105906	0.935874	0.003038
Adipose-like	3.282542	1.507330	15.767665	0.041399	0.000838
Fibroglandular-like	11.414691	37.135940	11.790853	0.791765	0.001324
Lesion-like	11.466832	44.401661	11.732783	0.895538	0.001480

Scenario-specific dielectric and conductivity descriptors were used to construct the corresponding scenario-specific tissue curves before fitting. A separate parameter set was obtained for each of the four tissue classes in each of the 120 scenarios, yielding 480 scenario–material fits. The maximum NRMSE across these fits was 0.03985, below the prespecified limit of 0.05. The fitted Debye law was assigned according to the cellwise tissue label.

Appendix B.3. Grid, Time Stepping, Excitation and Boundary Treatment

The operational solver used a uniform spatial grid with

$$\Delta x = \Delta y = 0.200 \text{ mm} \tag{A39}$$

The finer numerical comparator used

$$\Delta x_{\mathrm{fine}} = \Delta y_{\mathrm{fine}} = 0.180 \text{ mm} \tag{A40}$$

The time step was

$$\Delta t = 0.95 \frac{\Delta x}{c_0 \sqrt{2}} \tag{A41}$$

where c_0 is the speed of light in vacuum. For the 0.200 mm operational grid, this produced a time step of approximately 0.448 ps and a two-dimensional Courant number of approximately 0.6718. The total simulated duration was 12 ns, and production fields were stored in single precision.

Excitation used a Gaussian-modulated sinusoid,

$$s(t) = \exp\left[-\frac{1}{2}\left(\frac{t - t_0}{\sigma_t}\right)^2\right]\cos[2\pi f_0(t - t_0)] \tag{A42}$$

where

$$\sigma_t = \frac{\sqrt{2[-\ln(a_e)]}}{\pi B_p}, t_0 = 6\sigma_t \tag{A43}$$

In this study, $f_0 = 5.5$ GHz, $B_p = 5$ GHz and $a_e = 0.10$. The waveform was normalised before application, and the nominal line-current amplitude was 1 μA. The source spectrum was required to retain a relative amplitude of at least 0.03 at every requested frequency. Complex responses were evaluated at the requested frequencies within 3–8 GHz using the same frequency-domain transformation for all simulations.

An 18 mm free-space margin was retained outside the sensing circle before the absorbing boundary. The domain was terminated using an unsplit convolutional perfectly matched layer. The CPML used polynomial order 3, maximum $\kappa = 8$, an α-scaling fraction of 0.05 and a target-reflection parameter of 10^{-8}. The operational implementation used 18 CPML cells, corresponding to a nominal absorbing-layer thickness of 3.6 mm for the 0.200 mm grid.

Appendix B.4. Reconstruction, Localisation and Numerical Controls

Candidate-specific reconstructions used the same imaging domain and processing sequence. Complex differential responses were restricted to the candidate's declared frequency band and channel family. Images were evaluated on a Cartesian grid with a spacing of 2 mm within an imaging radius of 58 mm. The discrete image maximum was retained as a secondary diagnostic. The primary reconstructed position was obtained by fitting a concave log-quadratic surface to the 3×3 neighbourhood surrounding the discrete maximum. The subpixel estimate was accepted only when the fit was finite, concave and well conditioned; the estimated position also had to remain within 1.25 image pixels of the discrete maximum and within the 58 mm imaging radius. Otherwise, the discrete maximum was retained.

The stabilising quantity ϵ appearing in the peak-margin and spectrum-change definitions was a machine-precision safeguard used during double-precision post-processing. It was not an empirically selected scientific threshold. Before held-out evaluation, the implementation was required to satisfy source-spectrum support, free-space phase-velocity, homogeneous-dielectric phase-velocity, reciprocity, late-time-decay and Debye constitutive-reconstruction controls. CPU–GPU agreement used a relative tolerance of 5×10^{-4}. Numerical-comparison tolerances of 10^{-8} mm for distances and 10^{-12} for dimensionless quantities were applied only to prevent binary floating-point boundary effects and did not alter the prespecified scientific criteria.

The numerical-reference criteria, ambiguity definition, candidate-qualification rule, response-spectrum comparison and shortlist-level endpoints are defined in Section 2.6. The complete results are reported in Table A3.

Appendix C. Numerical-Reference Qualification, Candidate-Support Audit and Held-Out Transfer Results

Table A3. Numerical-reference qualification and analytical–FDTD candidate-support audit. (**A**) Operational-grid qualification; (**B**) analytical–FDTD candidate-support audit.

(**A**)			
Assessment	**Prespecified Criterion**	**Result**	**Interpretation**
Assessment design	Three independent scenarios evaluated at both stand-offs	Scenarios 85, 91 and 10; 15 and 20 mm; six units	Complete
Grid comparison	0.200 mm operational grid evaluated against a finer numerical comparator	Finer grid: 0.180 mm	Finer grid not treated as exact truth
Reference shortlist	Fixed before numerical-reference evaluation	$K_{ref} = 3$	Applied to all six units
Response-spectrum change	$\leq$5% in every unit	Maximum: 1.3035%	Criterion met
Best-in-qualified-shortlist regret	$\leq$2 mm in every unit	Maximum: 0 mm	Criterion met
Qualified top-three overlap	$\geq$2/3 in every unit	Minimum: 0.6667	Criterion met
Candidate availability	At least three qualified candidates on each grid	Minimum: 12 on each grid	Criterion met
Joint shortlist qualification	At least one operational-grid shortlisted candidate also qualified on the finer grid	Minimum: 3	Criterion met
Tissue-map consistency	Fibroglandular-fraction change $\leq$0.5 percentage points	Criterion met in all six units	Criterion met
Numerical and accounting controls	Candidate accounting, CPU–GPU agreement, canonical controls and CPML checks required to pass	All controls passed	Criterion met
Overall assessment	Every primary criterion required in all six units	Six of six units met all criteria	0.200 mm grid used for held-out shortlist evaluation
Secondary numerical sensitivity	Retained but not used as a primary criterion	Raw top-one comparator failed; maximum selected-maximum displacement: 2.3384 mm	Exact top-one preservation was not supported
(**B**)			
Assessment	**Prespecified Requirement**	**Result**	**Interpretation**
Audited support	All restricted FDTD candidates assessed within their corresponding biological scenarios	12 scenarios; 40 FDTD candidates per scenario; 480 rows	Complete
Analytical comparison library	Matching against the 80 analytical candidates from the same scenario	960 analytical scenario–candidate rows	Complete
Exact-matching rule	Agreement across all declared design descriptors using a scaled tolerance of 10^{-9}	Matching performed separately within each scenario	Applied as prespecified
Complete support alignment	Every restricted FDTD row required one unique analytical counterpart	Required: 480 of 480 rows	Not achieved

Table A3. *Cont.*

(**B**)			
Assessment	**Prespecified Requirement**	**Result**	**Interpretation**
Unique exact matches	Rows with exactly one analytical counterpart	12 of 480 rows (2.5%)	Insufficient for a complete direct matched analytical-to-FDTD comparison
Rows without a unique match	No exact match or more than one exact match	468 of 480 rows (97.5%)	Not used for direct matched comparison
Approximate substitution	Nearest-neighbour or approximate matching not permitted	No substitution performed	Candidate differences were not represented as exact matches
External comparison	Insufficient for a complete matched analytical-to-FDTD comparison	Complete alignment was unavailable	Frozen predictions were compared with FDTD outcomes within the restricted candidate pool
Attribution boundary	Physical-model and candidate-support effects separable only under exact alignment	Effects could not be separated	Transfer gap cannot be attributed solely to solver fidelity

Note for (A): The numerical-reference scenarios were excluded from both surrogate development and held-out transfer evaluation. The 0.180 mm grid was evaluated as a finer numerical comparator and was not treated as exact or fully grid-converged electromagnetic truth. CPU–GPU agreement used a relative tolerance of 5×10^{-4}. The CPML implementation passed its prespecified checks and used 18 cells at the operational grid; Note for (B): Matching was based on stand-off, angular offset, normalised ring offset, frequency count, lower and upper frequencies, centre frequency, bandwidth and channel-family indicators. Candidate identifiers were not assumed to represent identical designs across libraries. "Same-support" refers to the comparison of surrogate predictions and FDTD outcomes over the same restricted 20-candidate pool within each scenario–stand-off unit; it does not indicate that the restricted FDTD library was an exact subset of the original analytical library.

Table A4. Complete held-out surrogate-to-FDTD transfer results for the 24 scenario–stand-off units at $K = 5$. (**A**) Candidate-selection and regret outcomes. (**B**) Ranking agreement and ambiguity diagnostics.

(**A**)								
Scenario	**Stand-Off (mm)**	**Qualified FDTD Oracle Candidate ID**	**Oracle Error (mm)**	**Best Qualified Candidate in (K = 5)**	**Shortlist Error (mm)**	**Penalised Regret (mm)**	**Oracle Captured**	**Regret $\leq$ 2 mm**
5	15	8	20.4765	19	25.9350	5.4585	No	No
5	20	26	26.9568	39	28.5243	1.5675	No	Yes
20	15	16	11.0485	16	11.0485	0.0000	Yes	Yes
20	20	37	25.2363	36	25.3421	0.1057	No	Yes
40	15	10	11.8082	19	13.7232	1.9150	No	Yes
40	20	23	11.2186	30	11.2186	0.0000	No	Yes
42	15	7	6.3339	10	9.1545	2.8206	No	No
42	20	40	1.5417	40	1.5417	0.0000	Yes	Yes
51	15	9	7.8864	20	9.9879	2.1015	No	No
51	20	27	9.3009	39	17.4189	8.1180	No	No
67	15	7	3.6438	16	8.7313	5.0875	No	No
67	20	26	9.9733	30	12.6222	2.6490	No	No
82	15	17	7.4885	16	7.5053	0.0168	No	Yes
82	20	37	6.8681	36	6.9032	0.0351	No	Yes
89	15	4	18.4484	20	26.8933	8.4449	No	No

Table A4. *Cont.*

(A)								
Scenario	**Stand-Off (mm)**	**Qualified FDTD Oracle Candidate ID**	**Oracle Error (mm)**	**Best Qualified Candidate in (K = 5)**	**Shortlist Error (mm)**	**Penalised Regret (mm)**	**Oracle Captured**	**Regret ≤ 2 mm**
89	20	37	8.6731	36	9.0578	0.3847	No	Yes
104	15	6	18.0857	19	19.4038	1.3182	No	Yes
104	20	21	21.8629	39	23.3678	1.5050	No	Yes
111	15	1	15.8234	20	19.3787	3.5553	No	No
111	20	30	5.6418	30	5.6418	0.0000	Yes	Yes
115	15	16	2.3508	16	2.3508	0.0000	Yes	Yes
115	20	37	0.8176	36	1.7228	0.9052	No	Yes
118	15	8	4.1337	20	31.9709	27.8372	No	No
118	20	28	9.7698	36	13.0491	3.2794	No	No

(B)						
Scenario	**Stand-Off (mm)**	**Qualified Top-Three Overlap**	**Qualified Top-Five Overlap**	**Qualified Spearman Association**	**Ambiguous Candidates in (K = 5)**	**Qualified Candidates in (K = 5)**
5	15	0.0000	0.2000	−0.5663	1	4
5	20	0.0000	0.4000	−0.0505	1	4
20	15	0.3333	0.2000	−0.1048	2	3
20	20	0.3333	0.2000	−0.2615	1	4
40	15	0.3333	0.6000	0.3423	2	3
40	20	0.0000	0.2000	0.0429	1	4
42	15	0.3333	0.6000	0.0536	2	3
42	20	0.3333	0.4000	0.1123	1	4
51	15	0.6667	0.4000	0.1592	1	4
51	20	0.0000	0.0000	−0.4821	1	4
67	15	0.0000	0.4000	−0.1466	2	3
67	20	0.3333	0.4000	0.0933	1	4
82	15	0.3333	0.4000	−0.3890	2	3
82	20	0.3333	0.2000	−0.4857	1	4
89	15	0.0000	0.2000	−0.3481	2	3
89	20	0.3333	0.4000	−0.0857	1	4
104	15	0.3333	0.4000	−0.1341	3	2
104	20	0.0000	0.2000	−0.2397	2	3
111	15	0.0000	0.2000	−0.2431	2	3
111	20	0.0000	0.6000	−0.0330	1	4
115	15	0.6667	0.4000	0.1393	2	3
115	20	0.3333	0.4000	−0.1036	1	4
118	15	0.0000	0.2000	−0.4643	2	3
118	20	0.3333	0.2000	−0.2703	2	3

Notes: The qualified FDTD oracle was the candidate with the lowest localisation error among the non-ambiguous candidates in each stand-off-specific library of 20 candidates. The best shortlisted candidate was the qualified candidate with the lowest localisation error among the five candidates retained by the prospectively fixed random forest. All 24 shortlists contained at least one qualified candidate; penalised regret therefore equalled the observed best-in-shortlist regret in every evaluation unit. Oracle capture refers to candidate identity, whereas regret reflects localisation error. Zero regret without oracle capture was possible when different candidates produced the same minimum error, as occurred for Scenario 40 at a stand-off of 20 mm. Stand-off-specific results are descriptive because the two stand-offs associated with each biological scenario constitute paired observations.

References

1. Wei, Z.; Zhou, Z.; Wang, P.; Ren, J.; Yin, Y.; Pedersen, G.F.; Shen, M. Fully Automated Design Method Based on Reinforcement Learning and Surrogate Modeling for Antenna Array Decoupling. *IEEE Trans. Antennas Propag.* **2023**, *71*, 660–671. [CrossRef]
2. Zou, H.; Zeng, S.; Li, C.; Ji, J. A Survey of Machine Learning and Evolutionary Computation for Antenna Modeling and Optimization: Methods and Challenges. *Eng. Appl. Artif. Intell.* **2024**, *138*, 109381. [CrossRef]
3. Wu, Q.; Chen, W.; Yu, C.; Wang, H.; Hong, W. Machine-Learning-Assisted Optimization for Antenna Geometry Design. *IEEE Trans. Antennas Propag.* **2024**, *72*, 2083–2095. [CrossRef]
4. Khan, M.R.; Zekios, C.L.; Bhardwaj, S.; Georgakopoulos, S.V. A Deep Learning Convolutional Neural Network for Antenna Near-Field Prediction and Surrogate Modeling. *IEEE Access* **2024**, *12*, 39737–39747. [CrossRef]
5. Qashlan, A.M.; Aldhaheri, R.W.; Alharbi, K.H. A Modified Compact Flexible Vivaldi Antenna Array Design for Microwave Breast Cancer Detection. *Appl. Sci.* **2022**, *12*, 4908. [CrossRef]
6. Slimi, M.; Jmai, B.; Dinis, H.; Gharsallah, A.; Mendes, P.M. Metamaterial Vivaldi Antenna Array for Breast Cancer Detection. *Sensors* **2022**, *22*, 3945. [CrossRef] [PubMed]
7. Wang, L. Microwave Imaging and Sensing Techniques for Breast Cancer Detection. *Micromachines* **2023**, *14*, 1462. [CrossRef] [PubMed]
8. Wang, L. Holographic Microwave Image Classification Using a Convolutional Neural Network. *Micromachines* **2022**, *13*, 2049. [CrossRef] [PubMed]
9. Wang, L. Recoverability-Guided Reduced-Target Inversion for Microwave Imaging: A Synthetic Breast-Imaging Study. *Biomed. Phys. Eng. Express* **2026**, *12*, 047002. [CrossRef] [PubMed]
10. Asok, A.O.; Tripathi, A.; Dey, S. Breast Tumors Detection Using Multistatic Microwave Imaging with Antipodal Vivaldi Antennas Utilizing DMAS and it-DMAS Techniques. *Int. J. Microw. Wirel. Technol.* **2024**, *16*, 690–703. [CrossRef]
11. Yıldız, Ş.; Kurt, M.B. Breast Cancer Detection Using a High-Performance Ultra-Wideband Vivaldi Antenna in a Radar-Based Microwave Breast Cancer Imaging Technique. *Appl. Sci.* **2025**, *15*, 6015. [CrossRef]
12. Liu, B.; Akinsolu, M.O.; Song, C.; Hua, Q.; Excell, P.; Xu, Q.; Huang, Y.; Imran, M.A. An Efficient Method for Complex Antenna Design Based on a Self Adaptive Surrogate Model-Assisted Optimization Technique. *IEEE Trans. Antennas Propag.* **2021**, *69*, 2302–2315. [CrossRef]
13. Liu, Y.; Liu, B.; Ur-Rehman, M.; Imran, M.A.; Akinsolu, M.O.; Excell, P.; Hua, Q. An Efficient Method for Antenna Design Based on a Self-Adaptive Bayesian Neural Network-Assisted Global Optimization Technique. *IEEE Trans. Antennas Propag.* **2022**, *70*, 11375–11388. [CrossRef]
14. Li, Z.; Wu, B.; Wang, R.; Li, H.; Gong, M. Surrogate-Assisted Evolutionary Multi-Objective Antenna Design. *Electronics* **2025**, *14*, 3862. [CrossRef]
15. Sahu, K.C.; Koziel, S.; Pietrenko-Dabrowska, A. Surrogate Modeling of Passive Microwave Circuits Using Recurrent Neural Networks and Domain Confinement. *Sci. Rep.* **2025**, *15*, 13322. [CrossRef] [PubMed]
16. Noakoasteen, O.; Christodoulou, C.; Peng, Z.; Goudos, S.K. Physics-Informed Surrogates for Electromagnetic Dynamics Using Transformers and Graph Neural Networks. *IET Microw. Antennas Propag.* **2024**, *18*, 505–515. [CrossRef]
17. Leon, C.; Scheinker, A. Physics-Constrained Machine Learning for Electrodynamics without Gauge Ambiguity Based on Fourier Transformed Maxwell's Equations. *Sci. Rep.* **2024**, *14*, 14809. [CrossRef] [PubMed]
18. Karahan, E.A.; Liu, Z.; Gupta, A.; Shao, Z.; Zhou, J.; Khankhoje, U.; Sengupta, K. Deep-Learning Enabled Generalized Inverse Design of Multi-Port Radio-Frequency and Sub-Terahertz Passives and Integrated Circuits. *Nat. Commun.* **2024**, *15*, 10734. [CrossRef] [PubMed]
19. Schmeing, L.; Pioch, F. Advancements in Physics-Informed Neural Networks for Solving Maxwell's Equations: A Systematic Literature Review. *Electronics* **2026**, *15*, 2424. [CrossRef]
20. Yang, G.; Xiao, Q.; Zhang, Z.; Yu, Z.; Wang, X.; Lu, Q. Exploring AI in Metasurface Structures with Forward and Inverse Design. *iScience* **2025**, *28*, 111995. [CrossRef] [PubMed]
21. Hossain, K.; Sabapathy, T.; Jusoh, M.; Lee, S.-H.; Ab Rahman, K.S.; Kamarudin, M.R. Negative Index Metamaterial-Based Frequency-Reconfigurable Textile CPW Antenna for Microwave Imaging of Breast Cancer. *Sensors* **2022**, *22*, 1626. [CrossRef] [PubMed]
22. Syed, A.; Sheikh, M.; Islam, M.T.; Rmili, H. Metamaterial-Loaded 16-Printed Log Periodic Antenna Array for Microwave Imaging of Breast Tumor Detection. *Int. J. Antennas Propag.* **2022**, *2022*, 4086499. [CrossRef]
23. Syed, A.; Sobahi, N.; Sheikh, M.; Mittra, R.; Rmili, H. Modified 16-Quasi Log Periodic Antenna Array for Microwave Imaging of Breast Cancer Detection. *Appl. Sci.* **2022**, *12*, 147. [CrossRef]
24. Gabriel, S.; Lau, R.W.; Gabriel, C. The Dielectric Properties of Biological Tissues: II. Measurements in the Frequency Range 10 Hz to 20 GHz. *Phys. Med. Biol.* **1996**, *41*, 2251–2269. [CrossRef] [PubMed]
25. Gabriel, S.; Lau, R.W.; Gabriel, C. The Dielectric Properties of Biological Tissues: III. Parametric Models for the Dielectric Spectrum of Tissues. *Phys. Med. Biol.* **1996**, *41*, 2271–2293. [CrossRef] [PubMed]

26. Lazebnik, M.; McCartney, L.; Popovic, D.; Watkins, C.B.; Lindstrom, M.J.; Harter, J.; Sewall, S.; Magliocco, A.; Booske, J.H.; Okoniewski, M.; et al. A Large-Scale Study of the Ultrawideband Microwave Dielectric Properties of Normal Breast Tissue Obtained from Reduction Surgeries. *Phys. Med. Biol.* **2007**, *52*, 2637–2656. [CrossRef] [PubMed]
27. Lazebnik, M.; Popovic, D.; McCartney, L.; Watkins, C.B.; Lindstrom, M.J.; Harter, J.; Sewall, S.; Ogilvie, T.; Magliocco, A.; Breslin, T.M.; et al. A Large-Scale Study of the Ultrawideband Microwave Dielectric Properties of Normal, Benign and Malignant Breast Tissues Obtained from Cancer Surgeries. *Phys. Med. Biol.* **2007**, *52*, 6093–6115. [CrossRef] [PubMed]