KMANPUB | Health Nexus Vol. 4 No. 3 (2026): 1-19

Article history:
Received 14 February 2026
Revised 15 June 2026
Accepted 23 June 2026
Published online 01 July 2026

Latency-Aware Breast Cancer Detection in Mammography Images Using a Fog-Cloud Framework Based on Stacked Transfer Learning and PSO-Optimized XGBoost

Zahraa Abdulmajeed Ibrahim Al-Mohammed[1], Esmaeil Bagheri[2*], Ameer Hussein Mohammed Ali[3], Mehdi Hamidkhani[4]

1. PhD Student, Department of Computer Engineering, Isf.C., Islamic Azad University, Isfahan, Iran
2. Department of Engineering, Deh.C., Islamic Azad University, Isfahan, Iran
3. Al-Najaf Technical Institute, Al-Furat Al-Awsat Technical University, Najaf, Iraq
4. Department of Electrical Engineering, Isf.C., Islamic Azad University, Isfahan, Iran

*** Corresponding author email address:** bagheri@iau.ac.ir

Article Info

Article type:
Original Research

How to cite this article:
Al-Mohammed, Z. A. I., Bagheri, E., Ali, A. H. M., & Hamidkhani, M. (2026). Latency-Aware Breast Cancer Detection in Mammography Images Using a Fog-Cloud Framework Based on Stacked Transfer Learning and PSO-Optimized XGBoost. *Health Nexus, 4*(3), 1-19.
https://doi.org/10.61838/kman.hn.5698

ABSTRACT

Breast cancer is one of the most important public health challenges worldwide, and the quality of its early detection is directly associated with reduced mortality, lower treatment costs, and improved care pathways. Mammography remains one of the principal screening tools; however, the interpretation of mammographic images is affected by human error and infrastructural limitations in the presence of tissue density, noise, similarities between benign and malignant lesions, and variations in imaging protocols. This paper presents an automated breast cancer detection framework focused on reducing processing latency-aware deployment and increasing diagnostic reliability within a fog-cloud architecture. At the fog layer, mammographic images are standardized through resizing, intensity normalization, noise reduction, and lightweight preparation to reduce data-transmission volume and initial response time. At the cloud layer, three transfer-learning models—ResNet50, DenseNet121, and EfficientNetB7—are employed to extract deep features. The probabilistic outputs of these models are integrated in a three-level stacked ensemble architecture, with XGBoost used as the final meta-classifier. To reduce dependence on manual tuning and control overfitting, the key hyperparameters of the base models and the meta-classifier are optimized using particle swarm optimization. Evaluation was conducted on the public CBIS-DDSM dataset using stratified cross-validation. The results showed that the proposed model achieved an accuracy of 97.5% on the test subset, balanced performance across the two classes, a recall of 98.0% for the benign class, and 96.9% for the malignant class. The findings indicate that combining near-source preprocessing, transfer-based feature extraction, stacked ensembling, and metaheuristic optimization can provide a promising decision-support framework for distributed healthcare environments; however, clinical deployment requires external validation, explicit latency benchmarking, and prospective assessment.

Keywords: *breast cancer, mammography, fog computing, transfer learning, stacked ensemble, XGBoost, particle swarm optimization, CBIS-DDSM*

1. Introduction

Breast cancer, as one of the most common cancers among women, continues to impose a substantial clinical and economic burden on healthcare systems. A World Health Organization report indicates that breast cancer caused approximately 670,000 deaths worldwide in 2022 and was the most common cancer among women in most countries (1). The significance of the problem is not limited to its high prevalence; delayed diagnosis can also make the treatment pathway more invasive, more costly, and less favorable in terms of clinical outcomes. Therefore, the development of accurate, rapid, and deployable screening tools for diverse healthcare settings constitutes both a technological and clinical need.

Mammography is one of the most widely used breast cancer screening tools and serves as the starting point for clinical decision-making in many diagnostic programs. Nevertheless, mammogram interpretation is prone to error because of limited tissue contrast, overlapping anatomical structures, variations in breast tissue density, and the presence of small or atypical lesions. Studies of artificial intelligence in mammography screening have shown that false-positive and false-negative findings remain part of the diagnostic challenge even in structured screening systems (2). This situation highlights the need for computer-aided detection and diagnosis systems whose role is not to replace physicians, but to reduce the decision-making burden and improve the consistency of assessment.

Deep learning, particularly convolutional neural networks, has assumed an important role in medical image analysis in recent years. The principal advantage of these networks is their ability to extract hierarchical image representations automatically, without relying on manual feature engineering. However, the application of deep learning in medicine faces serious limitations, including shortages of labeled data, training costs, the risk of overfitting, weak generalizability to out-of-domain data, and the difficulty of deployment in healthcare centers lacking adequate computational infrastructure. Transfer learning mitigates some of these limitations by transferring knowledge from pretrained networks, but the use of a single transfer-learning model is generally insufficient to achieve stable performance (3).

On the other hand, cloud-only architectures are not always suitable for real-time medical systems. Transferring large volumes of medical images to a central cloud can consume bandwidth, increase response latency, and heighten data-privacy concerns. By moving part of the processing closer to the data source, fog computing operates between imaging devices and the cloud and can perform lightweight operations—such as normalization, noise reduction, and input preparation—closer to the patient (4). Therefore, combining fog and cloud computing for automated breast cancer detection can provide a practical trade-off between the accuracy enabled by cloud computation and the speed of edge processing.

To address this gap, this paper presents a hybrid framework for classifying mammographic images into benign and malignant classes. The novelty of the study lies in its simultaneous focus on four dimensions: lightweight preprocessing at the fog layer, deep feature extraction using three complementary transfer-learning architectures, decision fusion through a stacked ensemble, and hyperparameter tuning using PSO.

2. Theoretical Foundations and Review of Related Research

Research on mammogram analysis can be categorized into three main streams: handcrafted-feature methods, end-to-end deep-learning methods, and hybrid frameworks. In classical methods, an image is first enhanced using spatial- or frequency-domain filters; features such as texture, shape, brightness intensity, and boundary irregularity are then extracted and supplied to classifiers such as support vector machines, random forests, or decision trees. These methods offer the advantage of simplicity for small datasets, but they depend on segmentation quality and feature-engineering expertise and may fail to capture hidden tissue patterns completely.

Deep networks have reduced this limitation by learning features directly from images. By introducing residual connections, ResNet stabilized the training of deeper networks and alleviated the vanishing-gradient problem (5). DenseNet strengthened feature reuse and gradient flow through densely connected layers (6). EfficientNet sought to improve the accuracy-to-computational-cost ratio by jointly scaling network depth, width, and image resolution (3).

Health Nexus
E-ISSN: 2981-2569

These three architecture families are complementary in their design logic, and combining them can increase the diversity of feature representations.

Transfer learning is particularly important in medical imaging because of limited data availability. In this approach, weights learned from large public datasets are used as a starting point, and the final layers or a subset of the network are fine-tuned for the medical domain. Mammography studies have shown that transfer-learning models, when combined with preprocessing and data augmentation, can be more stable than complete training from scratch (7, 8). Nevertheless, the domain gap between natural and medical images can limit knowledge transfer; therefore, careful evaluation and the use of multiple complementary models are necessary.

Recent breast cancer detection research has moved toward hybrid models. Atrey et al. (2023) used a multimodal mammography-and-ultrasound approach for classification (9). Cruz-Ramos et al. (2023) investigated the fusion of deep and handcrafted features to distinguish benign from malignant tumors (10). Khourdifi et al. (2024) reported a deep-learning-based ensemble approach for early breast cancer detection from mammograms (11). Maruf et al. (2025) also proposed the integration of radiomics and transfer learning to improve diagnosis (12). A common conclusion across these studies is that a single model is insufficient for a problem characterized by high imaging heterogeneity.

Compared with simple voting, a stacked ensemble can learn the relationships among the predictions of the base models. In this structure, the outputs of the base models are passed as probability vectors or intermediate labels to classifiers at the next level, and the final meta-classifier uses the models' error patterns for decision-making. Because it employs gradient boosting, built-in regularization, heterogeneous-data handling, and computationally efficient implementation, XGBoost is a suitable option for the final level of such a structure (13).

In addition to model selection, hyperparameter optimization is a decisive factor. Manual adjustment of the learning rate, batch size, dropout rate, tree depth, and base-model weights can produce unstable results. PSO, as a population-based metaheuristic algorithm, explores a multidimensional search space without requiring direct gradients and is appropriate for problems in which objective-function evaluation is costly but computable (14). In this study, PSO was used to reduce dependence on manual trial and error and to identify a stable parameter combination.

Fog computing, as a layer between devices and the cloud, also addresses the problem of practical deployment. In medical systems, response speed and privacy protection cannot be achieved merely by increasing model accuracy. If a model is accurate but every image must be transferred in full to the cloud and incurs a long response time, its use in high-throughput screening will be limited. Accordingly, the present study integrates a fog-cloud architecture with the modeling pipeline so that both accuracy and operational feasibility are considered.

In recent studies, Murty et al. (2024) achieved an accuracy of 97.50% by combining WBCD and CBIS-DDSM, while Sandhu et al. (2024) reported an accuracy of 95.60% using a CNN-based model on CBIS-DDSM (15, 16).

Table 1

Position of the proposed method relative to selected approaches

Study	Data	Method	Key point
(2)	Multi-source mammography	AI screening system	Reduced reading errors compared with human readers
(9)	Mammography/ultrasound	Hybrid deep learning	Use of complementary information from two modalities
(11)	Mammograms	Deep-learning ensemble	Focus on early detection
(12)	CBIS-DDSM	Radiomics + transfer learning	Fusion of multi-source features
Present method	CBIS-DDSM	Fog-cloud + TL + stacking + PSO	Combination of accuracy and low-latency deployability

Health Nexus
E-ISSN: 2981-2569

3. Materials, Data, and Implementation Protocol of the Proposed Method

This section was developed to make the study reproducible. The implementation procedure includes data selection, image preparation at the fog layer, data splitting, deep feature extraction, training of the stacked structure, hyperparameter tuning, and final evaluation on unseen data. All stages were designed to maintain a clear separation among the training, validation, and test data and to avoid an overly optimistic estimate of model performance.

3.1. Dataset and Unit of Analysis

The dataset used in this study is CBIS-DDSM, a curated and standardized version of the DDSM database for computer-aided detection and diagnosis research in mammography. It contains mammographic images, regions of interest, bounding boxes, and pathological diagnoses. Because it provides valid labels for distinguishing benign and malignant lesions, it is suitable for evaluating CAD/CADe frameworks (17). In the present study, the decision unit is the binary classification of each image/sample into benign or malignant. Following initial quality control, each sample entered the preprocessing pipeline and was then assigned to one of the cross-validation folds.

The exact number of images retained after quality control, the benign/malignant distribution before and after exclusions, and whether classification was performed on full mammograms, cropped ROIs, or both should be explicitly reported. These details are necessary for reproducibility and for fair comparison with previous CBIS-DDSM studies.

During initial data quality control, samples with inconsistencies in format, brightness intensity, image readability, or labeling had to be identified before training, because a deep model can learn patterns even from artifacts or labeling errors. The objective of this stage was not to remove difficult cases, but to prevent unreliable data from entering the training process. Such control is important in medical imaging research because input-data quality directly affects false-positive and false-negative rates.

Table 2

Data characteristics and evaluation protocol used in the study

Component	Methodological description	Purpose in the study
Data	CBIS-DDSM with pathological labels and ROI information	Evaluation of benign/malignant diagnosis
Unit of analysis	Preprocessed mammographic sample	Input to the transfer-learning models
Data splitting	Stratified five-fold cross-validation	Preservation of class proportions and reduction of bias
Final test	Unseen subset in each fold	Reporting defensible performance

3.2. Preprocessing at the Fog Layer

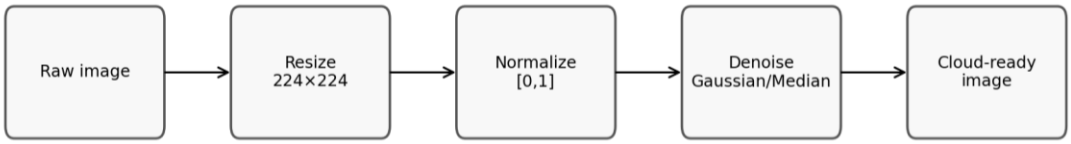
In the proposed architecture, preprocessing is performed at the fog layer to standardize image dimensions, intensity range, and noise before the images enter computationally intensive cloud models. First, all images were resized to 224 × 224 pixels to make their inputs compatible with ResNet50 and DenseNet121. The use of a higher resolution for EfficientNetB7 was examined during tuning; however, the base image size was kept identical across the model branches to control computational cost and permit a fair comparison.

After resizing, pixel intensities were mapped to the [0, 1] range. This operation reduces differences caused by imaging devices and heterogeneous intensity ranges and stabilizes model convergence. During noise reduction, a combination of Gaussian and median filters was used. The Gaussian filter smooths scattered noise, while the median filter is useful for reducing outliers and damages lesion boundaries less than simple averaging filters. The output of this stage is an image that retains the principal diagnostic details but is more stable for subsequent transmission and processing.

Health Nexus
E-ISSN: 2981-2569

Figure 1

Preprocessing chain at the fog layer



Algorithm 1: Preprocessing	
Input	Raw mammogram image I
1	Begin
2	// Step 1: Resizing
3	I_resized ← Resize(I, 224, 224)
4	// Step 2: Normalization
5	I_norm ← I_resized / 255
6	
7	// Step 3: Noise Reduction
8	I_denoise ← GaussianFilter(I_norm)
9	
10	// Step 4: Transmission to cloud
11	Prepare cloud-ready standardized image
12	Attach anonymized sample identifier and metadata required for inference
13	Transmit I_denoise to the cloud layer
14	Store only preprocessing log at the fog node
15	// No ensemble decision is generated at the fog layer
16	
17	Return preprocessing status and wait for cloud inference output
18	End
Output	Preprocessed image I_p and preprocessing status log; final decision is generated in the cloud layer

3.3. Data Splitting and Prevention of Data Leakage

The data were divided using stratified five-fold cross-validation. This method preserves the proportion of benign and malignant samples in each fold so that the model encounters a relatively similar class distribution during each training run. In each fold, the training data were used to optimize the weights, and the validation data were used to select the configuration. The test data remained unseen until final performance reporting.

For methodological clarity, the reported 198-sample test subset should be treated as the held-out test portion used for final reporting, not as an independent external validation cohort. If five-fold cross-validation is retained, fold-wise mean ± standard deviation for all metrics should also be reported to avoid an overly optimistic single-split interpretation.

The risk of data leakage is more serious in stacked structures than in single models. If the meta-classifier receives the outputs of the base models for the same samples on which those base models were trained, it may learn overfitted patterns. To reduce this risk, out-of-fold outputs were used to train the second and final levels. Consequently, the meta-classifier was trained on predictions that functioned as new data from its perspective, rather than data memorized by a base model.

In the corrected stacking protocol, each meta-classifier level must be trained only on out-of-fold predictions generated from models that have not seen the corresponding samples during training. Predictions from the same samples used to train the base models must not be used directly for training the higher-level classifiers.

3.4. Evaluation Metrics and Rationale for Reporting the Results

In breast cancer detection, relying on overall accuracy alone is insufficient because a false-negative error in the malignant class has more serious clinical consequences than many false-positive errors. Therefore, in addition to accuracy, precision, recall, and F1-score were reported for each class. Precision indicates the proportion of samples

Health Nexus
E-ISSN: 2981-2569

assigned by the model to a given class that were classified correctly; recall indicates the proportion of actual samples in a class that were identified; and F1 is the harmonic mean of precision and recall.

Table 3

Evaluation metrics used in the performance analysis

Metric	Formula	Importance in breast cancer detection
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Overall view of performance, but insufficient for imbalanced data
Precision	$\frac{TP}{TP+FP}$	Control of false-positive alerts
Recall	$\frac{TP}{TP+FN}$	Reduction of missed cases, particularly in the malignant class
F1-score	$\frac{2*(Precision*Recall)}{Precision+Recall}$	Balance between precision and recall

To reduce overfitting, transferred initial weights, restrictions on trainable layers, dropout, and early stopping were used. Transferred initial weights allow the model to begin with general representations learned from large datasets, avoiding the need for full training from scratch on limited medical data. Restricting the number of trainable layers prevents the destruction of prior knowledge and the learning of random patterns. Dropout also reduces the model's excessive dependence on specific pathways by randomly deactivating a proportion of units.

At the XGBoost level, overfitting was controlled by limiting tree depth, reducing the learning rate, using row and column subsampling, and applying L1 and L2 regularization. These decisions are consistent with the nature of the XGBoost input data, because the final-level input comprises the outputs of previous models and therefore consists of compact but potentially correlated features. Under such conditions, deep unregularized trees can rapidly become dependent on patterns specific to the training data.

3.5. Quality Control, Standardization, and Management of Data Heterogeneity

In mammography studies, data heterogeneity is not limited to image size. Differences in imaging devices, compression conditions, radiation intensity, breast tissue density, lesion type, and label quality can cause a model to learn incidental data cues rather than the true lesion pattern. Therefore, before training, the data should be examined for label completeness, image readability, absence of severe artifacts, and format consistency. This stage is reported as part of the proposed method so that reviewers can follow the data-preparation pathway.

In addition to facilitating convergence, pixel-intensity standardization prevents transfer-learning networks from encountering an unexpected input range. Pretrained models are generally trained on data with a specific intensity range and statistical distribution. Although medical images are not identical to natural images, normalization and control of the input range help prevent unstable gradients and excessive weighting of overall image brightness.

During noise reduction, the strategic principle is to preserve diagnostic information. Excessive filtering may blur microcalcifications, subtle mass boundaries, or important tissue patterns. Preprocessing should therefore be minimal but sufficient: noise and artifacts should be reduced without removing diagnostic signs.

Health Nexus
E-ISSN: 2981-2569

Table 4

Data-quality threats and the methodological controls applied

Source of heterogeneity	Potential effect on the model	Methodological control
Differences in image resolution	Incompatible network inputs	Standardized resizing
Differences in brightness intensity	Unstable convergence and intensity bias	Normalization to the [0, 1] range
Noise and artifacts	Increased false positives or loss of lesion boundaries	Controlled-strength Gaussian and median filtering
Class imbalance	Apparently high accuracy and low recall in the minority class	Stratified splitting and reporting of class-specific metrics
Data leakage	Overly optimistic performance estimation	Out-of-fold predictions in stacking

3.6. Formulation of the Optimization Objective Function

The PSO objective function should be consistent with the nature of the medical problem. If only accuracy is maximized, a model may be selected that performs well on the majority class but fails to identify a sufficient number of malignant cases. Therefore, accuracy, precision, recall, and F1 were considered simultaneously when defining the fitness function. In screening applications, assigning a greater weight to recall for the malignant class is clinically defensible; however, this must be accompanied by precision control to prevent an unacceptable increase in unnecessary referrals.

The general objective function was defined as Score = αAccuracy + βPrecision + γRecall + δF1. Unless a clinically weighted configuration is explicitly selected, the balanced research configuration should use $\alpha = \beta = \gamma = \delta = 0.25$. If the system is used for screening, γ may be increased for malignant-class recall, but this must be reported transparently together with the resulting false-positive rate.

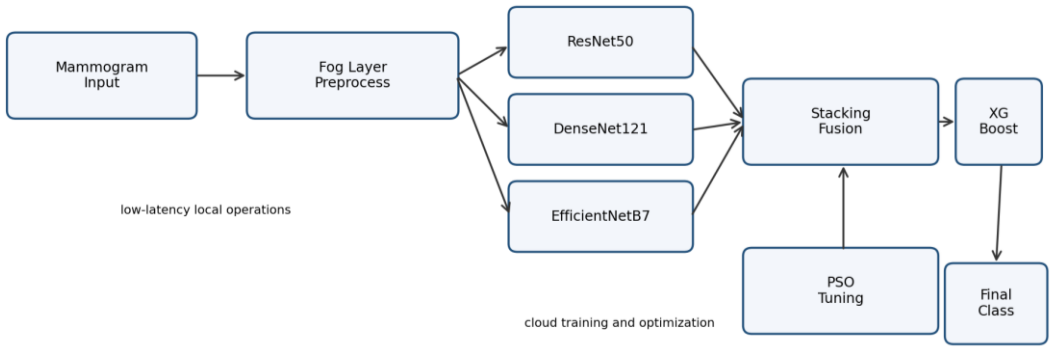
4. Proposed Method: Fog-Cloud Architecture and Stacked Ensemble

The proposed method is a distributed diagnostic pipeline extending from the data-generation source to the final decision. This pipeline is not merely a classifier; it is a fog-cloud architecture that performs lightweight, latency-sensitive, near-data operations in the fog and transfers computationally intensive processing, model training, and final decision-making to the cloud. Figure 2 presents an overview of the architecture.

To resolve the latency claim, the proposed framework should be interpreted as latency-aware rather than fully validated as low-latency. End-to-end latency must be benchmarked against a cloud-only baseline, including image-transfer time, fog preprocessing time, cloud inference time, and response-return time.

Figure 2

Overall proposed fog-cloud architecture for breast cancer detection



Health Nexus
E-ISSN: 2981-2569

4.1. *Fog Layer: Local Preparation and Reduction of Communication Load*

At the fog layer, the raw mammographic image enters the standardization stage after being received from the imaging device or PACS. This layer does not make the final decision because executing all three deep transfer-learning networks on a fog node may be costly in terms of memory and processing power. The main role of the fog is to reduce image heterogeneity, remove initial noise, generate a standardized input, and reduce the volume of data to be transmitted. This separation allows the architecture to exploit the proximity of the fog to the data source without allowing the fog node's hardware limitations to become the primary bottleneck.

4.2. *Cloud Layer: Transfer-Based Feature Extraction*

At the cloud layer, ResNet50, DenseNet121, and EfficientNetB7 were used as the three transfer-learning base learners. These architectures were selected because their learning mechanisms are complementary. ResNet50 uses residual connections to enable the training of deeper networks and improve gradient flow (5). DenseNet121, through dense connectivity, increases feature reuse and can produce more compact tissue representations (6). EfficientNetB7 uses compound scaling of depth, width, and resolution to achieve an appropriate balance between accuracy and computational cost (3).

The output of each base model was extracted as a two-class probability vector. For each image, ResNet50, DenseNet121, and EfficientNetB7 each generate two probabilities corresponding to the benign and malignant classes. These three outputs are converted into a six-dimensional vector. Using probabilities rather than only hard labels preserves information about the uncertainty of the base models and allows the second level to distinguish clear cases from borderline cases.

Table 5

Rationale for selecting the transfer-learning base models

Base model	Key mechanism	Role in the proposed method
ResNet50	Residual learning	Capture of deep features and mitigation of the vanishing-gradient problem
DenseNet121	Dense connections	Feature reuse and more compact representation of tissue patterns
EfficientNetB7	Compound scaling	Efficient use of network capacity for multiscale patterns

4.3. *Three-Level Stacked Ensemble*

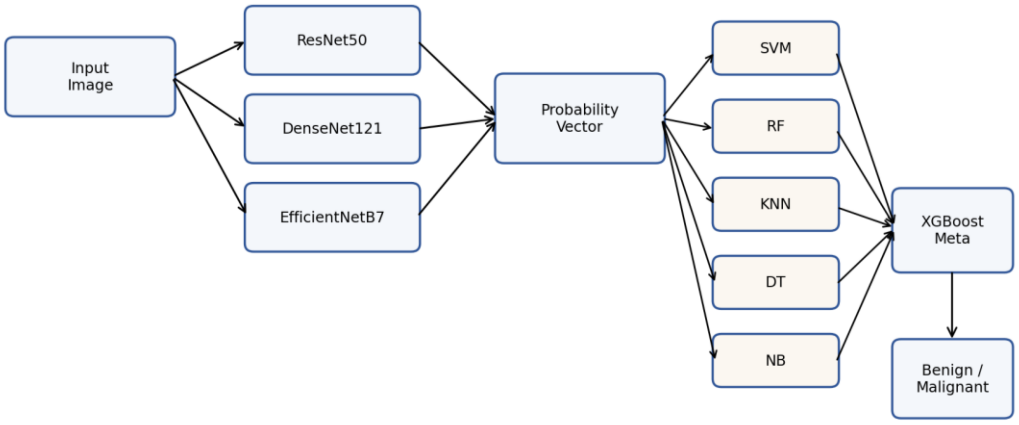
After the base-model outputs are extracted, the six-dimensional vector enters the second level. At this level, five machine-learning classifiers are used: SVM, random forest, KNN, decision tree, and naive Bayes. Multiple classifiers are used because each constructs a different decision boundary: SVM creates a margin-based boundary; KNN makes neighborhood-based decisions; random forest produces decisions through an ensemble of trees; decision tree generates hierarchical rules; and naive Bayes makes probabilistic decisions under the conditional-independence assumption. Thus, rather than repeating a single decision type, the second level provides several distinct perspectives on the outputs of the deep networks.

At the third level, the second-level outputs are provided to XGBoost. XGBoost does not perform simple voting; rather, it learns a pattern of trust in the intermediate classifiers. For example, when the outputs of the base networks are close, an intermediate classifier may behave more conservatively; however, if one base model produces a very high probability of malignancy, XGBoost can shift the decision weight toward that pattern. In this manner, the stacked structure moves beyond a mechanical combination of votes and models the relationships among intermediate decisions.

Health Nexus
E-ISSN: 2981-2569

Figure 3

Three-level stacked ensemble structure in the proposed method



4.4. Hyperparameter Optimization Using PSO

The PSO algorithm was used to search for an appropriate combination of hyperparameters. Each particle represents a candidate solution comprising the learning rate, batch size, dropout rate, number of trainable layers, and contribution weights of the base models. At each iteration, the performance of every particle is evaluated using the fitness function. The particles' velocities and positions are then updated according to their personal best and the global best experience so that the population moves toward promising regions of the search space (14).

For reproducibility, PSO should be reported with the number of particles, number of iterations, inertia weight, cognitive and social coefficients, random seed, stopping criterion, and the exact validation set used for fitness evaluation. Without these details, the PSO-optimized claim remains only partially reproducible.

The fitness function was defined as a combination of accuracy, precision, recall, and F1 so that the model would not be selected solely by maximizing overall accuracy. In medical applications, the weight assigned to recall for the malignant class is clinically important because a false negative can delay diagnosis. However, excessive weighting of recall can also increase false positives. The objective function must therefore balance diagnostic sensitivity against the control of false alarms.

Table 6

Principal search space for hyperparameter tuning

Parameter	Range/values examined	Expected effect
learning_rate	1e-5 to 1e-3 for deep models; 0.05 for XGBoost	Control of convergence speed and overfitting risk
batch_size	16, 32, 64	Balance between gradient stability and memory consumption
Dropout	0.2 to 0.5	Reduction of model dependence on specific pathways
max_depth	Intermediate values up to a depth of 6	Control of XGBoost tree complexity
ensemble weights	Normalized weights of the base models	Adjustment of the relative contribution of each transfer-learning branch

4.5. Rationale for Selecting XGBoost as the Meta-Classifier

XGBoost was selected because of its ability to model nonlinear relationships, support for regularization, computational efficiency, and strong performance on heterogeneous tabular data (13). In the proposed method, the final-level input is no longer a raw image; it consists of the compressed outputs of previous models. Such data are suitable for a boosted-tree meta-classifier because XGBoost

Health Nexus
E-ISSN: 2981-2569

can model interactions among the second-level outputs without requiring a new deep neural architecture.

The final XGBoost configuration comprised 300 trees, a learning rate of 0.05, a maximum depth of 6, min_child_weight = 3, subsample = 0.8, colsample_bytree = 0.7, gamma = 0.1, L1 regularization = 0.05, and L2 regularization = 1.0. This configuration is more conservative than using deeper trees and a higher learning rate and was selected to control overfitting.

Table 7

Final configuration of the XGBoost meta-classifier

XGBoost hyperparameter	Final value	Function
n_estimators	300	Number of boosting stages
learning_rate	0.05	Control of each tree's contribution
max_depth	6	Control of tree-pattern complexity
min_child_weight	3	Prevention of unstable splits
subsample	0.8	Reduction of correlation among trees
colsample_bytree	0.7	Column subsampling of features
gamma	0.1	Pruning of low-gain splits
reg_alpha / reg_lambda	0.05 / 1.0	L1 and L2 regularization

4.6. Data Flow in the Fog-Cloud Architecture

Data flow in the proposed architecture begins at the point of image generation. The raw image is first recorded at the imaging center or in the PACS and is then transferred to the local fog node. Low-cost and time-sensitive operations are performed at this node. After standardization, the prepared data or initial features are transmitted to the cloud. The cloud is responsible for executing the transfer-learning models, performing stacked fusion, periodically tuning the models, and storing the outputs. Finally, the model response is returned to the fog node or decision-support system.

This separation has a practical advantage. If the complete raw image is sent to the cloud without processing, bandwidth consumption and transmission time increase. If the entire deep model is executed in the fog, the fog node requires a powerful GPU and substantial memory. The proposed architecture lies between these two extremes: the fog is used for preparation and overhead reduction, whereas the cloud performs deep processing and final decision-making.

In multicenter hospitals, several fog nodes can be connected to a common cloud infrastructure. Each imaging center standardizes its data locally and sends a compatible output to the cloud. This design improves scalability because an increase in the number of centers does not necessarily require direct transfer of all raw data to a single central point. Model updates can also be performed centrally in the cloud, and the new version can be deployed across all centers.

Table 8

Data flow and execution location of each stage in the proposed architecture

Data-flow stage	Execution location	Stage output	Design rationale
Image acquisition	Device/PACS	Raw mammographic image	Clinical data source
Standardization	Fog node	Resized and normalized image	Reduction of heterogeneity and latency
Feature extraction	Cloud	Probability outputs of three transfer-learning models	Deep processing with more powerful resources
Intermediate decision	Cloud	Second-level output	Management of ambiguity in the base-model outputs
Final decision	Cloud and return to fog	Benign/malignant label and probability	Usable by a clinical decision-support system

Health Nexus
E-ISSN: 2981-2569

4.5. Probabilistic Output, Decision Threshold, and Management of Borderline Cases

The final system output should not consist solely of a hard label. In medical applications, the predicted probability is also important because two samples labeled malignant may have different confidence levels. A sample with a malignancy probability of 0.95 should receive higher review priority, whereas a sample with a probability of 0.52 should be interpreted as borderline.

The thresholding strategy can vary according to the application. In initial screening, reducing false negatives is the priority; therefore, the threshold for the malignant class can be lowered so that more suspicious cases are referred for human review. In confirmatory applications, controlling false positives is more important, and a more conservative threshold may be selected. This flexibility keeps the model in a decision-support role rather than presenting it as a definitive replacement for physician judgment.

Borderline cases should be identified in the system output. If the difference between the probabilities of the two classes is small, the system can produce a 'review required' output. From a risk-management perspective, this design is preferable to forcing the model to make a definitive decision for every sample. In breast cancer in particular, insisting on a hard decision for ambiguous samples may increase the rate of high-risk errors.

4.6. Runtime Inference Algorithm

After training is completed and the final model is stored, inference for a new image follows a shorter pathway. PSO is not executed at this stage, and the weights are not changed. The new image is first resized, normalized, and denoised at the fog node. The standardized image is then transmitted to the cloud layer. The three transfer-learning models generate probability outputs, the outputs are concatenated, the second-level classifiers produce intermediate decisions, and XGBoost provides the final label and probability.

This separation between training and inference has practical importance. If PSO or retraining were performed for every inference, the system would not be usable in a clinical environment. The computationally intensive cost therefore remains in the offline stage, while the online response time is reduced.

The inference algorithm should also manage low-confidence outputs. If the probability of the malignant or benign class is close to the threshold, the system should not provide only a hard label. In such cases, a 'review required' output can reduce the risk of an incorrect decision. This strategy is consistent with the system's decision-support role and avoids claiming that it can completely replace the physician.

Table 9

Sequence of online inference in the proposed framework

Inference step	Operation	Output
1	Receive the raw image at the fog layer	Input image
2	Resize, normalize, and reduce noise	Standardized image
3	Execute ResNet50, DenseNet121, and EfficientNetB7	Three probability vectors
4	Concatenate the outputs	Six-dimensional vector
5	Execute the second-level classifiers	Intermediate decisions
6	Execute XGBoost	Final label and probability
7	Apply threshold control and send the response	Report usable by the physician

5. Experimental Results and Performance Analysis

The experimental evaluation was conducted on an unseen test subset containing 102 benign and 96 malignant samples. For each sample, the final model output was compared with the ground-truth label, and the confusion matrix, classification metrics, and error analysis were reported. Because of the medical nature of the problem, result interpretation was not limited to overall accuracy; the error rate for the malignant class was analyzed separately.

Health Nexus
E-ISSN: 2981-2569

5.1. *Confusion Matrix and Calculation of Overall Accuracy*

The confusion matrix showed that, of the 102 benign samples, 100 were correctly classified as benign and 2 were incorrectly predicted as malignant. Of the 96 malignant samples, 93 were correctly classified as malignant and 3 were incorrectly reported as benign. Accordingly, the total number of correct diagnoses was 193 out of 198 samples, yielding an overall test accuracy of 97.5%.

Table 10

Numerical confusion matrix for the test set

Actual class / predicted class	Benign	Malignant	Total
Benign	100	2	102
Malignant	3	93	96
Total	103	95	198

5.2. *Class-Specific Classification Metrics*

For the benign class, precision was 97.1%, recall was 98.0%, and F1-score was 97.5%. For the malignant class, precision was 97.2%, recall was 96.9%, and F1-score was 97.0%. The similar precision values for the two classes indicate that the model is not strongly biased toward either class. However, the slight difference in recall is clinically important because lower recall for the malignant class corresponds to a greater number of missed malignant cases.

Table 11

Final performance of the proposed model by class

Class	Precision	Recall	F1-score	Interpretation
Benign	97.1%	98.0%	97.5%	Limited false positives and stable identification of benign cases
Malignant	97.2%	96.9%	97.0%	High sensitivity, but the three false negatives require threshold control
Mean	97.15%	97.45%	97.25%	Balanced performance across the two classes

From a clinical perspective, the two benign-to-malignant errors would probably lead to additional examination or unnecessary referral; however, the three malignant-to-benign errors carry greater risk because they could delay diagnosis and treatment initiation. Therefore, if the system is used for initial screening, the decision threshold for the malignant class should be adjusted to be slightly more sensitive, or borderline outputs should be marked as 're-evaluation required.' This interpretation allows the paper to clarify the consequences of the errors rather than merely reporting numerical values.

5.3. *Comparison of Training and Test Performance*

Training accuracy was reported as 100%, whereas test accuracy was 97.5%. The gap between training and test performance indicates that the model has high capacity and that overfitting control must be interpreted carefully. The model performed almost without error on the training data but showed a limited decline on unseen data. This pattern requires caution: on the one hand, the test decline is not severe and overall performance remains acceptable; on the other hand, 100% training accuracy may indicate excessive model capacity and a risk of overfitting. Reporting performance on an unseen test set and using out-of-fold outputs for the meta-classifier were therefore essential.

The 100% training accuracy should be interpreted as a warning sign for possible residual overfitting. Therefore, the reported test performance should be presented as preliminary until validated on an external dataset or a prospectively collected multicenter cohort.

Health Nexus
E-ISSN: 2981-2569

Table 12

Comparison of training and test performance

Data	Accuracy	Benign precision	Benign recall	Benign F1	Malignant precision	Malignant recall	Malignant F1
Training	100%	99.5%	100%	99.7%	99.6%	99.3%	99.4%
Test	97.5%	97.1%	98.0%	97.5%	97.2%	96.9%	97.0%

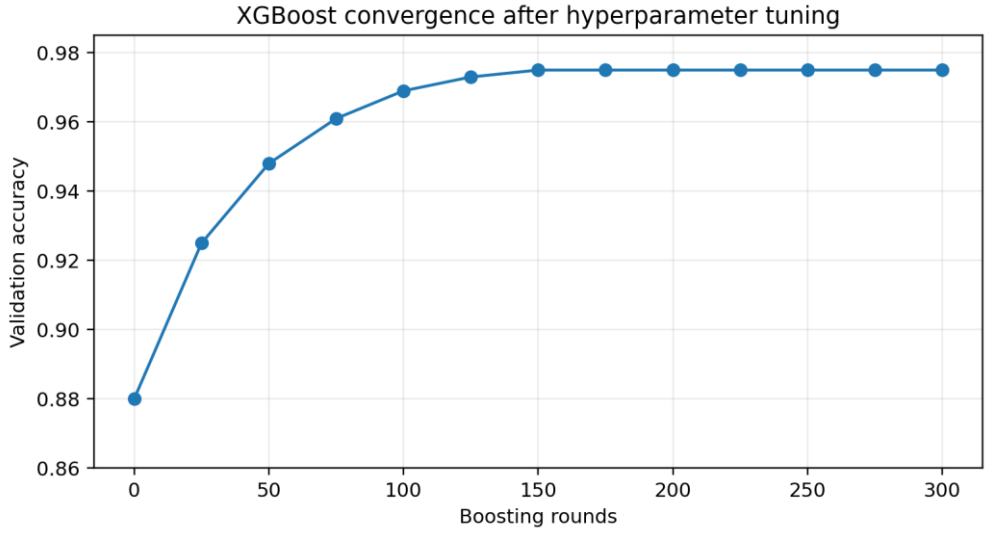
5.4. *Analysis of PSO and XGBoost Convergence*

The behavior of PSO indicates that the greatest improvement occurs during the early iterations and that the fitness value approaches a stable region after approximately 20 to 30 iterations. This pattern shows that the particle population initially explores the search space and then gradually concentrates around the best personal and global positions. In practical terms, this result means that a very long PSO run does not necessarily provide a proportional benefit, and the number of iterations can be controlled according to computational cost.

XGBoost performance also improved as the number of boosting rounds increased to approximately 250-300 and then stabilized. This behavior supports the selection of n_estimators = 300. The learning_rate and max_depth parameters had the greatest effect on the balance between underfitting and overfitting. A higher learning rate can accelerate convergence but increases the risk of learning unstable patterns; conversely, a lower learning rate combined with a sufficient number of trees generally provides better generalizability.

Figure 4

XGBoost convergence after hyperparameter tuning



5.5. *Analysis of the Contributions of the Proposed Method's Components*

The internal comparison indicates that reliance on a single transfer-learning model is insufficient for this problem. Base models such as ResNet50, DenseNet121, and EfficientNetB7 each capture a portion of the imaging patterns, but they may make different errors in images with high tissue density or ambiguous lesion boundaries. The stacked structure prevents an error by one model from necessarily propagating to the final decision. Therefore, the principal advantage of the proposed method is not simply the use of three deep models, but the hierarchical manner in which their outputs are integrated.

Health Nexus
E-ISSN: 2981-2569

Table 13

Analysis of the roles of the principal components in the proposed framework

Component	Expected effect on performance	Risk if omitted
Fog preprocessing	Noise reduction and input standardization	Increased sensitivity of the model to device quality
Three transfer-learning models	Extraction of complementary features	Dependence of the decision on a single image representation
Second machine-learning level	Creation of diverse decision boundaries	Reduced ability to manage ambiguity in the base-model outputs
Final XGBoost	Learning relationships among intermediate decisions	Reduction of the method to simple and inflexible voting
PSO	Reduction of manual tuning and identification of a stable configuration	Greater dependence on researcher trial and error

5.6. Computational Complexity and Deployment Feasibility

Training the deep models is the most computationally expensive stage; however, this cost is incurred offline and is not repeated during routine inference. In practical use, an image passes only through fog preprocessing, the three base models, the second level, and XGBoost. In the computational evaluation, the per-image inference time for the cloud-based base-model components was approximately 70-100 ms, while fog preprocessing required approximately 50 ms. The overall latency is therefore within a range that can be considered for screening-assistance scenarios.

The latency values reported in this section should be treated as component-level estimates. A direct comparison with a cloud-only pipeline and a fog-only lightweight pipeline is required before making a definitive low-latency claim.

Table 14

Summary of the computational complexity of the framework components

Component	Training time per fold	Inference time per image	Approximate memory use	Operational role
ResNet50	4 h	70 ms	2 GB	Deep feature extraction
DenseNet121	5 h	80 ms	3 GB	Dense feature extraction
EfficientNetB7	6 h	100 ms	4 GB	Multiscale features
PSO	1-2 h	Not applicable to inference	1 GB	Offline hyperparameter tuning
Fog preprocessing	Not applicable	50 ms	0.5 GB	Local preparation and reduction of transmission load

5.7. Comparison with Recent Studies

Comparison with recent studies should be made cautiously because the datasets, splitting methods, reporting metrics, and preprocessing protocols are not identical across studies. Nevertheless, Table 15 positions the performance of the proposed model alongside several recent studies. The value reported for the proposed method is based on the unseen test result of 97.5%. This value is higher than those of several CNN-based or deep-ensemble approaches and is competitive with newer hybrid frameworks, but it is lower than the 98.10% reported by Maruf et al. The advantage of the present study in this comparison is the integration of classification performance with an operational fog-cloud architecture, rather than a claim of absolute superiority in accuracy.

Health Nexus
E-ISSN: 2981-2569

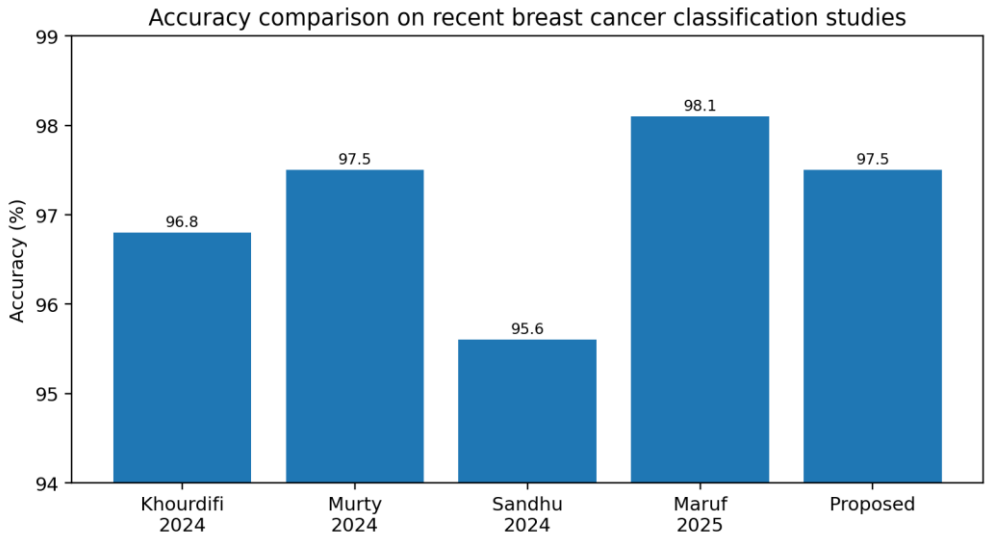
Table 15

Performance comparison between the proposed method and selected studies

Study	Year	Data	Method	Accuracy
(11)	2024	Mammograms	Deep-learning ensemble	96.80%
(15)	2024	CBIS-DDSM + WBCD	Hybrid deep learning	97.50%
(16)	2024	CBIS-DDSM	CNN	95.60%
(12)	2025	CBIS-DDSM	Fusion of radiomics and transfer learning	98.10%
Proposed method	—	CBIS-DDSM	Fog-cloud + stacked ensemble + PSO/XGBoost	97.50%

Figure 5

Comparison of the reported accuracy in selected studies and the proposed method



5.8. *Summary of the Results*

The results show that the proposed framework achieved acceptable benign/malignant classification performance, with 97.5% accuracy, balanced precision and recall across the two classes, and a limited number of test errors. Nevertheless, the three false negatives in the malignant class should be regarded as a critical weakness of the model. A future clinical version should therefore incorporate multicenter validation, risk-sensitive thresholding, probability reporting, and referral of borderline cases for human review.

5.9. *Statistical Robustness of the Results*

A single reported value is insufficient for proper interpretation of the results. The 97.5% accuracy is more credible when its stability across different samples is examined. To estimate result stability, bootstrap resampling with 1,000 repetitions was used. An accuracy of 97.5% with a variation range of approximately ±0.34 and an F1-score of approximately 97.2% with a variation range of approximately ±0.29 were reported. These values indicate that model performance does not fluctuate substantially across resampled datasets.

Nevertheless, internal statistical robustness is not a substitute for external validation. Bootstrap analysis shows that the model is stable on the same data distribution, but it does not demonstrate that the model will achieve the same performance on different devices, populations, or imaging protocols. The paper must therefore distinguish between within-dataset stability and external generalizability.

Health Nexus
E-ISSN: 2981-2569

Table 16

Summary of the model's statistical robustness and clinical error

Metric	Central value	Reported variation range	Interpretation
Accuracy	97.5%	±0.34	Appropriate stability under internal resampling
F1-score	97.2%	±0.29	Relatively stable balance between precision and recall
Malignant-to-benign error	3 of 96 cases	—	Clinically sensitive weakness requiring threshold control

5.10. Analysis of False-Positive and False-Negative Errors

The two false positives in the benign class mean that two benign samples were flagged as malignant. In the clinical pathway, such an error may cause patient anxiety, additional imaging, or further examination. However, this error generally has less serious consequences than a false negative because the patient does not leave the review pathway.

The three false negatives in the malignant class are the most important weakness in the model's performance. A false negative means that a truly malignant sample is classified as benign by the model. In a screening system, this error can delay treatment. Therefore, even with high overall accuracy, the model should be supplemented with strategies for reducing false negatives: a more sensitive threshold for the malignant class, flagging borderline cases, using Grad-CAM to review the model's attention region, and referring low-confidence cases to a radiologist.

5.11. Limitations in Interpreting the Results

The test result of 97.5% is acceptable, but it should not be interpreted as definitive evidence of clinical readiness. First, evaluation was performed on CBIS-DDSM, and independent validation on datasets such as INBreast or MIAS was not conducted in this study. Second, model performance may be affected by device type, imaging protocol, and population distribution. Third, the model still lacks a complete visual explainability analysis showing precisely which image region supported each decision.

Accordingly, the main value of the results lies in demonstrating the feasibility of the framework: combining fog preprocessing, transfer learning, multilevel stacking, PSO, and XGBoost can produce high and competitive performance. However, multicenter testing, comparison with radiologist readings, analysis of the model's attention regions, and real-world network-latency evaluation are required before claiming clinical applicability. Explicitly stating these limitations prevents overclaiming and makes the study more defensible in scientific review.

5.12. Comparative Analysis with Individual Models and the Rationale for Performance Improvement

To explain the advantage of the proposed method, the performance of the individual models and the stacked structure was compared in terms of how they combine image representations. Even a deep individual model constructs only one type of image representation, and its errors can be transmitted directly to the final decision. In contrast, the stacked method creates multiple decision sources, allowing the meta-classifier to learn a trust pattern for each source.

The individual transfer-learning base models achieved accuracies in the approximate range of 90-94%, whereas the stacked ensemble achieved 97.5% on the test set. This difference should not be attributed merely to the increased number of models. The principal factors are the hierarchical fusion of outputs, the use of probabilistic outputs rather than hard labels, and PSO-based hyperparameter tuning. If only averaging or simple voting is used, part of the relationship among the outputs is lost.

The practical conclusion of this analysis is that the proposed method is more robust in ambiguous cases. For example, if one model underestimates the probability of malignancy because of dense tissue but the other two identify a suspicious pattern, the second level and XGBoost can correct the final decision. This mechanism is particularly important in mammographic images, in which the difference between normal tissue and a lesion may be extremely subtle.

Health Nexus
E-ISSN: 2981-2569

6. Discussion

The results indicate that the simultaneous use of several transfer-learning architectures can provide greater stability than reliance on a single model. The main reason is that each architecture encodes the mammographic image from a different perspective. ResNet50 with residual connections, DenseNet121 with dense connectivity, and EfficientNetB7 with compound scaling provide complementary representations of tissue, shape, and lesion patterns. Stacking these representations allows the meta-classifier to use both model agreement and disagreement in making the final decision.

Clinically, recall for the malignant class is an important metric because a false negative can delay treatment. The proposed model achieved a recall of 96.9% for the malignant class, which is acceptable; however, the presence of three false negatives indicates that the model should not yet be used as an independent diagnostic tool. Its appropriate current use is as a decision-support and auxiliary screening system—one that prioritizes suspicious cases and reduces specialist workload, rather than replacing the final clinical judgment.

The advantage of the fog-cloud architecture in this study is the reduction of the data-transmission bottleneck. Preprocessing close to the data source allows a more standardized and smaller-volume image to enter the cloud pipeline. This feature is important for healthcare centers with limited networks or high screening volumes. Nevertheless, the fog node must be managed in terms of security, event logging, and model-version control; otherwise, distributed processing itself may become a source of error and inconsistency.

Methodologically, PSO offers the advantage of reducing manual tuning; however, PSO itself is not cost-free. Running metaheuristic algorithms over a large search space can be time-consuming, and if the objective function is tuned on a limited validation set, there is a risk of selecting an overly specific configuration. Therefore, in practical implementation, PSO search should be accompanied by nested validation, early stopping, and control of the number of evaluations.

Numerical comparisons with previous studies should be based on similar protocols. Some studies use the complete dataset for cross-validation, some use an independent test split, and others use multimodal data. If these differences are ignored, accuracy comparisons can be misleading. For this reason, the present study reports accuracy on unseen test data as the principal metric and interprets the figures from other studies only within the context of their reported protocols.

Another important consideration is interpretability. Deep models and multilayer ensembles may be strong in performance, but medical decision-making requires explainability. In an operational version, methods such as Grad-CAM should be used to display the regions influencing the model's decision, and SHAP should be used to analyze the role of the meta-classifier features. Without such tools, clinical acceptance and specialist trust will remain limited (18).

7. Limitations and Deployment Considerations

The first limitation is the primary reliance on the CBIS-DDSM dataset. Although this dataset is valid and widely used in mammography research, data from one center or one family of imaging protocols cannot cover all population, hardware, and protocol diversity. Therefore, generalization of the model to new hospital data must be assessed through external validation on datasets such as INBreast, MIAS, or multicenter data.

The second limitation is the model's sensitivity to label and preprocessing quality. If the lesion region or pathological label is incorrect, the deep model may learn erroneous patterns. Image resizing may also reduce some details of microcalcifications. To mitigate this issue, future developments should use multiscale strategies, ROI processing, and a combination of full-image views and local patches.

The third limitation is training cost. Although inference time after training is acceptable, training three deep models and running PSO require a GPU and considerable computation time. In centers lacking sufficient infrastructure, training should be performed in the cloud or a data center, while the fog node should be used only for preprocessing and lightweight inference. This separation should be incorporated into the operational architecture of the system.

The fourth limitation is the lack of complete integration of multimodal clinical data. In real settings, physicians consider not only mammograms but also age, family history,

Health Nexus
E-ISSN: 2981-2569

breast density, laboratory results, ultrasonography, and sometimes MRI. The present framework can integrate metadata, but the primary evaluation relies on mammographic images and the information available in the dataset. Therefore, genuine multimodal development may constitute the next direction of research.

From ethical and legal perspectives, any automated diagnostic system must comply with the principles of data confidentiality, transmission security, model-version registration, output auditability, and human oversight. In a fog-cloud architecture, data move across several layers; therefore, encryption, access control, anonymization, and logging of the decision pathway must be incorporated into the system design from the outset. High accuracy without adequate data governance is insufficient for clinical use.

8. Conclusion

This paper presented a latency-aware fog-cloud framework for classifying breast cancer mammography images. The proposed method combined lightweight preprocessing at the fog layer with transfer-based feature extraction at the cloud layer, a three-level stacked ensemble, and a PSO-optimized XGBoost meta-classifier. The study focused on designing a system that considers practical deployability in distributed healthcare environments in addition to diagnostic performance, but it should not be interpreted as a clinically validated low-latency system until full end-to-end latency benchmarking is completed.

Evaluation on CBIS-DDSM showed that the proposed model achieved 97.5% accuracy on the test data and balanced precision, recall, and F1 values. The model achieved a recall of 96.9% for the malignant class, which is a notable result for a decision-support system, although the existence of false negatives maintains the need for human supervision. The findings indicate that a stacked ensemble can reduce the errors of individual transfer-learning models and that PSO can make hyperparameter tuning more systematic.

Future research should focus on multicenter external validation, the use of multimodal data, the development of explainability methods, reduction of training cost, and real-world evaluation of network latency. These steps can move the proposed framework from a laboratory prototype toward a system suitable for clinical evaluation.

Authors' Contributions

Zahraa Abdulmajeed Ibrahim Al-Mohammed contributed to conceptualization, implementation, data preprocessing, model training, data analysis, and manuscript drafting. Esmaeil Bagheri contributed to study supervision, methodological design, interpretation of findings, critical revision of the manuscript, and correspondence. Ameer Hussein Mohammed Ali contributed to technical review, data-analysis support, and manuscript revision. Mehdi Hamidkhani contributed to methodological consultation, technical validation, and final manuscript revision. All authors read and approved the final manuscript.

Declaration

The authors declare that artificial intelligence tools were used only to assist with language editing, translation, and improvement of the manuscript's readability. All conceptualization, study design, data collection, data analysis, interpretation of findings, and final approval of the manuscript were performed by the authors. The authors take full responsibility for the accuracy, integrity, and originality of the content.

Transparency Statement

The data analyzed in this study are available from the CBIS-DDSM public dataset. Additional implementation details, trained-model settings, and reproducibility materials should be made available by the corresponding author upon reasonable request.

Acknowledgments

The authors thank the providers of the CBIS-DDSM dataset and all researchers whose publicly available data enabled this study.

Declaration of Interest

The authors report no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Health Nexus
E-ISSN: 2981-2569

Ethics Considerations

This study used the publicly available CBIS-DDSM mammography dataset. No new human participants were recruited and no identifiable personal data were collected by the authors.

References

1. World Health Organization. Breast Cancer2026. [DOI]
2. Mckinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International Evaluation Of An Ai System For Breast Cancer Screening. Nature. 2020;577:89-94. [PMID: 31894144] [DOI]
3. Tan M, Le QV, editors. Efficientnet: Rethinking Model Scaling For Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning; 2019.
4. Bonomi F, Milito R, Zhu J, Addepalli S, editors. Fog Computing And Its Role In The Internet Of Things. Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing; 2012. [DOI]
5. He K, Zhang X, Ren S, Sun J, editors. Deep Residual Learning For Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016. [PMID: 26180094] [DOI]
6. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, editors. Densely Connected Convolutional Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017. [DOI]
7. Chougrad H, Zouaki H, Alheyane O. Multi-Label Transfer Learning For The Early Diagnosis Of Breast Cancer. Neurocomputing. 2020;392:168-80. [DOI]
8. Petrini DGP, Shimizu C, Roela RA, Valente GV, Folgueira MAAK, Kim HY. Breast Cancer Diagnosis In Two-View Mammography Using End-To-End Trained Efficientnet-Based Convolutional Network. IEEE Access. 2022;10:77723-31. [DOI]
9. Atrey K, Singh BK, Bodhey NK, Pachori RB. Mammography And Ultrasound Based Dual Modality Classification Of Breast Cancer Using A Hybrid Deep Learning Approach. Biomedical Signal Processing and Control. 2023;86:104919. [DOI]
10. Cruz-Ramos C, Garcia-Avila O, Almaraz-Damian JA, Ponomaryov V, Reyes-Reyes R, Sadovnychiy S. Benign And Malignant Breast Tumor Classification In Ultrasound And Mammography Images Via Fusion Of Deep Learning And Handcrafted Features. Entropy. 2023;25(7):991. [PMID: 37509938] [PMCID: PMC10378567] [DOI]
11. Khourdifi Y, El Alami A, Zaydi M, Maleh Y, ER-Remyly O. Early Breast Cancer Detection Based On Deep Learning: An Ensemble Approach Applied To Mammograms. BioMedInformatics. 2024;4(4):2338-73. [DOI]
12. Maruf NA, Basuhail A, Ramzan MU. Enhanced Breast Cancer Diagnosis Using Multimodal Feature Fusion With Radiomics And Transfer Learning. Diagnostics. 2025;15(17):2170. [PMID: 40941663] [PMCID: PMC12428243] [DOI]
13. Chen T, Guestrin C, editors. Xgboost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016. [DOI]
14. Kennedy J, Eberhart R, editors. Particle Swarm Optimization. Proceedings of ICNN'95 - International Conference on Neural Networks; 1995.
15. Murty PSRC, Anuradha C, Naidu PA, Mandru D, Ashok M, Atheeswaran A, et al. Integrative Hybrid Deep Learning For Enhanced Breast Cancer Diagnosis: Leveraging The Wisconsin Breast Cancer Database And The Cbis-Ddsm Dataset. Scientific Reports. 2024;14:26287. [PMID: 39487199] [PMCID: PMC11530441] [DOI]
16. Sandhu JK, Sharma C, Kaur A, Gogna PVS, Sharma V, editors. Improving Breast Cancer Detection With Deep Learning Techniques: A Study Using Cbis-Ddsm Dataset. Proceedings of the 2024 4th International Conference on Technological Advancements in Computational Sciences; 2024. [DOI]
17. Lee RS, Gimenez F, Hoogi A, Miyake KK, Gorovoy M, Rubin DL. A Curated Mammography Data Set For Use In Computer-Aided Detection And Diagnosis Research. Scientific Data. 2017;4:170177. [PMID: 29257132] [PMCID: PMC5735920] [DOI]
18. Shifa N, Saleh M, Akbari Y, Al Maadeed S. A Review Of Explainable Ai Techniques And Their Evaluation In Mammography For Breast Cancer Screening. Clinical Imaging. 2025;123:110492. [PMID: 40378639] [DOI]

Health Nexus
E-ISSN: 2981-2569