

Feature Contribution Analysis Across Multimodal Breast Cancer Diagnostic Models

Muhaisin Tiyyumba Nantogmah^{1*}, Abdul-Barik Alhassan^{2*}

* Department of Computer Science, Faculty of Physical Sciences, University for Development Studies, Tamale, Ghana
muhaisintiyumbana1848@uds.edu.gh¹, abarik@uds.edu.gh²

Article Info

Article history:

Received 2026-08-09

Revised 2026-09-15

Accepted 2026-10-03

Keyword:

Breast cancer,
Clinical decision support,
Explainable AI,
Multimodal learning,
SHAP.

ABSTRACT

This study systematically assesses and contrasts the interpretability and clinical usefulness of SHAP-based explanations in multimodal breast cancer diagnosis models. A cross-task meta-analysis aggregated findings from five state-of-the-art multimodal breast cancer diagnostic studies addressing benign-malignant differentiation, HER2 expression status, axillary lymph node metastasis, Ki-67 proliferation assessment and molecular subtype classification. Outputs across ultrasound, mammography, MRI, PET/CT and clinical data were harmonized to common feature families and compared using global feature-importance rankings and case-level analysis, with all quantitative values reported as published in the source studies. Peritumoral texture entropy and skewness were the SHAP-attributed leading variables for benign-malignant differentiation, and PET/CT metabolic parameters and tumor diameter for HER2 status (AUC 0.888 training, 0.844 internal test). MRI-derived shape and texture features, flatness and joint entropy, were most influential for lymph node metastasis, and surface-to-volume ratio and small-area emphasis for Ki-67 status, while two mammographic descriptors, indistinct margins and calcification, led molecular subtyping. Case-level SHAP explanations illustrated one reported false-negative metastasis case attributed to model over-reliance on tumor shape. In one of the five studies, radiologists assisted by a SHAP-explained model gained 8.3–15.9 percentage points in subtype-classification accuracy over unassisted readers, with the largest gain among less experienced readers. This study demonstrates that SHAP-based interpretability is a feasible tool for human-in-the-loop transparency and auditability in multimodal breast oncology decision-making, helping clinicians examine model reasoning and detect failure modes, while underscoring that SHAP importance reflects model behavior rather than established clinical or biological causality.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. INTRODUCTION

Breast cancer is the most common cancer in women globally and early and accurate diagnosis is a key aspect which improves patient outcomes. Artificial intelligence (AI) has transformed the diagnosis of breast cancer in recent years, with deep learning and radiomics technologies showing excellent performance in mammography, ultrasound, magnetic resonance imaging (MRI) and digital pathology [1], [2], [3]. This is significant and a paradigm change is taking place, shifting from unimodal pipelines, focusing on a single

data source to multimodal pipelines that combine complementary sources like multiple imaging modalities (mammography, MRI, ultrasound, PET/CT), histopathology, and structured clinical variables [4], [5]. This integration always results in better diagnostic accuracy, stability and applicability than with individual modalities [5].

However, even with these improvements, the complex multimodal models are yet to be validated for clinical use due to their “black-box” character. The lack of transparency also exacerbates clinician mistrust, hides the scientific premise of a diagnosis, and makes it more difficult to get regulatory

approval [6]. To overcome this challenge, Explainable Artificial Intelligence (XAI) has arisen and SHapley Additive exPlanations (SHAP) is one of the most popular methods [7]. SHAP is founded on the principles of cooperative game theory and provides a theoretically sound and consistent method for assigning the contribution of each input feature to an individual prediction, providing both local (case level) explanation and global (model level) insight.

In the recent years, there is ample evidence that SHAP has emerged as the most popular model-agnostic XAI method for breast cancer studies, generally for explaining tree-based ensemble models like XGBoost, LGBM and decision trees [8], [9]. Systematic testing of explainability across modalities, beyond case-by-case reporting, is also called for by complementary reviews of multimodal breast cancer diagnosis, which also point to models that are moving towards integrating multiple types of data and towards justifying their decisions [10], [11], [12]. To date, there is a growing number of task-specific works that have used SHAP to interpret the predictions of the model in the context of the problem: SHAP was used to identify the optimal peritumoral expansion width for automated breast volume scanning (ABVS) in the task of benign-malignant differentiation using an LGBM model [13]; it was used to explain a dual-center study that used XGBoost predictions to classify HER2 expression status using 18F-FDG PET/CT parameters and clinicopathologic features, and showed that using XGBoost with SHAP could measurably increase radiologists' accuracy in this task [14]; it was employed to interpret a dual-sequence MRI radiomics model to predict axillary lymph node metastasis (ALNM) in invasive breast cancer, and revealed that the surface-to-volume ratio is the most important parameter in this model [15]; it was used to interpret a DCE-MRI radiomics model for predicting Ki-67 proliferation status, showing that the surface-to-volume ratio is the most important parameter in this model [16]; it was used to identify BI-RADS descriptors as the drivers of an interpretable decision-tree model to classify molecular breast cancer subtypes [17]. In a related multi-center study, the authors integrated a deep-learning model with SHAP values to demonstrate that the model for the prediction of lymph node metastasis in mammography images is mainly interested in clinically relevant areas of the tumor and its surroundings, including the activation maps obtained by Grad-CAM [18].

Combined, these studies confirm the clinical utility of combining multimodal models and SHAP, but each of them is only singling out one model, one task, and one architecture. Not even the last few systematic reviews of this space bring some unity in the way in which SHAP works across tasks, modalities and architectures [10], [11]. Existing reviews catalogue individual SHAP applications without harmonizing the underlying feature-importance evidence into a common comparative framework, leaving two specific gaps that motivate the present study: first, whether feature-importance patterns identified by SHAP are genuinely comparable across differently trained, differently validated models that address distinct clinical endpoints, or whether apparent overlaps

largely reflect task-specific coincidence; and second, whether SHAP explanations translate into measurable clinical benefit within a human-in-the-loop workflow, beyond their established use as a retrospective, case-by-case interpretive aid. With these gaps in mind, this study pursues three goals: (1) to harmonize and quantify the feature importance reported by five archetypal multimodal breast cancer studies and identify features that recur across tasks, while explicitly flagging where such recurrence should not be read as evidence of a single unified diagnostic mechanism; (2) to illustrate, through case-level SHAP explanations, how such explanations can support personalized decision support and expose model failure modes; and (3) to synthesize the one available report of SHAP-assisted human-in-the-loop performance and assess its implications, and its limitations, for clinical decision support. This work aggregates evidence from five recent studies, moving SHAP from a one-model position as an interpreter to a comparative, trans-modal position for multimodal oncological inference; because the underlying studies differ in cohorts, validation protocols, and clinical endpoints (Section III.E), the resulting synthesis is best read as hypothesis-generating rather than as a pooled, statistically validated meta-analysis.

II. METHOD

In this study, the authors utilize a structured meta-analysis framework to systematically synthesize and compare explanations of the five archetypal multimodal breast cancer studies [13]-[17] that use SHAP. The methodology includes four phases as shown in Fig. 1.

A. Study Selection and Meta-Feature Curation

The selected studies [13]-[17] were representative by both methodological criteria and by using multimodal data, with imaging and clinical variables combined, as well as current machine learning algorithms such as extreme gradient boosting (XGBoost), light gradient boosting machine (LGBM) and decision trees, and by explicit use of SHAP for model interpretability. A standardized diagnostic task (e.g., benign versus malignant classification, HER2 versus non-HER2 classification), data modalities (ABVS radiomics, PET/CT metabolic parameters, DCE-MRI radiomics, and structured clinical variables), and model architecture were then standardized for each study in a structured meta-feature curation process. To facilitate a systematic analysis of model behavior, feature relevance and patterns of interpretability from a global-level feature-importance ranking and an instance-level contribution value, outputs from SHAP were systematically compiled across studies.

B. Harmonized SHAP-Based Feature Contribution Analysis

Two analyses were performed: an initial survey and an in-depth evaluation. The top features in each study were normalized for global interpretability and compared, and grouped into feature families (e.g., peritumoral radiomics,

intratumoral metabolism, pathological markers, clinical demographics); importance of each family was assigned using weighted average rank across the tasks where they were included. To gain local interpretability, individual SHAP

force and waterfall plots reported in the primary studies were analyzed to create archetypal case studies of correct and incorrect predictions, breaking down the feature contributions that led to each prediction.

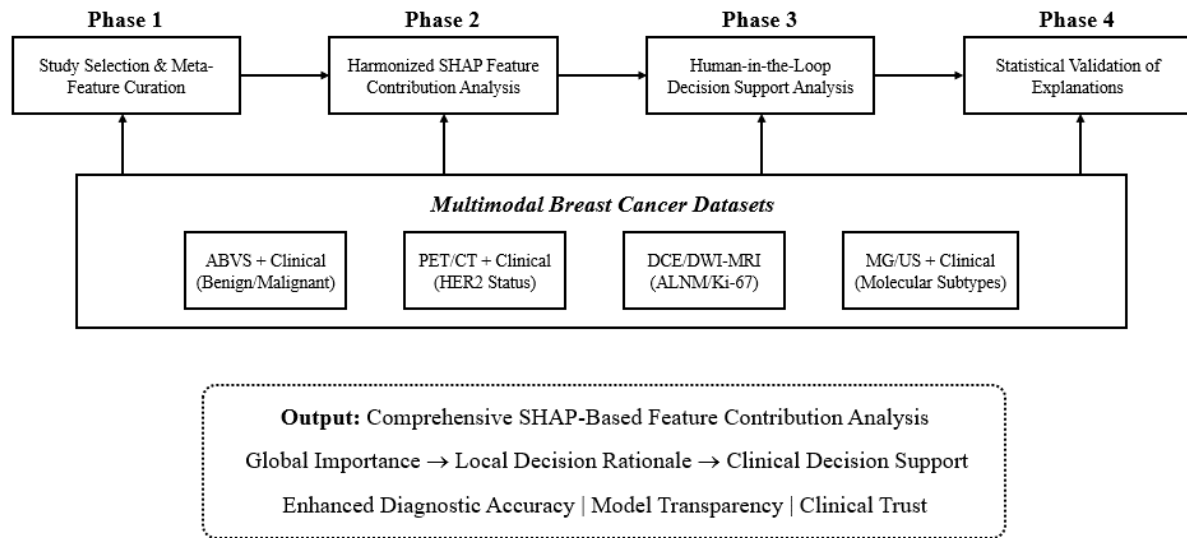


Fig. 1. Four-phase methodology workflow for the SHAP-based meta-analysis of multimodal breast diagnostic models.

C. Human-in-the-Loop Decision Support Analysis

Ma et al. [17] is the only of the five studies that reported an effect of model-assisted diagnosis for a clinical decision support system; this effect was analyzed to assess the utility of SHAP for such systems. This comparison of the diagnostic accuracy, sensitivity, and specificity of radiologists compared to radiologists using a decision-tree model developed with SHAP was for the distinction between triple-negative and non-triple-negative breast cancer and between luminal and non-luminal breast cancer subtypes.

D. Statistical Validation of Explanations

The robustness and stability of the SHAP explanations were evaluated by a qualitative synthesis of the different validation approaches used in each of the studies, including consistency of feature importance across cross validation folds, the standard deviation of the AUC value, and the use of DeLong tests for comparing models (indirectly validating the reliability of the SHAP-based feature rankings when the models have a stable and statistically significant performance across folds).

E. Study Characteristics and Interpretive Scope

This study does not use a new model but rather makes use of models from previous published studies; therefore, the design, cohort, and validation characteristics of the five source studies, as reported in their respective publications [13]–[17], are summarized in Table I. Unless explicitly reported in the source publication all quantitative values reported in Section III (AUC, accuracy and percentage-point

improvements) were not re-calculated, re-validated or re-tested by the present authors and are as published in the source document.

TABLE I
CHARACTERISTICS OF THE FIVE SOURCE STUDIES INCLUDED IN THE META-ANALYSIS

Ref.	Study Design & Center(s)	Sample Size (Train/Test/Ext.)	Validation Approach
[13]	Retrospective, single-center (Third Xiangya Hospital, Changsha, China)	N = 250	Internal train/test split
[14]	Retrospective, dual-center (Lianyungang First People's Hospital; Third Affiliated Hospital of Soochow University)	N = 241 (157, Center 1: 7:3 train/internal-test; 84, Center 2: external)	Internal split + independent external (cross-center) test
[15]	Retrospective, single-center (Affiliated Hospital of North Sichuan Medical College)*	N = 183 (128 train / 55 test, 7:3)	Internal train/test split
[16]	Retrospective, single-center (Affiliated Hospital of North Sichuan Medical College)*	N = 195 patients / 201 lesions (7:3 split)	Internal train/test split
[17]	Retrospective, single-center (Nanfeng Hospital); independent 4-radiologist reader study	N = 600 (450 train / 150 test)	Internal train/test split; separate reader-performance study

*Studies [15] and [16] were conducted at the same institution with overlapping authorship. The published reports do not state whether the two patient cohorts are independent; possible overlap could not be confirmed from the available materials and is carried forward as an open limitation in Section III.E.

There are two interpretive cautions throughout this synthesis. First, SHAP values measure the contribution of each feature to the output of a given model and are a description of model behavior – not verified clinical or biological causality – and a high SHAP rank should be interpreted as model associated evidence, not as a confirmed diagnostic mechanism. Second, these five tasks were synthesized to represent different clinical endpoints with different underlying biology, which were trained on model architectures, datasets, and validation protocols, none of which were the same (Table I). Recurring feature families across tasks (e.g., peritumoral texture or shape descriptors) are reported as an observed pattern in model behavior in the literature, but not as evidence that a single unified diagnostic mechanism is common to all five tasks.

The human-in-the-loop comparison presented in Section III. The accuracy improvements reported in the source study [17] are statistically significant and reported by four independent radiologists when comparing the accuracy of molecular-subtype and TNBC classification with and without access to the SHAP-explained decision-tree model output, split into two groups based on their experience. The published report on this study does not indicate whether the cases were presented in a counterbalanced (crossover) or random order, whether there was a washout period between the reading of the unassisted and assisted cases, and the exact confidence intervals around the reported accuracy gains; therefore, the present synthesis cannot exclude that part of the accuracy gain may have been due to a reader-learning or case-familiarity effect, and this is explicitly discussed as a limit in Section III.E.

III. RESULTS AND DISCUSSION

A. Global Feature Importance Across Diagnostic Tasks

The meta analysis shows that there are general task-specific patterns of feature contribution across the broad spectrum of multimodal diagnostics, thereby demonstrating the power of SHAP to break down the behavior of a model.

The set of global feature-importance rankings from the five studies, presented in Fig. 2 and summarized in Table II, is highly task-dependent, with the most predictive feature family being dominant in each study.

The most important features for ABVS radiomics were peritumoral texture features, a wavelet gray-level dependence-matrix dependence entropy feature and a wavelet first-order skewness feature, along with patient's age [13] highlighting the importance of the tumor microenvironment, which can be obtained from a 10 mm peritumoral expansion.

Tumor minimum diameter, SUVmean and CTmean were the most influential features for predicting HER2-zero versus HER2-low/positive expression from PET/CT data, with an AUC of 0.888 (training) and 0.844 (internal test); for the related task of distinguishing HER2-low from HER2-positive status, along with tumor maximum diameter, metabolic tumor volume (MTV) became the most influential features, with an AUC of 0.920 [14].

In predicting axillary lymph node metastasis (ALNM) from dual-sequence MRI, a shape flatness feature derived from the DCE and a gray-level co-occurrence joint entropy feature derived from the DWI were dominant, suggesting that tumor shape and micro-textural heterogeneity are important for making predictions; unimodal models performed poorly, while the combined model performed significantly better with an AUC of 0.864, and decision curve analysis showed that the combined model had the greatest net clinical benefit [15].

In the non-invasive assessment of Ki-67 proliferation status from DCE-MRI, an early-phase surface-to-volume ratio feature was most important, followed by a gray-level emphasis texture feature based on small areas with low gray-level, indicating that the morphological and textural characteristics of a tumor in the early phase after contrast wash-in are very predictive of the proliferation status [16].

Lastly, an interpretable decision-tree model explained by SHAP classified the molecular subtypes from mammography and ultrasound with an AUC of 0.971 [17] that is close to the known radiological knowledge and directly corresponds to SHAP explanations and clinical domain knowledge.

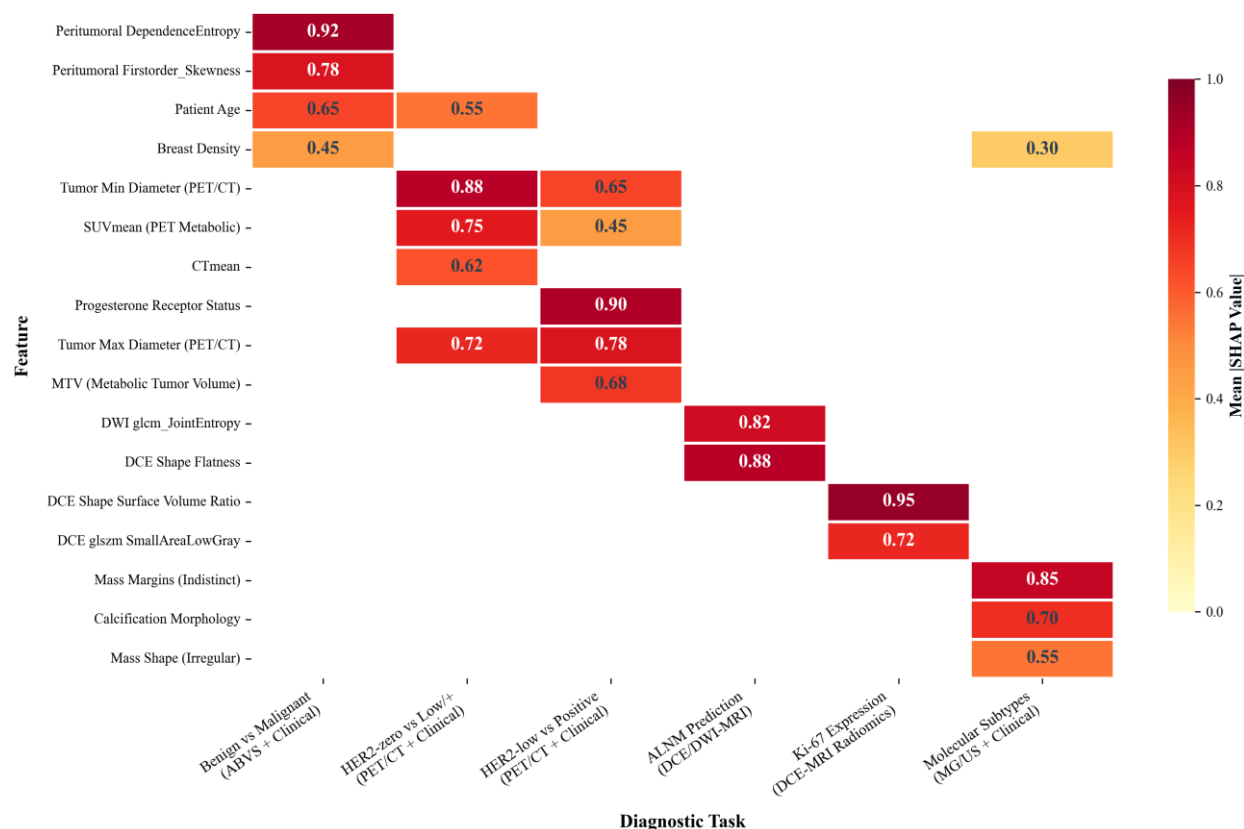


Fig. 2. Global SHAP feature importance across multimodal diagnostic tasks (mean absolute SHAP value; darker cells indicate greater contribution).

TABLE II
GLOBAL IMPORTANCE OF MULTIMODAL FEATURE FAMILIES ACROSS
DIAGNOSTIC TASKS AS REVEALED BY SHAP

Diagnostic Task	Modalities	Top SHAP Features (Family)
Benign vs. Malignant Mass	ABVS Radiomics + Clinical	Peritumoral Texture (Dependence Entropy, Skewness), Age
HER2-Zero vs. HER2-Low/Positive	PET/CT + Clinical	Tumor Min. Diameter, SUVmean (Metabolic), CTmean
Ki-67 (Low vs. High)	DCE-MRI Radiomics	Surface-to-Volume Ratio (Shape), Small-Area Emphasis (Texture)
ALNM Prediction	DCE/DWI-MRI Radiomics	Flatness (Shape), Joint Entropy (Texture)
TNBC vs. Non-TNBC	MG/US Imaging + Clinical	Mass Margins (Indistinct), Calcification

B. Local Interpretability and Individualized Decision Support

SHAP's local interpretability illustrates case-specific explanatory rationales for individual model predictions. The archetypes for forces presented in Fig. 3 are compiled from the studies analyzed. In a true-positive case from the Ki-67 prediction study, a high expression output is primarily caused

by a high surface-to-volume ratio and supported by supporting texture features, which gives a strong push towards the malignant class (Fig. 3A). The prediction toward the benign class is driven by a combination of the lack of calcification, younger age, low SUVmean and circumscribed margins in a true-negative case (Fig. 3B). In a false negative (FN) LN metastasis case, the same model, upon the application of SHAP analysis, shows that the relative contribution of the shape feature 'flatness' to the 'no metastasis' prediction was far more significant than the texture feature contribution, which was pointing towards malignancy (Fig. 3C). This single case report does not necessarily represent a systematic validation of failure mode but outlines one possible failure, namely over-reliance on a shape descriptor, and must therefore be treated as illustrative and not as a general description of the model's failure modes. This case-level analysis should not be interpreted as a general description of model failure because the source study does not report on how many similar cases were reviewed or how other cases are represented by this one. Despite this, the analysis of a misdiagnosis is directly applicable to feedback for a clinician and/or refinement of the model – transparency in the rationale for the decision (as opposed to just the decision) is vital for human-in-the-loop clinical workflow.

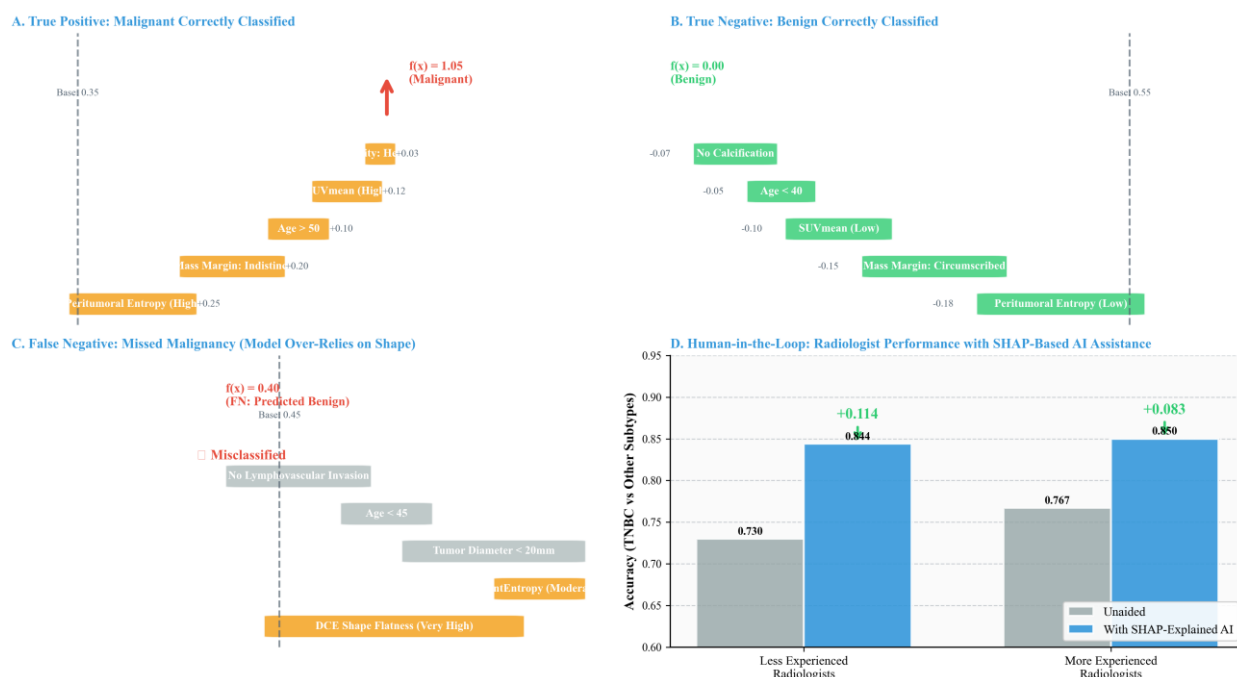


Fig. 3. Local, case-level SHAP interpretability: (A) true-positive malignant case; (B) true-negative benign case; (C) false-negative case showing model over-reliance on tumor shape; (D) human-in-the-loop radiologist performance with and without SHAP-explained AI assistance.

C. Enhancing Diagnostic Performance Through Explainable AI

The impact of an explainable model in a direct clinical application is reflected in the study that Ma et al. conducted [17] (Fig. 3D). The average accuracy of less experienced radiologists increased from 0.730 to 0.844 (+0.114), and the average accuracy of more experienced radiologists increased from 0.767 to 0.850 (+0.083), with the SHAP-interpretable decision tree model for differentiating TNBC from other breast cancer sub-types. Less experienced radiologists' average accuracy was improved from 0.617 to 0.776 (+0.159) for differentiating luminal from non-luminal subtypes. The results highlight the ability of an explainable model to improve the model output, adding the SHAP-based feature rationale as a teaching and decision-support tool that reduces the performance gap between junior and senior clinicians.

D. Principal Findings and Clinical Implications

This meta analysis is a cross-cutting validation of SHAP as an essential technology for multi-modal and interpretable AI systems in breast cancer diagnosis. The principal finding is that the explanations generated by SHAP are not uniform in nature; they are always task dependent and they also always give global importance to features that have a known biological or physiological basis across different modalities and different cancer domains [19], [20]. A recurring theme is the supremacy of peritumoral texture and shape parameters in the ultrasound-based and MRI-based studies (as well as metabolic parameters in PET/CT studies) for advanced

imaging tasks in comparison to standard clinical parameters [13]-[16]. These models are able to capture fine-grained, high-dimensional signatures of tumor heterogeneity and microenvironment that are not apparent to the human eye, but strongly related to pathology, and SHAP can help fill the gap in perception between an abstract radiomic feature like dependence entropy and a tangible clinical concept like tissue heterogeneity.

This synthesis follows the systematic reviews of explainable multimodal breast cancer diagnosis that recently have been conducted [8], which show that SHAP is currently the most prevalent explanation method in the space of breast cancer diagnosis, but do not go so far as to compare the attributions it yields across tasks. The present analysis, which unifies the results of the five heterogeneous studies, demonstrates that this behavior is systematic and not incidental, the central contribution of this work in relation to past reviews.

One of the most interesting and convincing cases for clinical relevance of XAI comes from analysis of human-in-the-loop decision support. The marked improvement in the performance of radiologists, especially less experienced ones, backs a key concept of explainable AI: that it is not only a tool for model developers but also a cognitive support that can make the expert attainable for everyone and boost the diagnostic accuracy on a large scale [17]. SHAP helps clinicians better grasp what the model is learning and why, thereby improving their confidence and decisions, building trust with the model, and facilitating safe oversight of AI

systems by revealing failure modes like when it just relies on a single feature in a false-negative case [15][18].

The clinical implications are significant. A future clinical workstation might not only provide a “malignant” diagnosis along with a probability score, but also a SHAP force plot, indicating that the classification is based on a combination of indistinct mass margins (radiological), elevated peritumoral entropy (AI-derived texture), and low HER2 expression (pathology). The degree of transparency is essential in ensuring that AI systems and their work can enter into a meaningful relationship with the clinical team, shifting from the black box, automated diagnosis to an explanatory, collaborative, and trustworthy decision support system.

E. Limitations and Future Directions

In addition to the qualitative nature of this synthesis, two additional limitations on generalizing the findings of this synthesis exist. First, the five diagnostic tasks reviewed here are clinically distinct endpoints with different underlying biology and different modality combinations (Table I); the recurring feature families observed across tasks (Fig. 2, Table II) are a descriptive pattern in model behavior; and the comparisons across tasks of the magnitude of SHAP is qualitative only because the underlying models differ in architecture, calibration, and outcome scale. Second, the patient cohorts, the number of centers, the train-test split, and external validation practice differed among each source study (Table I); two of the five studies were done at the same institution that overlapped in authorship and patient-level independence between the patient cohorts was not confirmed from the published reports. The reported AUC and accuracy of a source study may be overly optimistic in the absence of independent, external validation on the source population; no attempt has been made to adjust for that here.

Another drawback is the results of the actual task of showing interpretability achieved by SHAP in this synthesis. Explanation quality and explanation usefulness are orthogonal: While technically available (as a SHAP value is always computable), an explanation may be of poor quality because it does not align with the actual decision-making process of the model or because it is not consistent across retraining iterations, or because it is not understandable and clinically actionable by the practicing radiologist. The five source studies differed in the properties they evaluated, with most reporting feature importance and cross validation accuracy, and only Ma et al. [17] measuring the impact of the explanation on decision and/or confidence. This synthesis is thus best interpreted as providing evidence that such explanations of why the model was created, or adapted into a model, exist and are fairly consistent with domain knowledge, on a task-by-task basis, but not as evidence that they are stable, faithful, or clinically useful, in the more strong sense of the words. Similarly, the stability of SHAP rankings with respect to model retraining, resampling or different random seeds was not systematically evaluated from the source studies; when reported, stability was evaluated qualitatively and not with a common stability measure, so it is unclear what

stability implies for the ‘top’ features in Table II, relative to their robustness as a representation of model behavior, vs. artifacts of a specific training run.

The meta-analytical approach has a few disadvantages. The primary studies had used different datasets, model architectures and preprocessing, so it was not possible to proceed in a straightforward way to a quantitative, statistical meta-aggregation; the synthesis presented here is qualitative and comparative, and can not be formally aggregated as effect sizes. Theoretically sound, SHAP is nonetheless dependent on assumptions of feature independence that are often not satisfied when the features are correlated, and may be time consuming for deep-learning models. Future work should also look at multimodal datasets that are harmonized and standardized with standardized SHAP benchmarking to quantify aspects of explanation fidelity, explanation stability across model retraining and resampling, and aspects of clinical utility, as well as prospective, randomized and counterbalanced reader studies to address the noted gap in the literature for using the ‘crossover design’ when working with datasets that combine multiple modalities, such as images, tabular data, and emerging work on vision-language explanation frameworks, which could translate saliency into natural language [21].

IV. CONCLUSION

In this study, SHAP is used as a common lens to compare the explanations of the decisions made by five of the most typical multimodal machine-learning models used for breast cancer diagnosis. The cross-cutting synthesis demonstrated that peritumoral texture was the SHAP-attributed leading feature in the benign–malignant differentiation, metabolic imaging parameters the leading features in the HER2 status, MRI-derived shape and texture features the leading features in axillary lymph node metastasis and Ki-67 assessment, respectively, and mammographic descriptors (indistinct margins and calcification) the leading features in molecular subtyping; due to the differences among the five source studies in terms of cohort, architecture, and validation methods (Table I), the results of the cross-cutting synthesis should be interpreted as evidence that is repeatedly seen, associated with the models, and not as a single and validated diagnostic mechanism. Adding SHAP explanations to a human in the loop process was correlated with a measurable performance improvement by radiologists, ranging from 8.3% to 15.9%, across different tasks and radiologist experience in the one study that tested it, but it remains unclear how much reader-learning effects contributed to this improvement (Section II.E). In this range, interpretability through SHAP could be a viable tool to help make the behaviour of opaque and high-dimensional models more explainable to clinicians and to identify candidate model failure modes that are useful for prospective audit-ready and clinically actionable AI systems in breast oncology before they are widely adopted or deployed. More research is needed to validate SHAP-based decision support in a harmonized, prospective manner, particularly with counterbalanced reader

studies and common metrics for the fidelity, stability, and clinical value of explanations.

REFERENCES

- [1] L. Luo *et al.*, “Deep Learning in Breast Cancer Imaging: A Decade of Progress and Future Directions,” *IEEE Rev. Biomed. Eng.*, vol. 18, pp. 130–151, 2025, doi: 10.1109/RBME.2024.3357877.
- [2] X. Zheng *et al.*, “Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer,” *Nat. Commun.*, vol. 11, p. 1236, 2020, doi: 10.1038/s41467-020-15027-z.
- [3] D. Mandair, J. S. Reis-Filho, and A. Ashworth, “Biological insights and novel biomarker discovery through deep learning approaches in breast cancer histopathology,” *npj Breast Cancer*, vol. 9, p. 21, 2023, doi: 10.1038/s41523-023-00518-1.
- [4] Z. Sun *et al.*, “A scoping review on multimodal deep learning in biomedical images and texts,” *J. Biomed. Inform.*, vol. 146, p. 104482, 2023, doi: 10.1016/j.jbi.2023.104482.
- [5] J. Lipkova *et al.*, “Artificial intelligence for multimodal data integration in oncology,” *Cancer Cell*, vol. 40, no. 10, pp. 1095–1110, 2022.
- [6] A. B. Arrieta *et al.*, “Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *Inform. Fusion*, vol. 58, pp. 82–115, 2020.
- [7] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
- [8] A. Ghasemi, S. Hashtarkhani, D. L. Schwartz, and A. Shaban-Nejad, “Explainable artificial intelligence in breast cancer detection and risk prediction: A systematic scoping review,” *Cancer Innov.*, vol. 3, no. 5, p. e136, 2024, doi: 10.1002/cai2.136.
- [9] W. Suryaputri and M. Rahardi, “Analysis of Gradient Boosted Trees Algorithm in Breast Cancer Classification,” *J. Appl. Informatics Comput.*, vol. 10, no. 1, pp. 605–618, Feb. 2026, doi: 10.30871/jaic.v10i1.11875.
- [10] F. Abdullakutty, Y. Akbari, S. Al-Maadeed, A. Bouridane, I. M. Talaat, and R. Hamoudi, “Histopathology in focus: A review on explainable multi-modal approaches for breast cancer diagnosis,” *Front. Med.*, vol. 11, p. 1450103, 2024, doi: 10.3389/fmed.2024.1450103.
- [11] D. Alzamil, B. Alkhamees, and M. M. Hassan, “A systematic review of multimodal fusion and explainable AI applications in breast cancer diagnosis,” *Comput. Model. Eng. & Sci.*, vol. 145, no. 3, pp. 2971–3027, 2025, doi: 10.32604/cmesci.2025.070867.
- [12] N. Sibanda and T. Ndlovu, “Explainable Transformer and Machine Learning Models in Predicting Tuberculosis Treatment Outcomes: A Systematic Review,” *J. Appl. Informatics Comput.*, vol. 10, no. 1, Feb. 2026, doi: 10.30871/jaic.v10i1.11846.
- [13] M. Dai, Y. Yan, Z. Li, and J. Xiao, “Machine-learning models for differentiating benign and malignant breast masses: Integrating automated breast volume scanning intra-tumoral, peri-tumoral features, and clinical information,” *Digit. Heal.*, vol. 11, p. 20552076251332736, 2025, doi: 10.1177/20552076251332738.
- [14] Z. Sun *et al.*, “Personalized prediction of breast cancer candidates for Anti-HER2 therapy using ¹⁸F-FDG PET/CT parameters and machine learning: A dual-center study,” *Front. Oncol.*, vol. 15, p. 1590769, 2025, doi: 10.3389/fonc.2025.1590769.
- [15] D. Zhang, M. Shen, L. Zhang, X. He, and X. Huang, “Establishment of an interpretable MRI radiomics-based machine learning model capable of predicting axillary lymph node metastasis in invasive breast cancer,” *Sci. Rep.*, vol. 15, p. 26030, 2025, doi: 10.1038/s41598-025-10818-0.
- [16] L. Zhang, Q. Du, M. Shen, X. He, D. Zhang, and X. Huang, “Interpretable model based on MRI radiomics to predict the expression of Ki-67 in breast cancer,” *Sci. Rep.*, vol. 15, p. 13318, 2025, doi: 10.1038/s41598-025-97247-1.
- [17] M. Ma *et al.*, “Predicting the molecular subtype of breast cancer and identifying interpretable imaging features using machine learning algorithms,” *Eur. Radiol.*, vol. 32, no. 3, pp. 1652–1662, 2022, doi: 10.1007/s00330-021-08271-4.
- [18] D. Zhang *et al.*, “Deep learning on routine full-breast mammograms enhances lymph node metastasis prediction in early breast cancer,” *npj Digit. Med.*, vol. 8, p. 425, 2025, doi: 10.1038/s41746-025-01831-8.
- [19] A. P. Yudha and M. Rahardi, “Comparative Analysis of Random Forest and XGBoost Models for Cervical Cancer Risk Prediction using SHAP-based Explainable AI,” *J. Appl. Informatics Comput.*, vol. 9, no. 6, pp. 3198–3211, Dec. 2025, doi: 10.30871/jaic.v9i6.10357.
- [20] A. Kurniawan, Y. Hermawanti, and A. N. Safriandono, “Interpretable Machine Learning with SHAP and XGBoost for Lung Cancer Prediction Insights,” *J. Appl. Informatics Comput.*, vol. 8, no. 2, pp. 296–303, Dec. 2024, doi: 10.30871/jaic.v8i2.8395.
- [21] H. Nguyen *et al.*, “LangXAI: Integrating large vision models for generating textual explanations to enhance explainability in visual perception tasks,” in *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24), Demonstrations Track*, International Joint Conferences on Artificial Intelligence Organization, 2024, pp. 8754–8758, doi: 10.24963/ijcai.2024/1025.