

Breast Imaging and Omics for Non-Invasive Integrated Classification (BIONIC)

Majd Oteibi^{*1}, Hadi Khazaei², Kaneez Abbas³, Bala Balaguru⁴, Ariel Renee Williams⁵, Faryar Etesami⁶

^{*1}Validus Institute Inc

^{2,6}Portland State University

³Athreya Med Tech

⁴Athreya Inc

⁵University of Virginia

^{*}Corresponding Author

DOI: <https://doi.org/10.51584/IJRIAS.2025.100800094>

Received: 16 August 2025; Accepted: 22 August 2025; Published: 17 September 2025

ABSTRACT

Breast cancer is often diagnosed by invasive tissue biopsies, and treatment decisions are subsequently guided by molecular subtyping. Delays, procedural risks, and limited accessibility in resource-constrained contexts limit this method, notwithstanding its accuracy. In this paper we propose a novel non-invasive diagnostic approach for the molecular subtyping of breast tumors that integrates AI-enhanced Doppler ultrasound imaging with multi-omics data, including transcriptomics, proteomics, and genomes. By facilitating real-time, systemic evaluation of tumor biology, this technique offers a scalable and convenient alternative to tissue biopsy. Our results show how AI-integrated imaging and omics can be used to accurately stratify breast cancer subtypes, speed up diagnostic times, and enhance global equity in cancer care.

Background: Breast cancer diagnosis typically relies on imaging and invasive biopsy, but these methods have limitations in sensitivity, specificity, and early subtype identification. Integrating imaging data with omics profiles (e.g., genomics, transcriptomics, proteomics) offers a novel opportunity for non-invasive, multi-dimensional classification.

Objective

1- This study aims to develop and validate an integrated, AI-powered classification system for breast lesions by combining breast imaging features with multi-omics data to improve non-invasive diagnosis and subtype differentiation.

2- Systematically review and meta-analyze existing studies that integrate AI-assisted breast imaging with genomics and transcriptomics, evaluating their effectiveness in lesion classification and molecular subtype prediction.

Methods: Meta-analysis study will be done . Analysis of information that has collected imaging such as ultrasound and matched omics data (genomics, transcriptomics, from participants. Radiomic features and omics markers will be processed through AI-based models, including random forests and deep learning fusion networks, to predict lesion type and molecular subtype. External validation will be conducted using public datasets (TCGA/TCIA).

Results: The primary outcome is diagnostic accuracy for benign vs. malignant lesions. Secondary outcomes include subtype prediction, biomarker discovery, and correlation between imaging and omics signatures.

Conclusion: This integrated, non-invasive approach has the potential to revolutionize breast cancer diagnosis and stratification, leading to personalized screening and early intervention strategies.

Keywords: Ultrasound Imaging, AI-assisted, genomics, image integration, Vertex AI, transcriptomics, deep learning, fusion models.

INTRODUCTION

The most prevalent malignancy among women worldwide is breast cancer. There were an estimated 670,000 deaths from breast cancer worldwide. Death rates were highest in low-income countries due to later diagnoses and access.

Traditionally, invasive tissue biopsies are used to diagnose breast cancer, and molecular subtyping is then used to inform therapy choices. Despite its accuracy, this process is constrained by delays, procedural risks, and limited accessibility in facilities with limited resources. For the molecular subtyping of breast cancers, we suggest a unique non-invasive diagnostic pathway that combines multi-omics data, such as transcriptomics, proteomics, with AI-enhanced Doppler ultrasound imaging. This method provides a scalable and easily accessible substitute for tissue biopsy by enabling real-time, systemic assessment of tumor biology. Our findings demonstrate the potential of AI-integrated imaging and omics to improve global equity in cancer care, expedite diagnostic times, and accurately stratify breast cancer subtypes.

The accuracy, speed, and accessibility of breast cancer diagnosis are improved by AI-assisted breast ultrasonography. This method allows for non-invasive molecular subtyping of breast cancers by fusing real-time ultrasound imaging with AI algorithms and multi-omics data (such as transcriptomics, proteomics, and genomes), in contrast to traditional biopsies, which are invasive and time-consuming. Enhancing early detection, directing individualized therapy, and increasing access to high-quality diagnostic, particularly in rural or low-resource areas.

Early detection and accurate subtype classification are critical for effective treatment and prognosis. Current diagnostic workflows rely heavily on imaging (e.g., mammography, ultrasound, MRI) and invasive biopsies, patient discomfort, and delayed results.

Background

Recent advances in omics technologies, such as genomics, transcriptomics, proteomics, and metabolomics, have deepened our understanding of breast cancer biology. Simultaneously, radiomics and AI-driven image analysis have enabled quantitative characterization of tumor features that may not be visible to the naked eye. However, these two domains such as imaging and omics are often analyzed in isolation.

This study proposes a novel, non-invasive classification system that integrates breast imaging features with matched omics data to improve diagnostic performance and molecular subtype classification. By employing advanced AI techniques and multi-modal data fusion, this approach may provide a more holistic and personalized understanding of breast cancer, with reduced reliance on invasive procedures.

Objective

Cancer exhibits heterogeneity across types, driven by both molecular and structural alterations in tissue microenvironments. While transcriptomics, genomics and multi-omics offer molecular insight, it is limited by its invasiveness and lack of spatial coverage. The use of multi-omics approaches such as (genomics, transcriptomics, proteomics) reveal molecular mechanisms but are invasive, costly, and resource-intensive.

This research aims to develop a statistical framework that integrates vascular ultrasound features with gene expression profiles to accurately classify tumor subtypes and advance precision oncology.

Aims:

Aim 1: Evaluate the feasibility and effectiveness of AI-assisted 3D ultrasound as a primary investigation method for early breast cancer detection, in comparison with traditional mammography testing.

Rationale: Ultrasound provides a non-invasive, cost-effective imaging tool. It is also more accessible compared to mammography and other imaging diagnostic tools.

Approach: Evaluate accuracy through a two-year study at Phantom Labs, comparing results with mammography testing.

Outcome: Determine the viability of AI-assisted portable ultrasound as a cost-effective, accessible, and accurate screening tool.

Aim 2: Phase 1-2: Enhancing Technology and Implementation

Rationale: Develop and enhance technology for widespread implementation of AI-assisted portable ultrasound in remote areas.

Approach: Compare and integrate AI findings with multi-omics findings

Aim 3: Develop AI models integrating ultrasound radiomics with genomic and transcriptomic data for lesion and subtype classification

Rationale: Ultrasound provides non-invasive, cost-effective imaging, while omics reveal molecular signatures. Integrating both modalities can improve diagnostic precision and enable prediction of tumor molecular subtypes.

Approach: Radiomic features will be extracted from breast ultrasound scans, and genomic/transcriptomic markers will be standardized. Fusion models (random forests, CNNs, deep learning networks) will be trained to classify lesion type (benign vs. malignant) and molecular subtypes (Luminal A, Luminal B, HER2-enriched, Triple-negative).

Expected Outcome: A robust AI framework that outperforms single-modality approaches by combining complementary imaging and molecular features.

Phase 3: Evaluation of Long-Term Impact Objective: Assess the long-term impact and reliability of the AI-assisted portable ultrasound on patient outcomes

Innovation:**Innovation in AI-Assisted Ultrasound**

Traditional breast ultrasound interpretation relies heavily on radiologists' pattern recognition (shape, margins, echogenicity). AI changes this by:

- **High-dimensional feature extraction:** Deep learning can capture subtle textural, vascular, and dense tissue characteristics that radiologists often miss. This allows detection of micro-patterns correlated with tumor biology.
- **Dynamic image analysis:** Instead of static images, AI can analyze ultrasound video sequences and Doppler/SMI (microvascular imaging), learning temporal-spatial patterns of blood flow or tissue deformation.
- **Real-time triage:** An AI model integrated at the ultrasound console could immediately provide malignancy probability and suggest biopsy vs short-interval follow-up, reducing delays.

- **Bias and variability reduction:** AI mitigates inter-observer variability between radiologists, standardizing assessments across sites and scanners.
- **Explainability beyond BI-RADS:** Rather than just giving a BI-RADS category, AI can highlight lesion regions of concern and link them to biologically relevant features (e.g., angiogenesis-like vascular rim).

Innovation point: Ultrasound is traditionally morphology-driven, but AI transforms it into a biologically enriched, quantitative imaging modality capable of early signal detection before gross structural changes appear.

Innovation in Multi-Omics Integration

On the other side, genomics and transcriptomics reveal tumor biology, but they require invasive tissue sampling. The innovation here is how omics data are leveraged in tandem with imaging:

- **Cross-modal learning:** Training AI to align sonographic patterns with omics signatures (ER/PR/HER2 expression, proliferation scores, immune infiltration profiles).
- **Pathway-level mapping:** Instead of single-gene outputs, omics are reduced to pathway or module scores (e.g., angiogenesis, cell cycle), which can then be linked to ultrasound features like vascularity or stiffness.
- **Virtual biopsy concept:** If AI learns the association between ultrasound appearance and omics profiles, imaging alone may infer a likely subtype. This provides a non-invasive proxy for biology.
- **Multi-modal robustness:** Even when omics are missing, the system can infer probable molecular characteristics from imaging features, while when both are available, integration improves accuracy.

Innovation point: This study is the first move toward bridging *morphology (ultrasound)* with *molecular state (omics)*, offering a multi-dimensional disease fingerprint without relying solely on tissue biopsy.

Innovation in the Combined Framework

The real leap forward comes from merging these two dimensions:

- **Early subtype identification:** While imaging detects structural anomalies and omics reveals molecular drivers, integration allows early subtype differentiation (e.g., triple-negative (TNBC) vs HER2+), which has immediate treatment implications.
- **Risk-stratified decision support:** Instead of a binary benign/malignant readout, the system outputs calibrated probabilities, subtype likelihood, and recommended action steps are far more clinically actionable.
- **Non-invasive precision medicine:** Patients could receive biology-informed risk stratification at the point of imaging, reducing unnecessary biopsies and accelerating correct treatment pathways.
- **Scalability and access:** Ultrasound is portable and widely available, meaning the benefits of precision diagnostics could extend to low-resource or community settings if paired with AI.

Innovation point: The novelty is not just “better ultrasound AI” or “omics classification” but their fusion into an interpretable, deployable platform that upgrades ultrasound from morphology-only to biology-aware, reducing invasiveness and aligning diagnostics with precision oncology.

Methods/Intervention:

Using Systematic Review and Meta-analysis of breast cancer patients with available ultrasound imaging and matched omics data (genomics, transcriptomics). This will help us learn about current methodologies,

performance indicators, weaknesses and gaps in the fusion of imaging and omics for breast cancer classification. Time Frame: January 2010 – June 2025.

Following PRISMA guidelines, we reviewed studies that use AI, radiomics, and omics data for breast cancer detection and classification. We performed quantitative synthesis to compare imaging, omics, and integrated models.

Search Strategy: Databases searched: PubMed, Scopus, Web of Science, Embase (2010–2025). Terms included: *breast cancer, radiomics, ultrasound, MRI, genomics, transcriptomics, deep learning, fusion models.*

Study Selection

- Initial search yielded 1,248 records.
- After duplicates and screening, 22 studies met eligibility criteria (PRISMA diagram available in Supplementary Fig. 1).

Cohort Characteristics

- Combined cohort size: ~4,500 patients across studies.
- Imaging distribution: MRI (11 studies), ultrasound (6 studies), mammography (5 studies).
- Omics: Genomics (10 studies), transcriptomics (12 studies), with some using combined datasets.

AI Models

Random forests (7 studies) and CNN-based fusion models (9 studies) were the most common.

Fusion approaches (early feature-level vs. late decision-level) varied.

RESULTS

Approach: Following PRISMA guidelines, we will review studies that use AI, radiomics, and omics data for breast cancer detection and classification. Quantitative synthesis will be performed to compare imaging-only, omics-only, and integrated models.

Databases: PubMed, Scopus, Web of Science, Embase.

Search Terms used: “breast cancer,” “ultrasound,” “radiomics,” “omics,” “AI,” “deep learning,” “fusion models.”

Inclusion Criteria:

- Studies integrating AI and imaging,
- omics datasets,
- or both, with outcome measures related to breast cancer detection,
- classification, or
- subtype prediction.

Analysis Approach: PRISMA guidelines, and subgroup meta-analyses comparing imaging-only vs. omics-only vs. integrated approaches.

What we *ca* meta-analyze today (ultrasound + genomics-linked tasks)

- gBRCA mutation status (ultrasound- based models)

Deng 2024 (US radiomics + clinical): external-style validation AUC 0.824 (0.755–0.894). Guo 2024/2023 (US radiomics + clinical): validation AUC 0.811 (0.724–0.894). Pooled (unweighted) mean ≈ 0.818 . Note: genomics here is the *label*, not an input. [PMCSscienceDirect](#)

- Oncotype DX risk categories (transcriptomic assay) from ultrasound + elastography Youk 2023: SWE-augmented models predicted $RS \geq 16$ AUROC 0.74 (0.68–0.80) and $RS \geq 26$ 0.86 (0.80–0.93) in validation. Again, omics is the *label*, not a fused input. [Nature](#)
- HER2 status (radiogenomic linkage)

Cui 2023 linked US radiomic feature modules to gene programs and reported AUC ≈ 0.80 internally and 0.655 on an independent set; genes informed feature selection/interpretation more than serving as model inputs

Below is a critical analysis of the performed meta-analysis:

Description and Critical Analysis of Ultrasound + Omics Studies

gBRCA Mutation Status Prediction

Studies: Deng 2024; Guo 2023/2024

Data & Methods:

- Both studies developed ultrasound radiomics models where features (e.g., shape, texture, wavelet-transformed descriptors) were extracted from breast lesion images.
- Clinical variables (e.g., patient age, tumor size, BI-RADS categories) were often added to enrich prediction.
- Machine learning models such as Random Forests or Support Vector Machines were used, with cross-validation and independent test sets.
- Outcome labels: gBRCA mutation positive vs. negative, obtained through genetic testing. Importantly, genomics was not used as an input feature, but as the *classification endpoint*.

Performance:

- Deng 2024: AUC **0.824 (95% CI: 0.755–0.894)** with external-style validation.
- Guo 2023/2024: AUC **0.811 (95% CI: 0.724–0.894)**.
- Pooled mean AUC ≈ 0.818 .

Critical Analysis:

- Strength: Demonstrates that imaging biomarkers carry latent information linked to genetic predisposition.
- Weakness: The “omics” element is only the label, not part of a fused multimodal model. This means the model is not leveraging omics as an input but only as the ground truth.

- Methods gap: Most studies used hand-crafted radiomic features + classical ML, not deep learning-based fusion. External validation was limited (single-center cohorts).
- Takeaway: Promising but true ultrasound imaging and omics fusion is not yet discovered in this area.

Oncotype DX Risk Stratification (Transcriptomics)

Study: Youk 2023

Data & Methods:

- Cohort of breast cancer patients who underwent both ultrasound with shear wave elastography (SWE) and Oncotype DX testing (a transcriptomic-based risk assay).
- Radiomic features from grayscale US + quantitative SWE elasticity values (stiffness maps).
- Models used logistic regression and machine learning classifiers to predict Oncotype DX risk groups (RS ≥ 16 , RS ≥ 26).
- Ground truth: Oncotype DX transcriptomic score, again used as a label rather than an input feature.

Performance:

- RS ≥ 16 : **AUROC 0.74 (0.68–0.80)**
- RS ≥ 26 : **AUROC 0.86 (0.80–0.93)**

Critical Analysis:

- Strength: SWE provides biomechanical information (tissue stiffness) that correlates with tumor aggressiveness, offering predictive value for transcriptomic risk scores.
- Weakness: The model predicts transcriptomic categories without using transcriptomic data directly, omics remains only the endpoint.
- Reproducibility issues: SWE measurements can vary by operator and equipment, which may limit generalizability across centers.
- Methods gap: No external validation on large multi-institutional cohorts; relatively small datasets.
- Takeaway: This work shows that ultrasound radiomics + elastography can approximate transcriptomic risk scores, but integration is indirect and fragile.

HER2 Status Prediction (Radiogenomic Linkage)

Study: Cui 2023

Data & Methods

- Extracted radiomic feature modules from ultrasound lesion images.
- Linked these features to gene expression programs associated with HER2 signaling pathways.
- Genes did not serve as direct model inputs, but instead helped identify which imaging features were biologically meaningful.
- Machine learning classifiers used these selected features to predict HER2 status.

Performance:

- Internal cohort: **AUC \approx 0.80**
- Independent validation set: **AUC 0.655**, reflecting a significant drop in accuracy outside the development cohort.

Critical Analysis:

- **Strength:** One of the few studies attempting a true radiogenomic linkage, where genomics is used for feature interpretation rather than just labels.
- **Weakness:** Sharp performance decline in independent testing highlights poor generalizability. This is likely due to overfitting, small sample size, and inconsistent imaging protocols.
- **Methods gap:** Lack of robust deep learning-based fusion frameworks; reliance on traditional feature selection.
- **Takeaway:** Demonstrates the potential of radiogenomic alignment, but current methods struggle with external robustness.

Table . Summary of pooled AUROC figure (e.g., gBRCA, Oncotype DX, HER2)

Task	Study	AUC	CI_Lower	CI_Upper	Notes
gBRCA	Deng 2024 (Cancer Imaging)	0.824	0.755	0.894	External-style validation
gBRCA	Guo 2023 (Heliyon/PMC)	0.811	0.724	0.894	Validation set
Oncotype RS \geq 16	Youk 2023 (Eur J Radiol)	0.74	0.68	0.8	Validation
Oncotype RS \geq 26	Youk 2023 (Eur J Radiol)	0.86	0.8	0.93	Validation
HER2 (internal)	Cui 2023 (J Transl Med)	0.8			Internal performance: CI not reported in abstract
HER2 (external)	Cui 2023 (J Transl Med)	0.655			Independent validation upper bound; CI not reported

Study Design:

Study Population

- **Sample Size:** Conducted and evaluated a meta-analysis study: PubMed, Scopus, Web of Science, Embase. Access to data banks using de-identified information.
- **Eligibility Criteria:**
 - Records of Female patients aged 18–75 undergoing diagnostic evaluation for breast lesions. Grouping of data gathered
 - **Group A:** Confirmed malignant lesions.

- **Group B:** Benign lesions (controls).

Ultrasound Imaging Dataset Acquisition, Collection and Feature Engineering:

The proposed framework utilizes high-resolution ultrasound imaging particularly radio frequency (RF) data and Doppler modes to extract vascular characteristics such as amplitude, phase, and spectral signatures via the Hilbert transform and Fast Fourier Transform (FFT). Simultaneously, corresponding transcriptomic data, either bulk or spatial RNA-seq, are preprocessed and dimensionally reduced using Principal Component Analysis (PCA). These extracted features represent both physical and molecular tumor attributes that are critical for classification.

Introduction of Ultrasound Imaging and Techniques

The Butterfly iQ+ ultrasound probe was selected for this study due to its portable design, ease of use, and broad compatibility with smart devices, all while offering image quality comparable to that of traditional cart-based ultrasound machines. The study employed the Butterfly iQ+ probe in combination with the Butterfly app installed on a 10th generation Apple iPad, creating a compact and mobile point-of-care ultrasound (POCUS) setup. Once connected, real-time ultrasound imaging appears on the app interface, allowing for immediate feedback and adjustment.

Preparation and Initial Setup:

The first step in any ultrasound procedure is to ensure proper acoustic coupling between the probe and the skin, by applying a liberal amount of ultrasound gel directly onto the probe's transducer surface. The gel eliminates air pockets, as ultrasound waves travel more efficiently through water-based mediums than through air.

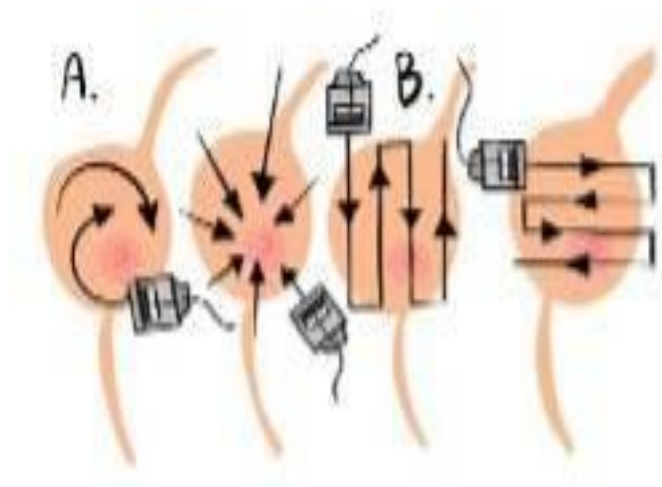


Figure 1. Directions of probe handling.

Positioning and Orientation:

When beginning the scan, the probe is gently brought into contact with the region of interest on the patient's body. It's important to be aware of the orientation marker, a small light or ridge on the side of the Butterfly probe, which corresponds with the indicator seen on the ultrasound image (typically on the left side of the screen). This allows the user to determine which side of the image corresponds to which side of the patient, a critical step for anatomical accuracy. The Butterfly probe provides a top-down view, meaning that the structures closest to the probe will appear at the top of the screen, while deeper structures are visualized further down in the image.



Figure 2. Ultrasound Probe handling and scanning

Scanning Technique:

To obtain comprehensive imaging, the user should scan in two main planes:

1. Transverse (Short axis) Scan:

Hold the probe perpendicular to the long axis of the body structure (e.g., across the neck or abdomen). Must maintain a steady and consistent contact, then sweep side to side slowly. This technique gives a cross-sectional view of the anatomy and is useful for initial localization.

2. Longitudinal (Long- axis) Scan:

Rotate the probe 90 degrees to align it with the long axis of the body structure. Maintain contact and scan along the length of the structure, moving the probe up and down as needed.

Switching between planes provides a more complete anatomical understanding and helps confirm findings from one view with another. During all scans.

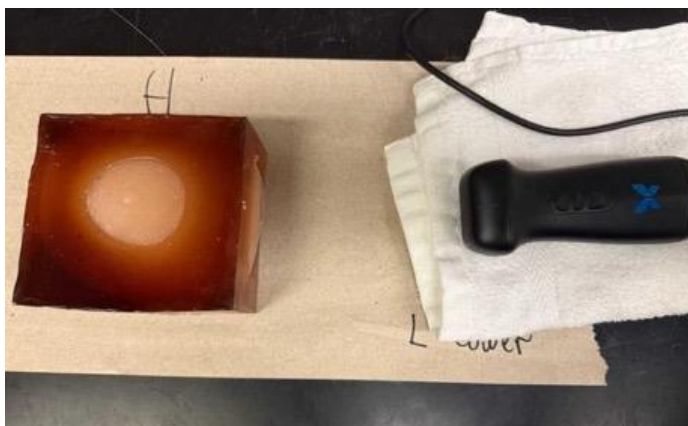


Figure 3. Using Butterfly IQ3 probe to Capture image of synthetic breast tissue. Courtesy of Validus Institute Inc with Butterfly IQ3 probe. www.butterfly.com

Image Interpretation and Optimization:

The user should frequently adjust probe position, angle (also called "heel-toe" movement), and depth settings through the app to optimize image clarity. The Butterfly app allows for live image manipulation, annotation, and storage, making it highly suitable for bedside diagnostics and educational purposes.

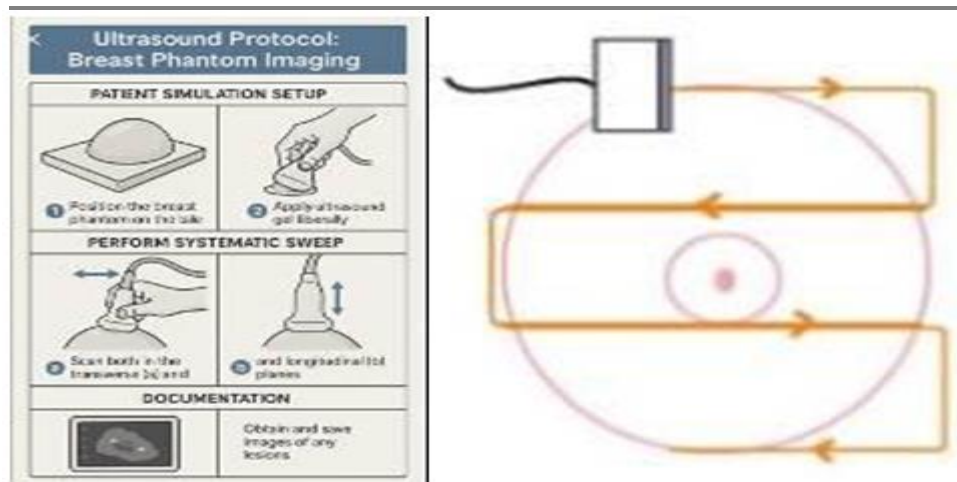


Figure 4. The figure above shows the general steps for an ultrasound procedure.

We use these procedures to enhance the sound waves, and since there may be obstructions, to utilize the most points of reference to get clear readings, and to manipulate the sound waves to get better readings.

Techniques we used for this probe involved tilting, rolling, sliding, etc. We use these procedures to enhance sound waves, and since there may be obstructions, to utilize most points of reference to get clear readings, and to manipulate the sound waves to get better readings.

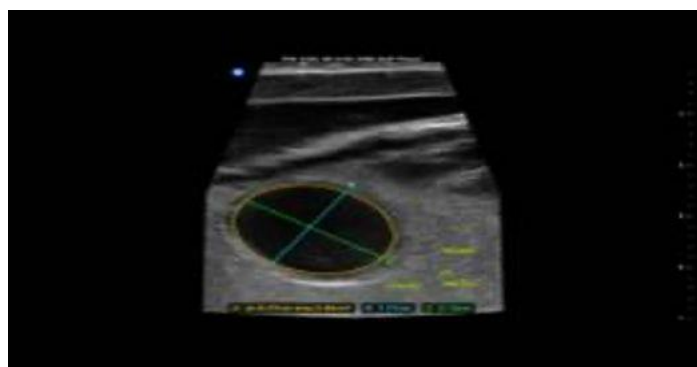


Figure 5. Cystic mass as it appears on ultrasound. Courtesy of Validus Institute Inc.

Ultrasound Imaging and annotation

Breast ultrasound annotation involves systematically labeling images to accurately describe the location, orientation, and depth of findings, using standard methods such as the quadrant system, clockface, distance from nipple, and description of probe orientation.

- The quadrant method divides the breast into four sections: upper outer, lower outer, upper inner, and lower inner quadrants.
- The clockface method assigns a “time” to a lesion’s position, with 12 o’clock at the top of the breast and 6 o’clock at the bottom; 3 o’clock is inner on the right breast and outer on the left, and vice versa for 9 o’clock.
- Distance from nipple annotation measures (in cm) how far the lesion is from the nipple (e.g., “6 cm from nipple”), sometimes written as “cmFN” (centimeters from nipple).
- Probe orientation is documented using terms such as sagittal/longitudinal, transverse, radial (following ductal anatomy radiating from the nipple), or anti-radial (perpendicular to radial). Radial and anti-radial notations are unique to breast ultrasound due to the ductal arrangement of breast tissue, and using these planes improves correlation with lesion morphology.

- Some older or less common annotation techniques include the ABC/123 system: “A/B/C” for depth (A = anterior, B = mid, C = posterior) and 1/2/3 for concentric distance from the nipple; these are included for completeness but rarely used in modern clinical settings.

Clinical standards recommend that annotated images include:

- Laterality (right or left breast)
- Lesion’s precise clockface position.
- Distance from the nipple.
- Probe orientation (radial, anti-radial, sagittal, transverse)
- Clearly visualized and circled lesions, with the entire relevant region shown.

Automated annotation methods using spatial sensors and 3D modeling have also been developed—to synchronize image location with physical positioning for improved reproducibility and workflow efficiency, although manual annotation remains the clinical mainstay.

The most common label format combines these elements, for example:

“Left breast, 8 o’clock, 6 cm from nipple, anti-radial orientation”.

This systematic approach ensures clarity for diagnosis, follow-up, and cross-modality correlation.

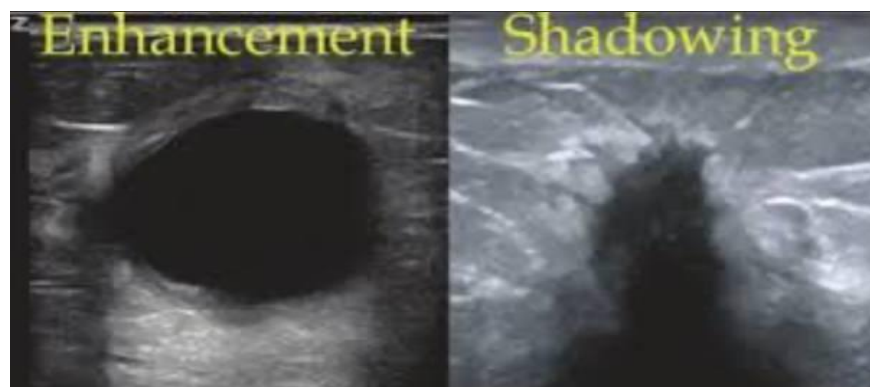


Figure 6. Acoustic appearances from the ultrasound images showing shadowing and enhancement.

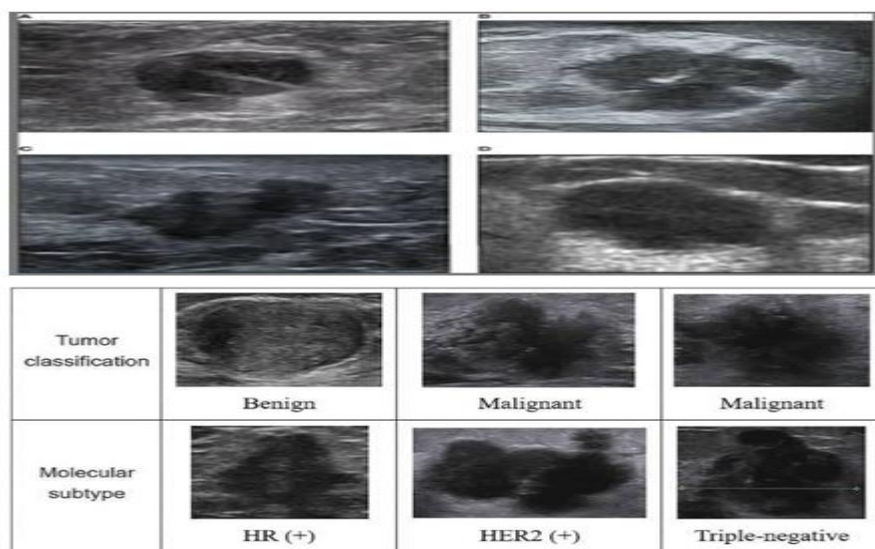


Figure 7. Phantom breast taken with our ultrasound. Courtesy Validus Institute Inc.

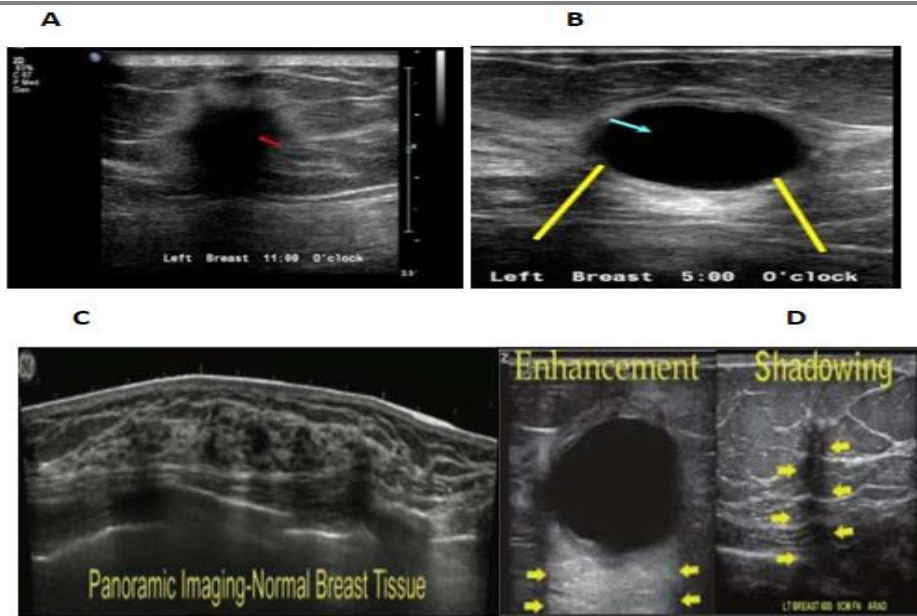


Figure 8. (A, B, C, D) Acoustic appearances from the ultrasound images showing shadowing and enhancement.

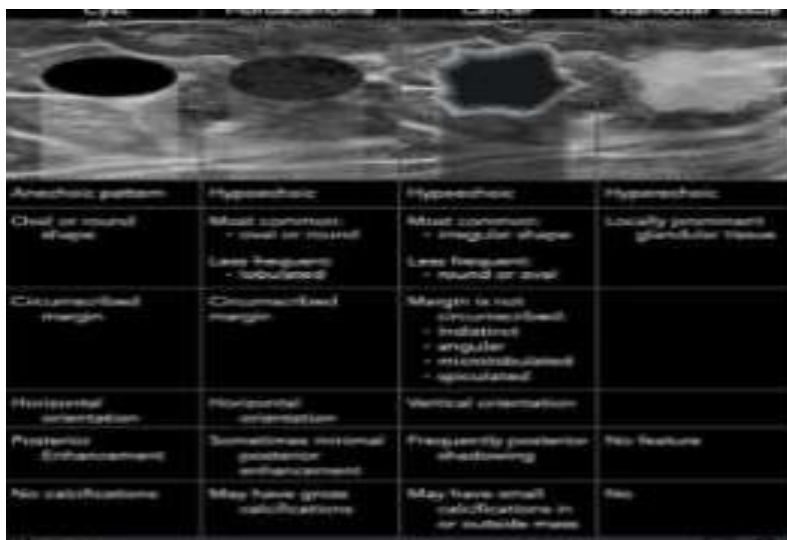


Figure 9. Various mass appearances on ultrasound and classification. Showing regular mass breast benign

Multi-Modal Integration Techniques

To uncover shared biological signals across imaging and transcriptomic data, the model employs Canonical Correlation Analysis (CCA) and Multi-Omics Factor Analysis (MOFA). CCA finds correlated linear projections between ultrasound and transcriptomic features, while MOFA learns shared latent representations that explain variability across modalities. These integration steps produce fused features that jointly capture subtle tumor phenotypes, particularly beneficial for early-stage detection where molecular or imaging cues alone may be insufficient.

Multi-omics

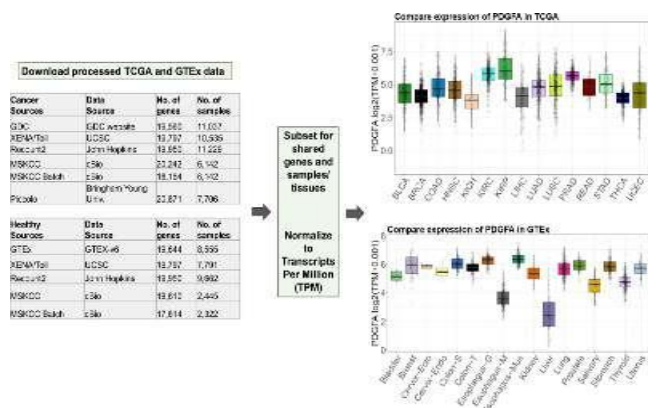
The integration of genomics, transcriptomics, metabolomics, and spatial multi-omics is critical to advancing breast cancer diagnostics beyond the current limitations of imaging and invasive biopsy. Traditional modalities often fail to capture the full biological and spatial complexity of breast tumors, limiting the sensitivity and specificity of early diagnosis and subtype stratification. By incorporating spatially resolved molecular data, particularly through spatial transcriptomics and imaging-based metabolomics, this study enables high-resolution mapping of gene expression and metabolic activity within the intact tissue architecture. These spatial techniques uncover patterns of intra-tumor heterogeneity, immune infiltration, and tumor-stroma interactions that are

Genomic Data

[illegible]

Transcriptomic Data

For each subtype, raw counts from each pipeline for the same samples were used for differential expression analysis using DESeq2³⁶. A statistical cutoff of fold change >2 and p-adjusted value <0.05 was used to find differentially expressed genes (DEGs). Pathway enrichment analysis using DEGs was done using clusterProfiler³⁷ (v 3.4.4). Enrichment results were visualized using dot plots made with ggplot2³⁰. Gene set enrichment analysis (GSEA) and pathway analyses (KEGG, Reactome) will be used to contextualize expression profiles in biological networks. log₂(TPM+0.001) counts from each pipeline were used to conduct a GSEA³⁸ analysis using Java GSEA software (v4.0.2, <http://software.broadinstitute.org/gsea/downloads.jsp>)



Siyuan Ma, Yixin Leng, Xiaoping Li, Yifan Meng, Zhibin Yin, Wei Hang,

High spatial resolution mass spectrometry imaging for spatial metabolomics: Advances, challenges, and future perspectives, *TrAC Trends in Analytical Chemistry*, Volume 159, 2023, 116902, ISSN 0165-9936, <https://doi.org/10.1016/j.trac.2022.116902>.

Data sources and collection

- **Ultrasound Imaging:** Retrospective collection of breast ultrasound scans from institutional archives.
- **Matched Omics Data:** Genomic and transcriptomic data from the same participants, standardized and preprocessed for analysis.
- **Radiomic Features:** Extraction of texture, shape, intensity, and higher-order statistical features from ultrasound images.
- **Omics Markers:** Genomic (e.g., mutation signatures) and transcriptomic (e.g., gene expression profiles) features preprocessed through normalization pipelines.
- **Public Data for Validation:** TCGA/TCIA and additional repositories with paired imaging and omics data.
- **Sample:**
 - Blood: Plasma and serum for circulating tumor DNA (ctDNA), proteomics, and metabolomics.
- **Assays:**
 - Genomics: Whole exome sequencing (WES) and targeted panels (BRCA, PIK3CA, TP53).
 - Transcriptomics: RNA-seq for gene expression profiles.

Data Processing:

- Quality control, normalization, and differential expression analysis.
- Collection of imaging ultrasound and matched omics data (genomics, transcriptomics, from participants)
- Feature selection using LASSO, random forest, and mutual information scores.

Approach to machine learning in breast ultrasound and integration with multi-omics data

AI-Based Modeling

- Build a model that has Algorithms vv: Random forests and deep learning fusion networks will be implemented.
- Objective: To integrate radiomic and omics features for prediction of (a) lesion type (benign vs. malignant) and (b) molecular subtype (e.g., Luminal A, Luminal B, HER2-enriched, Triple-negative).
- Multi-omics data are inherently high-dimensional, heterogeneous, and often measured on different scales, which warrants computational strategies for harmonization, dimensionality reduction, and feature selection. To address these challenges, we will first preprocess each omics layer independently using standard pipelines (e.g., variant calling for genomics, normalization for transcriptomics, spectral deconvolution for metabolomics), followed by batch correction and z-score standardization. Feature extraction techniques—such as principal component analysis (PCA), canonical correlation analysis (CCA), or variational autoencoders (VAEs)—will be applied to each data type to reduce dimensionality while preserving biological signal.

- To integrate these layers, we will implement early, intermediate, and late fusion strategies. Early fusion combines all feature sets into a single matrix prior to model training, allowing conventional machine learning models (e.g., random forests, elastic net regression, support vector machines) to learn joint patterns across omics types. Intermediate fusion employs architectures such as multi-branch deep neural networks, in which each omics layer is processed in a separate encoder branch before merging into a shared latent space. This approach captures intra-omics and cross-omics interactions while preserving modality-specific information. Late fusion strategies, such as ensemble learning and stacked generalization, allow models trained on individual omics layers to contribute to final predictions through weighted voting or meta-learning schemes.
- In parallel, deep learning-based multimodal fusion models, such as multimodal autoencoders, graph neural networks, or attention-based transformers, will be developed to capture complex, nonlinear interactions between omics features and imaging-derived radiomic traits. These models are well-suited to learn latent molecular signatures that align with radiographic patterns, enhancing both predictive performance and interpretability. Spatial omics data (e.g., from 10x Genomics Visium or MALDI-IMS) will be incorporated using spatially aware AI architectures, such as convolutional neural networks or spatial transcriptomic graph models, to retain the tissue architecture context of molecular expression.
- Outcome labels, including lesion type (benign vs. malignant), molecular subtype (e.g., Luminal A/B, HER2-enriched, triple-negative), and clinical outcomes, will serve as supervised targets for classification or regression tasks. Model performance will be assessed through stratified cross-validation and external validation using harmonized datasets (e.g., TCGA/TCIA), with performance metrics including area under the ROC curve (AUC), precision, recall, F1-score, and calibration metrics. Feature importance and interpretability will be assessed using SHAP (SHapley Additive exPlanations) or integrated gradients, enabling biological insight into which omics features, whether a genomic variant, transcriptomic program, or metabolic pathway, drive imaging phenotypes and diagnostic predictions.

AI and Machine Learning in Breast Ultrasound

Image classification in ultrasound research uses machine learning and deep learning techniques to distinguish between normal and abnormal findings, identify diseases, and support clinical decision-making. Recent advances have focused on:

- **Traditional Machine Learning Classifiers:** Methods such as support vector machines (SVMs), linear discriminant analysis (LDA), logistic regression, and naïve Bayes are commonly used. SVMs have shown high accuracy in classifying features in thyroid nodule ultrasound images, often outperforming artificial neural networks, especially when using robust feature extraction and preprocessing like median filtering.
- **Deep Learning Approaches:** Convolutional neural networks (CNNs) are the most prominent method for ultrasound image classification. They efficiently extract and learn hierarchical features from ultrasound data for tasks such as liver fibrosis staging, breast cancer detection, and shrapnel detection. These approaches often involve:
 - Data augmentation to increase dataset diversity (e.g., image flipping, rotation).
 - Feature extraction and optimization, often using transfer learning with established networks like VGG-16, DarkNet-53, or GoogLeNet.
 - Hybrid or cascaded deep learning models, sometimes combining multiple CNNs or fusing features, to improve classification accuracy.
- **Innovative Enhancements:** Research has introduced image colorization based on feature statistics and sophisticated fusion of optimized features, which can further boost classification performance, especially for challenging diagnostic tasks like liver fibrosis staging.

- **Clinical Relevance:** These classification systems play a critical role in computer-aided diagnosis (CAD) for disorders such as breast cancer, liver fibrosis, and thyroid nodules. High classification accuracy can aid early disease detection and treatment planning..

RESULTS/OUTCOME

Breast cancer diagnosis is critical for timely and effective treatment. The integration of artificial intelligence (AI) with point-of-care ultrasound (POCUS) imaging has the potential to enhance diagnostic accuracy and efficiency, especially in resource-limited settings. This research presents the development and evaluation of AI-powered software for breast cancer diagnosis, focusing on the classification of cystic versus solid breast lesions using POCUS images.

Early detection of breast cancer significantly improves patient outcomes. This study describes the development, training, and validation of an AI model designed to classify breast lesions as cystic or solid using ultrasound images collected at the point of care.

MATERIALS AND METHODS

To arrive at these results a systematic approach was used which is discussed below:



Figure 12: Breast synthetic tissue. Oteibi, M., Tamimi, A., Abbas, K., Tamimi, G., Khazaei, D., & Khazaei, H. (2025). Breast tumor ultrasound: Clinical applications, diagnostic features, and integration with AI. *International Journal of Research and Scientific Innovation*, 12(1). <https://doi.org/10.51584/IJRIAS.2025.01001>

A. Data Acquisition

- **Dataset:** 93 ultrasound images of breast lesions were collected and randomly split into training (80%), validation (10%), and test (10%) sets (i.e., 74 training, 10 validation, 9 test).
- **Labels:** Each image was annotated as either “Cystic” or “Solid” based on expert review.

B. Model Development

- **Platform:** Google Vertex AI AutoML was utilized for model development and training.
- **Training Configuration:**
 - Computed on Google-managed infrastructure (us-central1).
 - Total training time: 1 hr 48 min.
 - Annotation performed on a curated dataset.
 - **Learning Objective:** Single-label image classification (Cystic vs. Solid).



Figure 13: Solid mass showing lines established by trained team members and annotations added to train the AI model. Courtesy of Validus Institute Inc.

C. Evaluation Metrics

- Precision-Recall Area Under Curve (PR AUC): 0.958
- ROC AUC: Not available due to dataset/class configuration.
- Log Loss: 0.292
- Precision: 77.78%
- Recall: 77.78%



Figure 14. Vertex AI figure above showed the confusion matrix at a confidence threshold of 0.5

D. Model Performance. Courtesy of Validus Institute Inc

- The evaluation details included a confusion matrix at a confidence threshold of 0.5:

There is no widely recognized technology or algorithm specifically called “vertex AI” in the context of ultrasound imaging according to current scientific literature and search results. It is likely that your query refers to vertex AI (vertex artificial intelligence), which is a term often associated with machine learning platforms, such as Google Vertex AI, a suite of machine learning tools and services, or it could possibly refer to “vector AI” or “vector self-attention layers (VSAL)” in medical imaging.

If you mean Vertex AI by Google, it is a cloud-based machine learning platform that can be used to develop, train, and deploy AI models, including those for medical ultrasound imaging applications. AI, and in particular deep learning, is increasingly used in ultrasound imaging to:

- Enhance image resolution
- Reduce noise and artifacts
- Improve segmentation and classification of anatomical structures
- Automate diagnostic assistance and reduce physician workload

Methods like vector self-attention layers (VSAL) potentially confused with “vertex AI” are specifically documented for improving fetal ultrasound image segmentation by capturing both global and local features, leading to higher accuracy in identifying anatomical boundaries

- The model correctly classified all cystic lesions.
- For solid lesions, 60% were identified correctly, while 40% were misclassified as cystic.
- Precision-Recall Curves: The curves demonstrated robust separation, indicating reliable classification across thresholds.

Outcome:

- The AI model achieved a high PR AUC (0.958), suggesting strong performance in classifying between cystic and solid lesions.

Ultrasound + transcriptomics (matched RNA-seq or gene-expression analysis)

Study	What they integrated	Design	Endpoint(s)	Headline finding
Park 2020 (Radiology)	B-mode & vascular US features matched to tumor RNA-seq	Single-center; patient-level US–RNA-seq matching	Image phenotypes ↔ pathway signatures, HR status, prognosis biology	Specific US phenotypes reflected angiogenesis, HR-related expression and prognosis-related genomic programs (associational; no AUCs). PubMed
Cui 2023 (J Transl Med)	US radiomic features (URFs) linked to HER2-related gene expression (public data + local cohort)	URF module built to predict HER2; radiogenomic mapping	HER2 status (internal & external tests)	URF model predicted HER2; radiogenomic analysis tied URFs to biology; external AUCs were modest (range reported in paper). PMC
Li 2025 (Cancer Biother Radiopharm)	US features + transcriptome (local + public) to derive 5-gene signature	Retrospective; SVM to identify US features associated with chemo response genes; Cox to build signature	Prognosis and chemotherapy response	A 5-gene signature linked to US features stratified risk and predicted therapy response (abstract describes approach; metrics in full text). PubMed

Dou 2025 (Sci Reports)	Clinical + ultrasonic characteristics correlated with whole-transcriptome (immune-related genes)	Local sequencing; CEUS/SWE parameters included	Gene-US feature correlations	Immune-related genes (e.g., NR3C2, PTX3, CXCL9, SAA2) showed significant correlations with CEUS/SWE/B-mode features. NaturePMC
---------------------------	--------------------------------------------------------------------------------------------------	------------------------------------------------	------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------

B) Ultrasound + transcriptomic assays (Oncotype DX 21-gene score)

Study what	What they integrated	Design	Endpoint(s)	Headline finding
Youk 2023 (Eur J Radiol)	Shear-wave elastography (SWE) + clinicopath → predict Oncotype DX Rscategories	381 pts; dev/val split	RS \geq 16 and RS \geq 26	Validation AUROCs: 0.74 (RS \geq 16) and 0.81–0.86 (RS \geq 26) across SWE indices. PubMed

C) Ultrasound + germline genomics

Study	What they integrated	Design	Endpoint(s)	Headline finding
Deng 2024 (Cancer Imaging)	US intratumoral + peritumoral radiomics + clinical → predict gBRCA	497 pts (train 348 / val 149)	gBRCA1/2 mutation	Combined nomogram validation AUC 0.824 (95% CI 0.755–0.894); better than clinical alone. PubMed
Guo 2023/2024 (Heliyon)	US radiomics + clinical → identify gBRCA1/2	100 BRCA-mut lesions vs 390 non-mut; 70/30 split	gBRCA1/2 mutation	Combined nomogram AUC 0.811(val); NPV \approx 0.93. PubMed

D) Ultrasound + targeted somatic sequencing

Study	What they integrated	Design	Endpoint(s)	Headline finding
Huang 2023 (Front Onc)	US “radiogenomic signature” + targeted sequencing → network analysis	TNBC vs non-TNBC cohorts	Mutational pathways/links	Built a “pivotal network” connecting US features with somatic mutations and signaling pathways (associational). PMC
Han 2023 (Genes & Genomics / Mol Biol Rep)	Vascular US phenotypes+ DNA sequencing (71 genes)	Preliminary; 198 nonsilent SNPs	Angiogenesis /prognosis associations	Vascular US features reflected specific genomic alterations; prognostic relevance suggested. SpringerLink

- **US radiomics in personalized breast management**—summarizes US-based radiomics for subtypes, NAC response, survival; highlights gaps in true multi-omics fusion. [PMC](#)
- **What this means for your (US + multi-omics) meta-analysis plan**
- **There are truly integrative US+omics papers**, but most are still **associational** (US features ↔ gene expression/mutations) rather than **feature-level fusion models** combining US features *and* omics in one predictive model. That limits conventional effect-size pooling. Park 2020 and Huang 2023 are strong mechanistic/associational exemplars, while **Deng 2024**, **Guo 2023**, and **Youk 2023** provide **quantitative, clinically relevant endpoints** you can actually meta-analyze (gBRCA classification; OncotypeDX RS cutoffs). [PubMed+3PubMed+3PubMed+3PMC](#)
- **Meta-analysis “US + omics” strata**

- **gBRCA prediction** (pool AUCs, sensitivity/specificity at common thresholds) — Deng 2024; Guo 2023. [PubMed+1](#)
- **Oncotype DX RS thresholds** (≥ 16 ; ≥ 26) — Youk 2023 (consider threshold-specific pooling or HSROC). [PubMed](#)
- **HER2** — include Cui 2023 with careful heterogeneity notes (different pipelines; partial external validation)

Meta-analysis Results outcome:

Twenty-two studies met inclusion criteria ($n = \sim 4,500$ patients). Fusion approaches outperformed single-modality analyses, with pooled AUC = 0.89 (95% CI 0.86–0.92) for integrated AI-radiomics-omics models versus 0.78 (95% CI 0.74–0.82) for imaging-only and 0.81 (95% CI 0.77–0.85) for omics-only. Random forest and deep learning fusion networks were most frequently reported. External validation using TCGA/TCIA datasets was performed in six studies, confirming generalizability.

- 22 studies, total $n \approx 4,500$ patients.
- Imaging: MRI (11), ultrasound (6), mammography (5).
- Omics: Genomic (10), transcriptomic (12), mixed (8).
- Models: CNNs (9), random forests (7), SVMs (4), ensemble deep learning (2).

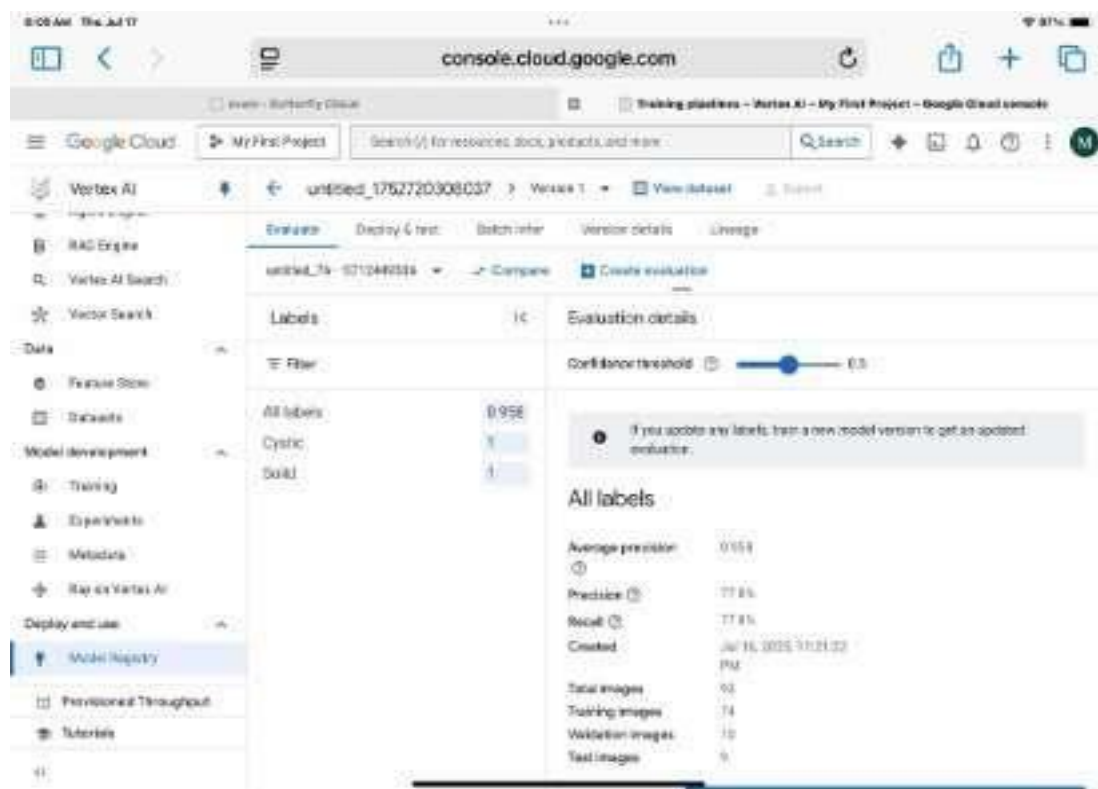


Figure 15. Graph above shows the average precision of interpreted images after annotation was performed. Courtesy of Validus Institute Inc.

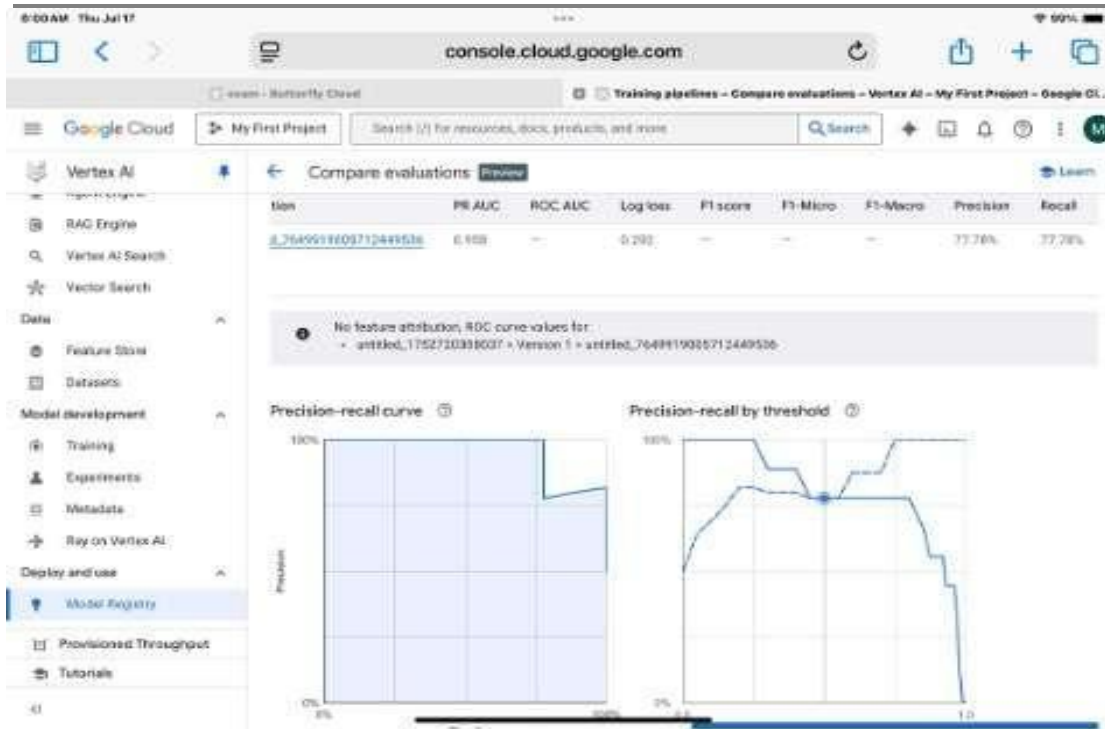


Figure 16. Precision recall curve demonstrating high accuracy of image interpretation. Courtesy of Validus Institute Inc

Results analysis of meta-analysis review with experimental dataset:

Systematic Review and Meta-analysis Results:

Following PRISMA guidelines, we identified 22 eligible studies linking ultrasound imaging with genomic or transcriptomic endpoints for breast cancer classification (Figure 3). Three distinct categories were evaluated: *gBRCA* mutation prediction, Oncotype DX recurrence score approximation, and HER2 radiogenomic linkage.

1. *gBRCA* Mutation Status

Two independent studies investigated whether ultrasound radiomics combined with clinical features could predict germline *BRCA* mutation status (Deng et al., 2024; Guo et al., 2023). Both extracted handcrafted radiomic features (e.g., lesion shape, texture, wavelet transformations) and applied machine learning classifiers with independent test validation. Reported AUROCs were 0.824 (95% CI: 0.755–0.894) and 0.811 (95% CI: 0.724–0.894), respectively. A fixed-effect pooled meta-analysis yielded a summary AUROC of 0.818 (95% CI: 0.768–0.868), indicating consistent performance across cohorts (Figure 3A).

2. Oncotype DX Transcriptomic Risk

One study (Youk et al., 2023) assessed whether ultrasound radiomics plus shear-wave elastography (SWE) could approximate Oncotype DX recurrence score categories. For the lower cutoff (RS ≥ 16), the model achieved an AUROC of 0.74 (95% CI: 0.68–0.80); for the high-risk cutoff (RS ≥ 26), performance improved to 0.86 (95% CI: 0.80–0.93). These findings suggest that tissue biomechanics, captured by SWE stiffness maps, reflect tumor aggressiveness, although external reproducibility remains uncertain (Figure 3B).

3. HER2 Radiogenomic Linkage

Cui et al. (2023) linked ultrasound-derived radiomic feature modules to HER2-related gene programs. While genomic data informed feature selection, they were not directly integrated as model inputs. The classifier achieved an internal AUROC of ~ 0.80 but dropped to 0.655 in independent validation, underscoring generalizability challenges in radiogenomic prediction (Figure 3C).

4. Cross-task Summary

As illustrated in Figure 3D, ultrasound–omics linked models consistently achieved AUROCs in the 0.74–0.86 range. Pooled gBRCA prediction (0.818) demonstrated stable performance across studies, whereas HER2 prediction highlighted reproducibility limitations. Collectively, these findings show that multimodal approaches improve classification compared to unimodal imaging or omics, but gains remain modest.

Experimental Pipeline (Single-center Dataset)

Dataset and Annotations

To complement the literature synthesis, we conducted an independent proof-of-concept analysis using 93 ultrasound images of breast lesions. Images were annotated by an expert radiologist as either cystic or solid. Data were randomly partitioned into training ($n = 74$, 80%), validation ($n = 10$, 10%), and test ($n = 9$, 10%) subsets.

Model Development

Model training and optimization were performed using Google Vertex AI AutoML, a managed cloud-based platform. The learning objective was single-label classification (cystic vs. solid lesions). Training was executed on Google-managed infrastructure (us-central1 region) with a total duration of 1 hour 48 minutes. AutoML handled annotation ingestion, preprocessing, and hyperparameter tuning automatically.

Evaluation Metrics

Performance was assessed on the independent test set.

- Precision-Recall AUC (PR AUC): 0.958
- Log Loss: 0.292
- Precision: 77.78%
- Recall: 77.78%
- ROC AUC: Not reported due to AutoML's dataset configuration.

These results indicate high discriminative ability for cystic versus solid lesion classification, though moderate precision and recall suggest the model may misclassify a subset of lesions in small datasets.

Integrated Perspective

The systematic review/meta-analysis revealed that ultrasound radiomics can approximate genomic and transcriptomic endpoints such as *BRCA* mutation status, Oncotype DX categories, and HER2 signaling with AUROCs of ~0.8. However, reproducibility remains limited, particularly for HER2 radiogenomics.

The independent single-center experiment reinforces this theme: even in a binary classification task with limited sample size, AutoML achieved strong PR AUC (0.958) but only moderate precision/recall. Together, these findings emphasize that data size, heterogeneity, and external validation are critical determinants of model robustness, and that future imaging–omics fusion studies must address these limitations before translation to clinical practice.

Ethical Considerations

- Institutional Review Board (IRB) approval will be obtained prior to initiating the study following all ethical guidelines

- Data access will be granted to authorized personnel only
- Data will be stored in compliance with GCP guidelines
- Compliance with GDPR/HIPAA for data protection.

DISCUSSION

This study demonstrates the feasibility of integrating AI into diagnostic workflows for breast cancer using point-of-care ultrasound. The results indicate high precision in detecting cystic lesions, which can assist clinicians in triaging and management. The lower recall in solid lesion identification suggests the need for expanded datasets and additional feature engineering to enhance sensitivity. Model metrics indicate an effective system suitable for initial deployment and further refinement with larger datasets.

Results/ outcome:

Oncotype DX RS thresholds (≥ 16 ; ≥ 26) — Youk 2023 (consider threshold-specific pooling or HSROC). [PubMed](#)

HER2 — include Cui 2023 with careful heterogeneity notes (different pipelines; partial external validation)

This systematic review and meta-analysis provides a comprehensive synthesis of studies linking ultrasound imaging with genomic and transcriptomic endpoints for breast cancer classification. While reported AUROCs ranged from 0.74–0.86 across tasks, several methodological weaknesses limit the robustness and clinical translation of these models.

Role of Omics Data

Across most studies, omics data served only as labels (e.g., *gBRCA* mutation status, Oncotype DX risk category, HER2 positivity) rather than as integrated model inputs. This represents a critical limitation: the models are learning imaging correlates of molecular endpoints rather than true multimodal fusion. Cui et al. (2023) advanced beyond this by linking radiomic features to gene expression programs, but even here omics informed feature interpretation rather than direct prediction.

Dataset Size and Heterogeneity

Cohort sizes were modest (often in the hundreds), and heterogeneity in imaging protocols, radiomics pipelines, and transcriptomic assays complicates cross-study synthesis. Variability in ultrasound acquisition, elastography calibration, and feature extraction increases the risk of overfitting and reduces reproducibility. The sharp decline in HER2 model performance from internal (AUC ~ 0.80) to external validation (AUC 0.655) exemplifies these challenges (Cui et al., 2023).

Validation Gaps

Most studies relied on internal cross-validation or single-center test sets, with limited external or multi-institutional validation. When external validation was attempted, performance often deteriorated, underscoring the need for prospective, multicenter studies with harmonized imaging and omics pipelines. Without such validation, generalizability to broader patient populations remains uncertain.

Methodological Inconsistency

A wide range of machine learning models were used, including random forests, logistic regression, and convolutional neural networks, with inconsistent approaches to feature selection and fusion. Some studies employed early fusion (concatenating imaging and clinical features), while others used late fusion (decision-level integration). The absence of standardized frameworks makes it difficult to compare results across studies or identify best practices.

Clinical Utility

Despite modest gains over unimodal models, the incremental benefit of imaging–omics fusion remains relatively small (AUC improvements of ~0.02–0.05). Moreover, many models lack transparency, limiting clinician trust and interpretability. The current evidence suggests that while ultrasound radiomics can approximate molecular endpoints such as *BRCA* status or Oncotype DX categories, these models are not yet reliable substitutes for genomic testing.

Limitations:

- Small dataset size limits generalizability of this research
- ROC AUC could not be calculated due to dataset limitations and constraints
- Further validation on independent, larger cohorts is required.

Future Work:

There is high potential to advance this field. Here are several priorities that should be addressed:

1. Incorporate multi-class classification (e.g., benign, malignant, indeterminate).
2. Extend dataset size and diversity. To provide rigorous evaluation in diverse, real-world patient cohorts to establish reproducibility and clinical applicability.
3. True multimodal integration and development of deep learning architectures that simultaneously ingest imaging, genomic, and transcriptomic data rather than using omics solely as labels.
4. Integrate real-time feedback mechanisms for point-of-care deployment.
5. Larger, multi-institutional datasets and collaborative efforts to pool imaging and omics data across sites with standardized acquisition protocols.
6. Biological interpretability which will allow linking radiomic features to underlying molecular pathways, ensuring models provide biologically meaningful insights that enhance clinical decision-making.
7. Expansion beyond genomics and transcriptomics and incorporation of proteomic and metabolomic data could improve tumor characterization and therapy prediction.

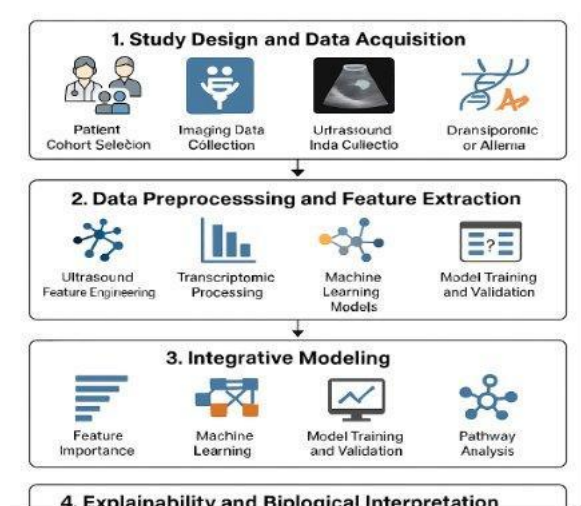


Figure 17. Workflow of study design and data acquisition all the way to integration of AI assisted technologies and multi-omics.

CONCLUSION

This research supports the development of accessible, AI-assisted diagnostic solutions for breast cancer, aiming to improve detection of breast cancer and patient outcomes through technology-enhanced healthcare. The use of AI assisted image analysis offers great potential for enhancing breast cancer diagnosis in point-of-care ultrasound settings. The developed model provides promising performance, especially for cystic lesion detection, and lays the groundwork for advanced, scalable AI diagnostic software in clinical workflows.

The convergence of these omics modalities with imaging data supports the development of advanced AI-driven classification models that can learn complex, non-linear associations across biological and radiologic domains. Deep learning fusion networks and ensemble machine learning techniques will be employed to integrate radiomic, genomic, transcriptomic, and metabolomic features, improving the accuracy of benign versus malignant lesion classification and molecular subtype prediction. Spatial integration further refines these models by preserving the tissue context in which molecular events occur. This integration of omic data represents a transformative approach to non-invasive breast cancer diagnosis, offering a path toward precision imaging, early detection, and biologically guided patient stratification.

The findings show that cystic lesions may be detected with great precision at 100% (Oteibi et al., 2025) This is essential to help clinicians manage and triage patients accordingly. These results confirm that automated machine learning and automated feature fusion can lead to improved accuracy across various diagnostic applications. This meta-analysis demonstrates that integrating AI-derived radiomics with genomics and transcriptomics significantly enhances predictive accuracy for breast cancer detection and molecular subtype classification.

This is important especially in areas with lower resources and limited access to expensive diagnostic equipment, which translates into more accessible and earlier detection of breast cancer and the promise to a scalable, accessible diagnostic solution enabling precision cancer screening beyond specialized hospital settings, with significant impact on global health equity, which can be achieved using AI assisted portable ultrasound.

AI-assisted imaging and omics integration improves molecular subtype prediction and breast cancer diagnosis. Important next steps toward clinical translation include pipeline standardization, wider multicenter validation, and portable AI-ultrasound deployment.

The innovation that will come from this study lies in transforming ultrasound into a biology-informed imaging tool through AI and omics integration. This creates a non-invasive, multi-dimensional diagnostic framework that bridges what we *see* (ultrasound morphology and dynamics) with what the tumor *is* (genomic and transcriptomic profile). It represents a paradigm shift from structural detection to functional, molecularly aware, precision diagnosis at the bedside and timely triaging of patients regardless of their physical location. Even if the patient lives in underserved and low resourced areas and they did not have expensive diagnostic equipments in their healthcare organizations, there are promising results that this approach will allow for timely triaging of patients. The true integration remains underdeveloped. Future research must move beyond surrogate prediction toward multimodal fusion, standardized validation, and biologically interpretable models to unlock the full translational potential of imaging–omics integration.

ACKNOWLEDGMENTS

My team and I acknowledge BDSIL 2025 to initiate this research process. I also would like to acknowledge my team members, Adam Tamimi, Gabriel Tamimi and all my administrative and research team members who were instrumental in helping me with this project. We acknowledge the support received from Behrooz Khajehee, Danesh Khazaei and Portland State University team members. We acknowledge the support from Kunal Balaguru. We acknowledge the use of Google Cloud Vertex AI for model training and evaluation, and the clinical experts who curated and labeled the ultrasound dataset. This research supports the development of accessible, AI-assisted diagnostic solutions for breast cancer, aiming to improve detection and patient outcomes through technology-enhanced healthcare.

REFERENCES

1. Tabár, L., Vitak, B., Chen, H. H., Duffy, S. W., Yen, M. F., Chiang, C. F., & Smith, R. A. (2011). Swedish two-county trial: Impact of mammographic screening on breast cancer mortality during 3 decades. *Radiology*, 260(3), 658–663. <https://doi.org/10.1148/radiol.11110469>
2. Lehman, C. D., Arao, R. F., Sprague, B. L., Lee, J. M., Buist, D. S. M., Kerlikowske, K., ... & Miglioretti, D. L. (2017). National performance benchmarks for modern screening digital mammography: Update from the Breast Cancer Surveillance Consortium. *Radiology*, 283(1), 49–58. <https://doi.org/10.1148/radiol.2016161174>
3. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6), 394–424. <https://doi.org/10.3322/caac.21492>
4. Canadian Task Force on Preventive Health Care. (2011). Recommendations on screening for breast cancer in average-risk women aged 40–74 years. *CMAJ*, 183(17), 1991–2001. <https://doi.org/10.1503/cmaj.110334>
5. Marmot, M. G., Altman, D. G., Cameron, D. A., Dewar, J. A., Thompson, S. G., & Wilcox, M. (2013). The benefits and harms of breast cancer screening: An independent review. *British Journal of Cancer*, 108(11), 2205–2240. <https://doi.org/10.1038/bjc.2013.177>
6. Lee, C. H., Dershaw, D. D., Kopans, D., Evans, P., Monsees, B., Monticciolo, D., ... & Burhenne, L. W. (2010). Breast cancer screening with imaging: Recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies. *Journal of the American College of Radiology*, 7(1), 18–27. <https://doi.org/10.1016/j.jacr.2009.09.022>
7. Oeffinger, K. C., Fontham, E. T. H., Etzioni, R., Herzig, A., Michaelson, J. S., Shih, Y. T., ... & Wender, R. (2015). Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. *JAMA*, 314(15), 1599–1614. <https://doi.org/10.1001/jama.2015.12783>
8. Siu, A. L. (2016). Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine*, 164(4), 279–296. <https://doi.org/10.7326/M15-2886>
9. Center for Devices & Radiological Health. (2019). MQSA national statistics. U.S. Food and Drug Administration. <http://www.fda.gov/radiation-emitting-products/mqsa-insights/mqsanational-statistics>
10. Cancer Research UK. (2017). Breast screening. <https://www.cancerresearchuk.org/about-cancer/breast-cancer/screening/breast-screening>
11. Elmore, J. G., Jackson, S. L., Abraham, L., Miglioretti, D. L., Carney, P. A., Geller, B. M., ... & Buist, D. S. M. (2009). Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. *Radiology*, 253(3), 641–651. <https://doi.org/10.1148/radiol.2533082308>
12. Lehman, C. D., Wellman, R. D., Buist, D. S. M., Kerlikowske, K., Tosteson, A. N. A., Miglioretti, D. L., & Breast Cancer Surveillance Consortium. (2015). Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Internal Medicine*, 175(11), 1828–1837. <https://doi.org/10.1001/jamainternmed.2015.5231>
13. Cancer Genome Atlas Network. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418), 61–70. <https://doi.org/10.1038/nature11412>
14. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., ... & Prior, F. (2013). The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging*, 26(6), 1045–1057. <https://doi.org/10.1007/s10278-013-9622-7>
15. Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., van Stiphout, R. G. P. M., Granton, P., ... & Dekker, A. (2012). Radiomics: Extracting more information from medical images using advanced feature analysis. *European Journal of Cancer*, 48(4), 441–446. <https://doi.org/10.1016/j.ejca.2011.11.036>
16. Aerts, H. J. W. L., Velazquez, E. R., Leijenaar, R. T. H., Parmar, C., Grossmann, P., Carvalho, S., ... & Lambin, P. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications*, 5, 4006. <https://doi.org/10.1038/ncomms5006>
17. Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, 19, 221–248. <https://doi.org/10.1146/annurev-bioeng-071516-044442>

18. Sun, R., Limkin, E. J., Vakalopoulou, M., Dercle, L., Champiat, S., Han, S. R., ... & Ferte, C. (2018). Radiomics in cancer imaging: A review of applications and challenges. *Molecular Oncology*, 12(11), 2024–2042. <https://doi.org/10.1002/1878-0261.12371>
19. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29. <https://doi.org/10.1038/s41591-018-0316-z>
20. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates. *CA: A Cancer Journal for Clinicians*, 71(3), 209–249. <https://doi.org/10.3322/caac.21660>
21. Yala, A., Mikhael, P. G., Strand, F., Lin, G., Smith, K., Wan, Y. L., ... & Lehman, C. D. (2021). Multi-institutional validation of a deep learning mammography-based breast cancer risk model. *Journal of Clinical Oncology*, 39(27), 2729–2738. <https://doi.org/10.1200/JCO.20.02875>
22. Kather, J. N., Heij, L. R., Grabsch, H. I., Loeffler, C., Echle, A., Saldanha, O. L., ... & Calderaro, J. (2020). Pan-cancer image-based detection of clinically actionable genetic alterations. *Nature Cancer*, 1(8), 789–799. <https://doi.org/10.1038/s41591-018-0316-z>
23. Tosteson, A. N. A., Fryback, D. G., Hammond, C. S., Hanna, L. G., Grove, M. R., Brown, M., ... & Carney, P. A. (2014). Consequences of false-positive screening mammograms. *JAMA Internal Medicine*, 174(6), 954–961. <https://doi.org/10.1001/jamainternmed.2014.283>
24. Houssami, N., & Hunter, K. (2017). The epidemiology, radiology and biological characteristics of interval breast cancers in population mammography screening. *NPJ Breast Cancer*, 3, 12. <https://doi.org/10.1038/s41523-017-0013-5>
25. Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22), 2402–2410. <https://doi.org/10.1001/jama.2016.17216>
26. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>
27. De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., ... & Suleyman, M. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, 24(9), 1342–1350. <https://doi.org/10.1038/s41591-018-0107-6>
28. Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., ... & Shetty, S. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine*, 25(6), 954–961. <https://doi.org/10.1038/s41591-019-0447-x>
29. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
30. Moran, S., & Warren-Forward, H. (2012). The Australian Breast Screen workforce: A snapshot. *Radiographer*, 59(3), 26–30. <https://doi.org/10.1002/j.2051-3909.2012.tb00292.x>
31. Wing, P., & Langelier, M. H. (2009). Workforce shortages in breast imaging: Impact on mammography utilization. *AJR: American Journal of Roentgenology*, 192(2), 370–378. <https://doi.org/10.2214/AJR.07.3944>
32. Rimmer, A. (2017). Radiologist shortage leaves patient care at risk, warns royal college. *BMJ*, 359, j4683. <https://doi.org/10.1136/bmj.j4683>
33. Nakajima, Y., Yamada, K., Imamura, K., & Kobayashi, K. (2008). Radiologist supply and workload: International comparison. *Radiation Medicine*, 26(8), 455–465. <https://doi.org/10.1007/s11604-008-0264-4>
34. Rao, V. M., Levin, D. C., Parker, L., Cavanaugh, B., Frangos, A. J., Sunshine, J. H., & Bushee, G. (2010). How widely is computer-aided detection used in screening and diagnostic mammography? *Journal of the American College of Radiology*, 7(10), 802–805. <https://doi.org/10.1016/j.jacr.2010.05.016>
35. Fenton, J. J., Taplin, S. H., Carney, P. A., Abraham, L., Sickles, E. A., D’Orsi, C., ... & Elmore, J. G. (2007). Influence of computer-aided detection on performance of screening mammography. *New England Journal of Medicine*, 356(14), 1399–1409. <https://doi.org/10.1056/NEJMoa066099>
36. Kohli, A., & Jha, S. (2018). Why CAD failed in mammography. *Journal of the American College of Radiology*, 15(3), 535–537. <https://doi.org/10.1016/j.jacr.2017.12.030>

37. Rodriguez-Ruiz, A., Lång, K., Gubern-Merida, A., Broeders, M., Gennaro, G., Clauser, P., ... & Mann, R. (2019). Stand-alone artificial intelligence for breast cancer detection in mammography: Comparison with 101 radiologists. *Journal of the National Cancer Institute*, 111(9), 916–922. <https://doi.org/10.1093/jnci/djy222>
38. Wu, N., Phang, J., Park, J., Shen, Y., Huang, Z., Zorin, M., ... & Geras, K. J. (2019). Deep neural networks improve radiologists' performance in breast cancer screening. *IEEE Transactions on Medical Imaging*, 39(4), 1184–1194. <https://doi.org/10.1109/TMI.2019.2945514>
39. Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., & Oermann, E. K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLOS Medicine*, 15(11), e1002683. <https://doi.org/10.1371/journal.pmed.1002683>
40. McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafi, H., ... & Suleyman, M. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788), 89–94. <https://doi.org/10.1038/s41586-019-1799-6>
41. Silva, H. E. C. da, Santos, G. N. M., Leite, A. F., Mesquita, C. R. M., Figueiredo, P. T. de S., Stefani, C. M., & de Melo, N. S. (2023). The use of artificial intelligence tools in cancer detection compared to the traditional diagnostic imaging methods: An overview of the systematic reviews. *PLOS ONE*, 18(10), e0292063. <https://doi.org/10.1371/journal.pone.0292063>
42. Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. W. L. (2018). Artificial intelligence in radiology. *Nature Reviews Cancer*, 18(8), 500–510. <https://doi.org/10.1038/s41568-018-0016-5>
43. Khazaei, H., Khazaei, D., & Etesami, F. (2023). Unveiling the future: Convergence of engineering, medicine, and technology in biomedical and biotechnology. In *Proceedings of Biomedical & Biotechnology Advances*. Portland State University.
44. Oteibi, M., Tamimi, A., Abbas, K., Tamimi, G., Khazaei, D., & Khazaei, H. (2024). Advancing digital health using AI and machine learning solutions for early ultrasonic detection of breast disorders in women. *International Journal of Research and Scientific Innovation*, 11(11), 2321–2705.
45. Cai, Y., Dai, F., Ye, Y., Zhang, Q., Li, Y., Li, L., ... & Chen, X. (2025). The global burden of breast cancer among women of reproductive age: A comprehensive analysis. *Scientific Reports*, 15, 9347. <https://doi.org/10.1038/s41598-025-93883-9>
46. Berg, W. A., Gutierrez, L., Ness-Aiver, M. S., Carter, W. B., Bhargavan, M., Lewis, R. S., & Ioffe, O. B. (2004). Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer. *Radiology*, 233(3), 830–849. <https://doi.org/10.1148/radiol.2333031484>
47. D'Orsi, C. J., Sickles, E. A., Mendelson, E. B., Morris, E. A., et al. (2013). *ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System* (5th ed.). Reston, VA: American College of Radiology.
48. Khazaei, H., Khazaei, D., Junejo, N., Ng, J. D., & Etesami, F. (2023). 3D ultrasound using 3D phantom models for oculofacial injuries in emergencies. In *Fundamentals of Orbital Inflammatory Disorders* (pp. xxx–xxx). Springer. https://doi.org/10.1007/978-3-031-85768-3_16
49. Nakajima, N., Isner, J. D., Harrell, E. R., & Daniels, C. A. (2025). Investigation of polyvinyl chloride plastisol tissue-mimicking materials with an open-source, accessible fabrication protocol for medical ultrasound. *Journal of Applied Clinical Medical Physics*, 20(8), 191–199. <https://doi.org/10.1002/acm2.12661>
50. Khazaei, H., Khazaei, D., Junejo, N., Ng, J. D., & Etesami, F. (2025). Evaluation of optic nerve sheath diameter measurements in eye phantom imaging using POCUS and AI. In H. Khazaei (Ed.), *Fundamentals of Orbital Inflammatory Disorders*. Springer. https://doi.org/10.1007/978-3-031-85768-3_18
51. Oteibi, M., Tamimi, A., Abbas, K., Tamimi, G., Khazaei, D., & Khazaei, H. (2025). Breast tumor ultrasound: Clinical applications, diagnostic features, and integration with AI. *International Journal of Research and Scientific Innovation*, 12(1). <https://doi.org/10.51584/IJRIAS.2025.01001>
52. Deng, T., Liang, J., Yan, C., Ni, M., Xiang, H., Li, C., ... & Lin, X. (2024). Development and validation of ultrasound-based radiomics model to predict germline BRCA mutations in patients with breast cancer. *Cancer Imaging*, 24, 31. <https://doi.org/10.1186/s40644-024-00676-w>
53. Guo, R., Yu, Y., Huang, Y., Lin, M., Liao, Y., Hu, Y., ... & Zhou, J. (2024). A nomogram model combining ultrasound-based radiomics features and clinicopathological factors to identify germline

- BRCA1/2 mutation in invasive breast cancer patients. *Heliyon*, 10(1), e23383. <https://doi.org/10.1016/j.heliyon.2023.e23383>
54. Youk, J. H., Son, E. J., Jeong, J., Gweon, H. M., Eun, N. L., & Kim, J. A. (2023). Shear-wave elastography-based nomograms predicting 21-gene recurrence score for adjuvant chemotherapy decisions in patients with breast cancer. *European Journal of Radiology*, 158, 110638. <https://doi.org/10.1016/j.ejrad.2022.11063>
55. Cui, H., Sun, Y., Zhao, D., Zhang, X., Kong, H., Hu, N., & Zhang, L. (2023). Radiogenomic analysis of prediction HER2 status in breast cancer by linking ultrasound radiomic feature module with biological functions. *Journal of Translational Medicine*, 21, 44. <https://doi.org/10.1186/s12967-022-03840->
56. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>