electronics

MDPI

Article

Stability-Driven Feature Extraction–Kolmogorov–Arnold Network-Driven Ensemble Framework for Reliable Breast Cancer Detection

Abdul Rahaman Wahab Sait [1,*] and Yazeed Alkhurayyif [2,*]

[1] Department of Documents and Archive, Center of Documents and Administrative Communication, King Faisal University, P.O. Box 400, Al-Ahsa 31982, Saudi Arabia
[2] Applied College, Shaqra University, Shaqra 11961, Saudi Arabia
* Correspondence: asait@kfu.edu.sa (A.R.W.S.); yalkhurayyif@su.edu.sa (Y.A.)

Abstract

Breast cancer screening is a fundamentally probabilistic diagnostic task that requires precise identification of complex imaging characteristics from diverse patient cohorts. Despite improvements in deep learning techniques, current automatic tools are typically trained on well-curated datasets and do not generalize to heterogeneous data, thereby limiting their application. This study aims to address these shortcomings by introducing a more effective and generalizable framework for breast cancer classification that focuses on the stability of features, the learning of complementary representations, and improved decision modeling. The proposed methodology incorporates stability-driven feature extraction (SDFE) with a multi-branch architecture that consists of EfficientNetV2 (Convolutional neural networks (CNNs)), EfficientFormer (Vision transformers (ViTs)), and multi-layer perceptron (MLP)-Mixer models to extract various feature representations. To improve non-linear decision boundaries, it uses a Kolmogorov–Arnold Network (KAN)-based classification head and selects the most credible prediction via an adaptive voting mechanism. This model is trained using patient-level splitting on the VinDr-Mammo dataset, evaluated using five-fold cross-validation, and subsequently externally validated on the CBIS-DDSM dataset. Experimental findings demonstrate the consistent performance of the proposed model, with accuracies of 94.5% in cross-validation, 93.3% on the VinDr-Mammo test set, and 94.6% on CBIS-DDSM, surpassing other recent state-of-the-art solutions. It demonstrates enhanced robustness and cross-dataset generalization, offering a scalable, consistent framework for breast cancer classification that supports the development of computer-aided diagnostic systems.

Keywords: breast cancer classification; mammography analysis; stability-driven feature extraction; Kolmogorov–Arnold networks; deep learning ensemble; computer-aided diagnosis

Check for updates

Academic Editor: Yoichi Tomioka

Received: 13 April 2026
Revised: 14 May 2026
Accepted: 19 May 2026
Published: 20 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Breast cancer screening is essentially a decision-making process under uncertainty, rather than a strictly deterministic process of image interpretation [1–3]. Radiologists in mass population screening need to produce high-stakes clinical decisions using only slightly and frequently ambiguous visual patterns observable in mammograms, which are inherently noisy and heterogeneous [3]. The implications of diagnostic errors are not symmetrical: false-negative results might defer diagnosis, negatively impact patient outcome,

and false-positive results might result in unnecessary biopsies, patient anxiety, and unnecessary healthcare expenditures [4]. The use of screening programs and imaging procedures has increased radiologists' workload, potentially affecting their diagnostic consistency [4].

Currently, mammography remains the effective modality in terms of breast cancer screening due to its cost-efficiency and mortality reduction effects [5]. However, mammogram interpretation is complex because malignant lesions may have low contrast and may even be hidden or mimicked by breast structures [6]. In addition, patient-to-patient variability in breast tissue density affects lesion detectability and interpretability, thus further complicating image interpretation. Notably, interpretation of mammograms requires considering multiple contextual factors, such as clinical expertise and the patient's risk profile, along with the natural acceptance of diagnostic uncertainty and the probabilistic nature of the decision-making process [7].

Despite all the advantages of mammograms and breast cancer screening techniques, the diagnostic capabilities of radiologists remain hindered by various human and systemic factors. Inter-radiologist variability in mammogram interpretation exists, especially when dense tissue and subtle changes need to be identified [7]. Borderline results classified as BI-RADS 3 and 4 involve significant uncertainty, which contributes to high rates of false positives. These issues emphasize the importance of decision-support solutions that provide consistent diagnoses with reduced sensitivity to variability [8]. Computer-assisted diagnosis (CAD) was first proposed as a tool to address weaknesses in mammogram interpretation by detecting suspicious regions or providing second interpretations [8,9]. These systems relied mainly on handcrafted features and the rules of clinical practice. The problem was that these algorithms tended to perform poorly across different conditions because they were prone to various imaging artifacts, preprocessing steps, and dataset characteristics.

Machine learning and deep learning solutions have revolutionized mammographic analysis by enabling the learning of relevant features directly from image data. Multilayer perceptron (MLP)-based approaches to breast cancer diagnosis are often combined with other methods using handcrafted features and radiomics [9]. They could easily become integral components of CAD algorithms due to their ability to capture nonlinear dependencies among input variables [10]. Nevertheless, traditional MLPs suffer from several inherent limitations stemming from their flattened input representations, which make them susceptible to various types of noise and inconsistencies, such as segmentation differences [11]. Notably, mammograms contain valuable spatial information about lesion morphology and tissue context; therefore, MLP-based approaches are unable to effectively leverage this information [11]. Convolutional neural networks (CNNs) and vision transformers (ViTs) have become the most widely used paradigms for image-based disease detection due to their ability to learn hierarchical spatial representations [12–14].

Despite these techniques delivering remarkable outcomes in detecting breast cancer via mammography, there are several challenges in the deployment of these methods in clinical settings. The majority of the current literature [13–15] aims to achieve higher classification accuracy on curated data using increasingly sophisticated CNNs or ViTs. However, these models heavily rely on saliency-driven or attention-based feature learning, which implicitly assumes that highly activated features are diagnostically reliable. Mammographic features are highly sensitive to noise, contrast variations, breast density, and acquisition protocols, resulting in unstable predictions and significant performance degradation during external validation [16].

The proposed approach is motivated by the limitations of existing breast cancer classification approaches, which commonly rely on either homogeneous feature extraction pipelines or static ensemble strategies. Another crucial gap is generalization assessment. Many studies report strong results on a single dataset, with cross-dataset and real-world

validation remaining limited. As a result, the reliability of these models under domain shift is still unclear. Mammographic abnormalities exhibit diverse visual characteristics, including subtle local structures, long-range contextual dependencies, and complex feature interactions [17–19]. Therefore, a single architectural paradigm may not sufficiently capture all diagnostically relevant information. In addition, mammographic images are affected by acquisition variability, differences in breast density, noise, and contrast changes, which can lead to unstable feature responses. Conventional fully connected classification heads may be insufficient for modeling the nonlinear decision boundaries between benign and malignant cases. Motivated by these gaps, this study focuses on reliability-aware learning by introducing stability-driven feature extraction, expressive non-linear classification via Kolmogorov–Arnold Networks (KAN), and adaptive ensemble voting, aiming to increase robustness, generalization, and clinical trustworthiness in mammography-based breast cancer detection.

The novelty of the proposed framework lies in three primary contributions.

(i) A Stability-Driven Feature Extraction (SDFE) mechanism is introduced to explicitly evaluate feature consistency under controlled perturbations and suppress unstable feature responses. This provides a robustness-oriented feature selection strategy specifically designed for heterogeneous mammographic data.

(ii) The study introduces a KAN-based classification head for branch-wise feature classification. Instead of using conventional fully connected layers, the Kolmogorov–Arnold Network models nonlinear feature–class relationships through learnable functional mappings, improving decision expressiveness for subtle benign–malignant discrimination.

(iii) A confidence-aware adaptive voting ensemble is proposed to dynamically select the most reliable branch prediction among CNN-, ViT-, and MLP-Mixer-based outputs. Unlike hard or soft voting, which treat model outputs uniformly, the proposed strategy performs case-specific decision selection based on prediction confidence, reducing the influence of uncertain branches and improving diagnostic reliability.

The rest of this study is structured as follows. Section 1 provides background on breast cancer detection and explains the motivation and contributions of the proposed study. Section 2 reviews the literature on deep learning-based mammography analysis, highlighting existing methods and identifying research gaps. Section 3 addresses the proposed methodology, which involves stability-driven feature extraction using CNNs, ViTs, and MLP architectures, with a KAN classification head. Additionally, it presents the proposed ensemble learning approach, which leverages an adaptive voting mechanism to classify mammograms into benign and malignant classes. Section 4 discusses the experimental setup and performance evaluation on VinDr-Mammo and CBIS-DDSM datasets. Section 5 discusses the obtained results and their implications for breast cancer diagnosis. Finally, Section 6 concludes the study and offers suggestions for future studies.

2. Literature Review

Mammography remains the most widely used imaging modality for breast cancer screening because it can detect early radiographic abnormalities in a non-invasive, cost-effective manner [20–22]. In contrast to histopathological imaging, which requires a tissue biopsy, it is primarily used for diagnostic confirmation, enabling mass population screening and detecting suspicious lesions, including masses, architectural distortion, and microcalcifications, before clinical symptoms become evident [22–24]. The availability of large annotated data sets, including VinDr-Mammo [25], CBIS-DDSM [26], and INbreast [27], has further boosted the development of automated detection systems. These datasets offer a variety of imaging conditions and clinically relevant variations that are required to train robust deep learning models. As a result, current artificial intelligence research efforts

are focused on mammography, aiming to improve screening accuracy, reduce radiologist workload, and enhance early breast cancer detection.

Numerous CNN-based models have demonstrated strong performance on curated benchmark datasets [28]. CNNs have been widely used because of their capacity to learn hierarchical representations of image structure and to capture fine-grained visual patterns related to malignant lesions [28]. Several studies have used pre-trained CNN backbones and transfer learning strategies to enhance classification and lesion detection performance on mammography datasets [29,30]. These approaches have shown promising results in detecting abnormalities such as masses, calcifications, and architectural distortions, especially when trained on large annotated datasets [31]. Nevertheless, CNNs have some limitations. Feature learning in CNNs may extract visually salient but diagnostically unreliable patterns. Additionally, CNNs can be sensitive to subtle variations in image intensity, contrast, or resolution, leading to instability under domain shift [32–35]. Increasing the depth of the network tends to improve performance on the training data, but substantially increases the computational overhead and the tendency to overfit.

Recently, ViTs have been introduced to capture long-range contextual relationships through self-attention mechanisms [35]. While ViTs have advantages for modeling global dependencies, they treat image patches equally during feature encoding and use attention weights to infer importance [35]. This approach does not explicitly account for feature reliability or stability, particularly in the presence of imaging artifacts or low-quality regions. Furthermore, ViTs tend to require large training datasets and significant computational resources, potentially limiting their use in routine screening workflows. To overcome these challenges, hybrid CNN-ViT architectures are developed to integrate the strengths of individual CNNs and ViTs in extracting crucial features associated with breast cancer. Across MLPs, CNNs, and ViTs, a linear or SoftMax layer is commonly used for final classification. These classifiers impose relatively simple decision boundaries and may be inadequate for modeling the complex, non-linear relationships in ambiguous mammographic findings. Subsequently, performance degradation is often observed when these models are evaluated on external datasets or deployed in real-world clinical settings. Table 1 highlights the strengths and limitations of existing breast cancer detection models.

Recent developments in artificial intelligence-based mammography have primarily focused on novel backbone architectures, particularly the adoption of CNNs and ViTs. While these architectures have demonstrated strong feature-learning capabilities, most existing methods use conventional linear or fully connected classification heads, which may limit the model's ability to learn complex nonlinear relationships in high-dimensional medical imaging data. Furthermore, while the concept of ensemble learning has been studied to boost predictive performance, most existing approaches use static fusion strategies (e.g., averaging, fixed-weight voting) that do not reflect the varying reliability of individual models across imaging conditions. Another important limitation is the lack of explicit mechanisms to ensure feature stability during representation learning. Mammographic images are often subject to varying contrast, noise, breast density, and acquisition conditions, which may contribute to instability in learned feature representations and thereby affect classification reliability. Despite the surge in large-scale mammography datasets, relatively few studies have addressed the problem of learning stable, perturbation-invariant features. Moreover, relatively little attention has been devoted to adaptive reliability-aware decision mechanisms that dynamically select the most confident predictions among heterogeneous models. These gaps underscore the need for frameworks that simultaneously address feature stability, nonlinear decision modeling, and adaptive ensemble decision-making, especially when trained on large heterogeneous datasets and evaluated across multiple mammography cohorts.

Table 1. Characteristics of Existing Studies.

Study	Modality/Dataset	Core Architecture	Features	Limitations
Khaled et al. [32]	Mammography (INbreast)	5-CNN Ensemble	Solves severe 12-class data imbalance.	Enormous computational and hyperparameter tuning cost.
Akselrod-Ballin et al. [36]	Mammography	Region-Based CNN	Unified segmentation and detection.	High computational overhead.
Chougrad et al. [37]	Mammography	CNN + Transfer Learning	Optimized fine-tuning prevents overfitting.	Performance is capped without global attention.
Dhungel et al. [38]	Mammography	Cascaded YOLO + CNN	Requires minimal user intervention.	Multi-stage pipeline is difficult to optimize.
Shen et al. [39]	Mammography (CBIS-DDSM)	All-Convolutional Network	Transitions to simpler image-level labels.	Still requires initial heavy annotations.
Ribli et al. [40]	Mammography (INbreast)	Faster R-CNN	High sensitivity, very low false positives.	Region proposals slow down inference.
Zeynali et al. [41]	Histopathology (BreaKHis)	Hybrid CNN-Transformer	Captures local textures & global context.	Generalizability risks across different scanners.
Sharma et al. [42]	Mammography	Pure Transformer	Lowers false positive/negative rates.	Data-hungry; prone to overfitting.
Murty et al. [43]	Mammography + Tabular	Hybrid CNN + SGD	Fuses image data with numerical attributes.	Slow convergence; scaling issues.
Firouzbakht et al. [44]	Mammography (CBIS-DDSM)	Neighborhood Attention Transformer + Shearlet	Denoises and extracts highly precise edges.	High computational overhead for frequency math.
Yoshitsugu et al. [45]	Mammography (VinDr, etc.)	ROI-Stratified Swin Transformer	Stronger benign vs. malignant distinction.	Reliant on precise, manual radiologist ROI masks.
Asif et al. [46]	Ultrasound (BUSI)	Channel Boosted CNN-Transformer	Suppresses ultrasound noise; learns morphology.	May struggle generalizing across ultrasound probe types.
Khan SU et al. [47]	Mammography	Ensemble of ViTs and CNNs	CNNs catch microcalcifications; ViTs catch distortion.	Massive memory and hardware resource requirements.

3. Materials and Methods

Breast cancer screening using mammography is a high-dimensional and clinically sensitive classification problem with subtle lesion patterns, heterogeneous tissue density, and domain variability across acquisition systems. Single-architecture models, such as CNNs, ViTs, and MLP-based classifiers, tend to develop complementary but partial representations of malignancy-related features. The CNN models are efficient for modeling local texture and mass morphology, while the ViT models are more efficient at modeling long-range contextual dependencies, and the MLP-based representations model global feature abstraction. However, reliance on a single backbone poses a risk of overfitting to architecture-specific inductive biases and limits robustness to distributional shifts. To overcome this limitation, the proposed framework (Figure 1) utilizes a heterogeneous

ensemble incorporating CNNs, ViTs, and MLP branches to enable complementary feature learning and reduce model variance. Rather than using a conventional linear classification layer, the framework introduces a KAN classification head, which replaces the global linear decision boundary with spline-based nonlinear transformations. This increases the expressiveness of the decisions and enables the modeling of complex non-monotonic feature interactions typical of ambiguous or dense breast cases. By leveraging the diversity of architectural combinations and nonlinear decision modeling, it improves discriminative capability, robustness, and generalization, making it especially suitable for large-scale and cross-domain breast cancer detection tasks.

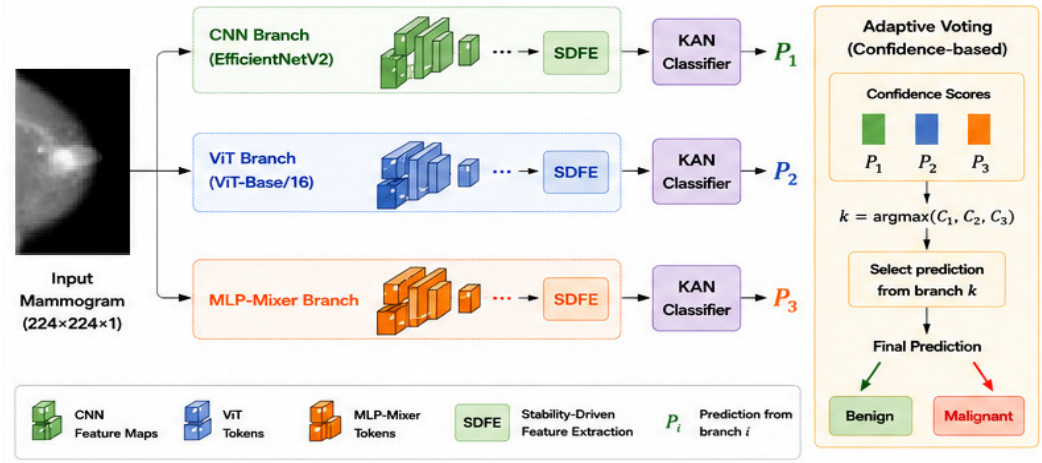


Figure 1. The Proposed Ensemble Learning-based Breast Cancer Detection.

Employing an individual KAN classification head for each feature-extraction branch is justified by the heterogeneity of the representations learned by CNNs, ViTs, and MLP architectures. Each model captures different characteristics of mammographic images, such as local spatial structures, global contextual dependencies, and high-level abstract feature interactions. Applying an independent KANs head allows the nonlinear decision function to adapt to the feature distribution of each branch, rather than requiring a shared classifier to work with heterogeneous representations. This design is highly expressive in the decision layer and can help better model complex feature-class relationships while retaining the complementary strengths of each feature-extraction model.

3.1. Dataset Description, Partitioning, and Preprocessing

3.1.1. Dataset Description

The VinDr-Mammo dataset [23,48] is a large-scale full-field digital mammography (FFDM) dataset from mammography examinations performed between 2018 and 2020 at two hospitals (HMUH and H108). A total of 5000 examinations, representing 20,000 images (four standard views per examination: L-CC, R-CC, L-MLO, R-MLO), were randomly sampled and de-identified. DICOM metadata includes patient age (with ages above 89 removed) and the imaging device manufacturer and model name were preserved. However, the personal identifiers were removed, including pixel-level identifiers, through corner masking and double manual review. Each evaluation was independently double-read by experienced radiologists using BI-RADS guidelines, with arbitration for disagreements. The dataset includes breast-level BI-RADS categories (1–5), breast density (A–D) and lesion-level bounding box annotations for suspicious findings (BI-RADS 3–5).

To strictly evaluate cross-domain generalization, the model was trained on the VinDr-Mammo dataset and then externally tested on the CBIS-DDSM dataset [24]. To further evaluate the robustness and generalization capability of the proposed framework, external

validation was performed using the CBIS-DDSM dataset under a patient-level evaluation protocol. Unlike VinDr-Mammo, which consists of full-field digital mammograms acquired from modern digital screening systems, CBIS-DDSM contains digitized film mammograms with substantially different imaging characteristics, including intensity distributions, contrast profiles, and acquisition-related artifacts. The CBIS-DDSM dataset comprises 2620 mammography studies from 1566 subjects and is a curated, standardized subset of the original DDSM collection. Importantly, the CBIS-DDSM dataset was used exclusively for external testing and was not involved in model training, hyperparameter optimization, or validation.

The lack of distributional agreement across these datasets introduces a realistic domain shift. Evaluating performance across different conditions enables a more robust assessment of model performance and reduces the risk of dataset-specific overfitting. This experimental design mirrors real-time deployment, evaluating the model's performance across institutions, populations, and imaging platforms.

3.1.2. Dataset Partition

To ensure stable evaluation and avoid data leakage, the dataset was split at the patient level, in line with the VinDr-Mammo dataset structure. Each examination in VinDr-Mammo has multiple mammographic views of the same patient. This overlap can leak data and lead to overly optimistic estimates of performance. To address this issue, the dataset owner labeled all images for each patient. The dataset was split into 80% for model development and 20% for a held-out test set. The development subset was further evaluated using 5-fold cross-validation, in which the 80% was divided into 5 folds. In each fold, four subsets were used for model training, and the remaining subset served as the validation set. This cross-validation strategy enables robust model evaluation while effectively utilizing training data. To further evaluate the generalization ability of the proposed framework, 20% of the CBIS-DDSM dataset (patient-level) was used as external evaluation data. Since CBIS-DDSM was not involved in the training process, testing the model on this dataset introduces domain variability. It presents a more realistic assessment of model robustness across different mammography acquisition environments.

3.1.3. Dataset Preprocessing and Augmentation

To prevent data leakage and minimize preprocessing-related biases, all dataset preprocessing steps were performed under a strict patient-level protocol. In particular, all mammographic images for the same individual were assigned to a single subset—no individual had their images distributed across multiple data splits, ensuring that information from the same subject was not shared between the training and test stages.

All preprocessing operations were deterministic and applied independently to each image. These included resizing to a fixed spatial resolution, min–max intensity normalization, and CLAHE-based contrast enhancement. These steps were used only to standardize image appearance and improve visibility of breast tissue structures; no label-specific or dataset-specific preprocessing parameters were learned from the validation, test, or external datasets. Therefore, preprocessing did not introduce information transfer between data partitions.

Data augmentation was applied only to the training subset after preprocessing. The validation set, VinDr-Mammo held-out test set, and CBIS-DDSM external test set were not augmented. The augmentation pipeline included RandAugment-based transformations, small affine changes, brightness and contrast variation, Gaussian noise injection, and random erasing. These transformations were constrained to preserve anatomical plausibility and avoid generating unrealistic lesion patterns. Importantly, augmentation

was performed online during training and was never used to create duplicate or modified samples in validation or testing subsets.

For cross-dataset evaluation, CBIS-DDSM was used exclusively as an external test dataset and was not involved in training, validation, hyperparameter tuning, augmentation policy selection, or model selection. As VinDr-Mammo consists of full-field digital mammograms and CBIS-DDSM consists of digitized film mammograms, the same preprocessing pipeline was applied to these datasets to reduce input-format variability without forcing dataset-specific adaptation, minimizing domain bias while preserving the natural distributional differences required for external generalization assessment.

3.2. Stability-Driven Feature Extraction

Deep learning models have shown great ability to learn discriminative representations from mammographic images. However, their performance may be influenced by variations in imaging conditions, including noise, contrast, and acquisition processes. These factors can introduce instability into the learned feature representations, making them less reliable for classification. To overcome this limitation, the proposed SDFE mechanism introduces a perturbation-aware feature-refinement stage that assesses the consistency of feature responses across multiple controlled image transformations. For each input mammogram, several perturbed versions are generated and passed through the same feature extraction backbone to obtain corresponding feature representations. The perturbation-wise variance is then computed to estimate feature stability. Features that exhibit consistent responses across perturbations receive higher stability weights, whereas perturbation-sensitive features are attenuated via exponential weighting.

Unlike conventional feature selection mechanisms that apply hard thresholding, SDFE performs soft stability-aware weighting, thereby preserving diagnostically relevant but naturally variable mammographic patterns before forwarding the refined feature vector to the KAN-based classification head. Figure 2 presents the proposed feature extraction using SDFE, capturing the crucial features associated with benign and malignant images.

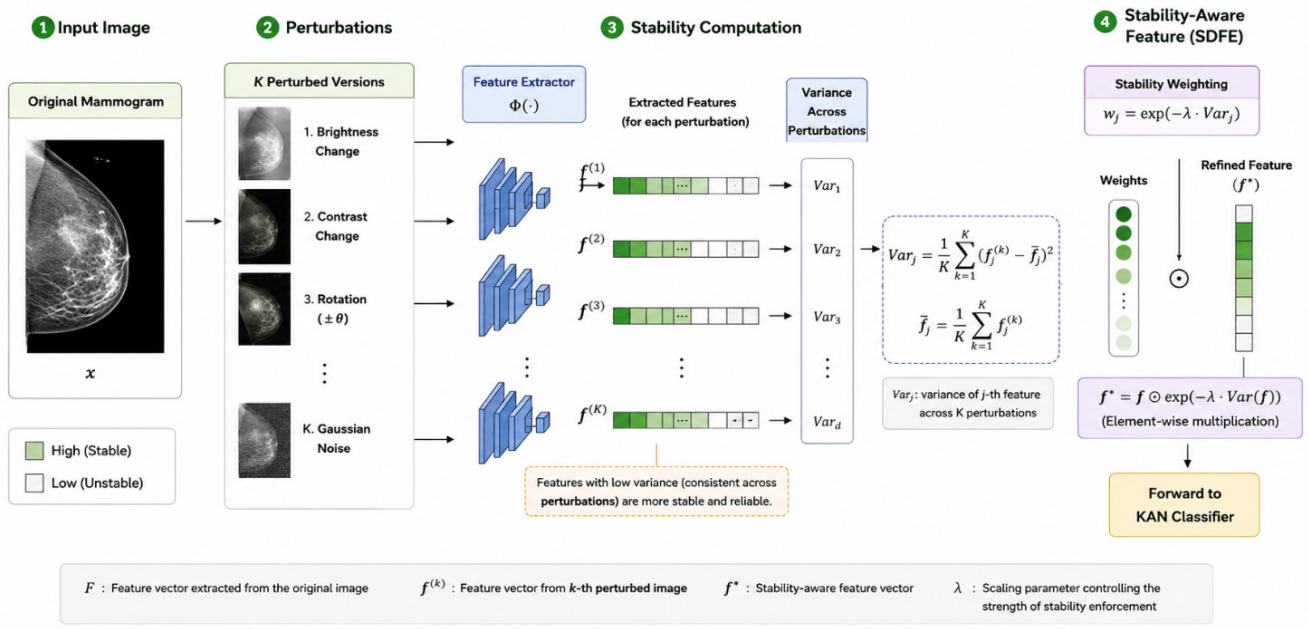


Figure 2. General architecture of the proposed SDFE mechanism.

Let x represents the input mammogram and $x^{(k)} = T_k(x)$ denotes its perturbed version generated through the transformation function, $T_k(.)$, where $k = 1, 2, \ldots, K$. Equation (1)

outlines the process of measuring variance across perturbations, quantifying the stability of the extracted features.

$$Var(f) = \frac{1}{K}\sum_{k=1}^{K}\left(f^{(k)} - \overline{f}\right)^2 \quad (1)$$

where $f^{(k)} = \Phi\left(x^{(k)}\right)$ is the feature extracted by the extraction function $\Phi(.)$ and $\overline{f}$ indicates the mean feature representation across the entire perturbations. Features with low variance are deemed stable and reliable, while those with high variance are considered noise-sensitive and less informative. To promote robustness, the feature vector is weighted using a stability weighting mechanism, as shown in Equation (2).

$$f^* = f \odot exp(-\lambda \times Var(f)) \quad (2)$$

where f^* is the stability-aware feature vector and λ is a scaling parameter controlling the strength of stability enforcement, and $\times$ denotes element-wise multiplication. The SDFE mechanism is incorporated into three complementary feature-extraction branches (EfficientNetV2, EfficientFormer, and MLP-Mixer), followed by a KAN classification head to model nonlinear decision boundaries. The detailed architecture and implementation of each branch are given in the following subsections.

It is important to note that SDFE does not completely discard high-variance features. Instead, the exponential stability component acts as a soft weighting mechanism that reduces the influence of responses susceptible to perturbations, while retaining their impact when they remain useful for making the classification decision. This aspect becomes especially vital in mammography, where a lesion, a low-contrast mass, or even an abnormality in a densely packed mammary gland may exhibit variability when subjected to intensity and contrast perturbations. Hence, the stability factor λ determines the extent of reduction to avoid excluding any pattern that is valuable from a clinical perspective but shows variance.

3.3. KAN-Based Classification

The benign and malignant features exhibit highly nonlinear, overlapping distributions. Traditional fully connected layers classify features using fixed-weight linear transformations followed by simple nonlinear activations, limiting their ability to model complex feature interactions and the ambiguous decision boundaries commonly observed in breast imaging. In addition, this limitation restricts their classification performance in identifying subtle lesions, low-contrast masses, architectural distortions, and dense tissue patterns. To overcome this limitation, KAN replaces static weight-based mappings with adaptive spline-based functional approximations. Following the successful SDFE refinement, the stability-ware feature vector is passed to the KAN-based classification layer. Equation (3) highlights the classification of benign-malignant classes using the KAN classifier.

$$p = \sigma\left(\sum_{q=1}^{Q}\varphi_q\left(\sum_{i=1}^{n}\varnothing_{q,i}(Z_i)\right)\right) \quad (3)$$

where $\varnothing_{q,i}(.)$ and $\varphi_q(.)$ are learnable spline-based univariate functions, Q denotes the number of functional mappings, and $\sigma(.)$ represents the activation function that generates the final probability output (p). The KAN-based method is advantageous in the application of breast cancer diagnosis compared to standard FC classifier heads. In particular, the KAN model can offer three key benefits in terms of breast cancer detection. Firstly, the proposed method improves the performance of the classifier in generating nonlinear models for mammograms that may exhibit subtle or ambiguous patterns. Secondly, the functional maps derived using splines enable easier adaptation to imaging heterogeneity

and perturbation awareness in the extracted features from SDFE. Lastly, the KAN model boosts the expressiveness of the classification function without increasing its size.

3.4. CNN-Based Feature Extraction Branch

The first branch uses EfficientNetV2 to extract the local spatial and morphological features of mammography images. EfficientNetV2 is selected for its optimized convolutional scaling strategy, which enables efficient boundary, tissue, and microcalcification extraction from fine-grained features while maintaining computational efficiency. EfficientNetV2 passes the image through a series of convolutional layers, nonlinear activations, and downsampling operations to produce hierarchical feature maps. These feature maps successively represent low-level visual patterns and high-level semantic structures. Figure 3 illustrates feature extraction using a CNN model.

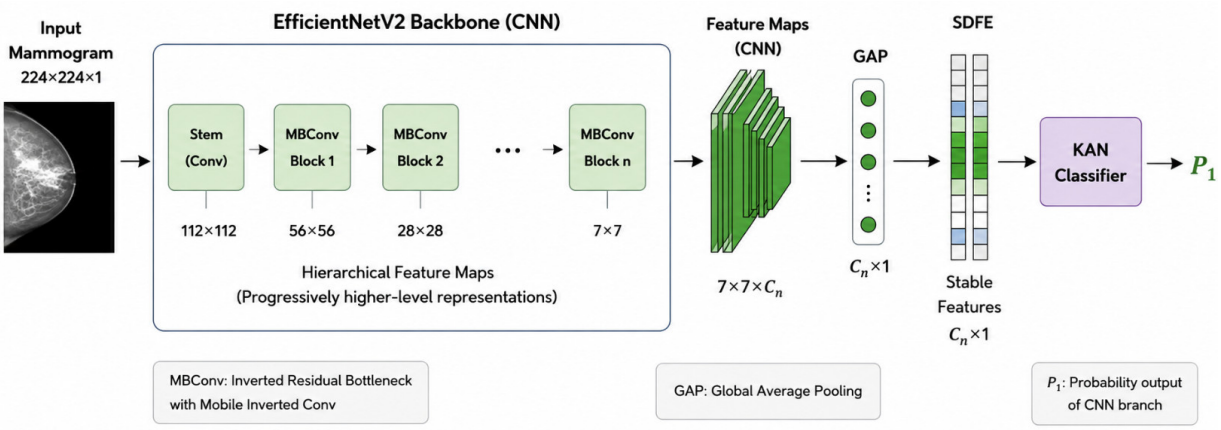


Figure 3. CNN-based Feature Extraction.

The final feature map is aggregated using global average pooling to obtain a compact feature vector. Equation (4) outlines the process of improving the robustness of the EfficientNet V2's features using the SDFE mechanism.

$$f^*_{cnn} = f_{cnn} \odot exp(-\lambda_{cnn} \times Var(f_{cnn})) \quad (4)$$

where λ_{cnn} controls the stability weighting for the CNN branch, suppressing unstable spatial activations while preserving consistent lesion-related features. Subsequently, the stability-aware feature vector (f^*_{cnn}) is passed to a KAN classification head, replacing the conventional fully connected layer. The classification head models nonlinear relationships between features and class labels, generating the probability outcome (p_1).

Early convolutional layers capture low-level visual patterns such as edges and intensity gradients, whereas deeper MBConv stages learn higher-level semantic representations associated with lesion morphology and contextual breast tissue characteristics. The proposed SDFE approach suppresses perturbation-sensitive feature responses caused by noise, contrast variation, and acquisition-related heterogeneity. This is particularly important in mammography, where dense tissue regions and low-contrast abnormalities may introduce unstable activations in conventional CNN models. The use of a KAN-based classification head enhances nonlinear decision modeling beyond traditional fully connected classifiers. The combination of hierarchical convolutional learning, perturbation-aware feature stabilization, and nonlinear functional approximation enables the proposed CNN branch to produce more discriminative and robust mammographic representations compared to conventional CNN-based breast cancer detection approaches.

Furthermore, to minimize computational cost, transfer learning is used by initializing EfficientNetV2 with pre-trained weights. The convolutional layers with generic low-level features, such as edges and textures, are frozen during training, while only the higher layers are fine-tuned to adapt to mammographic features. This strategy is to reduce the number of trainable parameters and increase the convergence speed. Additionally, EfficientNetV2's optimized compound scaling further improves computational efficiency compared to conventional CNN architectures.

3.5. ViT-Based Feature Extraction Branch

To capture global contextual relationships, the second branch leverages EfficientFormer, a light transformer variant that combines the efficiency of convolutional networks with a transformer-based attention mechanism. EfficientFormer is chosen for its ability to combine transformer-based global feature extraction with efficient token processing. Unlike traditional ViT, which has quadratic attention complexity, it uses efficient attention mechanisms and lower-dimensional token representations, resulting in a significant decrease in computational cost. It is especially suitable for modeling long-range dependencies across mammographic images while maintaining reduced computational complexity. Figure 4 presents the feature extraction process using the ViTs architecture.

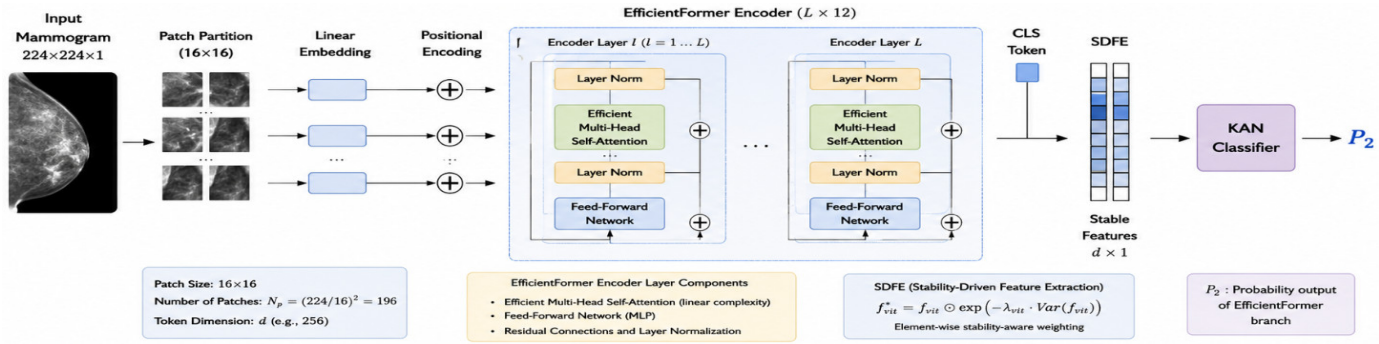


Figure 4. EfficientFormer-based feature extraction branch with SDFE refinement and KAN classification.

Initially, the architecture divides the mammogram (x) into patches of fixed size, converting these patches into token representations with additional positional information. Tokenized patches are fed through several EfficientFormer encoder blocks consisting of efficient multi-head attention and feed-forward networks. The self-attention mechanism enables the model to learn relationships between spatially distant regions of the breast, thereby improving the representation of tissue asymmetry, structural distortion, and contextual lesion distribution. This process allows the model to capture interactions between the distant regions of the breast, which are important in finding structural abnormalities. Equation (5) represents the feature enhancement process using the SDFE mechanism.

$$f^*_{vit} = f_{vit} \odot exp(-\lambda_{vit} \times Var(f_{vit})) \quad (5)$$

where f^*_{vit} is the stability-aware feature vector derived from the EfficientFormer's feature (f_{vit}). λ_{vit} regulates the stability constraint for the transformer branch. Moreover, pre-trained weights are used, and only the final blocks of the transformer model are fine-tuned, and earlier layers are frozen to overcome the training complexity. Using the KAN classification head, the probability output p_2 is produced by leveraging f^*_{vit}.

The stability of transformer-based representations is improved through SDFE. Instead of directly using attention-derived features, the perturbation-aware weighting mechanism evaluates feature consistency under controlled imaging variations, including contrast vari-

ations, brightness perturbations, and rotation transformations. This reduces sensitivity to acquisition-related variability and suppresses unstable attention responses that may arise in dense breast tissue or low-contrast mammograms. In addition, transfer learning with partial fine-tuning stabilizes optimization by preserving generalized low-level representations while adapting higher transformer layers to mammographic characteristics.

3.6. MLP-Based Feature Extraction Branch

The third branch uses MLP-Mixer to capture complex feature interactions with a pure MLP-based architecture. Unlike CNNs and ViTs, MLP-Mixer uses token-mixing and channel-mixing operations to capture relationships across spatial and channel dimensions, providing a complementary representation pathway. Figure 5 represents the proposed feature extraction mechanism using the MLP branch.

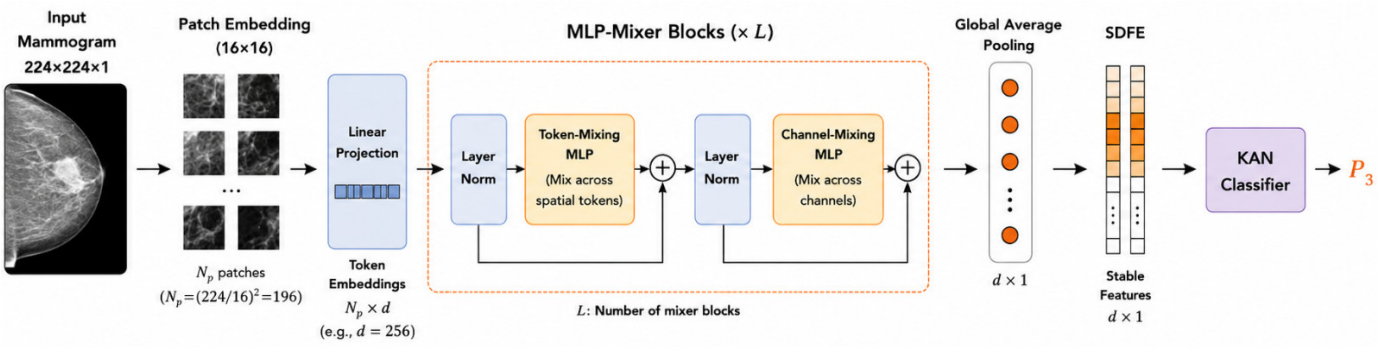


Figure 5. MLP-Mixer-based feature extraction branch with token-mixing, channel-mixing, SDFE refinement, and KAN-based classification.

The input image is split into non-overlapping patches and projected into embedding vectors. These embeddings are passed through alternating MLP layers, where token-mixing layers capture interactions between tokens at different spatial locations, and channel-mixing layers model relationships between tokens from different feature channels. This dual mixing strategy enables the branch to model high-order feature interactions associated with tissue organization, lesion distribution, and subtle mammographic patterns. Since the architecture relies primarily on matrix multiplication operations rather than convolution or quadratic attention, it achieves improved computational efficiency while preserving representation diversity. Equation (6) indicates the stability-ware feature vector (f^*_{mlp}) extraction using the MLP-Mixer feature (f_{mlp}).

$$f^*_{mlp} = f_{mlp} \odot exp\left(-\lambda_{mlp} \times Var\left(f_{mlp}\right)\right) \tag{6}$$

where λ_{mlp} controls the stability weighting for the MLP branch. The MLP-Mixer branch is intrinsically computationally efficient since it avoids both convolution and attention operations and relies on straightforward matrix multiplications to mix features. To further minimize computational cost, pre-trained weights are used, with only higher mixer layers fine-tuned while earlier layers are frozen. This strategy reduces the number of trainable parameters while maintaining the ability to model high-order feature interactions, making the MLP-Mixer branch suitable for efficient representation learning. Finally, f^*_{mlp} is processed through the KAN classification head, resulting in the probability output (p_3). The combination of token-channel interaction learning, perturbation-aware feature stabilization, and nonlinear classification allows the MLP-Mixer branch to generate stable and computationally efficient mammographic representations that complement the CNN and transformer branches within the proposed framework.

3.7. Ensemble Learning-Based Breast Cancer Classification

Following the SDFE and branch-specific (KAN) classifications, the proposed framework combines the predictions from the three branches using an adaptive voting-based ensemble mechanism. The ensemble stage combines the outputs of the CNN-KAN, ViT-KAN, and MLP-KAN classifiers. Each branch captures complementary representations of mammographic images: the CNN branch focuses on local spatial structures, the ViT branch captures long-range contextual dependencies, and the MLP branch models abstract nonlinear feature interactions. As a result, combining their predictions enables the system to use different views of the mammographic data, making the final diagnostic decision more reliable.

Let p_1, p_2, and p_3 be the probabilities estimated by the CNNs-KANs, ViTs-KANs, and MLP-KANs branches, respectively. Instead of traditional ensemble approaches such as simple averaging or fixed-weighted fusion, the proposed framework employs a confidence-aware adaptive voting mechanism. The rationale for this design is that each model may have a different predictive reliability depending on the characteristics of the input mammogram, selecting the most reliable prediction for an individual case. In probabilistic classification, the model outputs a probability indicating how likely a sample is to belong to a particular class. The decision boundary typically lies at a probability of 0.5, where the model is uncertain about the class. Predictions with probabilities near this boundary (e.g., 0.45 or 0.55) carry low confidence, since minor variations in the input can easily change the predicted class. On the other hand, probabilities close to 0 or 1 (e.g., 0.1 or 0.9) indicate high confidence, as the model assigns a single class with high probability. Therefore, the distance of the predicted probability from the decision boundary is a measure of certainty, with extreme values indicating reliable predictions. The probability of the final prediction is computed as in Equation (7).

$$P_{final} = max(p_1, p_2, p_3) \tag{7}$$

The adaptive voting mechanism is applied only at the final decision stage to select the branch with the highest confidence margin. Importantly, this mechanism functions as a soft reliability-aware selection strategy. Complementary information is captured during the preceding feature extraction and branch-wise classification stages, where the CNN, ViT, and MLP branches independently learn local, global, and interaction-based representations. To reduce instability in ambiguous cases, confidence is estimated using the separation between the highest and second-highest predicted probabilities, preventing weak-margin predictions from dominating the decision.

The proposed adaptive voting mechanism has several advantages over conventional ensemble strategies, including hard and soft voting. Traditional hard voting is based on majority-class selection, where all models are treated equally, and prediction confidence is ignored, leading to suboptimal decisions when weaker models have greater influence. Similarly, soft voting averages the models' prediction probabilities, assuming they are equally reliable, which may dilute strong predictions with uncertain ones. In contrast, the proposed approach enables dynamic model selection by prioritizing the most confident prediction for each case. This makes the decisions more robust by reducing the impact of low-confidence outputs while preserving the strengths of individual branches. Furthermore, the mechanism imposes very little computational overhead, grounding itself in confidence evaluation rather than requiring additional meta-learners or intricate stacking architectures, resulting in an efficient and scalable ensemble framework.

Furthermore, by integrating SDFE, branch-specific KAN classification heads, and a confidence-aware adaptive voting method, the proposed model constructs an efficient

ensemble architecture that leverages complementary feature representations to achieve reliable breast cancer classification.

3.8. Explainability Analysis

Explainability analysis is integrated into the proposed model through branch-specific visualizations that align with each branch's internal structure. To improve interpretability and address the black-box nature of the proposed architecture, branch-specific explainability analyses are embedded in the approach. As the proposed architecture comprises heterogeneous models, different methods for explanations are considered based on the type of representation utilized by each branch. Grad-CAM is used for CNNs branch to localize the discriminative spatial regions, attention rollout is used for ViTs branch to identify the influential image patches, and SHAP-based attribution is used for MLP branch to quantify the feature-level contributions. For consistency, all explanations are shown in the form of heatmap overlays on the input mammogram. In addition, the generated visualizations are reviewed by experienced radiologists to ensure that the highlighted regions match clinically relevant features. This combined quantitative and expert-guided evaluation framework ensures that the explainability component remains consistent with both the model's behavior and its decision-making. To quantitatively evaluate explanation quality, insertion and deletion score analysis is computed. Let $S(x)$ denotes the prediction confidence for image x. The insertion score evaluates whether highlighted regions positively contribute to the prediction by progressively inserting highly relevant regions into a blurred baseline image as shown in Equation (8).

$$IS = \frac{1}{N}\sum_{i=1}^{N} S\left(x_i^{ins}\right) \tag{8}$$

where x_i^{ins} represents the partially reconstructed image obtained after inserting the top-ranked explanatory regions. Likewise, Equation (9) outlines the deletion analysis that progressively removes crucial regions from the input image.

$$DS = \frac{1}{N}\sum_{i=1}^{N}\left[1 - S\left(x_i^{del}\right)\right] \tag{9}$$

where x_i^{del} represents the image after deleting highly relevant regions. Higher insertion and deletion scores indicate stronger alignment between the generated explanation maps and the model decision process.

3.9. Performance Evaluation

Evaluating breast cancer classification models requires a comprehensive set of performance metrics to ensure diagnostic accuracy and reliability. Accuracy is an overall measure of correct predictions; however, it might not be sufficient for imbalanced datasets. Therefore, precision and recall are of major importance: precision indicates the ability to reduce false positives, whereas recall (sensitivity) refers to the correct identification of malignant cases, which is important in clinical settings. The F1-score provides a balanced measure by combining the precision and recall. Furthermore, AUROC measures the model's discriminative power across all thresholds, whereas AUPRC is more informative in imbalanced settings by focusing on the positive (malignant) class. To ensure generalizability, statistical measures, such as standard deviation and confidence intervals, are incorporated to provide insights into performance consistency and reliability with different data splits.

4. Results

Table 2 explains the primary components of the pipeline: input preprocessing; feature extraction using EfficientNetV2, EfficientFormer, and MLP-Mixer; SDFE application; and the combination of KAN classification heads. In addition, it presents the ensemble strategy, training settings, and evaluation protocols. These configurations are carefully selected to balance computational efficiency and model performance, ensuring consistency with the proposed methodology and reliable implementation across a range of experimental environments.

Table 2. Configuration for implementing Breast Cancer Classification.

Component	Module	Configuration
Hardware Environment	Operating System	Windows 11 Pro 64-bit
	Processor	Intel Core i7+
	GPU	NVIDIA RTX-Series GPU with CUDA acceleration
	GPU Memory	12–24 GB VRAM
	System RAM	32 GB DDR5
Software Toolchain	Programming Language	Python 3.10
	Deep Learning Frameworks	PyTorch 2.1, TensorFlow 2.15
	CUDA/cuDNN	CUDA 12.1, cuDNN 8.9
	Scientific Libraries	NumPy 1.26.4, Pandas 2.1.4, Scikit-learn 1.3.2
	Imaging Libraries	OpenCV 4.8.1, PIL 10.1.0, Albumentations 1.4.0
	Visualization Libraries	Matplotlib 3.8.2, SHAP 0.44.1
Input Configuration	Image Resolution	224×224
	Image Format	Grayscale converted to 3-channel
	Normalization	Min–Max scaling to [0, 1]
Preprocessing	Contrast Enhancement	CLAHE (clip limit = 2.0, tile grid = 8×8)
	Resizing	Bilinear interpolation
Data Augmentation (Training Only)	Policy	RandAugment (N = 2, M = 9)
	Transformations	Rotation ($\pm 10^{\circ}$), Horizontal Flip, Brightness ($\pm 15\%$), Contrast ($\pm 15\%$)
Feature Extraction (CNNs Branch)	Backbone	EfficientNetV2-S (ImageNet pretrained)
	Output Feature Dimension	Global Average Pooling $\rightarrow$ 1280-d
	Frozen Layers	First 70% layers frozen
Feature Extraction (ViTs Branch)	Backbone	EfficientFormer-L1 (pretrained)
	Tokenization	Implicit patch embedding (~16×16)
	Output Feature Dimension	CLS token/pooled $\rightarrow$ 1024-d
	Frozen Layers	First 60% layers frozen
Feature Extraction (MLP Branch)	Backbone	MLP-Mixer-B/16 (pretrained)
	Patch Size	16×16
	Output Feature Dimension	Global pooling $\rightarrow$ 768-d
	Frozen Layers	First 60% layers frozen
SDFE	Number of Perturbations	(K = 3)
	Perturbation Types	Brightness, Contrast, Rotation
	Stability Parameter	$\lambda = 0.5$
	Application Stage	After feature extraction and before KAN classification

Table 2. *Cont.*

Component	Module	Configuration
KAN-Based Classification	Classification Head	KAN replacing conventional FC layer
	Basis Function	Cubic spline
	Grid Size	10
	Hidden Mapping Structure	Single hidden functional mapping layer
	CNNs Branch Output	(p_1)
	ViTs Branch Output	(p_2)
	MLP Branch Output	(p_3)
Adaptive Ensemble Learning	Ensemble Strategy	Confidence-aware adaptive voting
	Decision Rule	Dynamic selection using max-confidence branch prediction
Training Configuration	Optimizer	Adam
	Initial Learning Rate	1×10^{-4}
	Batch Size	16
	Epochs	50
	Loss Function	Binary Cross-Entropy
	Random Seed	42
Validation and Testing	Validation Strategy	5-fold cross-validation (80% training subset)
	Internal Testing	20% held-out VinDr-Mammo test set
	External Validation	Patient-level evaluation on 20% of the CBIS-DDSM dataset

Table 3 indicates that the proposed framework is robust and consistent with varying data splits, which indicates its high generalization performance. The resulting low standard deviation and high mean accuracy of 94.5 are pointers that the model is not only very accurate but also remains stable across different training conditions. This result is explained by the incorporation of SDFE, which focuses on invariant and reliable properties, and by the application of heterogeneous feature representations by the CNN, ViT, and MLP branches. Moreover, the classification head, based on KANs, facilitates modeling nonlinear decisions and better discriminates between benign and malignant cases. This consistent performance is especially important in breast cancer detection, where reliability and reproducibility are required for a real-time environment. The smaller value differences within folds indicate that the model is not sensitive to data differences, thus allowing its use in real-world screening and enhancing trust in automated diagnostic assistance systems.

Table 3. Five-Fold Cross-Validation Performance of the Proposed Model (Train Set).

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Fold 1	93.8	93.4	94.5	93.9
Fold 2	95.2	94.8	95.6	95.2
Fold 3	94.6	94.2	95.0	94.6
Fold 4	93.9	93.5	94.4	93.9
Fold 5	94.9	94.6	95.3	94.9
Mean ± SD	94.5 ± 0.5	94.1 ± 0.5	95.0 ± 0.4	94.5 ± 0.5
95% CI	93.9–95.1	93.5–94.7	94.5–95.5	93.9–95.1

Table 4 offers insights into why the proposed framework is superior to individual components and does not merely deliver incremental gains. Significantly, the difference in performance between single-branch models (CNNs, ViTs, MLPs) and their variants with combinations indicates that no single architecture can capture the full range of mam-

mographic properties. The local texture patterns are mainly described by CNNs, the global contextual relationships are depicted by ViTs, and the representations based on MLPs provide additional feature interactions. This feature diversity allows the ensemble to gradually improve performance, implying that the gain is not due to redundancy but rather to the complementarity of representations. More significantly, the shift from simple ensemble combinations to the introduction of SDFE shows that improvements in performance come from feature reliability, not their quantity. The observed stability gains under SDFE confirm that stabilizing properties under perturbation minimize noisy or misleading representations, which are essential in heterogeneous medical data. The implementation of the KAN-based classifier also shows that the bottleneck in current methods is not only in feature extraction but also in the ability to model decisions. The enhancement in incorporating KANs implies that nonlinear functional approximation is essential for handling delicate decision boundaries between benign and malignant cases. Lastly, the overall improved performance of the full model can testify that the suggested strategy is no mere aggregation strategy, but rather a systematic combination of complementary capabilities, stability-conscious learning, and adaptive decision modeling, which together would boost the reliability of diagnostic ability beyond the capabilities of individual components.

Table 4. Findings of Ablation Study (VinDr-Mammo Test Set).

Model Variant	Classifier/ Decision Head	SDFE	Ensemble	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	95% CI Accuracy	*p*-Value vs. Proposed
CNNs (EfficientNetV2)	FC	×	×	88.3	87.6	89.1	88.3	86.3–90.1	<0.001
ViTs (EfficientFormer)	FC	×	×	89.1	88.4	90.0	89.2	87.2–90.9	<0.001
MLP (MLP-Mixer)	FC	×	×	87.5	86.8	88.6	87.7	85.4–89.4	<0.001
CNNs + ViTs	FC	×	✓	90.4	89.8	91.5	90.6	88.6–92.1	0.004
CNNs + ViTs + MLP	FC	×	✓	91.3	90.7	92.4	91.5	89.6–92.9	0.012
CNNs + ViTs + MLP	Residual MLP Head	✓	✓	92.1	91.5	93.0	92.2	90.5–93.5	0.038
CNNs + ViTs + MLP	RBF-SVM Head	✓	✓	91.8	91.2	92.8	92.0	90.1–93.2	0.026
CNNs + ViTs + MLP	Kernel Ridge Head	✓	✓	91.6	91.0	92.5	91.7	89.9–93.0	0.019
CNNs + ViTs + MLP	KAN Head	✓	×	92.5	91.9	93.5	92.6	90.9–93.8	0.041
Proposed Model (Full)	KAN Head	✓	✓	93.3	92.8	93.9	93.3	91.8–94.7	—

Note: × = Not Enabled, and ✓ = Enabled.

Figure 6 shows the performance of the proposed model by class on the VinDr-Mammo test set and CBIS-DDSM external validation dataset. These findings show high accuracy, precision, recall, and F1-score for benign and malignant classes, indicating balanced classification. Interestingly, the model performs similarly across classes, suggesting it does not favor benign or malignant predictions, which is paramount in decision-making. Moreover, the minor improvement in the CBIS-DDSM dataset indicates strong generalization of the model across datasets of varying nature. The consistency of performance in internal and external assessments underscores the strength of the proposed framework. This can be attributed to the combination of heterogeneous feature representations, feature extraction stability, and adaptive decision modeling, which, in totality, enable reliable discrimination of breast cancer cases across different imaging conditions.

In order to evaluate the effectiveness of different ensemble techniques, additional experiments were conducted using hard voting, soft voting, weighted soft voting, and the proposed adaptive voting technique. Table 5 outlines the performance of these techniques. The findings indicate that hard voting produced the lowest performance because it considers only majority decisions and ignores each branch's confidence. Soft voting improved performance by averaging probability outputs. However, uncertain predictions from weaker branches may reduce the influence of highly confident predictions. Weighted

soft voting further improved performance by assigning branch-level importance, but the weights remain fixed across samples. In contrast, the proposed adaptive voting mechanism dynamically selects the most confident branch prediction for each mammographic case. This case-specific selection enables the model to reduce the influence of uncertain predictions. It improves overall classification stability, achieving the highest accuracy, precision, recall, and F1-score, confirming its effectiveness over conventional ensemble methods.

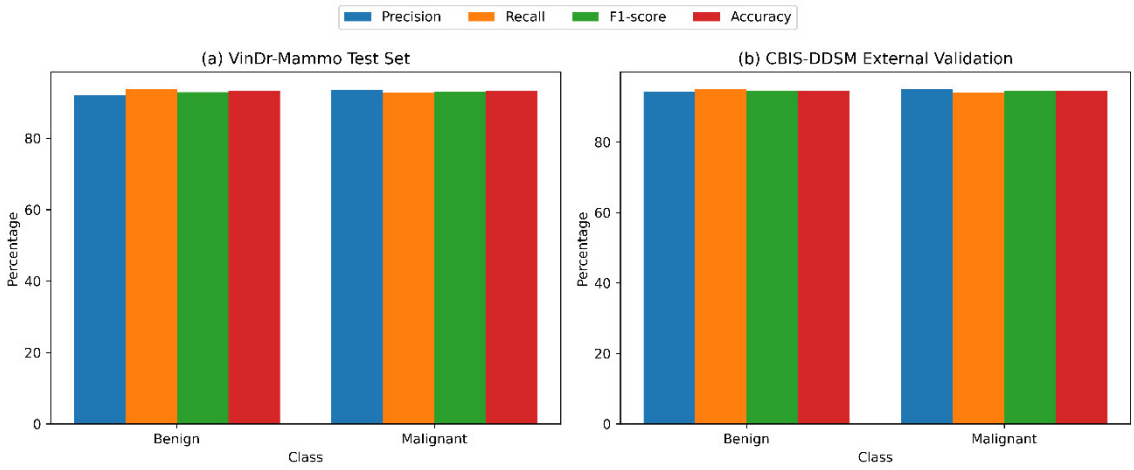


Figure 6. Class-wise Performance Comparison of the Proposed Model on (**a**) VinDr-Mammo Test Set and (**b**) CBIS-DDSM External Validation.

Table 5. Effect of Different Ensemble Strategies with Statistical Analysis on VinDr-Mammo Test Set.

Ensemble Strategy	Decision Rule	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	95% CI Accuracy	*p*-Value vs. Proposed
Hard Voting	Majority class decision	91.7	91.0	92.3	91.6	90.0–93.3	0.006
Soft Voting	Average branch probabilities	92.3	91.7	93.0	92.3	90.7–93.8	0.021
Weighted Soft Voting	Fixed weighted probability averaging	92.7	92.1	93.4	92.7	91.1–94.1	0.048
Adaptive Voting (Proposed Model)	Confidence-based branch selection	93.3	92.8	93.9	93.3	91.8–94.7	—

The performance of the proposed breast cancer classification framework in Figure 7a,b is demonstrated in terms of Area Under the Receiver Operating Characteristic (AUROC) and Precision–Recall curves. The AUROC curve shows high discriminative ability, with an AUROC score of 0.974, indicating that the model can differentiate between benign and malignant cases across different decision thresholds. The curve is close to the top-left corner, indicating a high true positive rate and very few false positives. Similarly, the Precision–Recall curve shows an AUPRC of 0.948, indicating that the model performs well at preserving high precision across a broad range of recall rates. This is especially important in medical diagnosis, where it is crucial to minimize false positives while maximizing sensitivity. The higher performance can be attributed to the proposed integration of stability-driven feature extraction, heterogeneous feature representations, and KAN-based nonlinear classification. Ultimately, these components make the features robust and reliable for decision-making, resulting in better classification performance than conventional approaches.

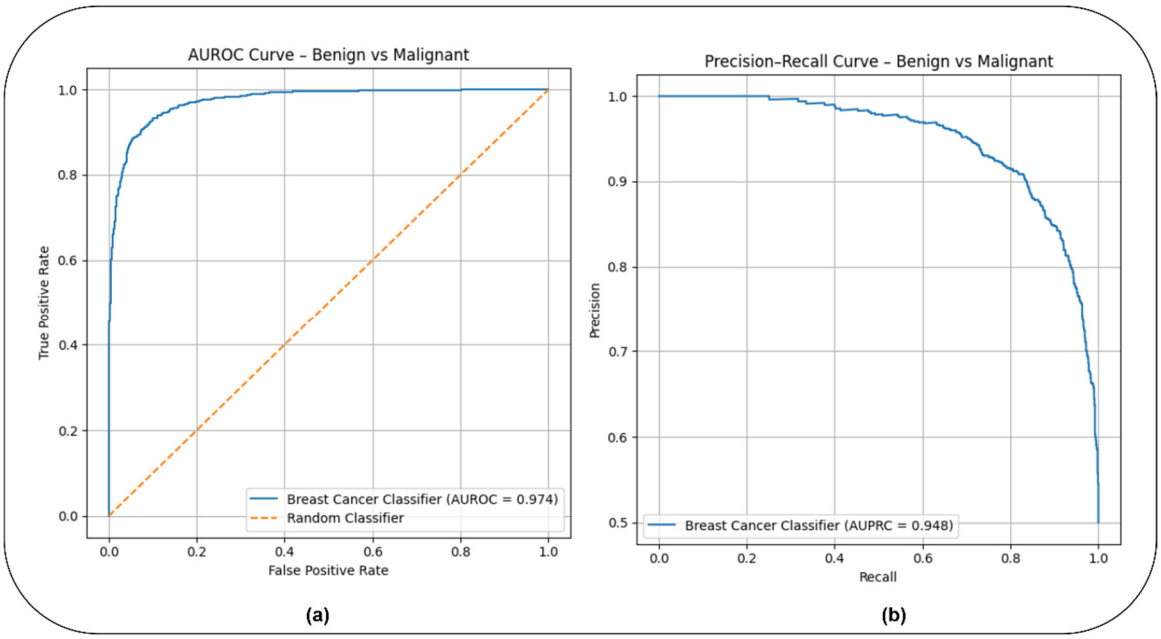


Figure 7. (**a**) AUROC and (**b**) Precision–Recall Curve.

The selection of Studies [41–48] is based on their methodological relevance and on recent developments in automated breast cancer detection. These studies examine prevailing architectural paradigms, such as the transformer paradigm, hybrid CNN-ViT, region-aware learning, and ensemble strategies. By incorporating various feature extraction algorithms—from convolutional feature representations to attention-based global modelling—they represent the current research trends in medical imaging. While these methods achieve better diagnostic performance, they also have drawbacks, such as computational complexity, improper feature fusion, and limited expressiveness of conventional classifiers. Consequently, their inclusion provides a rigorous criterion for evaluating the proposed framework and highlights its contributions to improved feature integration and effective nonlinear decision modeling.

A fundamental weakness of current strategies for mammography classification is evident in the comparative analysis in Table 6, especially when applied to large-scale, heterogeneous datasets such as those in VinDr-Mammo. The proposed model achieved the highest performance across all metrics, exceeding the baseline mean by 3.58% in accuracy, 3.45% in precision, 3.40% in recall, and 3.42% in F1-score. The proposed model also outperformed the best baseline by more than 2.5% across all metrics, indicating a clear performance advantage over existing methods. Although the existing models exhibit competitive performance, their representations are highly clustered, resulting in low discriminative capacity to capture subtle differences between benign and malignant cases. This can be explained primarily by their reliance on single-model architectures or traditional ensemble approaches, which do not adequately address feature variability and the uncertainty of real-world mammographic data.

Conversely, the proposed model is consistently positioned at the outer boundary of all evaluation metrics, indicating its high potential to achieve both high sensitivity and precision. This enhancement is not simply incremental but a sign of increased decision reliability. The addition of stability-based feature extraction ensures that only stable, reliable features are used in the classification task, thereby eliminating variability caused by noise. Moreover, the heterogeneous feature representations enable the model to learn both local structural information and global context, which are crucial for precise mammography interpretation.

Table 6. Comparative Performance with Statistical Analysis.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Zeynali et al. (2024) [41]	89.5	88.7	90.1	89.4
Sharma et al. (2024) [42]	88.2	87.5	89.0	88.2
Murty et al. (2024) [43]	90.1	89.6	90.8	90.2
Firouzbakht et al. (2025) [44]	89.0	89.3	90.5	89.9
Yoshitsugu et al. (2025) [45]	90.8	90.0	91.2	90.6
Asif et al. (2025) [46]	89.5	89.5	90.1	89.5
Khan et al. (2025) [47]	90.1	90.2	91.3	90.7
Khaled et al. (2025) [32]	90.6	90.0	91.0	90.5
Baseline Mean ± SD	89.72 ± 0.86	89.35 ± 0.89	90.50 ± 0.76	89.88 ± 0.83
95% CI of Baselines	89.13–90.32	88.74–89.96	89.97–91.03	89.30–90.45
Proposed Model	93.3	92.8	93.9	93.3
Improvement over Baseline Mean	+3.58	+3.45	+3.40	+3.42
Improvement over Best Baseline	+2.50	+2.60	+2.60	+2.60

Note: The "Improvement over Baseline Mean" represents the average performance gain of the proposed model compared with the mean performance of all existing baseline methods included in the comparative analysis. In contrast, the "Improvement over Best Baseline" indicates the performance gain relative to the best-performing method in the comparison.

Notably, the observed performance difference on the VinDr-Mammo test set indicates that the proposed framework is more applicable to real-time conditions, where data heterogeneity and ambiguity are common. The proposed method is more general, resulting in a more stable decision boundary, unlike current techniques, which can overfit to a particular feature pattern, thereby enhancing the technique's practical applicability in breast cancer screening.

The effectiveness of the proposed framework is further supported by the results in Figure 8, which are subject to external validation using the CBIS-DDSM dataset. Though some current methods exhibit competitive performance, their gains are not high and tend to level off across evaluation measures, suggesting limited adaptability to different data properties. These models usually rely on a single architectural paradigm or traditional ensemble approaches and may not fully leverage complementary feature representations or adapt well to dataset variability. Notably, the enhanced performance on the CBIS-DDSM indicates that the proposed model is not over-fitted to the VinDr-Mammo dataset but rather captures clinically relevant patterns that transfer effectively across datasets.

Table 7 shows that the proposed framework achieves better diagnostic performance with minimal computational overhead. While a number of recent approaches use large-scale transformers or stacked ensemble architectures with high parameter counts and inference latency, the proposed model employs a computationally efficient design by combining small-sized backbones (EfficientNetV2, EfficientFormer, and MLP-Mixer) with parameter-efficient KAN-based classification heads. Unlike conventional ensembles that aggregate predictions via additional meta-learning layers, the proposed adaptive voting mechanism incurs negligible overhead by operating directly on confidence scores. As a result, the model has a moderate computational footprint and a better classification accuracy. This balance between performance and efficiency underscores the practical applicability of the proposed approach in a real-time environment, where computational resources and inference time are critical considerations.

Although the proposed framework employs a multi-branch architecture, its computational scalability is controlled by lightweight, parameter-efficient backbone networks, including EfficientNetV2-S, EfficientFormer-L1, and MLP-Mixer-B/16. Unlike conventional large-scale ensemble frameworks that rely on heavyweight transformer backbones or stacked meta-learning modules, the proposed approach uses partially frozen pretrained

branches, thereby reducing the number of trainable parameters and optimization overhead during training. In addition, the three branches operate independently and can be executed in parallel during inference, improving scalability on GPU-based systems. The adaptive voting mechanism introduces negligible computational overhead, as it selects decisions based on confidence without requiring additional ensemble training stages. Consequently, the framework maintains moderate GPU memory consumption and inference latency despite combining heterogeneous feature extraction pathways. These characteristics enhance the practicality of the proposed approach for large-scale mammography analysis and for real-world deployment scenarios where computational efficiency and scalability are critical.

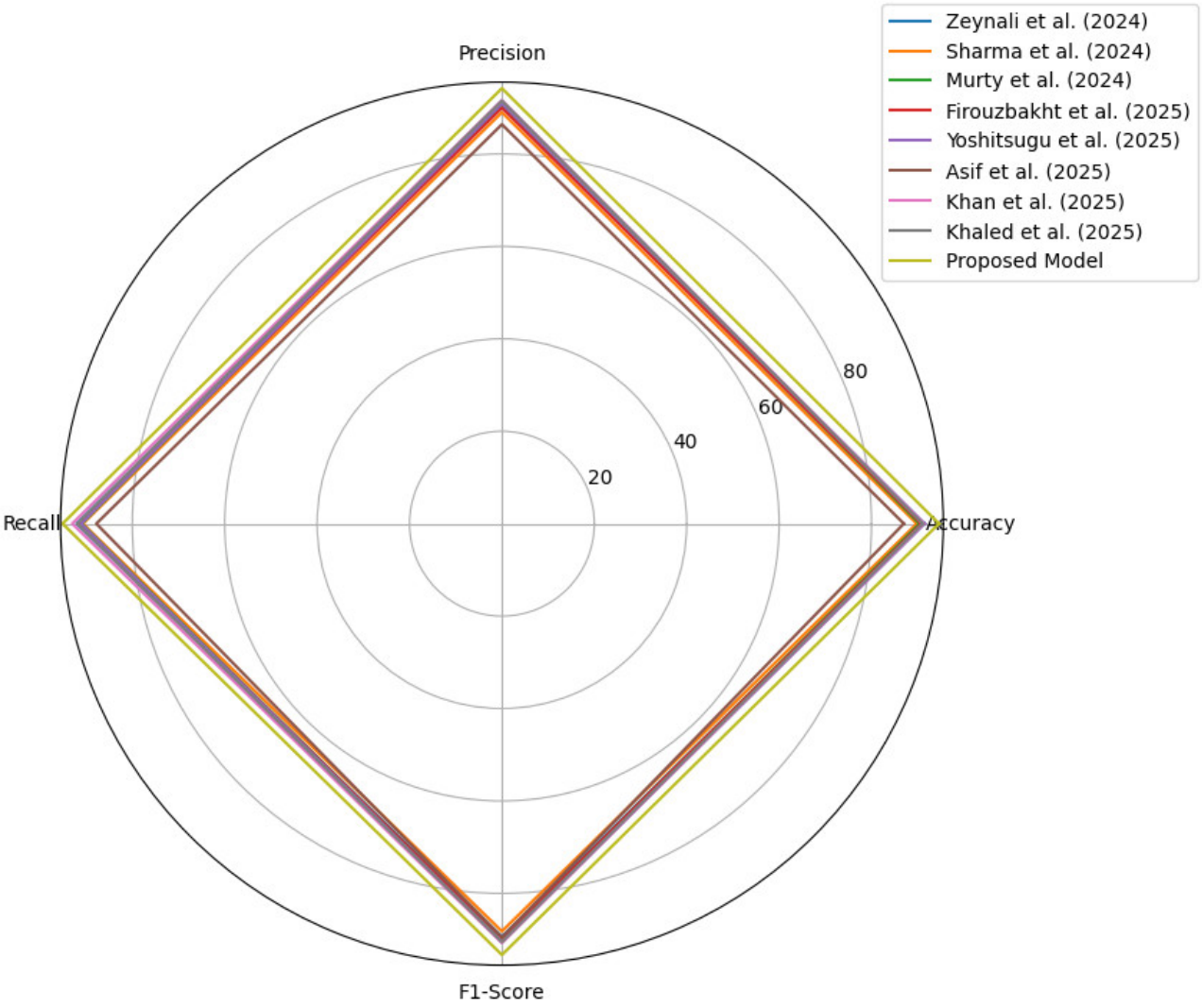


Figure 8. Findings of Comparative Analysis (CBIS-DDSM) [32,41–47].

Figures 9 and 10 demonstrate the interpretability of the proposed model through heatmap-based explanations for benign and malignant predictions. The overlay indicates the regions most important to the model's decision-making process, and the overlay images show the spatial correspondence between the influential regions and the original mammograms. In the benign case, the model is restricted to localized, well-defined regions with low spread, indicating stable, low-risk tissue characteristics. In contrast, the malignant case shows wider and higher levels of activation, and is especially activated in areas of irregular and dense tissue, which reflect crucial tumor characteristics. According to the Grad-CAM results, the CNN pathway mostly focuses on lesion borders, suspicious lesions, and local tissue deformities. The attention rollout from the EfficientFormer pathway reveals broader

tissue contexts, demonstrating the transformer model's ability to detect long-term structural relations between features. Similarly, SHAP attributions for the MLP-Mixer pathway also reveal feature interactions corresponding to diagnostic regions of mammograms.

Table 7. Computational Cost Comparison of Models.

Study	Model Type	Params (M)	FLOPs (G)	Inference Time (ms/Image)	GPU Memory (GB)	Training Time/Epoch (min)	Efficiency Level
Zeynali et al. (2024) [41]	CNNs + Transformer (Xception)	40	10	45	7.8	8.5	High cost
Sharma et al. (2024) [42]	Transformer (ViT-like)	90	18	75	11.6	14.2	Very high cost
Murty et al. (2024) [43]	Hybrid CNNs + Tabular	30	8	40	6.9	7.4	Moderate
Firouzbakht et al. (2025) [44]	CNNs + Shearlet	55	12	65	9.4	10.8	High cost
Yoshitsugu et al. (2025) [45]	CNNs	40	9	50	7.5	8.1	Moderate–high
Asif et al. (2025) [46]	CNNs	75	29	98	13.2	16.5	Very high cost
Khan et al. (2025) [47]	CNNs + ViT Ensemble	120	22	100	14.8	18.7	Very high cost
Khaled et al. (2025) [32]	CNNs + Ensemble	90	20	85	12.7	15.3	Very high cost
Proposed Model	EfficientNetV2 + EfficientFormer + MLP-Mixer + KAN	~28	8	38	6.4	7.1	Moderate/optimized

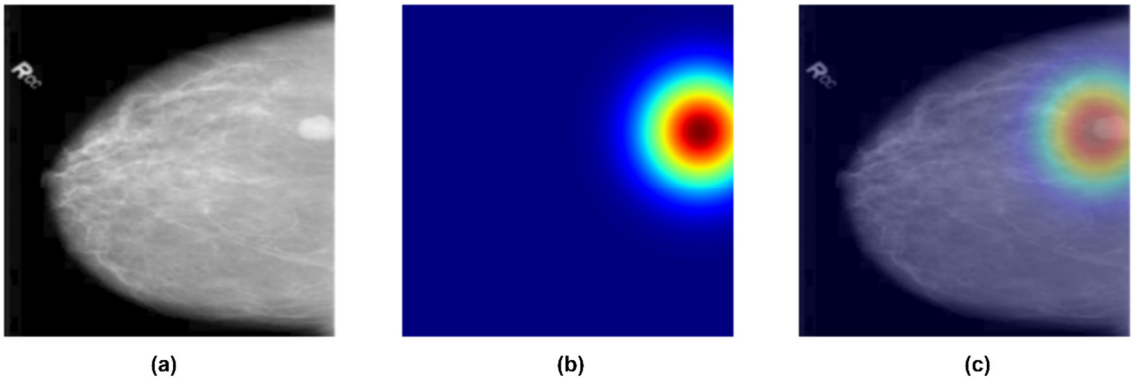


Figure 9. (**a**) Original Benign Image, (**b**) Feature-Importance Map (Predicted Benign Image), and (**c**) Heatmap-Overlay (Predicted Benign Image).

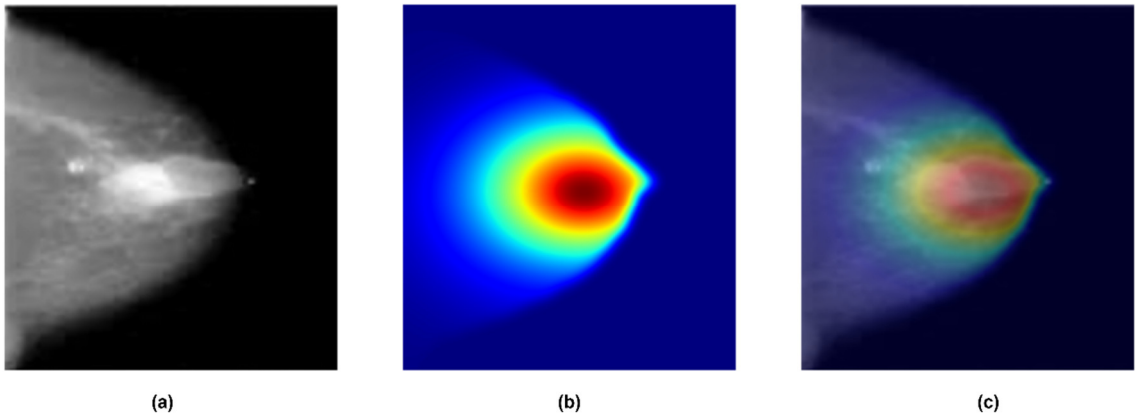


Figure 10. (**a**) Original Malignant Image, (**b**) Feature-Importance Map (Predicted Malignant Image), and (**c**) Heatmap-Overlay (Predicted Malignant Image).

To validate the reliability of the explanations, quantitative experiments (Table 8) for insertion and deletion analysis were also performed on selected benign and malignant mammogram images from both the VinDr-Mammo and CBIS-DDSM datasets. The proposed ensemble achieved the highest insertion and deletion scores, indicating strong correspondence between the highlighted regions and the model's prediction. Furthermore, the larger confidence drop observed after deleting the highlighted regions confirms that the generated explanation maps capture diagnostically influential areas rather than irrelevant background structures. These findings demonstrate that the proposed framework not only achieves strong classification performance but also provides interpretable and decision-consistent explanations, thereby improving transparency and trustworthiness in mammography classification.

Table 8. Quantitative Explainability Evaluation.

Branch	Explainability Method	Insertion Score ↑	Deletion Score ↑	Confidence Drop After Deletion ↑
CNN (EfficientNetV2)	Grad-CAM	0.86	0.81	0.42
ViT (EfficientFormer)	Attention Rollout	0.89	0.84	0.46
MLP-Mixer	SHAP Attribution	0.83	0.79	0.39
Proposed Ensemble	Combined Explanation	0.91	0.87	0.49

Note: ↑ indicates that a higher value is better.

5. Discussions

The proposed study enhances breast cancer classification by developing an advanced framework for representation learning based on stability and complementarity. Unlike prior approaches, which require either a homogeneously learned feature representation or a fixed ensemble model, the framework combines CNN-, ViT-, and MLP-based representations with an SDFE, a KAN-based nonlinear classifier head, and adaptive voting-based ensemble decision modeling. The empirical findings validate the proposed framework and support the underlying design rationale. The ablation study demonstrates that each proposed component contributes to progressive performance improvement, while comparative analysis on both VinDr-Mammo and CBIS-DDSM confirms superior generalization and balanced benign–malignant classification performance. Furthermore, the low variance observed in five-fold cross-validation validates the framework's robustness and stability under varying data distributions, highlighting its effectiveness for reliable mammography classification.

The proposed approach outperforms current mammography-focused artificial intelligence models, which have limitations in feature extraction, classification, and fusion. While the existing methods mostly focus on improving accuracy through deeper or larger models, the present study is more concerned with the reliability of the features, the expressiveness of the decisions, and the adaptability of the decision fusion. The achieved AUROC of 0.974 and AUPRC of 0.948 on a large, diverse dataset (VinDr-Mammo) demonstrate that the proposed system maintains a high degree of discriminative ability across a wide range of mammographic appearances. In particular, AUROC is a preferred metric for screening programs as it accounts for results across different possible operating decision thresholds and is less affected by class imbalance than accuracy. The reported accuracy of 93.3% on the VinDr-Mammo test data confirms that the model achieves a high degree of classification accuracy.

The increase in accuracy to 94.6% on the CBIS-DDSM during external validation is promising but remains within an acceptable range. The performance gain on the CBIS-DDSM dataset does not reflect overfitting or data leakage, but rather results from reduced data variability compared to the VinDr-Mammo dataset. While clinical datasets are usually

diverse, with varying imaging protocols, patient populations, and imaging features, the CBIS-DDSM dataset is homogeneous, leading to higher model performance in external validation. This highlights the stability of the proposed model and its ability to perform reliably across datasets with varying levels of complexity and distributional bias.

A limitation of most existing mammography-based artificial intelligence studies [41–48] is that they are based on relatively small or homogeneous datasets. Artificial intelligence models trained on small datasets are often impressive in one domain but degrade on other datasets. This undermines their transferability and practical utility for screening programs. In contrast, the proposed model was trained on a large and diverse dataset containing mammograms with varying image quality, density, lesion type, and acquisition conditions. This variability in the training helps the model to learn features that are less susceptible to dataset biases and artifacts. This results in more informative learned features. Exposure to this diversity in the training allows the model to learn representations that are less sensitive to spurious correlations and dataset-specific artifacts. As a result, the learned features are more representative of the pathological patterns.

Another important factor that contributes to better performance is the introduction of SDFE-based feature extraction. Existing CNN- and ViT-based approaches [41–48] typically rely on saliency-driven or attention-based mechanisms that favor features with high activations. While effective in diverse scenarios, such mechanisms do not explicitly guarantee the stability of selected features against perturbations typically encountered in clinical practice, such as noise, contrast variation, or scanner differences. By highlighting feature consistency under controlled perturbations, the proposed approach suppresses unstable or unreliable features and encourages representations that are more robust to domain shift. The use of a KAN-based classification head further helps the model distinguish between benign and malignant cases, especially near decision boundaries. Conventional linear classification layers enforce simple decision surfaces that may not be adequate for modeling the complex, nonlinear relationships that exist in mammographic data. By introducing a more expressive classifier while maintaining a computationally efficient architecture, the proposed framework improves discrimination without resorting to deeper or more computationally intensive architectures.

Lastly, the adaptive voting-based ensemble strategy plays a critical role in the improvement of reliability. A majority of existing ensemble methods are based on static fusion methods, and implicitly assume that each model contributes equally to all cases. In practice, individual models may differ in their strengths based on breast density, lesion visibility, or image quality. By dynamically weighting model predictions based on confidence and inter-model diversity, the proposed ensemble mechanism effectively reduces the influence of less reliable predictions on a case-by-case basis. This adaptive fusion strategy leads to more stable decision-making by prioritizing the most informative model outputs for each instance. Consequently, the model exhibits enhanced robustness and consistent performance across datasets with different characteristics, reinforcing its ability to generalize beyond dataset-specific patterns.

One challenge in the existing literature on breast cancer detection models is their limited applicability. Many studies [32,41–47] demonstrate promising performance metrics using datasets comprising a few hundred to thousands of samples, mostly originating from a single institution or scanner type. While methodological advances in breast cancer research allow for the development of effective models that facilitate detection, their performance in a real-world setting remains uncertain. In contrast, the present study proposes a technique that trains its model on a substantially large, heterogeneous dataset and explicitly validates it using an external benchmark. High AUROC and accuracy in

both datasets demonstrate the generalizability of the proposed framework compared to the existing approaches built and validated on small datasets.

The results of the current study have several important implications for real-time settings. The proven stability and strong generalization ability make the proposed approach a useful aid in radiologists' decision-making, especially during high-workload screening. A focus on reliability in addition to accuracy aligns well with the goals of clinical practice. By offering reliable, stable predictions, the proposed framework has the potential to decrease inter-observer variability and ensure consistent decision-making. Additionally, the proposed approach addresses screening needs by prioritizing model reliability over accuracy. Screening entails a different set of goals, as it aims to detect cancers rather than precisely quantify lesion characteristics. The high AUROC of the suggested model suggests it can be adjusted for screening to achieve desired levels of sensitivity or specificity, depending on clinical requirements. Moreover, because it is resource-efficient and uses a shallow architecture, the framework can be deployed in environments with limited computational power without any issues.

While the proposed approach has several strengths, it also has certain limitations. While the training dataset is sufficiently large and diverse, it might not capture all variations in the problem, including different screening protocols, populations, and lesion types. Although the proposed solution reduces some of the effects of label noise, its presence should be taken into consideration. It appears that high-quality annotated datasets are vital for achieving promising performance. Lesion-based localization enhances the model's decision-making. Retrospective clinical data from publicly available mammography databases were used. While VinDr-Mammo and CBIS-DDSM differ in imaging conditions, a larger-scale multi-center validation across geographically diverse screening populations is needed. The framework is tested only for binary classification between malignant and benign cases, without considering lesion localization, BI-RADS scale ratings, or multiple classification options. Finally, no clinical prospective validation, radiologist-reader studies, or workflow-level analyses have been performed; thus, no clear conclusions can be drawn about the system's applicability in routine clinical practice.

Several directions for future research are worth exploring. An important direction would be to explicitly estimate prediction uncertainties to provide more information for decision-making and support risk-aware decisions. Further investigation into multimodal clinical data, uncertainty-aware prediction, and lesion-localization mechanisms may enhance the practical applicability of the proposed framework for computer-aided breast cancer diagnosis. Using longitudinal data with prior examinations may improve the performance through temporal reasoning. Finally, the idea of designing a reliable decision-support tool may be extended to other screening techniques and medical domains, such as lung cancer or cardiovascular disease screening.

6. Conclusions

The study introduces a robust, highly transferable architecture for breast cancer classification that addresses the major weaknesses of current deep learning models, particularly in heterogeneous mammographic data and decision reliability. The proposed model combines SDFE, heterogeneous feature learning using CNNs, ViTs, and MLP architectures, and a KAN-based classification head with an adaptive voting mechanism. The combination of these contributions enables the model to represent complementary representations, focus on reliable features, and capture intricate nonlinear decision boundaries more effectively than traditional approaches. The experimental results show that the proposed framework outperforms existing approaches in both internal and external assessments. The five-fold cross-validation results, which have been high and stable, combined with the best results

on the VinDr-Mammo test set and the CBIS-DDSM datasets, validate the model's ability to generalize across a wide range of data distributions. The enhanced precision–recall further underscores its efficiency in accurately distinguishing between benign and malignant cases, which is essential for clinical decision-making. These findings emphasize that the proposed method not only increases classification accuracy but also improves diagnostic reliability in a real screening environment. Despite these developments, there are some limitations. The research mainly centers on binary classification and makes no explicit reference to multi-class BI-RADS classification or lesion-based localization. Furthermore, despite the focus on computational efficiency, additional optimization may be needed to be deployed in the resource-constrained clinical setting. Future research directions will investigate multi-class classification, the inclusion of multimodal clinical data, and the enhancement of explainability through unified ways of interpretation across model branches. These guidelines will further improve the clinical relevance and reliability of automated breast cancer diagnosis systems.

Author Contributions: Conceptualization, A.R.W.S. and Y.A.; methodology, A.R.W.S. and Y.A.; software, A.R.W.S. and Y.A.; validation, A.R.W.S. and Y.A.; formal analysis, A.R.W.S.; investigation, A.R.W.S. and Y.A.; resources, A.R.W.S. and Y.A.; data curation, A.R.W.S.; writing—original draft preparation, A.R.W.S. and Y.A.; writing—review and editing, A.R.W.S. and Y.A.; visualization, A.R.W.S. and Y.A.; supervision, A.R.W.S. and Y.A.; project administration, A.R.W.S. and Y.A.; funding acquisition, A.R.W.S. and Y.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. KFU262733].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets are available in the repositories: VinDr Dataset, https://vindr.ai/datasets/mammo, accessed on 2 March 2026, and CBIS-DDSM, https://www.kaggle.com/datasets/orvile/cbis-ddsm-dataset, accessed on 2 March 2026.

Acknowledgments: The authors would like to thank the Deanship of Scientific Research at Shaqra University for supporting this work. The authors thank King Faisal University's Deanship of Scientific Research for its financial support [Grant No. KFU KFU262733].

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Siegel, R.L.; Miller, K.D.; Wagle, N.S.; Jemal, A. Cancer statistics, 2023. *CA Cancer J. Clin.* **2023**, *73*, 17–48. [CrossRef]
2. Chong, A.; Weinstein, S.P.; McDonald, E.S.; Conant, E.F. Digital breast tomosynthesis: Concepts and clinical practice. *Radiology* **2019**, *292*, 1–4. [CrossRef]
3. Jaglan, P.; Dass, R.; Duhan, M. Breast cancer detection techniques: Issues and challenges. *J. Inst. Eng. (India) Ser. B* **2019**, *100*, 379–386. [CrossRef]
4. Arnold, M.; Morgan, E.; Rumgay, H.; Mafra, A.; Singh, D.; Laversanne, M.; Vignat, J.; Gralow, J.R.; Cardoso, F.; Siesling, S.; et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040. *Breast* **2022**, *66*, 15–23. [CrossRef]
5. Shah, T.A.; Guraya, S.S. Breast cancer screening programs: Review of merits, demerits, and recent recommendations practiced across the world. *J. Microsc. Ultrastruct.* **2017**, *5*, 59–69. [CrossRef] [PubMed]
6. Bevers, T.B.; Anderson, B.O.; Bonaccio, E.; Buys, S.; Daly, M.B.; Dempsey, P.J.; Farrar, W.B.; Fleming, I.; Garber, J.E.; Harris, R.E.; et al. Breast cancer screening and diagnosis. *J. Natl. Compr. Cancer Netw.* **2009**, *7*, 1060–1096. [CrossRef]
7. Singh, V.; Saunders, C.; Wylie, L.; Bourke, A. New diagnostic techniques for breast cancer detection. *Future Oncol.* **2008**, *4*, 501–513. [CrossRef]

8. Abdul Halim, A.A.; Andrew, A.M.; Mohd Yasin, M.N.; Abd Rahman, M.A.; Jusoh, M.; Veeraperumal, V.; Rahim, H.A.; Illahi, U.; Abdul Karim, M.K.; Scavino, E. Existing and emerging breast cancer detection technologies and its challenges: A review. *Appl. Sci.* **2021**, *11*, 10753. [CrossRef]
9. Smith, A.P.; Hall, P.A.; Marcello, D.M. Emerging technologies in breast cancer detection. *Radiol. Manag.* **2004**, *26*, 16–24.
10. Abimouloud, M.L.; Bensid, K.; Elleuch, M.; Ammar, M.B.; Kherallah, M. Advancing breast cancer diagnosis: Token vision transformers for faster and accurate classification of histopathology images. *Vis. Comput. Ind. Biomed. Art* **2025**, *8*, 1. [CrossRef] [PubMed]
11. Demiroğlu, U.; Şenol, B. Benchmarking Deep Learning Models for Breast Cancer Detection: A Comparison of Vision Transformers and CNNs. *Acad. Platf. J. Eng. Smart Syst.* **2025**, *13*, 108–119. [CrossRef]
12. Qin, G.; Liu, H.; Li, W.; Zhang, X.; Guo, Y.; Xia, X.G. UFPF: A Universal Feature Perception Framework for Microscopic Hyperspectral Images. *IEEE Trans. Image Process.* **2025**, *34*, 5513–5526. [CrossRef]
13. Bechar, A.; Elmir, Y.; Medjoudj, R.; Himeur, Y.; Amira, A. Transfer learning for cancer detection based on images analysis. *Procedia Comput. Sci.* **2024**, *239*, 1903–1910. [CrossRef]
14. Qin, G.; Liu, H.; Zhang, X.; Li, W.; Guo, Y.; Guo, C. Semi-Supervised Medical Hyperspectral Image Segmentation Using Adversarial Consistency Constraint Learning and Cross Indication Network. *IEEE Trans. Image Process.* **2025**, *34*, 5414–5428. [CrossRef] [PubMed]
15. Hayat, M.; Ahmad, N.; Nasir, A.; Tariq, Z.A. Hybrid deep learning EfficientNetV2 and vision transformer (EffNetV2-ViT) model for breast cancer histopathological image classification. *IEEE Access* **2024**, *12*, 184119–184131. [CrossRef]
16. Ayana, G.; Choe, S.W. Vision transformers-based transfer learning for breast mass classification from multiple diagnostic modalities. *J. Electr. Eng. Technol.* **2024**, *19*, 3391–3410. [CrossRef]
17. Othman, E.M. Breast Cancer Multi-Class Classification Using ViT Model. *Int. J. Comput. Appl.* **2024**, *186*, 13–18. [CrossRef]
18. Patheda, V.R.; Laxmisai, G.; Gokulnath, B.V.; Ibrahim, S.S.; Kumar, S.S. A robust hybrid CNN+ ViT framework for breast cancer classification using mammogram images. *IEEE Access* **2025**, *13*, 77187–77195. [CrossRef]
19. Gutierrez-Cardenas, J. Breast Cancer Classification Through Transfer Learning with Vision Transformer, PCA, and Machine Learning Models. *Int. J. Adv. Comput. Sci. Appl.* **2024**, *15*. [CrossRef]
20. Alruily, M.; Mahmoud, A.A.; Allahem, H.; Mostafa, A.M.; Shabana, H.; Ezz, M. Enhancing Breast Cancer Detection in Ultrasound Images: An Innovative Approach Using Progressive Fine-Tuning of Vision Transformer Models. *Int. J. Intell. Syst.* **2024**, *2024*, 6528752. [CrossRef]
21. Fatima, T.; Soliman, H. Application of VGG16 transfer learning for breast cancer detection. *Information* **2025**, *16*, 227. [CrossRef]
22. Abdikenov, B.; Rakishev, D.; Orazayev, Y.; Zhaksylyk, T. Enhancing breast lesion detection in mammograms via transfer learning. *J. Imaging* **2025**, *11*, 314. [CrossRef] [PubMed]
23. Hendry, A.; Felicia, E.; Natanael, Y.; Rumagit, R.Y. Implementation of Explainable AI on CNN-and ViT-Based Models for Classifying Breast Cancer. *Procedia Comput. Sci.* **2025**, *269*, 968–978. [CrossRef]
24. Figueiredo, I.S.; de Oliveira, J.P.; da Silva Galindo, J.M.; Vieira, J.V. A progressive fine-tuning strategy for robust breast lesion detection across multiple mammography datasets. *Cad. Pedagóg.* **2026**, *23*, e22794. [CrossRef]
25. VinDr Dataset. Available online: https://vindr.ai/datasets/mammo (accessed on 2 March 2026).
26. CBIS-DDSM. Available online: https://www.kaggle.com/datasets/orvile/cbis-ddsm-dataset (accessed on 2 March 2026).
27. Available online: https://www.kaggle.com/datasets/tommyngx/inbreast2012 (accessed on 2 March 2026).
28. Lonia, G.; Ciraolo, D.; Fazio, M.; Villari, M.; Celesti, A. Comparing CNNs and ViTs for Medical Image Classification Leveraging Transfer Learning. In *2024 IEEE Symposium on Computers and Communications (ISCC)*; IEEE: New York, NY, USA, 2024; pp. 1–7.
29. Jain, K.; Bansal, A.; Rangarajan, K.; Arora, C. MMBCD: Multimodal breast cancer detection from mammograms with clinical history. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer Nature: Cham, Switzerland, 2024; pp. 144–154.
30. Yerramaneni, S.; Reddy, S.K. A review on breast cancer detection using machine learning techniques. *Int. J. Data Min. Model. Manag.* **2025**, *17*, 142–164.
31. Sarıateş, M.; Özbay, E. A classifier model using fine-tuned convolutional neural network and transfer learning approaches for prostate cancer detection. *Appl. Sci.* **2024**, *15*, 225. [CrossRef]
32. Khaled, M.; Touazi, F.; Gaceb, D. Improving breast cancer diagnosis in mammograms with progressive transfer learning and ensemble deep learning. *Arab. J. Sci. Eng.* **2025**, *50*, 7697–7720. [CrossRef]
33. Madani, M.; Behzadi, M.M.; Nabavi, S. The role of deep learning in advancing breast cancer detection using different imaging modalities: A systematic review. *Cancers* **2022**, *14*, 5334. [CrossRef] [PubMed]
34. Tamo, J.B.; Nnamdi, M.C.; Lesbats, L.; Shi, W.; Zhong, Y.; Wang, M.D. Uncertainty-aware ensemble learning models for out-of-distribution medical imaging analysis. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*; IEEE: New York, NY, USA, 2023; pp. 4243–4250.

35. Abimouloud, M.L.; Bensid, K.; Elleuch, M.; Ammar, M.B.; Kherallah, M. Vision transformer based convolutional neural network for breast cancer histopathological images classification. *Multimed. Tools Appl.* **2024**, *83*, 86833–86868. [CrossRef]
36. Akselrod-Ballin, A.; Karlinsky, L.; Alpert, S.; Hasoul, S.; Ben-Ari, R.; Barkan, E. A region based convolutional network for tumor detection and classification in breast mammography. In *International Workshop on Deep Learning in Medical Image Analysis*; Springer International Publishing: Cham, Switzerland, 2016; pp. 197–205.
37. Chougrad, H.; Zouaki, H.; Alheyane, O. Deep convolutional neural networks for breast cancer screening. *Comput. Methods Programs Biomed.* **2018**, *157*, 19–30. [CrossRef] [PubMed]
38. Dhungel, N.; Carneiro, G.; Bradley, A.P. A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *Med. Image Anal.* **2017**, *37*, 114–128. [CrossRef] [PubMed]
39. Shen, L.; Margolies, L.R.; Rothstein, J.H.; Fluder, E.; McBride, R.; Sieh, W. Deep learning to improve breast cancer detection on screening mammography. *Sci. Rep.* **2019**, *9*, 12495. [CrossRef]
40. Ribli, D.; Horváth, A.; Unger, Z.; Pollner, P.; Csabai, I. Detecting and classifying lesions in mammograms with deep learning. *Sci. Rep.* **2018**, *8*, 4165. [CrossRef]
41. Zeynali, A.; Tinati, M.A.; Tazehkand, B.M. Hybrid CNN-transformer architecture with xception-based feature enhancement for accurate breast cancer classification. *IEEE Access* **2024**, *12*, 189477–189493. [CrossRef]
42. Sharma, S.; Singh, Y. Enhancing breast cancer detection using A transformer-based model. In *2024 International Conference on Information Science and Communications Technologies (ICISCT)*; IEEE: New York, NY, USA, 2024; pp. 264–269.
43. Murty, P.S.; Anuradha, C.; Naidu, P.A.; Mandru, D.; Ashok, M.; Atheeswaran, A.; Rajeswaran, N.; Saravanan, V. Integrative hybrid deep learning for enhanced breast cancer diagnosis: Leveraging the Wisconsin Breast Cancer Database and the CBIS-DDSM dataset. *Sci. Rep.* **2024**, *14*, 26287. [CrossRef]
44. Firouzbakht, M.; Amirmazlaghani, M. Breast cancer detection in mammography images using Neighborhood Attention transformer and Shearlet Transform. *Comput. Biol. Med.* **2025**, *198*, 111239. [CrossRef]
45. Yoshitsugu, K.; Kishimoto, K.; Takemura, T. Improving Benign and Malignant Classifications in Mammography with ROI-Stratified Deep Learning. *Bioengineering* **2025**, *12*, 885. [CrossRef]
46. Asif, S.; Zhu, L.; Yan, D.; Xu, L.; Huang, Z.; Xu, H.; Yan, R.; Cai, L.; Zheng, C.; Lin, J.; et al. BREAST-RANKNet: A fuzzy rank-based ensemble of CNNs with residual learning for enhanced breast cancer detection from ultrasound and mammogram images. *J. Big Data* **2025**, *12*, 194. [CrossRef]
47. Khan, S.U.; Asif, S.; Bilal, O. Ensemble architecture of vision transformer and CNNs for breast cancer tumor detection from mammograms. *Int. J. Imaging Syst. Technol.* **2025**, *35*, e70090. [CrossRef]
48. Nguyen, H.T.; Nguyen, H.Q.; Pham, H.H.; Lam, K.; Le, L.T.; Dao, M.; Vu, V. Vindr-mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography. *Sci. Data* **2023**, *10*, 277. [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.